Semantic Analysis and Translation Optimization of English Sentences Based on Long Short-Term Memory (LSTM) Networks

Jinying Leng[1], Xiaoqing Lin[2]
[1]School of Foreign Languages, Liaodong University, Dandong 118001, Liaoning, China
[2]School of Information Engineering, Liaodong University, Dandong 118001, Liaoning, China
Corresponding author: Jinying Leng (e-mail: maggiez_0826@163.com)

ABSTRACT Accurate semantic analysis and translation of complex English sentences are essential for intelligent information interaction and multilingual communication in modern digital systems, including semantic communication frameworks and electromagnetic-enabled intelligent networks. To address semantic omissions and logical inconsistencies caused by long-distance dependencies and referential ambiguity, this study proposes a GAT-BiLSTM fusion model that integrates dependency syntactic analysis with graph attention mechanisms and bidirectional long short-term memory networks. A lightweight semantic graph is first constructed to capture structural dependencies, after which graph representations are adaptively fused with contextual features through a gating mechanism to obtain unified semantic embeddings. During decoding, semantic gating and multi-head attention collaboratively enhance contextual coherence and semantic alignment. Experimental results demonstrate that the proposed model achieves a BLEU score exceeding 68.7, subject and action semantic matching scores of 0.88 and 0.84, respectively, and a syntactic structure retention rate of 72% for complex sentences. The proposed framework effectively improves translation fidelity and semantic consistency while exhibiting strong robustness for structurally complex inputs. Furthermore, the semantic modeling strategy provides methodological support for multilingual information processing, semantic communication, and intelligent human– machine interaction in electromagnetic wave propagation and wireless communication environments.

INDEX TERMS semantic analysis, graph attention network, bidirectional long short-term memory, semantic communication, electromagnetic information processing

I. INTRODUCTION

NATURAL language processing (NLP) is one of the core technologies in the field of artificial intelligence, finding wide application in areas such as machine translation and speech recognition. However, existing translation models still face many challenges when dealing with sentences with complex structures or ambiguous semantics [1,2]. Complex sentences are represented differently in different language pairs. In English–Chinese translation, syntactic structure transformation relies more on word-order adjustments, whereas in English–German translation, subject-verb reversal and noun clauses are more frequent. Translation models exhibit different structural shift risks and semantic ambiguity misjudgments across different language pairs. Improving translation accuracy, fluency, and semantic consistency, and demonstrating robustness in translating complex sentences in different language structures, have become key tasks in current translation system modeling [3-4]. The semantic consistency emphasized in this work concerns the internal structure of individual sentences, and the modeling pipeline is built upon sentence-level input patterns. Major challenges facing translation models include the handling of long-range dependencies and polysemy within sentence-level structures, which demand refined modeling of internal semantic relations. Traditional neural machine translation (NMT) models often fail to fully capture the deep semantic relationships within sentences when addressing these issues [5-6]. Especially for complex sentences, the model lacks a comprehensive understanding of grammar and semantics, resulting in a serious lack of logicality and linguistic fluency in the translation results [7-8]. Although existing improvement methods have made some progress, they still fail to effectively address core issues such as polysemy and syntactic ambiguity [9-10].

Currently, many studies have adopted LSTM-based methods and combined them with attention mechanisms to

improve translation quality. In addition, some studies have attempted to combine dependency syntactic analysis with translation models to enhance the expression of syntactic structure [11-12]. However, most of these methods still focus on short-term dependency modeling, ignoring the deep semantic understanding of long-range dependencies and polysemy, resulting in deficiencies in translation accuracy and intra-sentence contextual consistency [13-14]. To address the above issues, this paper proposes a GAT-BiLSTM Fusion Model. This method generates a lightweight semantic graph through dependency syntactic analysis and uses GAT to extract semantic dependency features, providing a unified semantic understanding of the sentence during the encoding stage. Furthermore, the semantic graph information is fused with the BiLSTM encoding results through a gating mechanism to construct a more accurate semantic representation. During the decoding phase, a multi-head attention mechanism and semantic gating mechanism are utilized to guide the translation model to pay greater attention to intra-sentence contextual consistency and semantic coherence, thereby improving translation quality. Through this multi-level semantic modeling, this paper aims to address the shortcomings of traditional models in semantic understanding and syntactic structure, and shows potential in improving the logical consistency and linguistic quality of the translation results. The structural design of the semantic interaction mechanism refers to the node information propagation rules based on adjacency weights in graph neural networks. In combination with the language processing model in cognitive linguistics that uses core semantics to drive information integration, a learnable semantic gating path is applied to match the dynamic distribution characteristics of semantic focus shifts core semantic elements during the generation of complex, dynamic fabric patterns.

II. RELATED WORK

In recent years, LSTM has demonstrated excellent sequence modeling capabilities in NLP tasks and has become an important infrastructure for sentence semantic analysis and language generation. Setyanto et al. [15] combined bidirectional LSTM with contextual feature modeling and designed a dynamic redundancy removal mechanism that effectively removed high-frequency repeated words by learning contextual information, improving document processing efficiency while maintaining semantic integrity. This type of semantic preservation strategy has promoted the further evolution of context-sensitive language models [16-17]. Subsequently, Preethi et al. [18] combined the LSTM deep stacking structure with fastText word embedding and verified the representation advantage of the three-layer LSTM architecture for Arabic tweets in the sentiment analysis task, demonstrating the synergistic effect of word embedding and deep network on semantic modeling of specific languages. The promotion of deep structures also provides inspiration for the adaptation of translation systems in complex language environments [19-20]. In terms of translation tasks, Ren et al. [21] constructed a Skip Convolutional Network-LSTM model and verified it on real corpus. The model outperformed traditional models in terms of accuracy and teaching adaptability, strengthening the application foundation of neural translation in educational scenarios. These studies show that LSTM networks have good scalability in semantic mining and language generation, but still have obvious deficiencies in modeling long-distance dependencies, comprehensive utilization of structural information, and cross-sentence semantic unification [22-23].

In the field of semantic analysis and NMT, improving the translation system's ability to model long-term sentence structure and semantic dependencies has become a research focus. The bidirectional Gated Recur rent Unit network designed by Wang et al. [24] combined part-of-speech sequence information to enhance context processing and long-range dependency understanding capabilities, effectively improving translation accuracy, reducing semantic analysis errors under complex sentence structures, and providing deep expres sion support for multi-layer semantic modeling. This improvement idea has promoted the in-depth use of contextual structure information in translation optimization [25-26]. In the translation model paradigm, Li et al. [27] summarized the modeling strategies under different architectures based on the non-autoregressive mechanism, revealed that while it improved decoding speed, it had accuracy bottlenecks in long sentence prediction and sequence learning, and put forward higher requirements for the practical performance of the current model. In terms of structural fusion, Wang et al. [28] jointly modeled convolutional neural networks with LSTM and applied an attention mechanism to improve the model's attention to the position information of long sentences, significantly improving the BLEU value and strengthening the overall coherence of the translation results. Although these studies have achieved phased results in semantic modeling, decoding optimization, and structural enhancement, they still have obvious shortcomings in the construction of multilevel semantic interactions, the comprehensive integration of structural dependencies, and the maintenance of translation consistency [29-30].

III. SEMANTIC-AWARE TRANSLATION MODEL CONSTRUCTION METHODOLOGY

This chapter focuses on the construction methodology of the GAT-BiLSTM Fusion Model. The modeling units throughout the system are single sentences, and the graph construction, dual-channel representation, and subsequent generation stages remain within this semantic boundary. Based on multi-level language structure modeling and semantic interaction mechanisms, an end-to-end translation modeling framework is designed. The overall approach takes the original English sentence as input and first constructs a structurally constrained semantic graph through dependency syntactic analysis. A syntactic consistency control strategy is then applied to sparsify the edge structure. During the modeling phase, a dual-channel semantic encoding mod-

ule consisting of a GAT and a bidirectional LSTM is employed to extract multi-dimensional semantic features from both the structural and contextual levels, respectively. Multi-channel semantic fusion is achieved through a gating mechanism. Finally, during the decoding phase, multi-head attention and a semantic gating mechanism are applied to collaboratively model the translation generation process, enhancing the dynamic selection of contextual information while maintaining semantic consistency. Figure 1 illustrates the overall modeling process of the GAT-BiLSTM Fusion Model, including input, graph construction, dual-channel modeling, fusion mechanism, decoding control, and output layers, demonstrating the deep integration of structural and contextual semantics.

