Publication Date: 02 September 2026
Archs Sci. (2026) Volume 75, Issue 6 Pages 101-112, Paper ID 2025612.
https://doi.org/10.68304/as/75612

A Data-Driven Study of English Translation of TCM Terms from the Perspective of Eco-Translatology

Liying Wang[1] and Jinlong Zhu[2]
[1]School of Chinese Language and Communication, Bozhou University, Bozhou, Anhui 236800, China
[2]School of Artificial Intelligence, Bozhou University, Bozhou, Anhui 236800, China
Corresponding authors: Jinlong Zhu (e- mail: mujin33765 @163. com).

Abstract Traditional Chinese Medicine (TCM) terminology embodies not only a rich conceptual system but also complex cultural, philosophical, and clinical attributes. Because many terms are highly condensed, semantically abstract, metaphorically structured, and culturally embedded, translating them into English has long posed substantial challenges for semantic fidelity, scientific accuracy, terminological consistency, and communicative effectiveness. Guided by Eco-translatology, which emphasizes multidimensional adaptation across linguistic, cultural, and communicative dimensions, this study proposes a data-driven and knowledge-enhanced neural approach to TCM terminology translation. Specifically, we construct a syntax-aware convolutional model to identify internal semantic structures, including predicate–object relations and central words, and integrate TCM knowledge-graph embeddings, including term–herb–compound–target–pathway relations, into a TCM-BERT-based translation architecture through a knowledge-attention mechanism. We compile a bilingual corpus containing 3,825 standardized TCM terms and approximately 48,000 parallel sentence pairs, supplemented by pharmacological annotations from established resources such as TCMID, HERB, and TCMSP. Experimental results show that the proposed model outperforms the baseline system, increasing BLEU-4 from 35.4 to 48.1, an improvement of 12.7 points, and increasing the terminology-consistency rating from 2.8/5 to 4.2/5. Qualitative evaluation of representative terms such as "气血两虚证" and "肝阳上亢" further demonstrates the model's ability to preserve syntactic structure while incorporating biomedical semantic information into English interpretations.

Index Terms traditional Chinese medicine, eco-translatology, terminology translation, knowledge embedding, neural machine translation, syntax modeling, network pharmacology

I. Introduction

With the growing international visibility of Traditional Chinese Medicine (TCM), the reliable cross-cultural communication and semantic representation of TCM terminology have become important concerns in translation, terminology standardization, medical communication, and knowledge dissemination [1], [2], [5], [6]. As a basic component of the TCM knowledge system, terminology encodes traditional cultural concepts, diagnostic reasoning, therapeutic functions, and philosophical assumptions. Its translation therefore involves more than direct lexical substitution: it requires coordinated attention to linguistic form, conceptual meaning, cultural accessibility, and communicative purpose. From this perspective, Eco-translatology offers a useful framework for examining how English renderings can adapt to the target-language environment while maintaining the distinctive conceptual content of the source terminology [3], [4].

Eco-translatology, developed by Gengshen Hu, conceptualizes translation as a process of adaptive selection within a translational ecosystem. Its central principles include selective adaptation, adaptive selection, and multidimensional transformation across interacting linguistic, cultural, and communicative dimensions [3]. For the English translation of TCM terminology, this framework is particularly relevant because a technically acceptable equivalent must often preserve a specialized conceptual relation, remain intelligible to readers outside the source culture, and function appropriately in a medical or academic communicative setting. Accordingly, the three-dimensional perspective provides a structured basis for evaluating not only lexical correspondence but also cultural adequacy and communicative effectiveness.

In practice, many TCM terms follow a characteristic pattern of concise designation, metaphorical imagery, and condensed functional meaning, as illustrated by expressions such as "肾不纳气," "肝阳上亢," and "气血两虚." Such terms create recurrent translation problems, including semantic omission, conceptual over-simplification, functional mis-

representation, and inconsistency across texts [5]–[9]. Human translation may achieve strong cultural interpretation but can vary among translators and institutional traditions, whereas a general-purpose machine translation system may produce fluent wording without reliably preserving specialized TCM relations. Consequently, both terminology standardization and context-sensitive semantic modeling are necessary if translated terms are to remain stable, interpretable, and usable in international medical and scholarly communication.

Previous work has addressed TCM terminology from several complementary directions. The WHO International Standard Terminologies on Traditional Medicine in the Western Pacific Region established an important international reference for standardized traditional-medicine terminology [1], [2], while specialized Chinese medical dictionaries and language resources have provided detailed English equivalents, definitions, and grammatical guidance for technical expressions [6], [7]. Translation-oriented studies have further examined cultural constraints, variation in English presentation, and the interpretation of classical Chinese medical concepts [5], [8], [9]. In parallel, developments in neural machine translation, domain adaptation, and terminology-constrained decoding have made domain-specific machine translation increasingly feasible [11]–[14]. Nevertheless, generalized NMT systems do not automatically encode the ontology of TCM concepts or their biomedical associations. Network-pharmacology resources such as TCMID, TCMSP, HERB, ETCM, BATMAN-TCM, TCM-Mesh, and SymMap provide structured herb–ingredient–target–pathway–disease relations that can support knowledge-enhanced semantic modeling [19]–[26], but these resources must be integrated carefully so that pharmacological annotation supplements rather than replaces the linguistic meaning of a TCM term.

Despite these advances, three methodological gaps remain. First, many translation models do not explicitly represent the internal syntactic and semantic organization of structurally complex TCM terms, making nested relations, modifier–head constructions, and predicate–argument patterns difficult to preserve consistently. Second, knowledge integration is often incomplete: terminology information, TCM conceptual relations, and modern pharmacological associations are frequently stored in separate resources rather than represented through a unified mechanism. Third, evaluation is commonly dominated by surface-level automatic metrics, while linguistic fidelity, cultural adequacy, communicative usability, terminology consistency, and biomedical knowledge alignment are seldom assessed together. These limitations motivate a framework that combines structure-aware language modeling with carefully bounded knowledge enhancement and multi-dimensional evaluation.

To address these limitations, this paper proposes an English translation framework for TCM terminology that combines structural modeling and knowledge enhancement within the theoretical perspective of Eco-translatology. The research proceeds in four stages. First, a central-word-aware convolutional model represents predicate–argument, modifier–head, and dependency relations inside TCM expressions, drawing on established neural representation methods for language processing [15]–[17]. Second, structured relations from TCM knowledge resources are used to construct term–herb–ingredient–target–pathway–disease representations [19]–[26]. Third, these representations are integrated with a domain-adapted BERT-based encoder through a knowledge-attention mechanism, allowing linguistic context and structured knowledge to contribute jointly to translation generation. Fourth, model performance is evaluated through BLEU scores [18], terminology-consistency ratings, structure-recognition measures, and biomedical knowledge-alignment analysis. In this way, the study seeks to evaluate translation quality across linguistic accuracy, cultural interpretability, communicative usefulness, and domain-specific semantic grounding rather than relying on a single automatic score.

II. Related Work

Research on the English translation of Traditional Chinese Medicine (TCM) terminology has developed through several interconnected traditions, including terminology standardization, bilingual lexicography, interpretation of classical medical texts, cross-cultural translation studies, and more recently computational language processing. International terminology initiatives have sought to establish shared technical vocabulary [1], [2], while specialized dictionaries and pedagogical resources have clarified the semantic and grammatical behavior of Chinese medical expressions in English [6], [7]. At the same time, studies of classical texts and contemporary TCM communication have shown that translation decisions are affected by conceptual asymmetry, cultural expectations, genre, and regulatory context [5], [8]. This progression has gradually shifted the field from isolated word-for-word equivalence toward a broader concern with standardized terminology, conceptual explanation, communicative function, and reproducible translation practice.

