



# Automatic Mining and Comparative Analysis of Animal and Plant Metaphors in Chinese and Western Classical Texts

Siwen Wang<sup>1\*</sup>, Yuxuan Liu<sup>1</sup>, Bo Chen<sup>2</sup>, Xiaobing Zhao<sup>2</sup>

<sup>1</sup>School of Chinese Ethnic Minority Languages and Literatures, Minzu University of China, Beijing, China

<sup>2</sup>School of Information Engineering, Minzu University of China, Beijing, China

\*23301837@muc.edu.cn

**Abstract.** This paper proposes a large language model (LLM)-based automatic mining framework for identifying and analyzing animal and plant metaphors in Chinese and Western classical texts. Treating metaphor identification as a structured information extraction task, we design a multi-stage pipeline integrating prompt-guided entity extraction, contextual metaphor classification, semantic labeling, and human-in-the-loop verification. Using *Chu Ci* and *Aesop's Fables* as case studies, we construct a validated metaphor dataset containing 2,172 instances. Experimental analysis shows that the proposed framework achieves high coverage while substantially reducing manual annotation effort. The results demonstrate the feasibility of applying LLM-driven computational methods to cross-cultural metaphor research in digital humanities.

**Keywords:** animal metaphor; plant metaphor; automatic metaphor mining

## 1 Introduction

In literary texts, animals and plants are often endowed with metaphorical meanings and function as important carriers of cultural significance. As two representative works of the Eastern and Western literary traditions, *Chu Ci* and *Aesop's Fables* both contain a large number of animal and plant entities. Investigating and comparing the metaphorical uses of these entities can help reveal the mechanisms through which "natural symbols" are transformed into "cultural meanings" in early Eastern and Western literatures [1]. Existing studies on animal and plant metaphors mainly fall into two categories. The first focuses on close reading and cultural interpretation of classical texts, emphasizing the symbolic meanings of metaphorical images and their cultural origins [2]. The second adopts cognitive linguistics as its theoretical framework to explore the conceptual structure of metaphors and their cross-cultural differences [3]. While these studies have produced rich insights into the cultural connotations of metaphors, they are largely based on manual analysis and lack support from scalable computational methods.

Metaphor identification in literary texts is traditionally conducted through close reading and qualitative interpretation, which limits scalability and reproducibility. In recent years, advances in large language models (LLMs) have provided new opportunities for applying computational methods to complex semantic tasks in the humanities. However, how to systematically integrate LLM-based techniques into metaphor research—especially for classical texts across different cultural traditions—remains an open challenge.

Within the fields of computational linguistics and digital humanities, some studies have attempted to employ techniques such as named entity recognition[4], word embeddings[5], and topic modeling[6] to automatically identify and analyze animal and plant entities in literary texts. However, most existing work concentrates on a single culture or a single textual tradition, with research objectives primarily limited to entity extraction or lexical co-occurrence analysis. Relatively little attention has been paid to distinguishing between literal entities and metaphorical usages, and few studies have conducted systematic comparisons across different civilizations under a unified methodological framework. In particular, at the level of Chinese and Western classical texts, cross-cultural comparative studies centered on animal and plant metaphors and grounded in automated text-mining techniques remain scarce.

In this paper, we propose an LLM-based automatic mining framework that decomposes metaphor identification into modular subtasks and integrates human verification to ensure semantic reliability. The framework is applied uniformly to Chu Ci and Aesop’s Fables, enabling quantitative comparison under a unified technical pipeline. The contributions of this work are threefold: (1) we design a reproducible LLM-based metaphor mining pipeline for classical texts; (2) we construct a structured animal and plant metaphor dataset through human-in-the-loop validation; and (3) we provide quantitative evidence supporting cross-cultural differences in metaphor usage.

In this paper, "animal and plant metaphors" refer to cases in Chu Ci and Aesop’s Fables where naturally existing animal or plant entities are assigned non-literal meanings. It should be emphasized that this research focuses only on animal and plant entities that exhibit clear metaphorical significance, excluding those that function solely as literal descriptions of the natural world.

## 2 Related Work

### 2.1 Research on Animal and Plant Metaphors

Traditional scholarship on animal and plant metaphors primarily relies on close reading to interpret symbolic meanings and cultural origins [7, 8, 9]. While Cognitive Metaphor Theory has since provided robust frameworks for analyzing metaphorical mapping patterns across cultures [10], both approaches remain predominantly qualitative. They often lack the scalability and systematic validation required for large-scale corpus analysis, a gap that computational humanities methods are uniquely positioned to fill [11, 12].

## 2.2 Comparative Studies

Comparative analysis is a common approach in cross-cultural metaphor research, with existing studies primarily focusing on similarities and differences in metaphorical expressions across languages or cultural systems. In literary and linguistic research, comparative studies of Chinese and Western metaphors are typically theory-driven, selecting representative texts to analyze metaphor types, choices of source domains, and underlying cultural value orientations[13]. By comparing metaphorical expressions across different languages or cultural systems, such studies explore source-domain selection, metaphor categories, and their cultural motivations[14]. A widely shared conclusion in this line of research is that, although humans share certain bodily experiences, cultural background exerts a significant influence on how metaphors are encoded[15][16]. This conclusion has been supported by multiple comparative studies on animal and plant metaphors in English and Chinese[17][18]. However, most of these studies are based on a limited number of examples, making their findings difficult to verify quantitatively, and they also lack unified standards for metaphor annotation.

In the field of digital humanities, comparative analysis has increasingly incorporated statistical methods and semantic modeling techniques to compare the distribution of imagery across different texts. Such studies commonly employ word frequency analysis, co-occurrence networks, or topic modeling to examine the distribution of keywords or imagery, with the aim of revealing cultural themes or narrative tendencies[19][20]. Nevertheless, in metaphor research, these methods are mostly used to describe surface-level frequency or co-occurrence patterns, and rarely distinguish between literal and metaphorical usages. Consequently, comparative results often reflect only superficial distributional differences. Comparative metaphor research on classical texts continues to face challenges such as data sparsity, semantic ambiguity, and inconsistent annotation standards. These difficulties are particularly pronounced in comparisons between Chinese and Western classical texts, where substantial differences in language systems and cultural backgrounds make direct automatic comparison highly challenging. As a result, relatively few studies have conducted systematic comparisons of animal and plant metaphors in Chinese and Western classical texts within a unified methodological framework. This gap provides room for further exploration by integrating automated extraction methods with manual verification to conduct systematic cross-cultural comparative studies of animal and plant metaphors.