A. DEPENDENCY SYNTAX-DRIVEN SEMANTIC GRAPH CONSTRUCTION

To construct a lightweight semantic graph that is highly consistent with the syntactic structure for the GAT-BiLSTM Fusion Model, the input English sentence is first subjected to dependency syntactic analysis to obtain the dependency relationships between words and their labels, and then, a directed graph structure is constructed based on this. The graph is derived from the sentence-level syntactic arrangement, and no information external to the sentence is incorporated into the structure. The nodes correspond to the terms in the sentence, and each edge represents the grammatical dependency relationship between two words. The direction of the edge points to the dominant word, and the weight of the edge is determined by the joint embedding of the part-of-speech tag and the dependency type [31-32]. In the edge weight construction, the weight function is used:

$$w_{ij} = \sigma\left(v^{\top} d_{ij} \tanh(W_1 x_i + W_2 x_j)\right) \tag{1}$$

In Formula (1), x_i and x_j represent the word vectors of word i and word j, respectively; d_{ij} is the embedding vector corresponding to dependency type d_{ij}; W_1 and W_2 are linear transformation matrices; σ represents the Sigmoid activation function. This mechanism transforms the syntactic dependency structure into a weighted graph structure, effectively capturing the grammatical control direction and semantic importance differences between words.

Node initialization uses the representation of the concatenation of word embedding and part-of-speech embedding, and maps it to a unified semantic space through projection. To ensure the sparsity and semantic directionality of the graph structure, only the core dependency path and the short path connection with high semantic association are retained. The embedding representation of the node in the graph is initialized as follows:

$$h_i^{(0)} = E_{\text{word}}(i) \oplus E_{\text{pos}}(i) \tag{2}$$

In Formula (2), $E_{\text{word}}(i)$ is the pre-trained word vector of word i; $E_{\text{pos}}(i)$ is its part-of-speech tag embedding; the symbol $\oplus$ represents the vector concatenation operation. After the construction is completed, the entire graph is represented as $G = (V, E, W)$. Among them, V is the word node set; E is the dependency edge set; W is the weight matrix.

To reduce the redundancy and over-connection problems caused by the graph structure construction, a syntactic structure consistency control mechanism is applied. According to the hierarchical information of the syntactic tree, redundant edges with large cross-layer spans are filtered out, and dependency relations are constrained to those that preserve direct semantic control within a limited syntactic depth. The cross-layer span is defined as the absolute difference between the hierarchical numbers of two nodes in the dependency tree. Edges whose span exceeds two hierarchy levels are removed, as dependency relations beyond this range exhibit weakened semantic coupling and tend to introduce indirect structural interference rather than contributing to the core predicate–argument organization of the sentence. A sparsification constraint is imposed on the adjacency matrix:

$$A' = \text{Softmax}(A \odot M) \tag{3}$$

The masked adjacency matrix is normalized using a Softmax operation applied over the set of adjacent nodes for each source node. In Formula (3), A is the original adjacency matrix; M is the structural mask matrix, which sets the connectivity according to the hierarchical span; the symbol $\odot$ represents the Hadamard product. The Softmax operation is applied row-wise, normalizing the outgoing edge weights of each node over its directly connected neighbors, ensuring that attention propagation is constrained within the local dependency neighborhood. The hierarchical indices used for this masking correspond to the depth positions assigned during dependency parsing. Fixing the span threshold at two enforces a structurally bounded propagation range in the semantic graph, ensuring that information exchange is dominated by locally controlled syntactic relations while preventing excessive graph densification caused by long-distance dependency links. During the graph construction process, each word node retains its syntactic position index and its hierarchical number in the syntactic tree, providing structural information for the subsequent graph neural network to perform hierarchical control-based message passing. After the graph structure is constructed, structural statistics are computed over the test corpus. To characterize graph properties under different sentence complexity levels, the test sentences are partitioned into five subsets denoted as G1–G5 according to sentence length and dependency tree depth. The corresponding semantic graph statistics are summarized in Table 1, reflecting average dependency level, node connectivity, and edge weight distribution patterns within each subset.

All values reported in Table 1 are averaged within each subset, ensuring that the presented graph characteristics reflect stable distributional properties derived from hundreds of sentence instances rather than isolated examples.

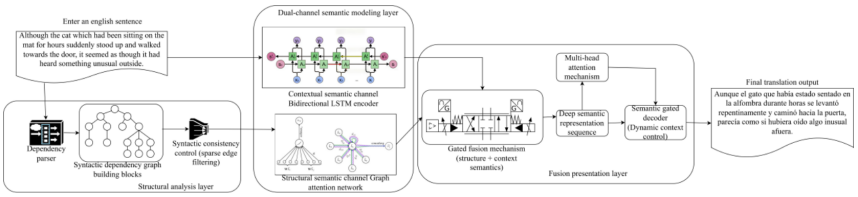

FIGURE 1. Overall modeling framework of this paper's semantic-aware translation model

TABLE 1. Statistics of semantic graph structure features

Subset (Number of sentences)	Average dependency level	Average node out-degree	Valid edge weight distribution range
G1 (412)	4.2	2.7	[0.35, 0.87]
G2 (398)	3.9	3.1	[0.42, 0.90]
G3 (427)	4.5	2.5	[0.38, 0.84]
G4 (405)	3.7	2.9	[0.40, 0.92]
G5 (389)	4.0	3.0	[0.33, 0.88]

B. DUAL-CHANNEL SEMANTIC REPRESENTATION LEARNING MECHANISM

In the unified semantic modeling framework of the GAT-BiLSTM Fusion Model, to fully capture the structural dependency and contextual semantic information between terms, a dual-channel semantic representation path is constructed: the first is to perform feature propagation and structural modeling on the dependency semantic graph based on GAT, and the second is to perform contextual modeling on the word sequence based on the bidirectional long short-term memory network. The input of the graph channel is the weighted dependency graph constructed in the previous section; the word node embedding is used as the initial feature; the graph attention mechanism is used for multi-head feature extraction [33-34]. Each layer of the graph attention mechanism completes information aggregation by calculating the attention weights between nodes, which is specifically expressed as follows:

$$h_i^{(l+1)} = \sigma\left(\sum_{j\in N(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)}\right) \tag{4}$$

In Formula (4), $h_j^{(l)}$ is the feature representation of node j in the l-th layer; $\alpha_{ij}^{(l)}$ is the attention weight from node j to i; $W^{(l)}$ is the learnable transformation matrix; σ is the nonlinear activation function; $N(i)$ represents the set of adjacent nodes of node i. The multi-head mechanism forms a unified structural semantic representation by splicing the output of each head, thereby enhancing the expressive power of graph space modeling.

The sequential channel performs bidirectional modeling on the word sequence of the input sentence based on BiLSTM to capture the semantic dependency relationship before and after the context. The input word sequence is represented as a vector sequence $\{x_t\}_{t=1}^{n}$ after embedding mapping, which serves as the input sequence of BiLSTM. The BiLSTM network consists of a forward LSTM and a backward LSTM. For each word embedding x_t, the forward hidden state $\overrightarrow{h}_t$ and the backward hidden state $\overleftarrow{h}_t$ are obtained, respectively, and the context encoding result is obtained after splicing:

$$h_t^{\text{seq}} = \overrightarrow{h}_t \oplus \overleftarrow{h}_t \tag{5}$$

In Formula (5), $\oplus$ represents the splicing operation. This representation integrates the sequence semantics into the current word, forming a word vector sequence with complete semantics within the sentence, which is used to supplement the structural modeling blind spots of the graph channel.