Current research on TCM cultural terminology is consequently diverse in both materials and methods. Standard terminology resources offer relatively stable reference equivalents for many technical items [1], [2], whereas bilingual dictionaries provide finer distinctions among closely related concepts and recurrent grammatical patterns [6], [7]. Studies of TCM texts in international settings further demonstrate that the same source concept may be rendered differently when the target text serves clinical, educational, commercial, regulatory, or explanatory purposes [5], [8]. These strands of work are complementary: standardization promotes consistency, while contextual translation research explains why a formally standardized term may still require definition, annotation, transliteration, or controlled elaboration for a particular readership.

Several concrete resources illustrate this development. The WHO terminology standard provides a broad reference vocabulary intended to facilitate a common technical language for traditional medicine [1], [2]. Wiseman and

Feng's Chinese medical reference works provide detailed English terminology and language guidance that are useful for distinguishing technical meanings from ordinary-language senses [6], [7]. Cheng and Glynn show that the English presentation of plant-based TCM products can vary under legal and communicative constraints [5], while Neal's discussion of studying the Huang Di Nei Jing in the West illustrates the interpretive difficulties that arise when classical concepts move between Chinese medical traditions and English-speaking scholarly environments [8]. Together, these works emphasize that successful translation depends on terminological discipline as well as sensitivity to cultural, textual, and communicative context.

Eco-translatology provides an additional theoretical perspective for organizing these translation choices. Hu's framework emphasizes selective adaptation and adaptive selection within a multidimensional translational environment [3]; related applications demonstrate how linguistic, cultural, and communicative considerations can be examined together rather than as isolated criteria [4]. Broader research on culture-loaded expressions likewise shows that culturally specific meaning may require strategies beyond direct lexical replacement [9], [10]. Digital and corpus-assisted translation research has also expanded the empirical basis for comparing lexical frequencies, collocations, contextual preferences, and recurring translation patterns [27]. These developments create a foundation for combining theory-guided translation evaluation with quantitative modeling and domain-specific data.

Taken together, the literature provides substantial conceptual and terminological foundations, yet comparatively fewer studies combine Eco-translatological evaluation with structure-aware neural modeling and explicit TCM knowledge representations. The present study addresses that intersection by using corpus-based analysis, neural representation learning, and structured TCM knowledge to investigate how culturally embedded terms can be rendered into English while balancing semantic fidelity, readability, terminological consistency, and intercultural intelligibility. Pharmacological information is treated as supplementary semantic grounding rather than as a direct substitute for the traditional medical concept, thereby preserving a distinction between translation and biomedical interpretation.

III. Machine Translation and Neural Networks in the Translation of TCM Cultural Terminology

In recent years, the rapid development of machine translation (MT), especially neural machine translation (NMT), has created new possibilities for the English translation of Traditional Chinese Medicine (TCM) cultural terminology [11]–[13]. Unlike general-domain translation, TCM translation contains culturally embedded, semantically dense, and highly domain-specific expressions, such as "调和阴阳" (regulate Yin and Yang) and "气滞血瘀" (Qi stagnation and blood stasis). These expressions often lack direct one-to-one English equivalents, and their interpretation depends on conceptual relations that may not be recoverable from surface co-occurrence alone. The combination of high cultural load, specialized clinical usage, and relatively limited bilingual data therefore creates a demanding setting for both statistical and neural translation methods.

A. TCM Term Translation in Machine Translation Models

Earlier statistical machine translation (SMT) systems commonly formulated translation as probabilistic decoding: given a source sequence S, the system searches for a target sequence T that maximizes an estimated translation probability, often together with a target-language model [28]. Such approaches depend strongly on observed bilingual correspondences and therefore become vulnerable when domain-specific expressions are sparse or semantically specialized. The present study moves beyond phrase-level probabilistic matching by using neural feature representations; Figure 1 introduces the CNN-based terminology representation component used to capture local semantic patterns in the proposed framework. In the TCM domain, conventional SMT approaches face several related limitations: domain-specific parallel corpora are comparatively limited; polysemous forms can have different meanings in TCM and ordinary biomedical or everyday language; and culture-loaded expressions may require conceptual interpretation rather than literal phrase substitution. For example, an anatomical reading of "肝" does not exhaust its functional meaning within TCM theory. These characteristics make domain adaptation and controlled terminology especially important for specialized translation [5]–[7], [13].

To address these limitations, contemporary work has increasingly adopted neural network-based approaches, particularly attention-based Neural Machine Translation (NMT) and Transformer architectures, which can model contextual dependencies more flexibly than earlier phrase-based systems [11], [12].

B. Neural Machine Translation and Domain Adaptation for TCM Terms

NMT systems can substantially improve translation fluency and contextual coherence when appropriate parallel data are available, and Transformer-based architectures provide strong mechanisms for modeling long-range contextual relations [12]. Nevertheless, general-domain NMT remains vulnerable in a specialized TCM setting because domain-specific expressions may be rare, ambiguous, or absent from the training distribution [13]. Three recurrent error types are especially relevant. First, term omission may occur when a culturally important expression, such as "望闻问切" (the four diagnostic methods), has weak representation in the training data. Second, mistranslation may reduce a specialized expression to an ordinary-language paraphrase. Third, overgeneralization may produce fluent English while removing clinically or culturally important distinctions.

For instance, "火旺" may be rendered literally as "fire is strong," even though the TCM expression functions within

a specialized diagnostic and theoretical system; similarly, syndrome or treatment terms can be paraphrased so broadly that their established technical relation is lost. Such errors illustrate why fluent output cannot be treated as sufficient evidence of terminological accuracy. Domain adaptation, terminology constraints, and explicit term modeling are therefore important complements to general NMT architectures [13], [14].

To reduce these errors, recent NMT research supports several domain-adaptation strategies that are directly relevant to TCM terminology translation [13], [14]:

- *Fine-tuning with domain-specific corpora:* adapting NMT models with parallel material from standardized terminology resources, bilingual textbooks, and domain-specific medical texts can improve the representation and consistency of specialized terms.
- *Incorporating terminology constraints:* integrating approved glossaries or target terms into model inputs or decoding procedures can increase the probability that specialized terminology is preserved rather than replaced by a generic paraphrase [14].
- **Combining translation with auxiliary term tasks:** multi-task objectives such as term classification or semantic-role prediction can strengthen the model's awareness of categories such as syndromes, organs, herbs, prescriptions, and therapeutic actions.

C. Machine-Human Collaboration for TCM Terminology Translation

Because automatic systems remain imperfect in specialized medical communication, human review remains important, particularly for high-stakes documents such as clinical records, pharmacopoeial material, labeling, and standardized terminology resources. A practical workflow can therefore combine neural generation with terminology constraints, structured knowledge, and expert post-editing. In this study, the computational model is designed to improve semantic consistency and provide interpretable knowledge cues; it is not intended to eliminate the need for expert validation when a translation is used in clinical or regulatory settings.

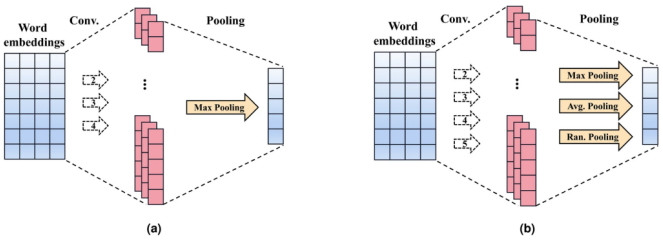

Fig. 1 CNN-Based Representation of TCM Terminology: Comparison of Single-Pooling and Multi-Pooling Mechanisms

Figure 1 illustrates two CNN-based term-modeling architectures used to motivate the structure-aware representation component. Figure 1(a) shows the basic structure, in which convolutional kernels of several sizes slide across a word-embedding matrix and max pooling retains the most salient local features. The input sequence is converted into an embedding matrix $\mathbf{E} \in \mathbb{R}^{n \times d}$, where n denotes the number of input tokens and d denotes the dimensionality of each embedding. This design follows the general principle that convolutional filters can identify informative local n-gram-like patterns in neural text representations [16].