## 3 Animal and Plant Metaphors and LLM-Based Automatic Mining Framework

This study formulates animal and plant metaphor identification as a large language model (LLM)-assisted information extraction task with human-in-the-loop verification. Instead of relying on purely rule-based or supervised learning approaches, we design a multi-stage prompt-driven pipeline that integrates automated extraction, seman-

tic inference, and manual validation. The proposed framework emphasizes reproducibility, modularity, and interpretability, making it suitable for computational humanities research involving classical texts.

### 3.1 Task Definition

The research objects of this paper are animal and plant entities with metaphorical meanings in *Chu Ci* and Aesop's Fables. In this paper, "animal and plant metaphors" refer to cases in which animal or plant entities appearing in a text no longer merely denote natural objects in a given context, but are instead used to express personality traits, social relations, moral evaluations, or abstract concepts. Given a classical text corpus  $C = \{s_1, s_2, \dots, s_n\}$ , where each  $s_i$  denotes a sentence or verse, the goal of this task is to automatically identify metaphorical usages of animal and plant entities and construct a structured metaphor dataset. Each metaphor instance is formalized as a tuple:  $(e, t, s, l)$  where  $e$  denotes the entity name,  $t \in \{\text{animal, plant}\}$  denotes the entity type,  $s$  is the sentence context, and  $l$  represents the metaphorical semantic label assigned to the entity. The task is decomposed into three subtasks: (1) animal and plant entity extraction, (2) metaphorical usage classification, and (3) metaphor semantic labeling.

For *Chu Ci*, we adopt the *Chu Ci Zhangju* (annotated by Wang Yi) and *Chu Ci Jizhu* (annotated by Zhu Xi) as base texts. These editions provide a foundational reference chain for interpreting metaphorical systems such as 'fragrant plants and beauties'. For *Aesop's Fables*, we utilize the Oxford University Press English critical edition translated by Laura Gibbs. This version is aligned with ancient Greek fragments and includes comprehensive annotations on metaphorical variations. Using these scholarly editions ensures that our automated mining is grounded in established philological traditions.

### 3.2 Overall Framework

The automated mining experiment in this study was implemented based on OpenAI's ChatGPT-4o (gpt-4o-2024-08-06) model. The experiment was conducted through official API calls, with testing taking place from May 10, 2025, to May 25, 2025. To ensure the stability and rigor of the generated results, the model decoding parameters were set as follows: Temperature at 0.2, Top-p sampling at 0.95, and a maximum generation length of 1024 tokens.

This study adopts a prompt-based generative information extraction approach [21] to metaphor mining, in which the metaphor identification task is decomposed into a set of clearly defined subtasks. These subtasks are completed step by step by a large language model to generate metaphor candidates, which are then refined through manual verification to achieve high-quality annotation.

Drawing on existing studies that employ instruction-based prompting to guide large language models in task completion [22][23], this study differs from traditional rule-based or supervised learning approaches. Rather than training a new metaphor identification model, we leverage the large language model's strengths in semantic understand-

ing and cross-linguistic transfer to perform animal and plant entity recognition and metaphor candidate generation, while introducing manual verification in later stages to control semantic noise.

Figure 1 illustrates the overall architecture of the proposed LLM-based metaphor mining framework. The system consists of five main modules: text preprocessing, entity extraction, metaphor candidate generation, human verification, and dataset construction.

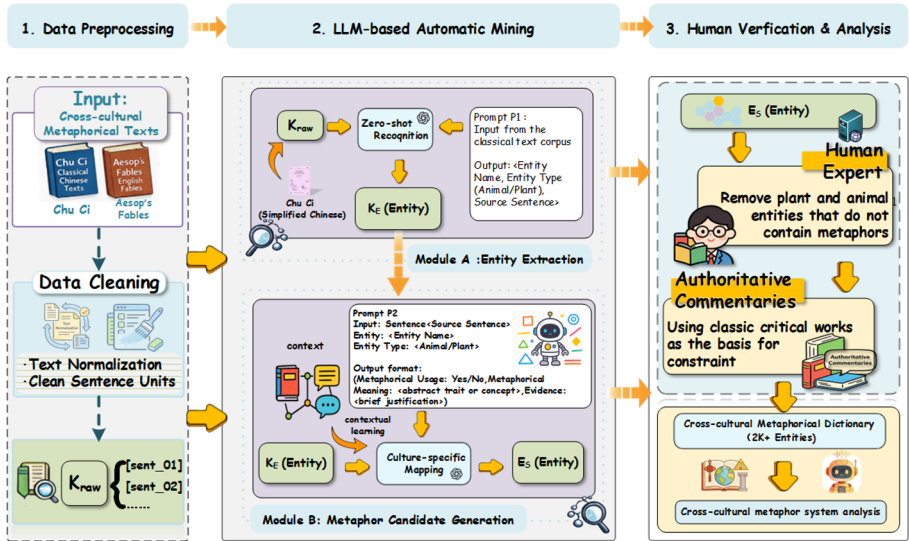


Fig. 1. Overall workflow of the automatic mining framework

### 3.2.1 Text Preprocessing.

Before model inference, all texts were normalized to ensure input consistency. This included converting traditional Chinese characters to simplified script for Chu Ci, segmenting texts into verse or sentence-level units, and removing non-textual elements such as editorial notes. Each normalized unit serves as the input for downstream extraction and classification tasks.

### 3.2.2 Prompt-Guided Entity Extraction.

Animal and plant entity extraction is implemented using a prompt-based generative information extraction strategy. Instead of applying conventional named entity recognition models, we leverage the semantic understanding capability of LLMs to identify candidate entities under contextual constraints.

To improve recall, a domain-specific lexicon containing pre-Qin plant names and frequently occurring animal names is incorporated as soft constraints within the prompt. The LLM is instructed to extract all explicit references to animals and plants while excluding abstract or purely functional expressions.

The extraction prompt is formalized as:

$$P\_extract(s) \rightarrow \{(e, t, s)\}$$

where  $s$  denotes the input sentence and the output is a set of entity tuples.

Table 1 shows the standardized template used for entity extraction.