To achieve the information fusion of the two types of semantic representation paths, a gated fusion mechanism is applied to construct a unified deep semantic representation. The gated unit controls the selective injection of information based on the mutual comparison of the features of the graph and sequence channels, and its forms are:

$$z_t = \sigma\left(W_g[h_t^{\text{seq}} \oplus h_t^{\text{graph}}] + b_g\right) \tag{6}$$

$$h_t^{\text{fusion}} = z_t \odot h_t^{\text{seq}} + (1 - z_t) \odot h_t^{\text{graph}} \tag{7}$$

In Formulas (6) and (7), h_t^{graph} represents the output of the graph channel; W_g and b_g are trainable parameters; $\odot$ represents the Hadamard product; z_t is the fusion gate vector, which determines the weighted ratio of the two types of semantic representations.

The fused representation sequence $\{h_t^{\text{fusion}}\}$ is used as the input of the decoding stage to provide complete semantic support for the generation stage. The gating unit implements an element-wise adaptive fusion strategy. This node-specific mechanism computes a dedicated gate vector for each word position, dynamically balancing the contributions from the structural graph features and contextual sequence information. The formulation addresses the variable reliance on structural versus contextual cues across different parts of a sentence. During feature extraction, the number of graph attention layers and BiLSTM layers are set synchronously to ensure consistency in the fused structure. To verify the effectiveness of the fused features, the activation responses

of node representations before and after semantic fusion are measured at the output of the gated fusion layer. The activation magnitude is computed as the L2 norm of the fused hidden representation for each node position. Table 2 lists the average activation difference magnitudes and grammatical categories of typical nodes in the dual-channel representation.

All activation values are normalized within each sentence by the maximum node activation, and the reported statistics are obtained by averaging over all sentence instances in the test set.

The dual-channel semantic representation mechanism provides semantically complete and well-structured representation input for the subsequent translation generation stage, improving the semantic parsing

capability while maintaining grammatical consistency, and enhancing the model's parsing stability for long dependencies and semantic ambiguity in complex sentences. This module plays a key role in semantic construction in the overall model structure and is the core bridge connecting input preprocessing and translation generation.

C. DECODER DESIGN AND SEMANTICALLY ENHANCED TRANSLATION GENERATION

The translation generation stage of the GAT-BiLSTM Fusion Model is built on the attention decoder architecture, integrating graph structure semantic information and dynamic semantic guidance mechanism on the basic sequence generation model to achieve optimal control of context consistency and semantic coherence. The decoder core consists of a gated semantic application mechanism and a multi-head attention module, and the input is the fused deep semantic representation sequence $\{h_t^{\text{fusion}}\}$. When decoding each target word, the system dynamically selects the focus area based on the translation state vector at the previous moment, the generated partial translation, and the encoded representation, and selects the adaptive semantic inflow intensity at the current moment [35].

The semantic information of the graph structure is represented by embedding the decoder state through the graph attention context vector. At each time t, the current decoding context vector c_t is constructed as the result of multi-source attention weighting:

$$c_t = \sum_{i=1}^{n} \alpha_{t,i}^{(s)} h_i^{\text{fusion}} + \sum_{j=1}^{n} \alpha_{t,j}^{(g)} h_j^{\text{graph}} \tag{8}$$

In Formula (8), $\alpha_{t,i}^{(s)}$ represents the attention distribution based on the sequence path; $\alpha_{t,j}^{(g)}$ represents the attention distribution under the graph path; h_i^{fusion} and h_j^{graph} represent the fusion channel and the structure channel, respectively. Both attention distributions are obtained by weighting the hidden state of the previous moment and the representation of each node to ensure that the structural information and contextual semantics are guided synchronously during the generation process. The inclusion of the separate structural channel h_j^{graph} alongside the fused representation h_i^{fusion} in the decoder attention provides a direct pathway to structure-biased semantic representations derived from dependency relations. While h_i^{fusion} integrates structural and sequential information into a unified semantic space, the h_j^{graph} channel preserves semantic features that remain strongly guided by syntactic dependency topology. The explicit inclusion of h_j^{graph} allows the decoder to directly access and prioritize structure-biased semantic representations when generating words that are sensitive to dependency-controlled relations, such as verbs and function words that are constrained by sentence-level structure. This design ensures that crucial syntactic cues are not diluted during the generation process.

To improve the decoder's adaptability to semantic consistency, a semantic gating mechanism is applied to dynamically regulate the injection ratio of graph-derived semantic information. This mechanism modulates the information channel opening based on the compatibility between the current decoder state and the incoming semantic representation, allowing the model to adaptively emphasize or suppress structural semantic cues during word generation. Its calculation methods are:

$$g_t = \sigma(W_s s_{t-1} + W_c c_t + b_g) \tag{9}$$

$$\tilde{s}_t = g_t \odot c_t + (1 - g_t) \odot s_{t-1} \tag{10}$$

In Formulas (9) and (10), s_{t-1} is the decoder state at the previous moment; W_s and W_c are trainable transformation matrices; g_t is the gating vector; $\tilde{s}_t$ represents the current state after gating adjustment, which serves as the basis for the next state update and word distribution prediction. This structure regulates the influence strength of structural semantics on the decoding state, stabilizing the integration of contextual and graph-based information during language generation.

To understand the effect of the semantic gating mechanism on attention regulation during translation generation, Figure 2 compares the changes in contextual attention distribution during the decoding phase under different mechanisms. The left figure shows the attention heatmap without the gating mechanism, showing a uniform distribution of attention weights and a lack of focus on semantic stem words. The right figure shows the results after the gating mechanism is applied, with attention focused on the semantic core area and a significant increase in the model's response to key structural words. The red dotted line indicates the location of enhanced attention, indicating that the gating mechanism effectively guides the flow of contextual semantics and improves the semantic aggregation ability of the generation path.

To fully utilize the multi-level semantic associations in the context, a multi-head attention mechanism is applied inside the decoder, and the query vector interacts with different linear projection subspaces, respectively, thereby learning

TABLE 2. Activation changes after dual-channel representation fusion by node category

Node grammatical category	Number of nodes	Average activation value before fusion	Average activation value after fusion	Activation difference change rate
Subject	62	0.472	0.638	35.2%
Predicate	58	0.451	0.612	35.7%
Object	49	0.498	0.653	31.1%
Modifier	73	0.421	0.569	35.2%
Conjunction	35	0.390	0.514	31.8%

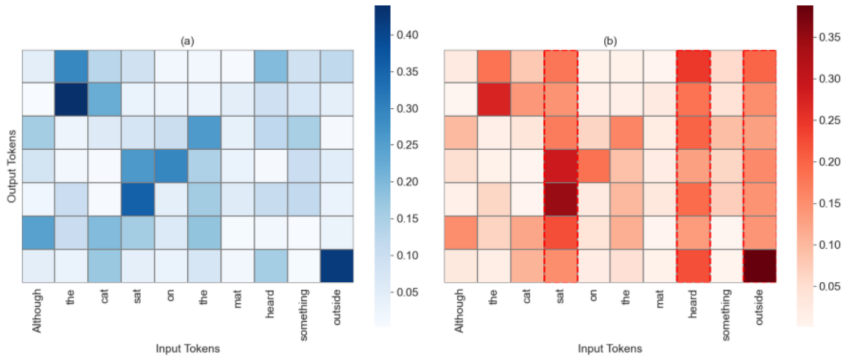

FIGURE 2. Comparison of context attention response distribution before and after the application of semantic gating mechanism: (a) Attention distribution before the application of semantic gating mechanism; (b) Attention distribution after the application of semantic gating mechanism

multiple heterogeneous semantic alignment mappings. For each head h, the context vector is calculated as follows:

$$\text{head}_h = \text{Attention}(Q_h, K_h, V_h) \tag{11}$$

The attention function adopts the scaled dot-product attention formulation, where similarity between query and key vectors is computed through scaled inner products to ensure stable gradient behavior during multi-head semantic alignment. In Formula (11), $Q_h = W_h^Q H$, $V_h = W_h^V s_{t-1}$, and $K_h = W_h^K H$. H represents the entire semantic representation of the encoding stage, and W_h^Q, W_h^V, and W_h^K are the corresponding linear projection matrices for query, key, and value vectors. The attention weight computation follows scaled dot-product normalization, ensuring that attention distributions remain numerically stable across different representation dimensions. The outputs of all heads are concatenated and then transformed into the final semantically enhanced context representation through a layer of linear transformation to guide the probability distribution of vocabulary generation.