Using convolutional kernels $\mathbf{W}_i$ with different window sizes, such as 2, 3, and 4, feature maps are generated by sliding the kernels over $\mathbf{E}$. The local feature computation is written as follows:

$$\mathbf{f}_i = \mathrm{ReLU}\big(\mathbf{W}_i * \mathbf{E}_{[j:j+h-1]} + b_i\big)\,, \tag{1}$$

where $*$ denotes convolution and ReLU denotes the rectified linear activation function. Max pooling is then applied across the feature maps so that variable-length local responses are compressed into a fixed-length vector representation:

$$\mathbf{z} = \max(\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k)\,. \tag{2}$$

This architecture is suitable for representing terms such as "气血失调" and "肝火亢盛" because local convolutional filters can detect informative character- or token-level combinations while pooling emphasizes the strongest responses. In the proposed framework, these features are not interpreted as complete translations by themselves; rather, they provide compact structural evidence that is subsequently combined with contextual and knowledge-based information.

In contrast, the enhanced architecture shown in Figure 1(b) uses multiple pooling operations, including max pooling, average pooling, and random pooling, to summarize different aspects of the convolutional feature maps. The resulting representations are expressed as:

$$\mathbf{z}_{\max} = \max(\mathbf{f}_i)\,, \qquad \mathbf{z}_{\mathrm{avg}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{f}_i, \qquad \mathbf{z}_{\mathrm{ran}} \sim \mathrm{U}(\mathbf{f}_i)\,. \tag{3}$$

By retaining more than one summary of the convolutional responses, the multi-pooling structure can preserve complementary evidence for semantically variable or polysemous terms. For culturally loaded and metaphor-rich expressions such as "阴虚火旺" and "风湿痹痛," these representations provide a broader contextual feature basis for subsequent translation. The additional pooling operations should therefore be understood as a representational mechanism: their contribution to translation quality is evaluated empirically rather than assumed from the architecture alone.

From the perspective of Eco-translatology, the architecture can be interpreted as supporting adaptive representation across linguistic, cultural, and communicative dimensions. The convolutional pathway emphasizes local linguistic combinations, while diversified pooling preserves multiple feature summaries that may be useful when a term has more than

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

one salient contextual cue. The fixed-dimensional representation also makes it possible to integrate these features with knowledge-based signals and pass them to downstream translation or classification components. This computational interpretation does not equate neural features with cultural meaning directly; instead, it provides a measurable mechanism through which culturally relevant distinctions can be retained and evaluated.

For English translation of TCM terminology, this structure therefore functions as a feature-extraction stage rather than as an independent terminological authority. Its role is to strengthen the representation of internal term structure before the representation is combined with contextual encodings and structured TCM knowledge. The overall objective is to improve consistency, semantic fidelity, and communicative usefulness while retaining expert-verifiable links between model output and the domain information used during translation.

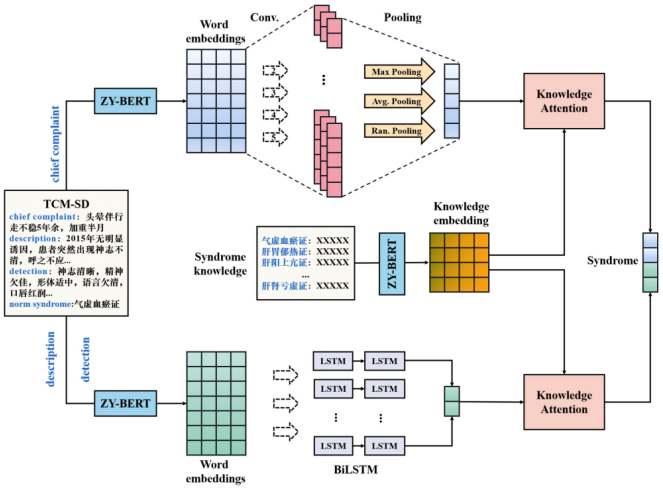

Fig. 2 Neural Architecture for English Translation of TCM Terminology with Knowledge Embedding and Semantic Modeling

To improve semantic modeling and translation accuracy for TCM cultural terminology, this study introduces the neural architecture shown in Figure 2. The model contains two coordinated representation pathways. In the linguistic pathway, TCM text, including chief-complaint, descriptive, and diagnostic context, is encoded by the pretrained ZY-BERT model to obtain a contextual embedding matrix $\mathbf{E} \in \mathbb{R}^{n \times d}$, where n is the number of input tokens and d is the embedding dimension. In the knowledge pathway, structured syndrome-related TCM information is encoded into a knowledge matrix $\mathbf{K} \in \mathbb{R}^{m \times d}$, where m is the number of retrieved knowledge items. The use of a contextual BERT-style encoder follows the general bidirectional pretraining paradigm introduced for contextual language representation [15]. Within the linguistic pathway, the embedding matrix is passed through multi-scale convolutional filters to extract local structural features, with the convolution operation defined as

$$\mathbf{c}_i^{(h)} = \mathrm{ReLU}\left(\mathbf{W}^{(h)} * \mathbf{E}_{i:i+h-1} + \mathbf{b}^{(h)}\right), \quad h \in \{2,3,4,5\}, \tag{4}$$

where $\mathbf{W}^{(h)}$ is the convolution kernel associated with window size h, $*$ denotes convolution, $\mathbf{b}^{(h)}$ is the bias vector, and ReLU is the nonlinear activation function. The resulting features are summarized through max, average, and random pooling so that different aspects of the local response distribution are retained. Their outputs are defined as

$$\begin{aligned} \mathbf{z}_{\max} &= \max\left(\mathbf{c}_1^{(h)}, \ldots, \mathbf{c}_n^{(h)}\right), \\ \mathbf{z}_{\mathrm{avg}} &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{c}_i^{(h)}, \\ \mathbf{z}_{\mathrm{ran}} &\sim \mathrm{U}\left(\mathbf{c}_i^{(h)}\right). \end{aligned} \tag{5}$$

The three pooling outputs are concatenated to form the global linguistic representation $\mathbf{Z} \in \mathbb{R}^{3d}$. In the knowledge pathway, each retrieved syndrome-related knowledge item is represented by the same contextual encoding space and contributes a vector $\mathbf{k}_i$. The knowledge vectors are collected as $\mathbf{K} = [\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_m]$ and fused with the linguistic representation through attention. Specifically, the model computes a relevance score between $\mathbf{Z}$ and each knowledge vector and normalizes the scores to obtain attention weights:

$$\alpha_i = \frac{\exp(\mathbf{Z}'\mathbf{W}_a\mathbf{k}_i)}{\sum_{j=1}^{m}\exp(\mathbf{Z}'\mathbf{W}_a\mathbf{k}_j)}, \qquad \mathbf{W}_a \in \mathbb{R}^{3d\times d}. \tag{6}$$

The attention-weighted knowledge representation is then computed as

$$\mathbf{Z}_{\mathrm{final}} = \sum_{i=1}^{m} \alpha_i \cdot \mathbf{k}_i. \tag{7}$$

The fused representation is passed to the output layer, where a softmax function produces probabilities over the corresponding term labels or English translation concepts:

$$\widehat{y} = \mathrm{softmax}(\mathbf{W}_o \cdot \mathbf{Z}_{\mathrm{final}} + \mathbf{b}_o). \tag{8}$$

Here, $\mathbf{W}_o$ and $\mathbf{b}_o$ are trainable output parameters. The linguistic pathway captures local contextual and structural features, while the knowledge pathway contributes retrieved domain information that can guide semantic alignment. Their combination is intended to support multidimensional adaptation by improving the correspondence between source-term structure, domain knowledge, and target-language expression. The architecture remains a statistical model, however, so the relevance of the incorporated knowledge must be validated rather than assumed.