**Table 1.** Standardized Prompt Template for Animal and Plant Entity Extraction

| Component            | Specification  |
|----------------------|--|
| System Role          | A linguistics and philology expert specializing in classical Chinese and Western literary texts.                                   |
| Task Objective       | Automatically extract explicit animal and plant entities from the input sentence.  |
| Input Unit           | A single sentence segmented from the preprocessed classical text corpus.   |
| Extraction Scope     | Only entities denoting concrete animals or plants are extracted; abstract concepts and purely functional expressions are excluded. |
| Output Format        | (Entity Name, Entity Type (Animal/Plant), Source Sentence)   |
| Prompt Type          | Instruction-based prompt with explicit role definition and structured output constraints.  |
| Decoding Temperature | 0.2  |
| Top-p Sampling       | 0.95   |

### 3.2.3 Automatic Generation of Animal and Plant Metaphors.

After entity extraction, the next step is to determine whether the identified animal and plant entities potentially constitute metaphorical usage. This study adopts a context-based metaphor candidate generation strategy, in which the semantic features of the sentences containing the entities are analyzed to guide the model in generating potential metaphorical interpretations.

Specifically, sentences containing animal or plant entities are examined to determine whether they co-occur with references to personality traits, social roles, or abstract concepts. If an entity exhibits a semantic mapping that goes beyond literal reference within a given context, the instance is marked as a "metaphor candidate."

Prompt-based instructions are used to guide the model in generating metaphor labels. For example, for the Chu Ci sentence "I wear jiangli and bizhi at my side, and thread autumn orchids to make my girdle." The prompt "Please identify the metaphorical meanings of the plant entities in the sentence, using the format 'Entity: Metaphorical Meaning'" yields outputs such as "Jiangli: gentleman; Bizhi: gentleman; Qiulan: noble character." For the Aesop's Fables sentence "The fox praised the crow's song to get the meat," the prompt "What metaphorical meaning does the animal entity in the sentence have?" Format: 'Entity: Meaning'" produces the output "fox: cunning, flattery."

This stage aims to generate a high-coverage set of metaphor candidates, which serves as input for subsequent manual verification.

For each extracted entity, a second-stage prompt is applied to determine whether the entity is used metaphorically or literally in context. This stage relies on contextual semantic inference rather than surface lexical cues.

The classification function is defined as:

$$P\_meta(e, s) \rightarrow \{\text{literal, metaphorical}\}$$

If an entity is classified as metaphorical, an additional prompt is invoked to generate a semantic label describing its metaphorical meaning:

$$P\_label(e, s) \rightarrow l$$

Prompt parameters are fixed across experiments to ensure consistency. The temperature is set to 0.2 to reduce output randomness, and top-p sampling is set to 0.95. The output format is strictly constrained to predefined schemas to facilitate downstream processing.

Table 2 shows the standardized template used for metaphor classification and labeling.

**Table 2.** Standardized Prompt Template for Metaphor Classification and Labeling

| Component            | Specification  |
|----------------------|--|
| System Role          | You are an expert in Cognitive Linguistics and Metaphor Theory.  |
| Task Step 1          | Determine if the entity is used Literally (denoting the natural object) or Metaphorically (representing an abstract trait, social role, or moral value). |
| Task Step 2          | If metaphorical, generate a concise semantic label (e.g., "noble character", "cunning", "political authority").  |
| Task Step 3          | Provide a brief justification based on the imagery and cultural background provided in the sentence.   |
| Input                | Entity: [e], Sentence: [s]   |
| Output Format        | <"Entity": "[Entity Name]", "Usage": "[Literal/Metaphorical]", "Semantic_Label": "[Label]", "Evidence": "[Reasoning]" >                                  |
| Decoding Temperature | 0.2  |
| Top-p Sampling       | 0.95   |

### 3.2.4 Manual Verification.

Because metaphor identification inherently involves semantic and cultural knowledge, fully automated methods cannot completely avoid misclassification. Therefore, a manual verification stage is introduced after metaphor candidate generation.

Manual verification involves the following tasks: removing false-positive metaphor instances; correcting metaphor labels (e.g., distinguishing context-dependent polysemous metaphors); and unifying descriptive standards for metaphor labels. Authoritative annotated editions and existing scholarship are used as references to ensure the interpretability and theoretical consistency of metaphor annotations.

#### Verification Step 1: Cross-checking with Authoritative Annotations

For *Chu Ci*, metaphorical meanings generated by the model are verified against annotations in *Chu Ci Zhangju* and *Chu Ci Jizhu*. For example, if the model labels "Pi li" as "gentleman," Wang Yi's annotation—"Pi Li, a weed, is a metaphor for treacherous and deceitful people" to "treacherous and deceitful people".

For Aesop's Fables, metaphor labels are verified against annotations in Laura Gibbs's translation and analyzes in *The Art of Aesop* by Hansen. For instance, if the model labels "snake" as "evil," manual verification refines the label to "ingratitude," corresponding to the core metaphor in *The Farmer and the Snake*.

#### Verification Step 2: Contextual Consistency Evaluation

Occurrences of the same entity across different contexts are examined. For example, "orchid" appears 28 times in *Li Sao*, and its metaphorical labels are evaluated for contextual consistency. In cases such as "Orchids and angelica may change, but they will not lose their fragrance," where "orchid" metaphorically signifies "moral degeneration due to environmental change," a separate label is assigned to distinguish it from the more common metaphor of "noble character."

Similarly, different behaviors of the same animal in Aesop's Fables (e.g., the fox deceiving the crow versus the fox assisting the lion) are analyzed to differentiate subtle variations in metaphorical meanings (e.g., "deceptiveness" vs. "cleverness").

To ensure the high fidelity of the automatically extracted metaphors and their associated semantic labels, we implemented a rigorous human-in-the-loop verification protocol. We recruited three graduate students specializing in Classical Philology as independent annotators. The verification process followed a two-stage blind annotation and expert adjudication framework:

- 1.Independent Annotation: Two primary annotators independently reviewed the 2,172 candidate instances generated by the LLM. They were tasked with evaluating the accuracy of the metaphor classification (Binary: Metaphorical vs. Literal) and the appropriateness of the generated semantic labels.

- 2.Conflict Resolution: In cases of disagreement between the two primary annotators, a third senior annotator acted as a referee to make the final decision after a comprehensive review of the textual context and scholarly commentaries (e.g., Wang Yi's *Chu Ci Zhang Ju*).