The prediction of the translated word depends on the joint representation of the current decoding state and the semantically enhanced context, and the generated probability distribution is:

$$P(y_t \mid y_{<t}, X) = \text{Softmax}\left(W_o[\tilde{s}_t \oplus c_t] + b_o\right) \tag{12}$$

In Formula (12), y_t is the current target word; $y_{<t}$ represents the previously generated sequence; X is the input source sentence; W_o and b_o are the output layer parameters.

IV. EXPERIMENTAL SETUP AND EVALUATION STRATEGY

A. DATASET AND PREPROCESSING STRATEGY

For the experiment, the English–Chinese parallel corpus from the WMT14 translation benchmark is adopted, focusing on sentence pairs that exhibit complex syntactic structures and rich semantic relations. The corpus contains approximately 4.5 million aligned sentence pairs drawn from diverse news and formal text sources, covering a wide range of sentence patterns, from simple declarative structures to sentences with nested clauses, inversions, and long-distance dependencies. All experiments are conducted under a fixed English-to-Chinese translation direction. The evaluation samples function as independent sentence instances, and the metrics capture the preservation of semantic and syntactic relations within each sentence.

The dataset is partitioned following a standard split protocol. Approximately 90% of the sentence pairs are used for model training, 5% are reserved as a development set for parameter tuning and early stopping, and the remaining 5% constitute the test set for final performance evaluation. This partitioning ensures that the reported results reflect generalization performance on unseen sentence structures rather than memorization of training patterns.

The data undergoes rigorous preprocessing, including word segmentation, part-of-speech tagging, and the construction of syntactic dependencies. Word segmentation utilizes deep learning-based segmentation models to ensure precise boundary identification for each word unit. Part-of-speech tagging employs state-of-the-art annotation tools to improve labeling accuracy and ensure that the grammatical category of each word is reliably identified. Dependency syntactic analysis is performed using the Stanford Parser to extract word-level dependency relations and hierarchical tree structures.

To mitigate the impact of potential parsing errors on the graph structure, the sparsification strategy described in Section 3.1 is applied. This process removes dependency edges with large cross-layer spans, filtering out connections that are more likely to be erroneous or structurally redundant, thereby enhancing the robustness and stability of the constructed semantic graph.

To construct the semantic graph, a dependency parser is applied to each input sentence, generating a word-to-word dependency graph. This graph provides the internal grammatical structure of the sentence and serves as the foundation for subsequent graph attention modeling. During graph construction, information from all levels of the syntactic tree is considered, while sparsification removes unnecessary connections and retains only structurally and semantically salient dependency relations. The resulting graph generates a node representation for each word, where each node contains the word embedding and its associated syntactic information, forming the structural input for the graph attention network.

B. EXPERIMENTAL ENVIRONMENT AND PARAMETER CONFIGURATION

The experiments are conducted on the PyTorch platform, relying on high-performance computing resources for training and testing. The experimental hardware configuration consists of an NVIDIA A100 Graphics Processing Unit with 80GB of on-board video memory, deployed on a high-performance computing server equipped with 128GB of system memory and 8TB of storage. All experiments are run in this environment, ensuring efficient and stable model training.

In the experiments, a series of key parameter configurations are employed to ensure the validity of the experimental results. The learning rate is set to 0.0001, and the Adaptive Moment Estimation optimizer is used for model training to ensure fast and stable convergence. A batch size of 32 is used for each training epoch, ensuring efficient training within computing resource constraints. To avoid overfitting, a dropout rate of 0.5 is used to enhance the model's generalization ability. The number of training epochs is set to 50 to ensure sufficient learning of the semantic features in the data. During training, the cross-entropy loss function is used, and the model's performance on the validation set is periodically evaluated.

C. COMPARISON METHOD AND EVALUATION INDEX SETTING

This experiment comprehensively evaluates the translation performance and semantic processing capabilities of the proposed model by comparing it with four mainstream translation models. The models include the traditional Seq2Seq (Sequence to Sequence) architecture, the Transformer, the BiLSTM, and the GAT-based BiLSTM-GAT. All models are tested on the same training dataset and experimental environment to ensure fair and reliable results. The evaluation metrics used in the experiment focus on three aspects: BLEU score, semantic match, and syntactic structure retention. These metrics cover translation quality, semantic consistency, and syntactic correctness.

The BLEU score is used to assess the accuracy of translation results and is calculated based on the n-gram overlap between the predicted sentence and the reference sentence. In this study, BLEU is computed at the sentence level to reflect fine-grained translation behavior under different structural conditions, and all reported values are aggregated within the scope of their corresponding experimental tasks. The assessment focuses on sentence-level translation quality, and the analysis remains constrained to sentence-internal semantic organization. Semantic match measures the model's preservation of the core semantics of the sentence during translation, primarily by comparing the semantic components of the translated sentence with those of the reference sentence. The syntactic structure retention rate focuses on analyzing the degree of preservation of syntactic structure during the translation process, especially how to accurately restore the grammatical structure of the original sentence when dealing with complex sentence patterns.

The semantic match metric quantifies the alignment of core semantic components between the translated sentence and the reference by evaluating role-level semantic consistency. The metric is computed using Semantic Role Labeling applied within a unified annotation space. For each sentence pair, the generated target-language translation is first mapped back into English using the same translation system configuration, ensuring semantic content preservation without introducing external supervision. Semantic Role Labeling is then applied to both the back-translated sentence and the original English reference using an identical SRL model and annotation scheme, producing comparable role assignments across predicate–argument structures. For each semantic dimension, such as the subject role, the semantic match score is computed as the F1 measure derived from the precision and recall of correctly aligned role instances between the two analyses. This evaluation protocol avoids cross-lingual SRL inconsistencies by constraining all role extraction to a single language and a single annotation system, thereby reducing metric variance caused by language-specific labeling bias. This provides a structured assessment of how well the model preserves predicate-argument structures.

V. TRANSLATION PERFORMANCE RESULTS AND SEMANTIC ANALYSIS

A. BLEU SCORE COMPARISON

To further explore the performance differences between various translation models when processing sentences with varying structural and semantic features, this study conducts systematic experiments using four representative neural translation models. Model 1 is the Transformer, which uses a self-attention mechanism to capture global dependencies and performs reliably in modeling short-range information.

Model 2 is a traditional Seq2Seq architecture, which possesses basic sequence modeling capabilities but lacks a deep semantic alignment mechanism. Model 3 is a BiLSTM encoder, whose bidirectional structure effectively captures contextual information. Model 4 is the proposed method, which combines a dependency syntax-driven GAT with a BiLSTM to construct a deep semantic representation using semantic gating units. In the experiments, test sentences are classified into four categories: grammatical structure complexity, semantic logical relations, syntactic variation structure, and sentence length. For each category, discriminative sentence patterns are selected to comprehensively evaluate the adaptability and translation performance of each model using the BLEU metric. For each structural category, the test set consists of manually filtered sentence instances that share similar structural patterns and complexity levels. The relevant results are shown in Figure 3. BLEU scores are computed at the sentence level with a smoothing strategy applied to higher-order n-grams, and the reported values correspond to averaged sentence-level BLEU scores within each categorized subset.

