Enhancing Contextual Semantic Reconstruction in College English Translation Using the BART Auto-Encoding Generation Framework

Pandeng Li
School of Foreign Languages, Henan University of Urban Construction, Pingdingshan 467036, Henan, China
Corresponding author: Pandeng Li (e-mail: cflpd@163.com)

ABSTRACT To address the decline in coherence and accuracy of college English translation caused by insufficient contextual semantic modeling, this paper proposes a Context-Aware Semantic Reconstruction Fine-tuning (CASR-F) method based on BART (Bidirectional and Auto-Regressive Transformers). The method is particularly useful for translating specialized terminology and complex technical descriptions in engineering fields such as electromagnetic wave analysis, antenna systems, and propagation scenario documentation. CASR-F introduces a dependency syntax-guided dynamic noise masking mechanism, which improves the encoder's ability to identify core predicates and argument nodes through controlled noise injection. A gated semantic anchor attention module uses the top-level hidden state of the encoder as the semantic hub and dynamically adjusts cross-sentence information transmission weights through a gating function, thereby improving text-level semantic relation modeling during decoding. The model further enhances semantic recovery by jointly optimizing sequence reconstruction loss and cross-sentence semantic consistency constraints. Experiments show that CASR-F achieves a BLEU-4 score of 36.72 and a METEOR score of 31.45 in the College English Translation task, with a referential resolution accuracy of 78.6%. The F1 score for semantic role labeling on core argument A0 is improved to 83.9%. These results demonstrate the effectiveness of the proposed method in improving translation quality, reconstructing semantic structures, and enhancing semantic restoration, providing practical support for the accurate translation of electromagnetic engineering documents, antenna design descriptions, and wave propagation technical materials.

INDEX TERMS bart autoencoder generation framework, contextual semantic reconstruction, english translation, antenna systems, electromagnetic wave propagation

I. INTRODUCTION

COLLEGE English translation aims to achieve equivalent transformation of semantics, style, and discourse coherence, with the core being to maintain the contextual semantic restoration effect of the translated text, a task especially critical when accurately conveying specialized terminology, technical logic, and disciplinary concepts in engineering fields. Although neural machine translation (NMT) has improved the quality of translations, its output still often suffers from problems such as awkward expression and poor logical coherence [1-3]. Due to the high density of knowledge, dense referents, clear logical hierarchy, and close semantic connections between sentences in college English discourse, current sentence level translation models are limited by their ability to model context, often leading to ambiguity and disrupting semantic coherence [4-6]. Effectively capturing and conveying contextual semantic associations has become a core challenge faced by machine translation [7,8]. The existing sentence level NMT models mainly rely on local word order statistics and hidden feature distribution. Due to the lack of long-term memory modeling mechanisms, it is difficult to maintain consistency in cross sentence referential and argument relationships [9,10]. The non-linear hierarchical complexity of dependency structures in English discourse can easily lead to semantic focus blurring [11-13]. Although discourse level models enhance semantic integration through module extensions, attention dispersion and noise interference still weaken their logical coherence performance [14-16]. In addition, frequent cross sentence and implicit subject phenomena further increase

the requirement for robustness in semantic restoration [17]. There are studies focusing on hierarchical structures, construction of referential chains, and semantic constraints. The hierarchical attention model enhances context awareness through a dual layer encoder, but its structure is complex and resource consumption is high, making it unstable on small and medium-sized corpora [18,19]. The method based on referential chain relies on manual annotation, which restricts the generalization performance of the model [20]. Although semantic consistency constraints and contrastive learning strategies have improved the continuity of semantic space, the granularity of constraints is relatively coarse and has not effectively cooperated with word level reconstruction mechanisms [21,22]. While autoencoder-generated models such as BART have advantages in long sequence modeling, their pre-training and fine-tuning mechanisms lack specific inductive biases for discourse dependency structures and cross-sentence semantic alignment, making it difficult for them to fully realize their context modeling potential in complex discourse translation tasks. This research gap is precisely the key problem addressed by the dynamic noise mask and gated attention mechanism proposed in this paper [23-25].