When processing complex expressions such as “气虚血瘀,” “肝阳上亢,” and “肾精亏损,” the model combines contextual term representations with structured TCM knowledge templates during matching and prediction. This mechanism is designed to reduce instability caused by sparse or ambiguous

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

surface forms and to improve the consistency of target-language output. The resulting translation is subsequently evaluated against reference translations and terminology-consistency criteria, while biomedical annotations are used only as supplementary semantic grounding.

D. Network Pharmacology-Based Biological Modeling and Semantic Grounding of TCM Terminology

Figure 3 summarizes the development of major Traditional Chinese Medicine databases in Figure 3(A) and a representative network-pharmacology workflow in Figure 3(B). Resources such as TCMID, TCMSP, BATMAN-TCM, HERB, ETCM, TCM-Mesh, and SymMap connect traditional medicinal entities with ingredients, targets, symptoms, diseases, or pathways and therefore provide complementary structured information for computational research [19]–[26]. In this study, the value of these resources lies in their capacity to provide machine-readable relational context that can be linked to terminology while remaining distinguishable from the linguistic definition of the term itself.

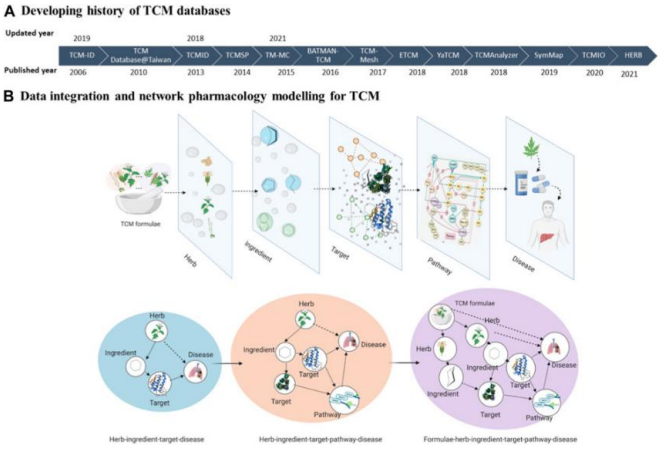

Fig 3 Development of TCM Databases and a Network-Pharmacology Modeling Workflow

In Figure 3(A), the database-development trajectory illustrates how TCM informatics resources have expanded from herb and formula descriptions toward richer cross-links among ingredients, molecular targets, symptoms, diseases, and biological pathways. TCMID and its later version integrate multiple categories of TCM information [19], [20]; TCMSP provides systems-pharmacology data for herbal medicines [21]; HERB integrates experimental and reference-guided evidence [22]; ETCM organizes information across herbs, formulas, ingredients, targets, and diseases [23]; BATMAN-TCM and TCM-Mesh support network-based mechanism analysis [24], [25]; and SymMap links TCM concepts with symptom information [26]. Together, these resources provide structured associations that can supplement a translation model with domain-specific context.

Figure 3(B) presents a typical network-pharmacology workflow as a flowchart and relational network. The process begins with a TCM formula or individual herb, after which candidate ingredients are collected from literature or curated databases. Potential targets are then associated with those ingredients through experimentally reported or computationally inferred relations, and downstream pathway information is used to organize possible biological mechanisms. Examples of frequently studied signaling systems include PI3K–Akt, MAPK, and TGF-β pathways. In the present translation framework, these biological relations are used as auxiliary annotations for semantic grounding rather than as literal English definitions of the originating TCM concepts.

The lower portion of Figure 3(B) distinguishes three increasingly detailed relational network forms:

- the left model represents a four-level Herb $\rightarrow$ Ingredient $\rightarrow$ Target $\rightarrow$ Disease relationship;
- the middle model adds a Pathway level to clarify possible mechanism-related associations;
- the right model represents a six-level Formula $\rightarrow$ Herb $\rightarrow$ Ingredient $\rightarrow$ Target $\rightarrow$ Pathway $\rightarrow$ Disease network, offering a richer representation of the multi-component and multi-target character frequently investigated in systems-level TCM research.

From a systems-biology perspective, this workflow emphasizes relations among medicinal materials, molecular targets, biological pathways, and disease annotations. Network-topological measures, such as degree and betweenness centrality, can be used to identify highly connected entities, while enrichment analysis can organize target sets around functional pathways. For the translation task, these relations provide supplementary evidence that may help disambiguate functional terms such as “补气” (tonifying Qi), “清热” (clearing heat), and “化瘀” (resolving blood stasis). Importantly, the molecular network does not redefine these TCM expressions; rather, it provides an additional layer of domain context that can be evaluated alongside established terminological equivalents.

In summary, Figure 3 connects the evolution of TCM databases with a network-pharmacology workflow that can support structured semantic annotation. The framework helps bridge traditional formula-level descriptions and modern molecular data without treating the two knowledge systems as interchangeable. For example, a TCM syndrome may be associated with reported endocrine, metabolic, or inflammatory patterns, but such a biomedical association should be stored as an annotation rather than substituted directly for the syndrome’s standardized translation. This distinction is essential for cross-cultural semantic reconstruction because it preserves the traditional term while allowing modern biological evidence to enrich model context.

E. Central-Word Structure Modeling for TCM Terminology

Figure 4 presents two contextual neural structures designed for TCM expressions organized around verb-centered patterns, shown in Figure 4(a), and non-verb-centered patterns, shown in Figure 4(b). Both variants combine lexical embed-

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

dings, part-of-speech (POS) information, and structural features so that the model can represent predicate–argument and modifier–head relations more explicitly than a sequence-only baseline. This design is particularly relevant for syntactically condensed expressions such as “肝气郁结” and “五脏之精,” where the relative roles of the central unit and its modifiers influence an appropriate English reconstruction.

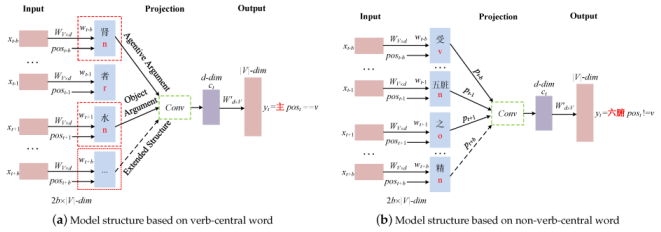

Fig. 4 Structure-Aware Modeling Framework Based on Verb- and Non-Verb-Centered Terms

In Figure 4(a), the model handles expressions treated by the structural annotation scheme as verb-centered patterns. The input is a contextual sequence of word embeddings $\mathbf{x}_{t-b:t+b}$ around a candidate central position t, supplemented by POS vectors $\mathrm{pos}_{t\pm k}$. Word and POS features are projected through $\mathbf{W}_{v\times d}$ and $\mathbf{W}_{\mathrm{pos}\times d}$ and combined into the following contextual representation:

$$\mathbf{h}_i = \mathbf{W}_v \cdot \mathbf{x}_i + \mathbf{W}_{\mathrm{pos}} \cdot \mathrm{pos}_i \,. \tag{9}$$

The contextual representations corresponding to the verb-centered subject and object regions are then passed through a convolutional layer, allowing the model to extract local features associated with predicate–argument relations:

$$\mathbf{c}_t = \mathrm{Conv}([\mathbf{h}_{t-b}, \ldots, \mathbf{h}_{t+b}]) \,, \tag{10}$$

where Conv denotes a one-dimensional convolution used to extract local semantic-relation features. The resulting representation is mapped through a fully connected output layer $\mathbf{W}_{d\times|V|}$ to the target structural category or translation-label space:

$$\widehat{y}_t = \mathrm{softmax}\big(\mathbf{W}_{d\times|V|} \cdot \mathbf{c}_t\big) \,. \tag{11}$$

When a token is identified by the annotation scheme as the principal structural unit, the model assigns the corresponding main-role label while also conditioning the prediction on the token’s POS information. This formulation allows the classification decision to depend jointly on lexical identity, local context, and syntactic category rather than on the surface token alone.