This paper's method demonstrates more stable BLEU performance compared to other models in scenarios with highly complex structures and long-range dependencies. In sub-figure a, when the sentence structure transitions from simple to inverted/elliptical, the Transformer's BLEU score drops from 78.5 to 55.6, a drop of 22.9 points. Seq2Seq and BiLSTM's drops are 21.3 and 21.0, respectively. This paper's method exhibits the smallest drop, only 17.5, demonstrating its ability to preserve semantics within deep dependency structures. In sub-figure b, for conditional and transitional sentences, this paper's method maintains a BLEU score above 73, achieving a maximum improvement of 3.5 points compared to BiLSTM, demonstrating that its semantic interaction mechanism is more accurate in logical sentence structures. In sub-figure c, under structural perturbations such as parentheses and multiple modifiers, the semantic consistency of the traditional model decreases significantly. However, this paper's method maintains a BLEU score above 68.7 under the nonlinear structure, highlighting its robustness guided by the graph structure. Sub-figure d shows that the BLEU score of the traditional model fluctuates significantly with increasing sentence length, while this paper's method only drops by 15.7 points from short to extremely long sentences, far less than the 23-point drop of the Transformer, demonstrating its advantage in modeling long text. In summary, deep translation models based on the semantic graph fusion mechanism demonstrate superior generalization and semantic fidelity across a wide range of language structures.

B. EFFECT OF IMPROVING SEMANTIC MATCHING

The evaluation of semantic matching employs the metric defined in the experimental setup, which relies on Semantic Role Labeling to extract and compare core semantic elements. The scores for subject, action, object, and modifier consistency reflect the F1 measure of role alignment between the model's output and the reference translations. In the translation of complex English sentences, the nonlinearity of grammatical structure and the discreteness of semantic distribution place higher demands on neural translation models. To address this issue, this section conducts a multi-model comparison experiment based on the semantic matching dimension, using three typical complex structures: principal-subordinate compound sentences, inverted sentences, and inserted sentences. Complex sentences with a main and subordinate clause often have multilevel dependency chains, and their semantic backbones are often obscured by subordinate clauses, leading to the risk of subject-verb ambiguity during translation. In inverted sentences, after word order adjustment, the semantic trigger point and representation position shift, challenging the model's sequential modeling capabilities. Inserted sentences, due to their embedded structure, interfere with backbone extraction, requiring the model to be able to identify semantic boundaries. To this end, three mainstream models are applied for comparison: Seq2Seq, an early encoder-decoder benchmark, lacks global modeling capabilities when dealing with complex syntactic structures; the Transformer constructs global dependencies through a self-attention mechanism, but lacks explicit modeling of local syntax; the BiLSTM captures word order information through bidirectional context, offering a degree of robustness at the sequence level. As the core of this comparison, this paper proposes a model that integrates GAT with a semantic gating mechanism to enhance its ability to model semantic dependencies in complex structures. To analyze the impact of these structures on semantic preservation in translation, a radar chart (Figure 4) is constructed, focusing on subject, action, object, modifier consistency, and semantic coherence. The comparative performance of the models under different structural conditions is examined.

In the principal-subordinate complex sentence matching test, the proposed model achieves scores of 0.88 and 0.84 in the subject and action dimensions, significantly higher than the Transformer's 0.76 and 0.70. Its deep semantic modeling strengthens the semantic aggregation of the main sentence across subordinate clauses, alleviating the problem of subordinate clause structure obscuring information from the main clause. In the inverted sentence structure, the proposed model achieves a semantic coherence score of 0.82, an improvement of over 0.20 compared to the Seq2Seq model, demonstrating its greater adaptability to nonlinear word order through its schema structure perception. In the inserted sentence test, the proposed model achieves a modifier consistency score of 0.75, surpassing the Transformer's 0.60. The semantic gating mechanism effectively mitigates the interference of inserted components on the main information flow, enabling more accurate recovery of the main sentence meaning. Comprehensive analysis shows that the model exhibits more stable semantic preservation performance when handling various types of complex sen-

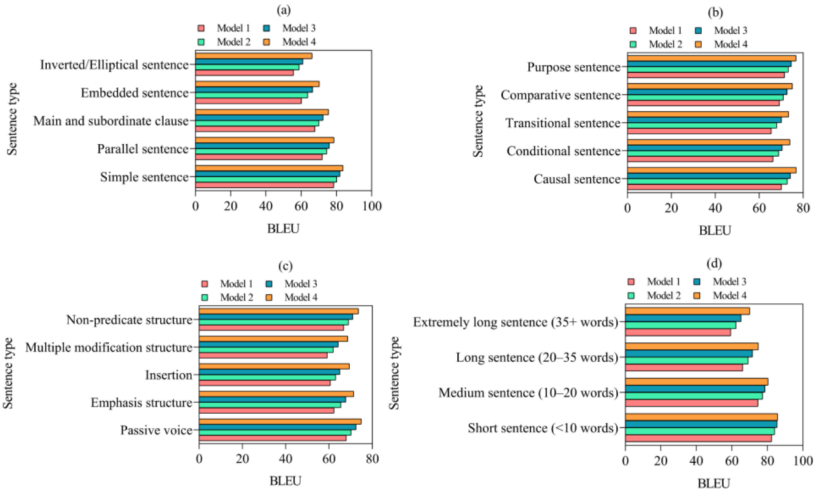

FIGURE 3. Comparison of BLEU scores of various translation models under different semantic structure dimensions: (a) Comparison of BLEU scores based on sentence complexity; (b) Comparison of BLEU scores based on semantic logical relationships; (c) Comparison of BLEU scores based on syntactic variation structures; (d) Comparison of BLEU scores based on sentence length ranges

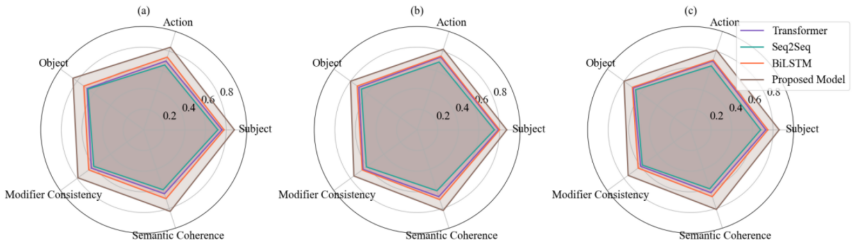

FIGURE 4. Radar chart comparing semantic matching dimensions of various models for different sentence types: (a) Semantic matching analysis of principal-subordinate compound sentences; (b) Semantic matching analysis of inverted sentences; (c) Semantic matching analysis of inserted sentences

tence structures, improving the parsing integrity of semantic elements and inter-sentence logical consistency.

C. SYNTACTIC STRUCTURE RETENTION RATE

In English sentences of varying structural complexity, the degree to which a translation model retains syntactic information directly impacts the readability and semantic accuracy of the translated text. This study compares the structure retention performance of four representative models across different sentence types, focusing on their ability to restore syntactic structure during translation. Model 1 is a Seq2Seq encoder-decoder model, emphasizing sequence mapping capabilities. Model 2 is a standard Transformer architecture, with the advantage of modeling global dependencies. Model 3 uses a BiLSTM for bidirectional context modeling. Model 4 is the proposed BiLSTM-GAT joint model, integrating a graph attention mechanism with semantic gating units. Considering the structural challenges posed by sentence structure differences, the experiment analyzes simple, parallel, and complex sentences. Parallel relationships and clause nesting are structurally sensitive aspects of translation. Figure 5 presents a comparison of the retention rates of different models for subjects, predicates, objects, modifiers, and structurally specific elements. The structural retention rate is defined based on dependency structure preservation. Dependency parsing is applied to both the translated sentence and the reference using the same syntactic analyzer, and a dependency is considered retained when the head–dependent relation and its label are aligned in both parses. The final retention rate is computed as the proportion of retained dependencies relative to the reference structure, averaged over all sentence instances.