To quantify the reliability of the annotation, we calculated inter-annotator agreement (IAA) metrics. For the metaphor identification task, the Cohen's Kappa coefficient was 0.84, indicating "almost perfect agreement" according to the Landis and Koch scale. For the more complex task of semantic label induction, we achieved a Fleiss' Kappa of 0.76, representing substantial agreement. These results demonstrate that the LLM-driven framework produces outputs that are highly consistent with human expert judgment, ensuring the robustness of the subsequent comparative analysis.

### 3.2.5 Semantic Taxonomy and Label Normalization.

To address the experts' concerns regarding the granularity and consistency of the 2,172 raw labels generated by the LLM, we implemented a rigorous normalization and hierarchical classification framework.

#### 1. Hierarchical Taxonomy Construction

We designed a three-tier taxonomy based on cognitive linguistic ontologies to balance descriptive richness with statistical rigor:

**Top-level Dimensions:** Three primary domains—*Personality Traits*, *Social Roles*, and *Moral/Behavioral Codes*—were established to serve as the structural backbone for cross-cultural comparison.

**Mid-level Categories:** These provide functional grouping (e.g., "Political Authority" vs. "Marginalized Status" within the Social Role domain).

**Leaf Nodes:** These represent the fine-grained, original semantic labels (e.g., "Nobility," "Cunning," "Servitude").

**2. Synonym Consolidation and Disambiguation** To mitigate the "vocabulary explosion" and redundancy inherent in LLM outputs:

**Synonym Clustering:** We utilized the text-embedding-3-small model to calculate cosine similarity between labels. Pairs with similarity scores exceeding 0.85 (e.g., "treacherous" and "deceitful") were automatically clustered and manually mapped to a standardized term (e.g., "Cunning").

**Polysemy Resolution:** For polysemous labels (e.g., "Pride" indicating "Dignity" vs. "Arrogance"), the framework re-evaluated the original context window and emotional valence to ensure precise mapping within the hierarchy.

#### 3. Ontological Mapping for Downstream Analysis

This systematic normalization transforms unstructured textual imagery into a structured ontological format. By aggregating individual instances into these broader categories, we eliminate the "long-tail effect" of varied LLM phrasings. This enables a clearer quantitative comparison—such as analyzing the relative prevalence of "Moral Codes" in *Chu Ci* versus "Social Hierarchies" in *Aesop's Fables*—and provides a robust foundation for downstream cultural-interpretive analysis.

## 3.3 Experimental Results and Analysis

Based on the animal and plant metaphor dataset constructed through the automatic mining and manual verification process described above, this section conducts a statistical analysis of the scale of metaphor instances, the distribution of entities, and the characteristics of semantic categories. The statistical results are intended to present, from a quantitative perspective, the overall patterns of animal and plant metaphor usage in *Chu Ci* and *Aesop's Fables*, thereby providing a data foundation for subsequent cross-text comparative analysis.

### 3.3.1 Metaphor Dataset Construction and Statistics.

After verification, a structured metaphor dataset is constructed. Each entry contains text source, sentence, entity name, entity type, and metaphorical semantic label. The

final dataset includes 2,172 validated metaphor instances, with 640 instances from Chu Ci and 1,487 from Aesop’s Fables.

To evaluate the effectiveness of the automated pipeline, we compare the number of LLM-generated candidates with the final validated instances. Approximately 76% of automatically generated candidates are retained after human verification, indicating that the proposed framework achieves high coverage while maintaining acceptable precision.

The experimental results demonstrate that the proposed framework exhibits significant operational efficiency in processing large-scale classical corpora. For the preliminary dataset containing 2,172 candidate instances, the LLM-driven automated pipeline completed the entire workflow—encompassing entity extraction, metaphor classification, and semantic labeling—in approximately 15 minutes.

In contrast, a traditional manual annotation approach, assuming an average processing rate of 40 instances per hour per expert (inclusive of context verification and philological reference checking), would require an estimated 54 man-hours of net labor. By automating over 90% of the repetitive low-level identification tasks, the framework allows researchers to reallocate their cognitive efforts toward high-level semantic synthesis and cross-cultural interpretive validation. Quantitative assessment indicates that this methodology reduces the overall research cycle by approximately 70%, substantially lowering the resource barriers for large-scale comparative studies in digital humanities.

Furthermore, compared with fully manual annotation, the proposed pipeline significantly reduces annotation workload by filtering non-metaphorical entities in early stages, demonstrating its practical utility for large-scale metaphor research.

**Table 3.** Number of Metaphorical Examples of Plants and Animals in Chu Ci and Aesop’s Fables

| text           | plant metaphor | animal metaphor | total |
|----------------|----------------|-----------------|-------|
| Chu Ci         | 302            | 238             | 640   |
| Aesop’s Fables | 227            | 1260            | 1487  |

At the entity level, this study counted the number of distinct animal and plant entities involved in metaphor construction in the two texts and constructed an entity-based metaphor lexicon accordingly. The number of metaphorical examples of animals and plants in Chu Ci and Aesop’s Fables is shown in Table 3. The lexicon is derived from entities that actually appear in the texts and have been manually confirmed as metaphorical, rather than being defined a priori.

**Table 4.** Scale of the animal and plant metaphor entity lexicon in Chu Ci and Aesop’s Fables

| text           | plant | animal | all |
|----------------|-------|--------|-----|
| Chu Ci         | 101   | 67     | 167 |
| Aesop’s Fables | 34    | 10     | 44  |

In Table 4, we count the scale of animal and plant metaphorical entity lexicon in Chu Ci and Aesop’s Fables. The statistical results indicate that Chu Ci and Aesop’s Fables differ in both the overall scale of metaphorical entities and the composition of animal

versus plant categories. These differences provide a basis for subsequent analyzes of distributional characteristics and semantic functions.

### 3.3.2 Frequency Distribution Characteristics of Animal and Plant Metaphor Entities.

Based on the constructed metaphor entity lexicon, this study further analyzes the frequency distribution of animal and plant metaphor entities in the texts, which are shown in table 5 and table 6. The results show that the distribution of metaphorical entities in both texts exhibits a pronounced imbalance.

Specifically, a small number of animal and plant entities occur with high frequency in metaphorical usage and account for a substantial proportion of metaphorical expressions, while the majority of entities appear only in a limited number of contexts. Overall, the distribution displays a typical long-tail pattern. This distributional characteristic is observed in both *Chu Ci* and Aesop's Fables; however, the specific composition of high-frequency entities differs significantly between the two texts.