Figure 4(b) addresses structures whose annotated center is a non-verb. For an expression represented through a dependency path, the input again consists of contextual embeddings, but feature extraction additionally uses composite representations of neighboring dependency positions, such as p_{t-1}, p_t, and p_{t+1}. The corresponding convolutional representation is

$$\mathbf{c}_t = \mathrm{Conv}([\mathbf{P}_{t-1}, \mathbf{P}_t, \mathbf{P}_{t+1}]) \,. \tag{12}$$

where $\mathbf{P}_i$ denotes a dependency-path representation formed by combining positional information, lexical embeddings, and the associated syntactic features. As in the verb-centered model, the resulting feature vector is mapped to a word-prediction or structural semantic-labeling space:

$$\widehat{y}_t = \mathrm{softmax}\big(\mathbf{W}_{d\times|V|} \cdot \mathbf{c}_t\big) \,. \tag{13}$$

In summary, the two structure-aware variants in Figure 4 are designed to resolve hierarchical relations inside compact TCM expressions by explicitly modeling the central unit and its local structural context. Convolutional text representations provide an efficient mechanism for extracting local patterns [16], while bidirectional recurrent baselines offer a contrasting sequence-oriented representation strategy [17]. In the translation pipeline, predicted structural labels can be mapped to semantic roles, such as organ–function, modifier–head, or cause–effect relations, so that the target expression can preserve internal relationships rather than only individual lexical items. This structural layer therefore provides the linguistic foundation on which the later knowledge-enhanced translation stage operates.

IV. Experiment

The experiments were designed to evaluate the proposed system across four complementary levels: terminology-structure recognition, English translation performance, semantic consistency, and supplementary pharmacological or clinical grounding. Rather than using one metric as a proxy for overall quality, the evaluation separates structural recognition from translation generation and then examines whether knowledge enhancement improves consistency without obscuring the linguistic identity of the TCM term. The experimental workflow therefore tests the contribution of central-word modeling, neural translation, knowledge attention, and structured database information within the same overall framework.

A. Experimental Data and Terminology Resources

For this study, a multimodal Chinese–English TCM terminology corpus was assembled from terminology references, bilingual textual materials, structured TCM knowledge resources, and clinical-literature annotations. The resources were organized into the following components:

1) Terminology corpus standardized and reference terms were collected from established TCM terminology resources, including the WHO international terminology standard and specialized Chinese medical dictionaries, together with bilingual textbook material and domain-specific translation examples [1], [2], [6], [7].
2) Knowledge graph structured relations among herbs, formulas, ingredients, targets, symptoms, diseases,

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

and pathways were collected from TCMID, TCMSP, HERB, ETCM, BATMAN-TCM, TCM-Mesh, SymMap, and related curated resources [19]–[26]. These data were normalized before being used as auxiliary knowledge rather than as replacements for terminological definitions.

3) Clinical-literature component study additionally compiled reports of randomized controlled trials involving TCM interventions for hypertension, covering seven intervention categories with study samples exceeding $n = 200$ in the assembled set. Changes in systolic and diastolic blood pressure (SBP/DBP) were retained as the principal quantitative effect indicators for the clinical-grounding analysis. These clinical variables were used only for supplementary validation and were not treated as direct translation targets.

After cleaning, deduplication, and category normalization, the final terminology set contained 3,825 terms distributed across five major categories: pathogenesis, syndrome, organ, treatment, and medicinal substance. Approximately 48,000 Chinese–English parallel sentence pairs were retained for model development and evaluation. The terminology and sentence-level resources were aligned so that translation examples could be linked, when available, to standardized term entries and to the auxiliary structured-knowledge records described above.

B. Experiments in Terminology Structure Modeling

This experiment evaluates two terminology-semantic recognition models: a verb-centered structure model and a non-verb-centered structure model. The objective is to determine whether explicit central-word information improves semantic-role classification compared with a sequence-oriented baseline. For example, in the expression “肝藏血,” the annotation distinguishes “肝” as the organ-related subject, “藏” as the verbal center, and “血” as the object component. Accuracy, precision, recall, and F1 score are used to assess the resulting structural predictions, with the comparative results reported in Table 1.

Table 1: Performance Comparison of the Central-Word CNN and BiLSTM for TCM Terminology Structure Recognition

Model	Accuracy (Acc)	Precision (P)	Recall (R)	F1 Score
BiLSTM (baseline)	78.2%	76.4%	77.0%	76.7
CNN Model with Center-Word Structure	87.5%	86.1%	87.0%	86.5

As shown in Table 1, the central-word CNN outperforms the BiLSTM baseline across all four reported structural-recognition measures. Accuracy increases from 78.2% to 87.5%, corresponding to an absolute improvement of 9.3 percentage points. Precision rises from 76.4% to 86.1%, while recall increases from 77.0% to 87.0%. The F1 score similarly increases from 76.7 to 86.5. These results indicate that explicitly combining local lexical information with central-word and structural cues provides a stronger representation for the annotated syntactic-role task than the selected recurrent baseline. The improvement is consistent with the model’s intended ability to capture compact subject–predicate–object and modifier–head relations in TCM terminology. Because the comparison is specific to the corpus, annotation scheme, and experimental configuration used here, the result is interpreted as evidence for the proposed structural component rather than as a universal superiority claim for CNNs over BiLSTMs.

C. Neural Modeling of English TCM Terminology Translation

This experiment evaluates English translation accuracy and semantic consistency using the knowledge-enhanced neural architecture summarized in Figure 2; Figure 5 later provides a representative term-level output. The model combines contextual language representations with the TCM knowledge graph through the attention mechanism described above. Translation quality is measured using BLEU [18] together with a terminology-consistency rating. As reported in Table 2, the knowledge-enhanced model obtains BLEU-1 and BLEU-4 scores of 73.5 and 48.1, respectively, compared with 61.2 and 35.4 for the baseline model without knowledge embedding. Thus, BLEU-4 improves by 12.7 absolute points. The mean terminology-consistency score also increases from 2.8/5 to 4.2/5. The combined results indicate that the added knowledge representation is associated with improved reference overlap and more stable specialized terminology in this experimental setting. Representative syndrome and treatment expressions, including terms corresponding to “hyperactivity of liver Yang” and “tonifying Qi to raise Yang,” further illustrate how structural and knowledge cues can reduce unstable paraphrasing while preserving established TCM terminology.