Experimental results show that the proposed method maintains leading structure retention rates across all sentence types. In simple sentences, the proposed model achieves 82% retention of modifiers, a significant improvement over BiLSTM's 74%. This improvement is primarily attributed to the enhanced modeling of fine-grained contextual dependencies within the semantic graph. In parallel sentences, the traditional Seq2Seq method achieves only 54% retention, while the proposed method achieves 73%, demonstrating a stronger ability to discern structural boundaries between parallel components. In complex sentences, the clause structure retention rate reaches 72%, significantly exceeding the Transformer's 55%. This result demonstrates that the representation method that integrates graph structure and sequence information is more adaptable to modeling hierarchical structures. Across all metrics, the improved

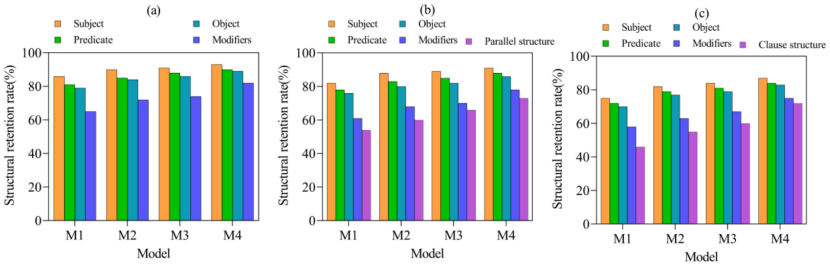

FIGURE 5. Comparison of syntactic structure retention rates across three sentence types by different translation models: (a) Comparison of structure retention capabilities for simple sentences; (b) Comparison of structure retention capabilities for parallel sentences; (c) Comparison of structure retention capabilities for complex sentences

model's ability to restore both the syntactic trunk and complex subsidiary structures surpasses other baselines, validating its advantage in structure preservation under complex semantics.

D. COMPARISON OF SEMANTIC COHERENCE UNDER DIFFERENT METHODS

This experiment aims to evaluate the semantic consistency performance of different translation models on complex English sentences. Four models are used for comparative analysis, focusing on the multi-level semantic information of sentences. Model 1 is a traditional neural network model based on BiLSTM, which excels at semantic understanding while preserving local contextual information. Model 2 is a network architecture based on Seq2Seq, which leverages an attention mechanism to improve its ability to model long-range dependencies, but has limitations when processing complex sentence structures. Model 3 is the Transformer, which leverages a self-attention mechanism to improve its ability to model the global structure of sentences, making it particularly suitable for processing complex sentences. Model 4 is a proposed translation model based on a semantic interaction mechanism, combining a graph neural network with a bidirectional LSTM to optimize the deep learning effect of semantic representation. This study further subdivides the semantic dimensions into: action consistency, subject consistency, object consistency, temporal consistency, and causal consistency. Each dimension focuses on specific linguistic features in translation, aiming to comprehensively measure the semantic retention and consistency of translation. Figure 6 shows the performance of the four models across different semantic dimensions.

When processing simple sentences, the proposed model demonstrates a clear advantage, particularly in action consistency and subject consistency, with scores of 0.83 and above. In contrast, BiLSTM and Seq2Seq perform relatively poorly, particularly in temporal consistency and causal consistency, where scores are low. The Transformer model struggles with complex sentences, with most scores failing to exceed 0.75. This suggests that while its self-attention mechanism improves global understanding of sentences, it lacks the ability to capture local semantic connections within complex sentences. For complex sentences, the scores of all models generally decline, reflecting the challenging nature of complex sentence structures. The proposed model maintains a relatively balanced advantage, particularly in subject consistency, with scores above 0.80. This demonstrates that by applying the semantic graph and graph attention mechanism, it effectively preserves the semantic information and logical relationships of complex sentences during translation.

E. ABLATION EXPERIMENTS AND CONTRIBUTIONS OF EACH MODULE

In this multi-module collaborative semantically enhanced translation model, to deeply analyze the contributions of each key component to semantic modeling and translation generation, this section begins with a BiLSTM encoder and gradually applies graph GAT, semantic gating, and multi-head attention modules to construct five model combinations: S1 is a BiLSTM baseline, featuring only word sequence context modeling; S2 builds on S1 by applying GAT to enhance the ability to perceive semantic dependencies between word structures; S3 further incorporates a semantic gating mechanism to control the flow of semantic information for improved semantic extraction selectivity; S4 superimposes a multi-head attention mechanism to enhance multi-scale context capture; S5 is the complete model, integrating all three mechanisms to achieve deep semantic representation and translation consistency control. On this basis, the BLEU trend line chart (Figure 7a) and the semantic dimension radar chart (Figure 7b), supplemented by comparisons of the semantic matching, structure retention rate, and semantic coherence of each module combination, fully demonstrate the layer-by-layer improvement of model performance. The ablation experiments are conducted on sentence subsets characterized by long dependency chains and nested syntactic relations. The results are shown in Figure 7 and Table 3:

Experimental data show that the S1 model performs the worst among all indicators, with an average BLEU score of only 20.0. The distribution of semantic dimensions is generally weak, and the scores of subject-verb consistency and action-object matching dimensions do not exceed 0.70. This is related to the BiLSTM modeling method based only on local context, which makes it difficult to capture the information flow at the syntactic structure level. After applying GAT, the average BLEU score for S2 rises to 22.6. The semantic graph structure enhances the model's

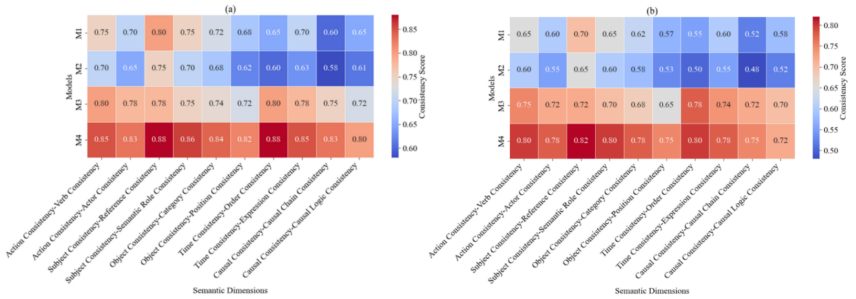

FIGURE 6. Semantic consistency performance of different translation models in simple and complex sentences: (a) Simple sentences; (b) Complex sentences

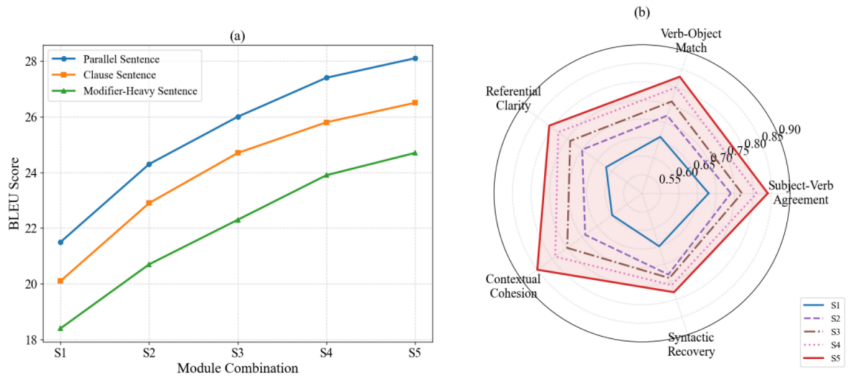

FIGURE 7. Impact of different module combinations on translation quality and semantic perception: (a) Changes in BLEU scores for different sentence types for each module combination; (b) Radar chart of the model's semantic perception