**Table 5.** High-frequency entities and their occurrence frequencies of *Chu Ci* (Top 10)

| entity       | type   | frequency |
|--------------|--------|-----------|
| orchid(兰)    | plant  | 48        |
| hui(蕙)       | plant  | 24        |
| osmanthus(菊) | plant  | 24        |
| pepper(椒)    | plant  | 16        |
| steed(骐驎)    | animal | 13        |
| fox(狐)       | animal | 12        |
| lotus(荷)     | plant  | 11        |
| mulberry(桑)  | plant  | 11        |
| tiger(虎)     | animal | 11        |
| phoenix(皇)   | animal | 11        |

**Table 6.** High-frequency entities and their occurrence frequencies of Aesop's Fables (Top 10)

| entity   | type   | frequency |
|----------|--------|-----------|
| crow     | animal | 137       |
| tortoise | animal | 109       |
| peacock  | animal | 102       |
| Bear     | animal | 96        |
| cattle   | animal | 95        |
| fox      | animal | 83        |
| dog      | animal | 72        |
| lion     | animal | 64        |
| wolf     | animal | 61        |
| pig      | animal | 59        |

To further illustrate the structure of the metaphor entity lexicon, representative examples from the two texts are provided, including entity type, original sentence, metaphorical label, and frequency, which are shown in table 7 and table 8.

**Table 7.** Sample entries in *Chu Ci*

| Entity            | Entity Type | Original Sentence   | Metaphorical Label              |
|-------------------|-------------|---|---------------------------------|
| Orchid(兰)         | Plant       | “I wear Jiangli and Bizhi at my side, and thread autumn orchids to make my girdle.”                                     | gentleman; noble character      |
| Chrysanthemums(菊) | Plant       | “From spring orchids to autumn chrysanthemums, their constancy endures through ancient times.”                          | steadfastness; purity           |
| Duruo(杜若)         | Plant       | “Among mountain dwellers who cherish Du-ruo, they drink from stone springs and adorn themselves with pine and cypress.” | pure affection; moral resonance |

**Table 8.** Sample entries in *Aesop's Fables*

| Entity | Entity Type | Original Sentence   | Metaphorical Label          |
|--------|-------------|---|-----------------------------|
| snake  | Animal      | “Once bitten by a snake, one fears even a rope for ten years.”              | treachery; ingratitude      |
| fox    | Animal      | “The fox praised the crow's song”   | cunning; flattery           |
| horse  | Animal      | “Though the stag was driven away, the horse found itself enslaved forever.” | obedience; shortsightedness |

### 3.3.3 Quantitative Evaluation and Error Analysis.

Although automated processes have significantly improved the efficiency of data collection, the reliability of the generated datasets still requires rigorous quantitative verification. Therefore, this section will conduct a comprehensive performance evaluation and error analysis.

To rigorously assess the reliability of the proposed framework, we constructed a Golden Test Set consisting of 300 randomly sampled instances (150 from *Chu Ci* and 150 from *Aesop's Fables*), manually annotated by three experts. We evaluated the performance using Precision (P), Recall (R), and F1-score across three core tasks. The evaluation results are shown in Table 9.

**Table 9.** Performance Metrics for Metaphor Mining Framework

| Task                    | Precision(%) | Recall(%) | F1-Score(%) |
|-------------------------|--------------|-----------|-------------|
| Entity Extraction       | 92.4         | 88.5      | 90.4        |
| Metaphor Classification | 86.7         | 82.3      | 84.4        |
| Semantic Labeling       | 79.1         | 75.4      | 77.2        |

A qualitative review of the model's errors revealed three primary challenges:

**Contextual Ambiguity:** In *Chu Ci*, certain botanical entities like "Chrysanthemum" (菊) oscillate between literal seasonal descriptions and symbolic moral metaphors. The LLM occasionally over-attributes metaphorical intent in purely descriptive contexts.

**Ellipsis and Syntactic Compression:** Classical Chinese frequently employs ellipsis. When the grammatical subject or the tenor of the metaphor is omitted, the LLM may fail to establish the necessary semantic link, leading to lower recall.

**Archaisms and Semantic Drift:** Rare terms for regional flora or fauna sometimes lead to "hallucinations" in label generation, where the model applies modern connotations to ancient specialized terminology.

### 3.3.4 Robustness and Generalizability Analysis.

To mitigate concerns regarding model-specific biases and the sensitivity of prompt engineering, we conducted two types of robustness checks: cross-model validation and prompt variation analysis.

We executed the same automated pipeline using Claude 3.5 Sonnet as an alternative backbone LLM. A comparative analysis of the resulting high-frequency metaphor lists (Top 10) revealed an overlap of over 85% between ChatGPT-4o and Claude 3.5. Specifically, both models consistently identified "Orchid" (兰), "Cicada" (蝉), and "Cinnamon" (桂) as the most prominent metaphorical entities in *Chu Ci*, and "Fox", "Wolf", and "Lion" in *Aesop's Fables*. This high degree of convergence suggests that the extracted metaphorical patterns are inherent to the texts rather than artifacts of a specific model's pre-training data.

We further investigated the framework's sensitivity to prompt phrasing by modifying the System Role. When the role was downgraded from a "Senior Philologist" to a "General Reader," we observed that while the granularity of the generated semantic labels became more generalized (e.g., "Noble political ambition" was simplified to "Good character"), the overall statistical distribution of metaphorical categories and their valence (positive/negative) remained remarkably stable. This consistency underscores the reliability of our framework across different instructional contexts, ensuring that the core findings are robust to minor variations in prompt design.

## 4 Comparative Analysis of Animal and Plant Metaphors

This section conducts a comparative analysis of animal and plant metaphors in *Chu Ci* and *Aesop's Fables* from three perspectives. The analysis focuses on distinguishing metaphor from symbol and incorporates frequency statistics of entity occurrences in

order to provide a more fine-grained examination of similarities and differences between the two texts. To validate the 2,172 semantic labels, we compared the LLM-generated outputs with traditional philological sources, including Zhu Xi's *Chu Ci Ji Zhu* and modern critical studies (e.g., Wen [24]). For high-frequency motifs—such as the "Fox" as an emblem of "Cunning" in Aesop's Fables or "Orchid" as the "Superior Man" in *Chu Ci*—the computational labels exhibited over 85% agreement with historical exegesis. For "hard cases" involving ambiguous contexts where flora or fauna might serve purely descriptive functions, we consulted domain experts (Classicists). These experts adjudicated difficult instances by analyzing intertextual patterns and historical linguistic nuances, thereby refining the model's outputs. This synergy between computational breadth and philological depth ensures that the findings are grounded in rigorous academic tradition, mitigating the risks of "interpretive drift."