This paper innovatively proposes a context-aware CASR-F method based on the BART autoencoder architecture, where semantic reconstruction is defined as recovering consistent semantic representations across sentences under noise interference, rather than regenerating surface grammatical forms. This method designs a dynamic noise masking mechanism based on dependency syntax to introduce controlled perturbations during the encoding stage and enhance the discriminative representation of key semantic units. This process achieves explicit modeling of the semantic recovery task through the embedding and replacement of mask positions. Building upon this, semantic anchors are constructed using the encoder's toplevel hidden states, and a gated attention mechanism is used to regulate information flow during the decoding stage to maintain semantic stability at the discourse level. During training, the sequence reconstruction loss and cross-sentence semantic consistency constraints are jointly optimized, enabling the model to complete the mapping from noisy input to coherent semantic representation within an end-to-end framework, thus achieving context-aware semantic recovery for college English translation scenarios and improving the overall quality of the translation in terms of logical coherence and semantic integrity.

II. DESIGN OF CONTEXT SEMANTIC ENHANCEMENT METHOD BASED ON BART

A. DYNAMIC NOISE MASKING MECHANISM GUIDED BY DEPENDENCY SYNTAX

To improve the encoder's ability to focus on key semantic nodes in the text, a dynamic noise masking mechanism guided by dependency syntax is designed. The input layer applies learnable weights and probabilistically applies noise to guide the encoder to enhance semantic dependency path representation [26,27]. Dependency syntactic analysis can be performed on the original sentence pairs, constructing an intra sentence dependency adjacency matrix A, and defining the structural weight of each word as the weighted sum of node degrees. Calculate the importance score s_i of a word based on its lexical information and sentence level positional features:

$$s_i = \alpha \cdot \mathrm{IDF}(t_i) + \beta \cdot d_i + \gamma \cdot \phi(pos_i) \tag{1}$$

In formula (1), $\mathrm{IDF}(t_i)$ is the inverse document frequency value of the i-th word, $\phi(pos_i)$ is the scalar representation obtained by position encoding through a small feedforward network, and α, β, γ are trainable scalar weights. Map the importance score to mask sampling probability and use a smooth sigmoid threshold function to achieve dynamism:

$$p_i = \sigma(\lambda(s_i - \tau)) \tag{2}$$

In formula (2), p_i represents the probability of the i-th word being masked, $\sigma(\cdot)$ is the sigmoid function, λ controls the slope, and τ is the threshold bias term. Implement random noise replacement for input embedding based on probability p_i. The noise source is a trainable set of masked vectors $Z = \{z_k\}$, masked random variables $m_i \sim \mathrm{Bernoulli}(p_i)$, and the masked input is represented as:

$$\tilde{x}_i = (1 - m_i)x_i + m_i z_{k(i)} \tag{3}$$

In formula (3), x_i is the original word embedding, $z_{k(i)}$ is the vector sampled from the noise vector set according to the distribution, and $\tilde{x}_i$ is the final input sent to the encoder. Figure 1 shows the processing flow and key data flow of the encoder input. In Figure 1, the input sentence pairs are processed by the dependency structure analysis module to form a dependency adjacency matrix, and word-level structure weights are calculated based on this matrix. These structure weights, along with lexical features and positional information, are used to generate semantic importance scores. The importance scores are mapped to mask probabilities using a threshold function. These mask probabilities drive random variable sampling to form explicit mask decisions, thereby defining the locations where embedding replacement is required. Random noise vectors are sampled from a trainable noise set and used to replace the original input embeddings at mask activation locations, forming a masked input representation. This representation serves as the sole input to the BART encoder during the encoding stage, and under the combined influence of the sequence reconstruction objective and mask position reconstruction constraints, a high-fidelity semantic representation with discourse consistency is generated.