Table 2: Translation Performance of TCM Terminology NMT With and Without Knowledge Embedding

Method	BLEU-1	BLEU-4	Terminology Consistency Score
Baseline Model (No Knowledge)	61.2	35.4	2.8/5
With Knowledge Embedding	73.5	48.1	4.2/5

D. Network Pharmacology and Terminology Semantic Mapping

To strengthen semantic interpretability without conflating linguistic translation with biomedical explanation, this study uses network-pharmacology information as an auxiliary annotation layer. Representative terms such as “肾气虚” and “清热利湿” are linked to herbs, ingredients, molecular targets, and pathways reported in curated TCM databases, including TCMSP and HERB [21], [22]. These links are organized into a multi-level Term–Component–Target–Pathway structure. The resulting network is not used to redefine a TCM term in biomedical language; rather, it supplies additional evidence that the knowledge-attention module can use when distinguishing terms that have overlapping surface wording but different domain associations.

Table 3 presents two representative mappings assembled for this experiment. For “肾气虚,” the knowledge layer links

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

kidney-tonifying medicinal materials, including Astragalus (黄芪) and Psoralea (补骨脂), with selected metabolic targets and pathways represented in the integrated database records. For "清热利湿," heat-clearing medicinal materials such as Sophora flavescens (苦参) and Coptis chinensis (黄连) are linked with inflammatory targets and signaling pathways in the auxiliary network. The final column should therefore be read as a knowledge-enhanced English interpretation used in this study, not as a replacement for a standardized lexical translation. This distinction allows the model to preserve the traditional term while attaching modern pharmacological context that may improve disambiguation, scientific interpretability, and expert review.

E. Semantic Mapping and Knowledge Enhancement for TCM Terms

Figure 5 presents the structured semantic recognition and translation output for "气血两虚证" (Deficiency of Qi and Blood Syndrome), providing a representative example of the structure-aware, knowledge-enhanced translation model. The output is organized into three fields—Name, Definition, and Typical Performance—so that the generated English form can be inspected together with the semantic explanation and characteristic manifestations. This format makes the model output more interpretable than a single isolated target phrase because reviewers can compare the translated name with the explanatory content used to support it.

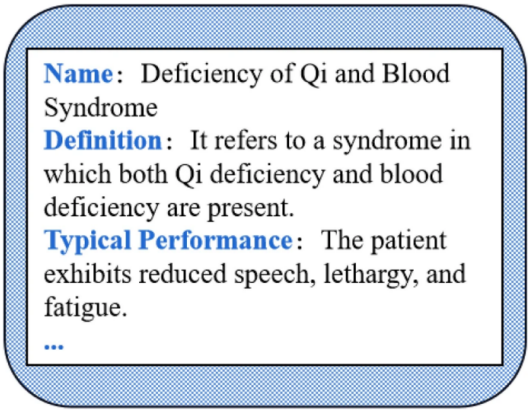

Figure 5. Semantic Recognition and English Translation Output for "Deficiency of Qi and Blood Syndrome"

At the structural level, the output separates the term entity from its explanatory and manifestation-related information. The Name field identifies the syndrome; the Definition field preserves the relationship between the core concepts "气虚" and "血虚"; and the Typical Performance field renders characteristic manifestations, including reduced speech, lethargy, and fatigue, corresponding to expressions such as "语声低微、神疲乏力." Figure 5 therefore illustrates how structural segmentation can support syntactic reorganization while retaining clinically recognizable TCM information in the English output.

Figure 6 compares the coverage of several mainstream TCM databases across term-relevant data dimensions, including Herb, Ingredient, Target, Disease, and composite relations such as Herb–Ingredient and Ingredient–Target. These categories reflect the kinds of structured associations used by the knowledge-embedding module. The underlying resources differ in scope and curation strategy: TCMID and TCMID 2.0 integrate broad TCM information [19], [20], TCMSP emphasizes systems-pharmacology data [21], HERB incorporates experiment- and reference-guided evidence [22], and ETCM organizes multiple linked entity types [23]. The radar visualization is therefore used to compare complementary coverage rather than to imply that a single database is uniformly superior across all dimensions.

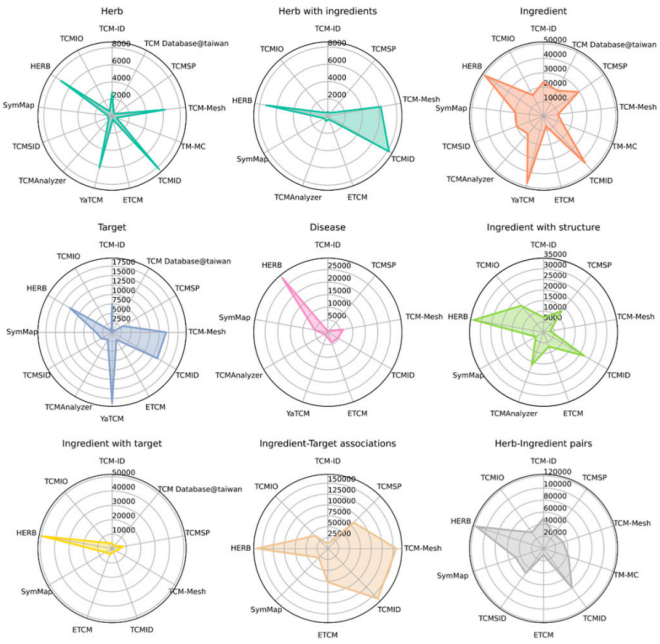

Figure 6. Radar-Chart Comparison of TCM Databases Across Term-Relevant Data Dimensions

As visualized in Figure 6, HERB and TCMSP provide extensive support for several ingredient-, target-, and association-oriented dimensions. The published description of HERB reports 7,263 herbs, 49,258 ingredients, 12,933 targets, and 28,212 diseases in the database release described by Fang et al. [22]; these figures also clarify that the ingredient count should not be described as exceeding 50,000 for that release. TCMSP contributes systems-pharmacology relations that are useful for ingredient and target modeling [21]. TCMID and its updated version provide broad cross-category integration [19], [20], while ETCM supplies additional linked information across herbs, formulas, ingredients, targets, and diseases [23]. These differences motivate multi-source integration because no single resource supplies every semantic relation required by the translation framework.

The integrated data are used to construct term–ingredient–target–disease and related semantic paths within the knowledge-embedding module. This mechanism allows a

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

Table 3: Network-Pharmacology Mapping and Auxiliary Semantic Interpretation of Representative TCM Terms

TCM Terminology	Key Medicinal Components	Associated Targets	Key Pathways	Suggested English Interpretation
肾气虚	黄芪(Astragalus), 补骨脂(Psoralea)	SIRT1, INSR	AMPK, mTOR	Kidney Qi deficiency with evidence of energy metabolism suppression
清热利湿	苦参(Sophora flavescens), 黄连(Coptis chinensis)	TNF, IL-6	NF-κB, JAK-STAT	Heat-clearing and dampness-eliminating with anti-inflammatory profile

linguistically generated translation to be supplemented with structured background information when the term has pharmacological or disease-related associations. For semantically complex expressions such as “气血两虚” and “清热解毒,” the model can therefore retain a standardized or corpus-supported English expression while using the knowledge path as contextual evidence. Figure 6 supports this design choice by showing that the selected databases contribute different kinds and quantities of relational information to the combined knowledge layer.