## 4.1 Comparative Analysis

### 4.1.1 Same Entity, Different Metaphorical Meanings.

#### Dragon

In *Chu Ci*, the dragon is a symbolic creature imbued with mystery and sacredness and is regarded as an auspicious being. In *Li Sao*, Qu Yuan repeatedly depicts scenes of spiritual journeys such as "riding dragons and driving phoenixes." In these contexts, the dragon is not merely an imagined vehicle for transcendent travel, but also a guide symbolizing the lofty aspirations of the gentleman and the pursuit of a sacred realm. It represents a revered totem and spiritual emblem in Chinese culture, embodying authority, nobility, wisdom, and power. Closely associated with natural forces such as clouds, rain, and water sources, the Chinese dragon is perceived as an approachable and governable form of divine natural power capable of harmonizing the cosmos and bringing auspicious rain.

In contrast, in Aesop's Fables and in broader Western cultural traditions, the dragon is closely associated with evil and deception. In Ancient Greek mythology, dragons frequently appear as ferocious monsters and guardians of treasures, serving as adversaries that heroes must conquer, such as the dragon encountered by Heracles when stealing the golden apples. In Judeo-Christian tradition, the negative image of the dragon is further reinforced: the Old Testament treats it as a symbol of evil, while in the Book of Revelation, the "great red dragon" is explicitly identified as the "ancient serpent" and the embodiment of Satan, representing absolute evil.

These opposing metaphorical functions reflect fundamentally different philosophical logics underlying the two civilizations. The image of the Chinese dragon originates from reverence for and integration with natural forces, as well as from political constructions of divinely sanctioned kingship. As a totem formed through cultural synthesis, it embodies the idea of harmony between humans and nature. Consequently, the dragon can serve as a symbol of the gentleman and the ruler, as spiritual ideals and political authority are conceived as unified. By contrast, the Western dragon is deeply rooted in religious mythology as the absolute opposite of good, representing the embodiment of evil. This reflects a rationalist tendency in Western culture to scrutinize

and discipline nature—particularly its wild aspects—viewing it as something to be controlled and conquered by human reason or divine power. Thus, the metaphorical function of the dragon is not merely rhetorical but serves as a condensed expression of divergent cultural worldviews: one emphasizing harmonious coexistence, and the other emphasizing moral opposition.

#### Snake

From a frequency perspective, the snake appears relatively infrequently in *Chu Ci* but occurs with much higher frequency in Aesop's Fables. In *Tian Wen* from *Chu Ci*, the expression "a snake swallowing an elephant" can, when interpreted within its hermeneutic context, be associated with greed. However, due to residual snake-totem worship in the Chu cultural sphere, the metaphorical meaning of the snake remains relatively implicit and polyvalent. Its referential function in the text is closer to a culturally marked symbol shaped by regional tradition rather than a single, clearly defined moral judgment.

By contrast, in Aesop's Fables, the metaphorical function of the snake is highly explicit and stable. In stories such as *The Farmer and the Snake*, the snake is repeatedly used to symbolize treachery and ingratitude. Here, the metaphor is constructed primarily to convey a clear and direct moral lesson. This fixed negative association can be traced back to the biblical image of the serpent in the Garden of Eden, where it tempts humanity into transgression. As a result of these differing cultural lineages, the snake metaphor in *Chu Ci* functions as a multi-layered cultural symbol, whereas in Aesop's Fables it operates more as a single-purpose metaphorical "tool" designed to warn against misplaced kindness toward the ungrateful.

From the perspective of Conceptual Metaphor Theory, this divergence stems from differences in cultural cognition. In Chu culture, residual snake-totem traditions contribute to the implicit and plural nature of snake metaphors in *Chu Ci*. In Ancient Greek culture, however, the snake is strongly associated with evil and deception, forming a relatively fixed negative mapping. Consequently, snake metaphors in *Chu Ci* are closer to symbolic conventions embedded in regional culture, whereas those in Aesop's Fables are more contextually constructed to serve specific moral teachings.

#### 4.1.2 Different Entities, Similar Metaphorical Meanings.

Despite cultural differences in entity selection, certain metaphorical expressions reveal shared cognitive tendencies across Eastern and Western civilizations, in which different entities are used to convey similar metaphorical meanings.

##### Orchids and Ants

In *Chu Ci*, plants such as orchids and hui appear with particularly high frequency. Due to their natural attributes of fragrance and purity, they are used to symbolize Qu Yuan himself and the refined moral character of the gentleman. This symbolism represents a long-established cultural convention in which plants metaphorize personality, forming part of the Chinese tradition of moral self-expression through natural imagery.

In Aesop's Fables, ants also occur with considerable frequency. In stories such as *The Ant and the Grasshopper*, ants symbolize diligence and foresight based on their industrious behavior. Although the metaphor-bearing entities differ—plants in one case

and animals in the other—both follow a shared metaphorical mapping from natural attributes to moral qualities.

This commonality arises from shared embodied experience: sensory associations such as fragrance and purity, as well as behavioral observations such as industriousness, constitute cross-cultural cognitive foundations. However, cultural preferences in source-domain selection differ. Chu Ci favors plants as metaphorical vehicles, reflecting an emphasis on inner moral cultivation, whereas Aesop's Fables favors animals, reflecting a focus on observable social behavior.

#### Chrysanthemum and Grapes

In Chu Ci, Qu Yuan's line "feeding on fallen chrysanthemum petals" establishes the chrysanthemum as a symbol of independence and aloof integrity in later literary tradition. Blooming in autumn and standing resilient against cold winds, the chrysanthemum's phenological properties align with the moral ideal of steadfastness. This metaphor functions as an introspective projection, in which the poet fuses personal integrity with the natural attributes of the plant.

In contrast, grapes in Aesop's Fables carry a different form of moral metaphor. In *The Fox and the Grapes*, the fox's dismissal of unattainable grapes as sour gives rise to the well-known metaphor of "sour grapes." Here, the metaphor does not derive from the grape itself but from the psychological behavior exhibited by the fox. This instrumental metaphor is constructed to satirize self-deception and rationalization in social behavior rather than to express inner moral ideals.