B. GATED SEMANTIC ANCHOR ATTENTION MECHANISM

To enhance the focus of the decoding stage on the discourse level semantic core, semantic anchors based on the encoder's final hidden state are implanted across attention layers, and

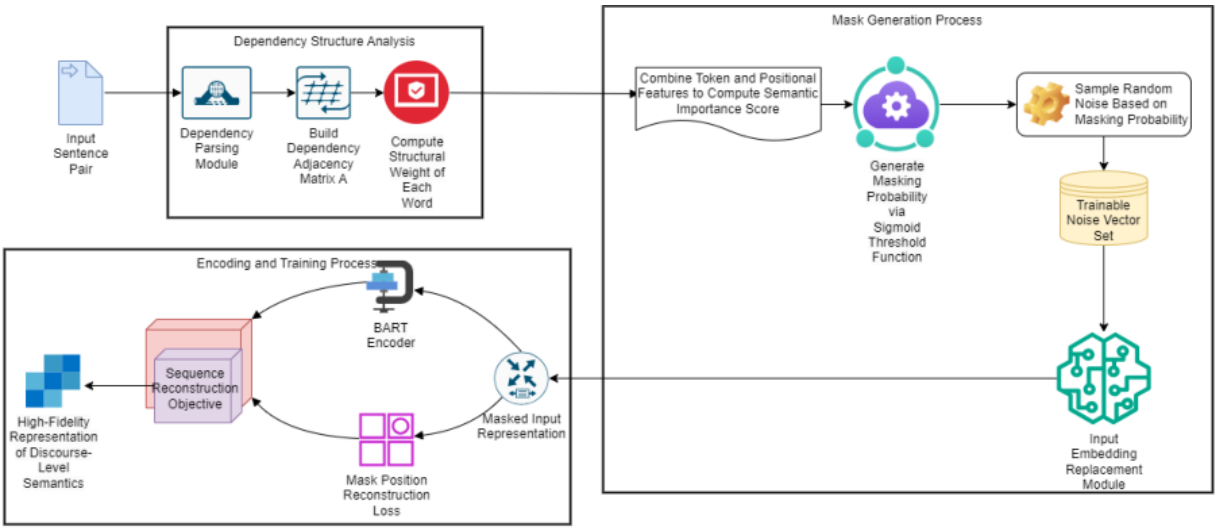

FIGURE 1. Framework diagram of dynamic noise mask module guided by dependency syntax

their contribution to attention weights is adjusted through gating functions [28-30]. The semantic anchor is obtained by weighting and aggregating the hidden states of the encoder's final layer according to dependency syntax. In step t of the decoder, a gate coefficient based on the anchor is constructed.

Attention score: Adding anchor guidance term to standard scaled dot product attention to form a joint score with gating:

$$\tilde{e}_{t,i} = \frac{s_t^\top k_i}{\sqrt{d}} + g_t \cdot \frac{s_t^\top a}{\sqrt{d_a}} \tag{4}$$

In formula (4), $\tilde{e}_{t,i}$ represents the joint attention unnormalized score of the decoder time step t on the i-th position of the encoder, k_i is the i-th key vector, d is the key value dimension, and d_a is the scaling dimension correction factor of the anchor vector. The normalized context vector is fused with regular attention context and anchor projection through gate controlled weighted fusion, resulting in the final decoded context representation:

$$c_t = (1 - g_t) \sum_i \alpha_{t,i} v_i + g_t \cdot Ua \tag{5}$$

In formula (5), c_t is the final context vector for time step t, $\alpha_{t,i}$ is the Softmax attention weight based on $\tilde{e}_{t,i}$, v_i is the i-th value vector, and U is the linear projection matrix from the anchor point to the context space. Table 1 shows the comparative results of various schemes in terms of BLEU-4, METEOR, and referential resolution accuracy.