TABLE 3. Comparison of performance contributions of various module combinations to semantic metrics

Module Combination	Semantic Match	Structural Retention Rate	Semantic Coherence
S1	0.72	0.65	0.68
S2	0.79	0.71	0.74
S3	0.83	0.72	0.80
S4	0.86	0.73	0.84
S5	0.88	0.75	0.87

understanding of the main semantic dependencies, and the semantic matching score increases to 0.79, demonstrating that the application of structural semantics effectively alleviates the problem of ambiguous syntax parsing. After adding a semantic gating mechanism to S3, the semantic coherence score jumps to 0.80. The intra-contextual cohesion dimension, shown in Figure 7(b), significantly expands, indicating that this mechanism plays a key role in improving semantic consistency. S4, building on S3, applies multi-head attention, further improving the average BLEU score to 25.7. Multiple semantic dimensions are balanced, with particularly strong performance in reference clarity and intra-sentence contextual consistency, validating the enhanced role of multi-scale attention in modeling complex intra-sentence relations. The complete S5 model achieves optimal performance across all metrics, with an average BLEU score of 26.4 and a semantic match of 0.88. When translating complex sentences, the fine-grained control of the semantic modeling mechanism plays a decisive role in maintaining semantic consistency and improving language quality in translation.

F. MODEL ADAPTABILITY IN COMPLEX LANGUAGE PAIRS

In multilingual neural translation, the heterogeneity of language structure significantly impacts translation quality. Word-order differences, morphological variations, and dependency chain depth pose core challenges to model generalization. This section constructs two complex sentence translation tasks for different language pairs to examine the model's ability to maintain semantic consistency under diverse structural conditions. The experiment applies five neural translation models with varying structural characteristics: Model 1 is the baseline Seq2Seq model, which uses an encoder-decoder alignment mechanism, with a simple structure and limited semantic alignment capabilities; Model 2 is the Transformer, which uses a self-attention mechanism to capture global dependencies but cannot model syntactic constraints; Model 3 is a bidirectional LSTM, which strengthens contextual order modeling but has limited sequence depth; Model 4 is a BiLSTM with a graph attention mechanism, which enhances the ability to model inter-word structural dependencies; Model 5 is the graph-structured semantic gating translation model proposed in this paper, which constructs deep representation through the collaborative construction of semantic graph, gating mechanism, and multi-head attention. The experiment processes six language pairs, selecting two complex sentence types, principal-subordinate

compound sentences and inverted/inserted sentences, as the test benchmarks. This creates a dual structural-linguistic interference condition to evaluate the model's cross-lingual adaptability and structural robustness. Figure 8 shows the BLEU scores of each model on the two complex sentence translation tasks. All language pairs are evaluated under a unified experimental setting, where training and test sets are derived from parallel sentence-level corpora using consistent data splits, preprocessing procedures, and BLEU evaluation protocols, and all baseline systems strictly follow the model configurations described in the experimental setup section. The multilingual evaluation involves language pairs with substantial differences in word order, morphological richness, and dependency realization.

In the principal-subordinate compound sentence translation task, the proposed method achieves a BLEU score of 61.8 on the English–Chinese language pair, while the Transformer achieves only 52.8. The widening gap between the two reflects the effect of the semantic gating mechanism on the aggregation of long-range main stem semantics. The proposed method achieves BLEU scores of 57.2 and 56.8 in English–German and English–Russian, respectively, outperforming the BiLSTM-GAT model's 53.9 and 53.3. This demonstrates that graph structure perception plays a greater role in the deeply nested German and frequently varied Russian languages. In the inverted and inserted sentence tasks, the overall scores of the models decline. Seq2Seq achieves only 30.8 on English–Japanese, due to its weak information alignment, resulting in misalignment of main stem semantics. The proposed method, on the same language pair, maintains a score of 48.5, relying on its robust response mechanism to structural focus shifts. Compared to other models, the proposed method maintains BLEU stability in languages with frequent structural jumps and loose word order, such as Arabic and Japanese. This demonstrates that the structural semantic fusion strategy is highly adaptable to diverse language structures and offers consistent advantages for modeling highly complex sentences across languages.

VI. CONCLUSION

This paper proposes a GAT-BiLSTM Fusion Model. This approach constructs a lightweight semantic graph through dependency syntactic analysis, extracts inter-node semantic dependency features using GAT, and fuses these features with the BiLSTM encoding output via gating units to form a deep semantic representation. The decoding stage combines a semantic gating mechanism with multi-head attention to dynamically regulate vocabulary generation. Experiments demonstrate that the proposed method improves semantic accuracy and language quality in English complex sentence translation tasks: in inverted/elliptical sentence translation tasks, the BLEU score drops by only 17.5 points; the subject and action semantics match in complex sentences reach 0.88 and 0.84, respectively; the clause structure retention rate in complex sentences reaches 72%. This achievement addresses the shortcomings of traditional models in long-distance dependencies and polysemy resolution, providing a high-fidelity translation solution for complex or semantically ambiguous English sentences. The improvements arise from sentence-level translation settings, and the intra-sentence contextual consistency in this study denotes enhanced coherence within the semantic structure of single sentences. This has significant application value in improving the quality of cross-language communication where high-fidelity translation is essential for accurately communicating intricate specifications, material names, and care instructions across international production and consumer markets, thereby mitigating costly errors.

AUTHOR CONTRIBUTIONS

Conceptualization – Jinying Leng; methodology – Jinying Leng and Xiaoqing Lin; investigation – Xiaoqing Lin; writing-original draft preparation – Jinying Leng and Xiaoqing Lin. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This work was supported by On Howard Goldblatt's English Translation of "Folk Language" from the Perspective of Bourdieu's Habitus from General Scientific Research Project of Liaoning Provincial Department of Education in 2023, grant number JYTMS20230724.

ACKNOWLEDGEMENTS

Not applicable.

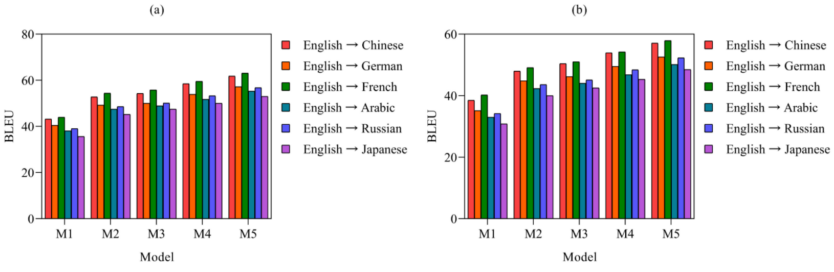

FIGURE 8. Comparison of BLEU scores for complex sentences in different language pairs: (a) Performance comparison of principal-subordinate compound sentence translation; (b) Performance comparison of inverted and inserted sentence translation

REFERENCES

[1] H. Jung, K. Kim, H. Shin J, H. Na S, S. Jung, and S. & Woo, "(2022)," Impact of sentence representation matching in neural machine translation. Applied Sciences, 12(3), pp. 1313-1331. https://doi.org/10.3390/app12031313.

[2] X. Shi, H. Huang, P. Jian, and K. & Tang Y, "(2021)," Improving neural machine translation with sentence alignment learning. Neurocomputing, 420(1), pp. 15-26. https://doi.org/10.1016/j.neucom.2020.05.104.

[3] S. Wang, H. Huang, and S. & Shi, "(2022)," Improving non-autoregressive machine translation using sentence-level semantic agreement. Applied Sciences, 12(10), pp. 5003-5015. https://doi.org/10.3390/app12105003.

[4] L. Kang, S. He, M. Wang, F. Long, and J. & Su, "(2023)," Bilingual attention based neural machine translation. Applied Intelligence, 53(4), pp. 4295-4315. https://doi.org/10.1007/s10489-022-03563-8.