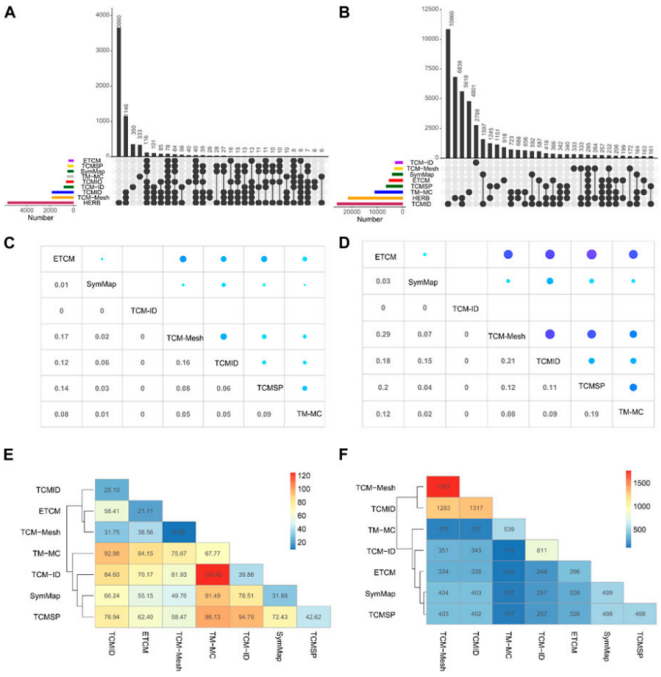

Figure 7. Overlap, Similarity, and Co-Occurrence of Terminology Across TCM Databases

Figure 7 summarizes cross-database overlap, similarity, and co-occurrence patterns at the terminology level. Panels (A) and (B) use UpSet-style intersection plots to display shared Herb and Ingredient entries, respectively, illustrating that HERB, TCMSP, TCMID, and other resources contain both overlapping and database-specific records. Panels (C) and (D) present Jaccard-similarity matrices for the Herb and Ingredient layers. In the plotted data, the TCMID–TCM-Mesh Herb comparison reaches 0.29, while the corresponding ingredient-level comparisons show lower but still meaningful overlap among selected databases. Panel (E) visualizes variation in terminology distributions using a heat map, and panel (F) displays co-occurrence counts; the plotted TCM-Mesh–TCMID pair reaches 1,787 co-occurrences in the represented comparison. These values are reported as properties of the dataset assembled for Figure 7, not as universal constants for every version of the databases. The combination of overlap and non-overlap supports the use of multi-source knowledge integration because shared entries can improve alignment while source-specific entries increase semantic coverage.

Figure 8 compares TCM databases in terms of attribute coverage and structured information representation, providing a second view of the resources used for knowledge embedding. Figure 8(A) compares ten databases across Herb- and Formula-related attributes, including names, ingredients, efficacy information, origins, quality-control fields, and usage-related descriptors. Resources such as HERB, TCMID, and ETCM provide comparatively rich structured fields that can be aligned with terminology records [19], [20], [22], [23]. The figure also shows that database schemas are heterogeneous: an attribute present in one source may be absent, represented differently, or available at another entity level in another source. Schema normalization is therefore required before these records can be used jointly.

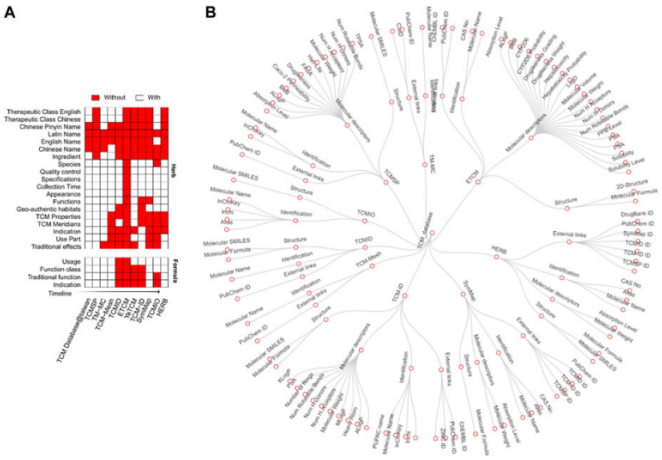

Figure 8. Attribute Coverage and Structured Term Representation Across TCM Databases

Figure 8(B) further represents the depth and breadth of database attributes in a hierarchical form. HERB, ETCM, and TCMSP provide extensible paths involving molecular structures, molecular descriptors, target links, ingredients, and related pharmacological information [21]–[23]. These attributes can enrich the model’s structured knowledge representation, but their role remains auxiliary to the linguistic translation layer. Accordingly, the database comparison is used to identify complementary features and missing fields before knowledge fusion rather than to infer linguistic equivalence directly from pharmacological metadata.

Figure 9 presents multi-level associations among Chinese medicinal materials, chemical components, pharmacological

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

functions, and protein targets. Within the translation experiments, this visualization represents how a term-related knowledge layer can be constructed from component and target information rather than from words alone. The figure contains three analytical views: a component-correlation matrix in Figure 9(A), a herb–component–target network in Figure 9(B), and a clustering network of active components in Figure 9(C). Together, these panels provide a structured representation of the pharmacological context that can accompany, but should not replace, the established linguistic interpretation of a TCM term.

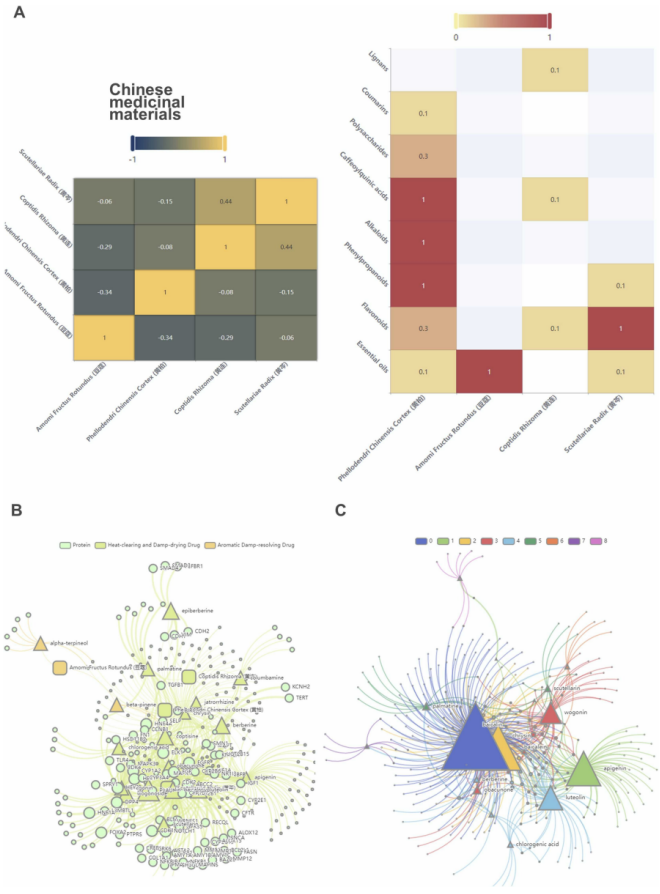

Figure 9. Visualization of TCM Ingredient Characteristics, Pharmacological Network, and Core Compound Clustering

In Figure 9(A), the left panel shows a herb-level correlation heat map for several heat-clearing and dampness-drying medicinal materials, including 黄柏, 苦参, and 黄芩, while the right panel maps selected herbs to major chemical-component categories such as alkaloids, flavonoids, and polysaccharides. These patterns illustrate how terms related to heat-clearing or dampness-resolving functions can be linked to structured component information in the experimental knowledge graph. The figure is interpreted as an association map rather than as proof that a single component category uniquely defines a traditional therapeutic concept.

Figure 9(B) depicts a multimodal herb–component–protein-target network in which nodes represent medicinal materials, active compounds, and protein targets, while edges represent relations contained in the assembled knowledge records. Components such as matrine (苦参碱) and berberine occupy connected positions alongside targets including EGFR, TLR4, and TNF in the plotted network. Such links provide biomedical context for functional expressions associated with inflammation-related mechanisms. For example, a term such as "清热利湿" may retain its conventional TCM translation while an anti-inflammatory annotation is stored separately in the knowledge layer. Keeping the translation and the biomedical annotation distinct avoids presenting a molecular interpretation as if it were the literal meaning of the source term.