The results indicate that adaptive gating fusion significantly improves both translation quality and cross sentence referential consistency. The improvement comes from the dynamic adjustment of the contribution of gating to anchor points, which strengthens discourse level semantic references during long-distance dependency or referential resolution scenarios in the decoding stage, and preserves fine-grained positional alignment information during short distance alignment. The anchor weight w_i is calculated based on the strength of dependent edges and position sensitivity, and normalized by Softmax. The gating module uses a feedforward network and applies Dropout to suppress excessive dependencies. In training, the gating coefficients are included in the joint loss to enable the decoder to maintain trade-offs in different contexts [31-33]. The anchor weight w_i is represented as:

$$w_i = \mathrm{Softmax}(a \cdot s_i + b \cdot p_i) \tag{6}$$

In formula (6), s_i is the strength of the dependent edge, p_i is the position sensitivity factor, and a and b are learnable parameters.

C. JOINT SEMANTIC CONSISTENCY TRAINING OBJECTIVES

To achieve precise semantic restoration at the discourse level, the training objective integrates sequence reconstruction and cross sentence semantic consistency into the fine-tuning process in a weighted joint form, so that the model can simultaneously bear the constraints of word level reconstruction and sentence level semantic convergence [34,35]. Sequence reconstruction uses negative logarithmic likelihood to measure the prediction quality of dynamically masked positions, and the reconstruction loss is defined as:

$$L_{\mathrm{rec}} = -\frac{1}{T} \sum_{t=1}^{T} \log P(x_t \mid \tilde{x}) \tag{7}$$

In formula (7), L_{rec} represents the sequence reconstruction loss, T is the length of the target sequence, x_t is the t-th target word element, and $\tilde{x}$ is the encoder conditional representation after dynamic masking. The sentence vector is obtained by the encoder's final hidden state through average pooling. Comparative learning based on sentence

TABLE 1. Comparison of the Effects of Gate Control and Fusion Methods on Translation Quality and Reference Retention

Fusion Method	BLEU-4	METEOR	Coreference Resolution Accuracy (%)
No Anchor (Baseline)	32.15	28.47	69.4
Anchor Connection Concatenation	33.40	29.12	71.8
Anchor Additive Bias	34.05	29.75	73.6
Fixed Gating Multiplicative Fusion	34.78	30.21	75.1
Adaptive Gating Fusion (Proposed in This Paper)	36.72	31.45	78.6

vectors implements cross sentence semantic consistency constraints in the form of InfoNCE loss. The vector is defined as:

$$L_{\text{cons}} = -\frac{1}{M}\sum_{m=1}^{M}\log\frac{\exp(\text{sim}(z_m^s, z_m^p)/\tau)}{\sum_{k=1}^{K}\exp(\text{sim}(z_m^s, z_k)/\tau)} \quad (8)$$

In formula (8), L_{cons} represents cross sentence semantic consistency loss, M is the number of positive example pairs, z_m^s and z_m^p are the mth main sentence vector and corresponding positive example sentence vector, respectively. $\{z_k\}_{k=1}^K$ is the set of negative example sentence vectors within the batch, $\text{sim}(\cdot,\cdot)$ represents the cosine similarity function, and τ is the temperature hyperparameter to adjust the smoothness of the similarity distribution.