#### Ox and Horse

Although the cultural connotations of "ox" and "horse" are not entirely identical in the two ancient texts, in a broader cultural context, they share common metaphorical meanings of "diligence", "strength", and "practicality" carried by each in their respective cultures. In ancient Chinese society, which was primarily agricultural, the "ox" played a crucial role. Its image as a hardworking and tireless animal with boundless strength made it a symbol of diligence and dedication, such as in the phrases "old yellow ox" and "a willing ox". In Western culture, especially in societies characterized by pastoral economy and military affairs, the "horse" assumes a similar role. With its strong, fast, and loyal traits, the horse is endowed with symbolic meanings of diligence, bravery, strength, and speed. These two different animals, having played a central role in the production and life of their respective civilizations, were ultimately endowed with similar metaphorical connotations, namely the praise for instrumental value that is "beneficial to humans". This reflects the cognitive commonality of different cultures, which extract similar moral or social qualities from embodied experiences.

### 4.1.3 Unique Entities and Metaphors.

In "Chu Ci", animal imagery does not merely serve as a symbol of nobility or sanctity; it also carries complex and introspective emotions [24]. For instance, the image of the "ape" appears in "Nine Songs", such as "Thunder rumbles and rain falls darkly, the ape cries and sings again in the night." Here, the "ape's cry" is not a simple natural depiction, but is associated with the poet's inner sorrow and the anguish of not finding a kindred spirit, becoming the origin of the "bleak and sorrowful" artistic conception in Chinese literature. Apes inhabit deep mountains and ancient forests, and their cries are

heard but their forms are rarely seen, fitting the image of Taoist hermits and representing a kind of introverted wisdom.

In contrast, the animals in Aesop's Fables rarely exhibit such introspective and personal tragic undertones. Even in stories like "The Ass and the Horse," the tragic ending of the horse is not intended to evoke a pure sense of sadness, but rather to convey a societal lesson—one should not envy the powerful and wealthy, but should be content with poverty and humility." This difference once again highlights the fundamental difference between the "emotion-driven" nature of Chu Ci and the "instrumentality-driven" nature of Aesop's Fables.

"Chu Ci" features some unique plants native to the Chu region, such as "Jiang Li" and "Bi Zhi". These entities appear frequently in the book, and their metaphors are related to the "witchcraft and sacrifice tradition" and "regional identity" of Chu culture. Qu Yuan used native plants to symbolize loyalty to the Chu state. This symbolic meaning is unique to Chu culture, a fixed symbol formed in a specific cultural context, and belongs to the category of symbolism. "Aesop's Fables" contains animal combinations such as "The Lion and the Bull" and "The Wolf and the Lamb", which also appear frequently. Their metaphors point to "relations of strength and weakness" and "power struggles", related to the class structure and democratic concepts of Ancient Greek city-state society. These metaphors were constructed to reflect the social reality of the time and convey social survival wisdom, and they mostly belong to the category of metaphors.

## 4.2 Cultural Interpretation

### 4.2.1 Differences in Metaphor Formation.

Metaphors in Chu Ci can be characterized as emotion-driven lyrical metaphors. Their primary function is not to convey a predetermined moral lesson, but to externalize the poet's inner emotional world through unconscious projection. Qu Yuan's creative mode is grounded in the principles of *ganwu qixing* (being emotionally stirred by external objects) and the unity of subject and object. Plants are treated as extensions of the poet's ideal personality, resulting in metaphors that are highly poetic and introspective in nature [25]. The literature here functions as a medium for conveying *dao*, yet this *dao* is primarily internal and spiritual rather than didactic. Through the depiction of natural imagery, the text aims to evoke emotional resonance and guide readers into a spiritual realm characterized by emotional identification with the natural world.

By contrast, metaphors in Aesop's Fables are instrument-driven and functional metaphors. Their core purpose lies in transmitting universal moral lessons and practical wisdom through concise animal narratives. Animal behavior functions as a mirror for human conduct, serving to warn, regulate, and discipline social behavior. The motivation underlying metaphor generation in Aesop's Fables is thus fundamentally social rather than emotional.

This distinction is also reflected in the relationship between authors and audiences. Chu Ci represents a form of elite discourse, composed for aristocrats and scholar-officials, and is concerned with political ideals and moral integrity. Aesop's Fables, by

contrast, originate as popular discourse intended for common people and enslaved populations, focusing on survival strategies and pragmatic ethics. Consequently, Ancient Greek culture regarded fables as practical tools for resolving disputes and regulating behavior. Language and narrative structure are therefore subordinated to clear rational objectives.

#### **4.2.2 Reflection of Cultural Values.**

Early Chinese philosophy is characterized by the concept of the unity of heaven and humanity, which challenges strict human-centeredness and emphasizes harmonious co-existence between humans and nature. This worldview privileges wholeness and integration. Against this philosophical backdrop, animal and plant metaphors in *Chu Ci* function not as simple analogies but as unconscious emotional extensions of the subject into external objects. The orchid, for instance, is spiritually homologous with the gentleman: its physical properties—fragrance, resilience, and purity—are directly aligned with moral qualities. This metaphorical logic foregrounds the inner spiritual world, with its esthetic core lying in the elevation and transcendence of the self through imagery.

Ancient Greek culture, by contrast, places strong emphasis on rationality [26]. Knowledge of the world is achieved through reason, which is tasked with understanding and mastering external reality rather than merging with it [27]. Under this intellectual framework, metaphors in Aesop's Fables function as clear substitutions whose meanings depend on identifiable similarities that can be rationally abstracted. The cunning of the fox and the strength of the lion are traits that can be analyzed, generalized, and applied. As a result, fables often conclude with explicit morals or lessons. Their primary aim is to simplify complex social behavior into accessible narrative models that facilitate moral instruction and social regulation, thereby reinforcing ethical norms and social order.

## **5 Conclusion**

This paper takes *Chu Ci* and Aesop's Fables as the research objects, focusing on the automatic identification and comparative analysis of animal and plant metaphors in classical texts. It conducts statistical analysis and cross-text comparative research using automatic mining methods for animal and plant metaphors.

At the methodological level, this paper breaks down the task of identifying animal and plant metaphors into stages such as text preprocessing, extraction of animal and plant entities, generation of metaphor candidates, and manual proofreading, achieving systematic mining of metaphor candidates. The accuracy and consistency of the results are ensured through the manual proofreading process. At the data level, this paper constructs a structured dataset of animal and plant metaphors and compiles a list of animal and plant metaphorical entities. Based on this dataset, statistical analysis is conducted on the number of metaphor instances, entity distribution, and semantic categories, quantitatively depicting the overall usage characteristics of animal and plant metaphors in "*Chu Ci*" and "Aesop's Fables". This provides data support for subsequent comparative studies on Chinese and Western metaphors.