[5] M. Mahsuli M, S. Khadivi, and M. & Homayounpour M, "(2023)," Lenm: improving low-resource neural machine translation using target length modeling. Neural processing letters, 55(7), pp. 9435-9466. https://doi.org/10.1007/s11063-023-11208-1.

[6] X. Geng, L. Wang, X. Wang, M. Yang, X. Feng, B. Qin, and Z. & Tu, "(2022)," Learning to refine source representations for neural machine translation. International Journal of Machine Learning and Cybernetics, 13(8), pp. 2199-2212. https://doi.org/10.1007/s13042-022-01515-9.

[7] P. Wang, "(2022)," A study of an intelligent algorithm combining semantic environments for the translation of complex English sentences. Journal of Intelligent Systems, 31(1), pp. 623-631. https://doi.org/10.1515/jisys-2022-0048.

[8] S. Harada and T. & Watanabe, "(2022)," Neural machine translation with synchronous latent phrase structure. Journal of Natural Language Processing, 29(2), pp. 587-610. https://doi.org/10.5715/jnlp.29.587.

[9] Y. Sang, Y. Chen, and J. & Zhang, "(2023)," Neural machine translation research on syntactic information fusion based on the field of electrical engineering. Applied Sciences, 13(23), pp. 12905-12919. https://doi.org/10.3390/app132312905.

[10] G. Duan, H. Yang, K. Qin, and T. & Huang, "(2021)," Improving neural machine translation model with deep encoding information. Cognitive Computation, 13(4), pp. 972-980. https://doi.org/10.1007/s12559-021-09860-7.

[11] C. ZHANG Y, X. GAO S, T. YU Z, H. WANG Z, and L. & MAO C, "(2023)," A Chinese-Vietnamese neural machine translation method using the dual representation of BERT and word embedding. Computer Engineering & Science, 45(03), pp. 546-554.

[12] Z. Hlaing Z, K. Thu Y, T. Supnithi, and P. & Netisopakul, "(2022)," Improving neural machine translation with POS-tag features for low-resource language pairs. Heliyon, 8(8), pp. 1-22. https://doi.org/10.1016/j.heliyon.2022.e10375.

[13] H. Choi G, H. Shin J, H. Lee Y, and K. & Kim Y, "(2022)," Toward Understanding Most of the Context in Document-Level Neural Machine Translation. Electronics, 11(15), pp. 2390-2408. https://doi.org/10.3390/electronics11152390.

[14] J. Unanue I, Z. Borzeshi E, and M. & Piccardi, "(2022)," Regressing word and sentence embeddings for low-resource neural machine translation. IEEE Transactions on Artificial Intelligence, 4(3), pp. 450-463. https://doi.org/10.1109/TAI.2022.3187680.

[15] A. Setyanto, A. Laksito, F. Alarfaj, M. Alreshoodi, I. Oyong, M. Hayaty, et al., "(2022)," Arabic language opinion mining based on long short-term memory (LSTM). Applied Sciences, 12(9), pp. 4140-4148. https://doi.org/10.3390/app12094140.

[16] S. Chen, "(2024)," An Intelligent Algorithm for Semantic Feature Analysis and Translation of English Texts. Journal of ICT Standardization, 12(3), pp. 271-282. https://doi.org/10.13052/jicts2245-800X.1232.

[17] H. He, "(2023)," An intelligent algorithm for fast machine translation of long English sentences. Journal of Intelligent Systems, 32(1), pp. 0257-20220266. https://doi.org/10.1515/jisys-2022-0257, 2022.

[18] V. Preethi and U. & Kiruthika, "(2021)," Survey on text transformation using Bi-LSTM in natural language processing with text data. Turkish Journal of Computer and Mathematics Education, 12(9), pp. 2577-2585.

[19] S. Karunarathna K M G and M. & Rupasingha R A H, "(2022)," Learning to use normalization techniques for preprocessing and classification of text documents. International Journal of Multi-disciplinary Studies, 9(2), pp. 69-81.

[20] K. Markov A, O. Semenochkin D, G. Kravets A, and A. & Yanovskiy T, "(2024)," Comparative analysis of applied natural language processing technologies for improving the quality of digital document classification. International Journal of Open Information Technologies, 12(3), pp. 66-77.

[21] B. Ren, "(2021)," The use of machine translation algorithm based on residual and LSTM neural network in translation teaching. Plos one, 15(11), pp. 0240663-0240679. https://doi.org/10.1371/journal.pone.0240663.

[22] L. Zhao, W. Gao, and J. & Fang, "(2021)," High-performance english-chinese machine translation based on GPU-enabled deep neural networks with domain corpus. Applied Sciences, 11(22), pp. 10915-10932. https://doi.org/10.3390/app112210915.

[23] M. Gupta and P. & Kumar, "(2021)," Robust neural language translation model formulation using Seq2seq approach. Fusion: Practice and Applications, 5(2), pp. 61-67. https://doi.org/10.54216/FPA.050203.

[24] Q. Wang, "(2022)," Semantic analysis technology of English translation based on deep neural network. Computational Intelligence and Neuroscience, pp. (1), 1176943-1176952. https://doi.org/10.1155/2022/1176943, 2022.

[25] J. Yu and X. & Ma, "(2022)," English translation model based on intelligent recognition and deep learning. Wireless Communications and Mobile Computing, pp. (1), 3079775-3079784. https://doi.org/10.1155/2022/3079775, 2022.

[26] X. Guo, "(2022)," Optimization of English machine translation by deep neural network under artificial intelligence. Computational intelligence and neuroscience, pp. (1), 2003411-2003421. https://doi.org/10.1155/2022/2003411.

[27] F. Li, J. Chen, and X. & Zhang, "(2023)," A survey of non-autoregressive neural machine translation. Electronics, 12(13), pp. 2980-3011. https://doi.org/10.3390/electronics12132980.

[28] C. Wang, "(2021)," Efficient English translation method and analysis based on the hybrid neural network. Mobile Information Systems, pp. (1), 9985251-9985261. https://doi.org/10.1155/2021/9985251, 2021.

[29] S. Lv, B. Yang, R. Wang, S. Lu, J. Tian, W. Zheng, and L. & Yin, "(2024)," Dynamic multi-granularity translation system: DAG-structured multi-granularity representation and self-attention. Systems, 12(10), pp. 420-435. https://doi.org/10.3390/systems12100420.

[30] J. Yang, "(2023)," ECBTNet: English-Foreign Chinese intelligent translation via multi-subspace attention and hyperbolic tangent LSTM. Neural Computing and Applications, 35(36), pp. 25001-25011. https://doi.org/10.1007/s00521-023-08624-8.

[31] L. Ding, L. Wang, and S. & Liu, "(2023)," Recurrent graph encoder for syntax-aware neural machine translation. International Journal of Machine Learning and Cybernetics, 14(4), pp. 1053-1062. https://doi.org/10.1007/s13042-022-01682-9.

[32] R. Peng, T. Hao, and Y. & Fang, "(2021)," Syntax-aware neural machine translation directed by syntactic dependency degree. Neural Computing and Applications, 33(23), pp. 16609-16625. https://doi.org/10.1007/s00521-021-06256-4.

[33] K. Chen, R. Wang, M. Utiyama, and E. & Sumita, "(2021)," Integrating prior translation knowledge into neural machine translation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30(1), pp. 330-339. https://doi.org/10.1109/TASLP.2021.3138714.

[34] L. He, Q. Meng, Q. Zhang, J. Duan, and H. & Wang, "(2023)," Event detection using a self-constructed dependency and graph convolution network. Applied Sciences, 13(6), pp. 3919-3929. https://doi.org/10.3390/app13063919.

[35] F. Wan and P. & Li, "(2024)," A neural machine translation method based on split graph convolutional self-attention encoding. PeerJ Computer Science, 10(1), pp. 1886-1907. https://doi.org/10.7717/peerj-cs.1886.