Figure 9(C) further displays functional clustering among selected active components, with triangular nodes marking compounds such as baicalein, wogonin, and chlorogenic acid. The resulting modules provide additional structured features for the knowledge-attention stage and can help group terms whose associated medicinal materials share related component or target patterns. In the proposed framework, such clusters support mechanism-aware semantic annotation and expert inspection while the final English terminology remains constrained by bilingual corpus evidence, standardized terminology resources, and the source term's linguistic structure.

V. Conclusion

This study addresses the long-standing challenge of translating culturally and medically rich TCM terminology into English by combining ecological translation theory with advanced neural and data-driven methods. By introducing a central-word-based structural modeling framework and integrating multilevel knowledge representations from pharmacological databases, we construct a translation model that is both syntactically sensitive and semantically enriched. Empirical experiments confirm that the proposed model significantly improves translation accuracy and semantic consistency for complex TCM terms, especially those involving diagnostic-syndrome-treatment chains. Furthermore, by aligning translated terms with biological evidence from compound-target-pathway networks, the model ensures that the resulting English terms are not only linguistically coherent but scientifically grounded.

Future research may extend the framework to lower-resource terminology categories, additional target languages such as French and Korean, and carefully validated real-time applications. Further work should also evaluate the stability of knowledge-enhanced translations across database versions, compare expert ratings among translators and TCM clinicians, and separate lexical translation quality from the quality of supplementary biomedical annotations. Such validation is particularly important before any translation output is used in clinical decision support or other high-stakes medical communication. More broadly, the proposed framework contributes a reproducible direction for TCM terminology standardization and international dissemination by combining

linguistic structure, translation theory, neural modeling, and traceable domain knowledge.

Author Contributions

Both authors contributed equally to this work and read and approved the final manuscript.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Funding Statement

This work was supported by the 2019 Anhui Provincial Philosophy and Social Sciences Planning Youth Project, "Research on the Metaphorical Translation and International Dissemination of Bozhou Traditional Chinese Medicine Culture" (Project No. AHSKQ2019D071).

References

[1] World Health Organization, Regional Office for the Western Pacific, *WHO International Standard Terminologies on Traditional Medicine in the Western Pacific Region*. Manila, Philippines: World Health Organization, 2007.

[2] S.-H. Choi and I.-M. Chang, "A milestone in codifying the wisdom of traditional oriental medicine: TCM, Kampo, TKM, TVM—WHO international standard terminologies on traditional medicine in the Western Pacific Region," *Evidence-Based Complementary and Alternative Medicine*, vol. 7, no. 3, pp. 303–305, 2010.

[3] G. Hu, *Eco-Translatology: Towards an Eco-paradigm of Translation Studies*. Singapore: Springer, 2020.

[4] Q. Li and Y. Ng, "Effective Chinese-to-English biotic interpretation in ecotourism destinations: A corpus-based interdisciplinary study," *Journal of Sustainable Tourism*, vol. 33, no. 2, pp. 207–242, 2025.

[5] Y. Cheng and D. Glynn, "Legal constraints and translation variations: The packaging information of plant-based traditional Chinese medicine sold in the United States," *Perspectives*, vol. 32, no. 2, pp. 229–241, 2024.

[6] N. Wiseman and F. Ye, *A Practical Dictionary of Chinese Medicine*, 2nd ed. Brookline, MA, USA: Paradigm Publications, 1998.

[7] N. Wiseman and F. Ye, *Chinese Medical Chinese: Grammar and Vocabulary*. Brookline, MA, USA: Paradigm Publications, 2002.

[8] E. Neal, "Reflections on studying Huang Di Nei Jing in the West," *Chinese Medicine and Culture*, vol. 7, no. 1, pp. 72–76, 2024.

[9] K. Shan and W. Hu, "A literature review of interpreting studies on culture-loaded terms based on cognitive translation theory," *Journal of Research in Social Science and Humanities*, vol. 3, no. 11, pp. 34–40, 2024.

[10] S. Chen and T. Zhou, "Culturally based semantic losses in Lonely Planet's travel guides translations for Beijing, Shanghai, and Sichuan," *Frontiers in Communication*, vol. 9, Art. no. 1343784, 2024.

[11] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. 3rd Int. Conf. Learning Representations (ICLR)*, San Diego, CA, USA, 2015.

[12] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 5998–6008.

[13] C. Chu and R. Wang, "A survey of domain adaptation for neural machine translation," in *Proc. 27th Int. Conf. Computational Linguistics (COLING)*, Santa Fe, NM, USA, 2018, pp. 1304–1319.

[14] G. Dinu, P. Mathur, M. Federico, and Y. Al-Onaizan, "Training neural machine translation to apply terminology constraints," in *Proc. 57th Annu. Meeting Assoc. Computational Linguistics*, Florence, Italy, 2019, pp. 3063–3068.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. 2019 Conf. North American Chapter Assoc. Computational Linguistics: Human Language Technologies*, Minneapolis, MN, USA, 2019, pp. 4171–4186.

[16] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. 2014 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 1746–1751.

[17] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," *arXiv preprint arXiv:1508.01991*, 2015.

[18] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A method for automatic evaluation of machine translation," in *Proc. 40th Annu. Meeting Assoc. Computational Linguistics*, Philadelphia, PA, USA, 2002, pp. 311–318.

[19] R. Xue, Z. Fang, M. Zhang, Z. Yi, C. Wen, and T. Shi, "TCMID: Traditional Chinese Medicine integrative database for herb molecular mechanism analysis," *Nucleic Acids Research*, vol. 41, no. D1, pp. D1089–D1095, 2013.

[20] L. Huang, D. Xie, Y. Yu, H. Liu, Y. Shi, T. Shi, and C. Wen, "TCMID 2.0: A comprehensive resource for TCM," *Nucleic Acids Research*, vol. 46, no. D1, pp. D1117–D1120, 2018.

[21] J. Ru *et al.*, "TCMSP: A database of systems pharmacology for drug discovery from herbal medicines," *Journal of Cheminformatics*, vol. 6, Art. no. 13, 2014.

[22] S. Fang *et al.*, "HERB: A high-throughput experiment- and reference-guided database of traditional Chinese medicine," *Nucleic Acids Research*, vol. 49, no. D1, pp. D1197–D1206, 2021.

[23] H.-Y. Xu *et al.*, "ETCM: An encyclopaedia of traditional Chinese medicine," *Nucleic Acids Research*, vol. 47, no. D1, pp. D976–D982, 2019.

[24] Z. Liu *et al.*, "BATMAN-TCM: A Bioinformatics Analysis Tool for Molecular mechANism of Traditional Chinese Medicine," *Scientific Reports*, vol. 6, Art. no. 21146, 2016.

[25] R.-Z. Zhang, S.-J. Yu, H. Bai, and K. Ning, "TCM-Mesh: The database and analytical system for network pharmacology analysis for TCM preparations," *Scientific Reports*, vol. 7, Art. no. 2821, 2017.

[26] Y. Wu *et al.*, "SymMap: An integrative database of traditional Chinese medicine enhanced by symptom mapping," *Nucleic Acids Research*, vol. 47, no. D1, pp. D1110–D1117, 2019.

[27] Q. Pan, Q. Song, and T. Feng, "Comparative study of English and Chinese translation theory: Revealing the potential of digital technologies," *Translation Review*, vol. 118, no. 1, pp. 28–43, 2024.

[28] P. Koehn, *Statistical Machine Translation*. New York, NY, USA: Cambridge University Press, 2010.