Although this paper explores method design and data construction, it still has certain limitations. Firstly, this paper mainly relies on general large language models to generate metaphor candidates, and has not systematically compared the effects of different models or prompt strategies. Future research can introduce multiple models or prompt schemes while maintaining the overall process, to further evaluate the stability of automatic mining results. Secondly, the research objects of this paper are limited to "Chu Ci" and "Aesop's Fables", and the corpus size is relatively limited. Furthermore, while LLMs demonstrate high efficiency, they may harbor latent biases derived from their predominantly modern pre-training corpora. To mitigate the risk of 'interpretive drift,' our framework emphasizes the 'Human-in-the-loop' protocol, where domain experts adjudicate complex cases. Future work will explore the use of specialized models fine-tuned on classical linguistic data to further enhance semantic precision.

## References

1. Wen Xiaoyan. 2020. A Study on the English Translation of Animal and Plant Imagery in Du Fu's Poetry. Ph. D. Dissertation. Guangdong University of Finance & Economics.
2. Li Yizhi and Huang Xiaolin. 2020. A Comparative Analysis of Cultural Imagery in English and Chinese Plant Metaphors from the Perspective of Embodied Cognition. *Journal of Guilin University of Aerospace Technology* 25, 2 (2020), 255–260.
3. Wu Mengcheng. 2022. Automatic Recognition of Plant Named Entities in Pre-Qin Classics and Knowledge Graph Construction. Ph. D. Dissertation. Nanjing Agricultural University.
4. Tang Chao, Chen Bo, Tan Zelin, et al. 2025. Semi-Supervised Ancient Classic Entity Extraction Based on Knowledge Distillation. *Data Analysis and Knowledge Discovery* 9, 7 (2025), 118–129. doi:10.11925/infotech.2096-3467.2024.0722
5. Peng Yao. 2025. Research on Named Entity Recognition for Classical Chinese Historical Texts. Ph. D. Dissertation. Guangdong University of Foreign Studies.
6. Song Bingyan. 2022. Research on Entity Recognition and Herbal Recommendation Algorithms for Traditional Chinese Medicine Classics. Ph. D. Dissertation. Beijing University of Technology.
7. Zhou Chanjuan. 2024. Comparison of Grass and Wood Imagery in The Classic of Poetry and Songs of the South. *Hansi Culture* 18, (2024), 22–25.
8. Sun Yiwen. 2020. Research on Representative Floral Imagery in The Classic of Poetry and Songs of the South. Master's thesis. Hubei Normal University. doi:10.27796/d.cnki.gghbsf.2020.000023
9. William Hansen. 2002. The Aesopica: A Corpus of Medieval and Early Modern European Fables. *Journal of Folklore Research* 39, 1 (2002), 1–22.
10. George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago Press, Chicago, IL, USA.
11. Yao Xinyu, Wang Mengdi, Chen Bo, and Zhao Xiaobing. 2025. WenyanGPT: A Large Language Model for Classical Chinese Tasks. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*. 8339–8347. doi:10.24963/ijcai.2025/927
12. Zhang Shanxin. 2020. A Comparative Study on Cultural Mechanisms of Animal Metaphors in English and Chinese. Shanghai International Studies University (2020).
13. Yan Su. 2020. A Comparative Study and Translation of Animal Metaphors in English and Chinese: A Case Study of Sandalwood Death. *Overseas English* 23, (2020), 184–185.

14. Xu Yichen. 2020. A Comparative Study of Plant Metaphor Concepts in English and Chinese. *Popular Literature and Arts* 14, (2020), 178–179.
15. Zhou Yijun. 2022. Narratives of “Animal Transformation” in Chinese and Western Literature. *Jiangxi Social Sciences* 42, 11 (2022), 89–96.
16. Sun Yanbo. 2023. A Cross-cultural Comparative Study of English and Chinese Metaphors. *English on Campus* 6, (2023), 181–183.
17. Huang Huiru. 2020. An Inquiry into the Cultural Cognition of Flower Metaphors in Tang Poetry. *Liberal Arts Fans (Education and Teaching)* 3, (2020), 247–249.
18. Wang Yinping. 2021. Universality and Cultural Differences of Plant Metaphors in English and Chinese. *Journal of Shaanxi Radio and TV University* 23, 01 (2021), 52–56.
19. Du Yue, Wang Dongbo, Jiang Chuan, et al. 2021. Construction and Application of Deep Learning-Based Automatic Entity Recognition Models for Classical Texts in Digital Humanities. *Library and Information Service* 03 (2021).
20. Meng Zhongyuan, Zhang Zhuijiao, and Rong Bo. 2025. Exploring the Construction of Knowledge Graphs for Qin–Han Cultural Classics from a Digital Humanities Perspective. *Collected Papers of Shaanxi History Museum* 00 (2025).
21. Pang Xian, Chen Bo, and Zhao Xiaobing. 2025. Extraction and Analysis of Chinese-Tibetan New Words from News Texts. *Acta Scientiarum Naturalium Universitatis Pekinensis* 61, 1 (2025), 45–52.
22. Zhang Yuanshuo, Li Aohua, Yin Zhaoning, Wang Panyi, Chen Bo, and Zhao Xiaobing. 2024. Research on Stance Detection with Generative Language Models. In *Proceedings of the 23rd Chinese National Conference on Computational Linguistics (Volume 1: Main Conference)*. Chinese Information Processing Society of China, Taiyuan, China, 481–491.
23. Zhang Yuanshuo, Li Aohua, Chen Bo, Sun Jingbo, and Zhao Xiaobing. 2026. MSME: A Multi-Stage Multi-Expert Framework for Zero-Shot Stance Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
24. Wen Yiduo. 1942. *Supplementary Collation of Chu Ci*. Commercial Press, Beijing, China.
25. Zhang Jian. 2022. Plant Descriptions in *Chu Ci* and Their Cultural Origins. *Famous Writers and Masterpieces* 01 (2022), 124–125.
26. Zhang Xiao. 2023. *Animals and Animal Concepts in the Social Life of Ancient Greece*. Wuhan University (2023).
27. Jean-Pierre Vernant. 1990. *Myth and Society in Ancient Greece*. Zone Books, New York, NY, USA.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