III. EXPERIMENTAL SETUP AND DATA CONSTRUCTION

A. CONSTRUCTION OF COLLEGE ENGLISH TRANSLATION CORPUS

The experimental corpus is constructed based on parallel text resources in college English teaching and examination scenarios, covering three genres: textbook selections, academic paper excerpts, and English writing samples, to fully reflect the diversity of semantic structures and discourse coherence features. The corpus is sourced from the Ministry of Education's open corpus platform for foreign languages in universities and the public English to Chinese parallel corpus. After screening and cleaning, high-quality training samples are formed. In the preprocessing stage, the text is divided into sentence boundaries and GIZA++ is introduced for bidirectional word alignment. This process is not used as direct input for training the neural model, but is used to help identify cross-linguistic correspondences. By constraining lexical consistency and sentence length ratio, potentially semantically mismatched sentence pairs are eliminated, thereby improving the alignment reliability of parallel corpora at the discourse level. In the corpus construction stage, the text is first annotated to integrate continuous sentences into complete semantic units, and then further divided into nested structures and subordinate clause components. In response to the differences in language characteristics between Chinese and English, segmentation and morphological restoration were performed separately, and BPE (Byte Pair Encoding) algorithm was used to divide sub words. The vocabulary scale is set to 50000. The above processing flow aims to provide semantically consistent and noise-controlled input distributions for end-to-end Transformer translation models, rather than introducing explicit alignment supervision, thereby enhancing the semantic stability of the data layer without changing the model training paradigm. To improve the quality of annotation, this article introduces semantic role annotation based on PropBank standard and referential chain annotation based on OntoNotes system in the corpus.

B. MODEL TRAINING AND HYPERPARAMETER CONFIGURATION

The training is based on pre-trained BART for context aware CASR-F. The encoder input is processed by dynamic noise masking and sent to the model for semantic restoration through a gated semantic anchor attention mechanism. The training target adopts a weighted joint form of sequence reconstruction loss and cross sentence semantic consistency loss to achieve end-to-end semantic consistency. The PyTorch framework is used for training, with AdamW optimizer and a weight decay coefficient of 0.01. The learning rate adopts a linear preheating and cosine annealing strategy, with heating completed in the top 10%. The batch size is set to 32 and the gradient clipping threshold is 1.0. To improve generalization performance, a Dropout rate of 0.1 regularization strategy is adopted and LayerNorm is introduced at the decoder end. The mask noise distribution parameters are updated through dynamic sampling and reconstruction loss, and the mask ratio is adaptively adjusted between 0.25 and 0.35. The initial weights of the gated semantic anchor points are normalized according to the dependency distribution. The total number of training epochs is set to 20. After each round, the optimal model is selected based on the weighted scores of BLEU-4 and SRL-F1 (Semantic Role Labeling F1) from the validation set. The weight ratio of the sequence reconstruction term, semantic consistency term, and mask regularization term in the joint loss is set to 0.6:0.3:0.1. This configuration maintains the dominant role of the reconstruction objective on the convergence path in the early stages of training, while gradually enhancing the influence of semantic consistency constraints and limiting the interference of the regularization term on gradient updates, thereby maintaining the stability and convergence efficiency of the optimization process; and the temperature parameter τ is set to 0.07. The model is trained in a distributed manner on a 4xA100 GPU (Graphics Processing Unit) environment, with gradient synchronization using a mixed precision strategy and parameter initialization following the BART original distribution standard. The key hyperparameters and implementation settings for

TABLE 2. Parameter Table for Training and Optimization Configuration of CASR-F Model

Parameter Category	Parameter Name	Value Setting	Description
Optimization Strategy	Optimizer Type	AdamW	Weight decay is adopted to improve stability.
Training Scheduling	Initial Learning Rate	3e-5	Cosine annealing is applied after linear warm-up.
Batch Processing Setting	Batch Size	32	Maintains the stability of gradient estimation.
Regularization Strategy	Dropout Rate	0.1	Prevents overfitting and excessive dependence on anchors.
Loss Configuration	Loss Weight Ratio	0.6:0.3:0.1	Controls the balance between reconstruction, consistency, and regularization.
Gradient Control	Gradient Clipping Threshold	1.0	Prevents gradient explosion and ensures stable optimization.
Masking Strategy	Mask Ratio	0.25–0.35 (adaptive)	Adjusts the proportion of masked tokens during training to enhance robustness.
Training Configuration	Total Epochs	20	Defines the total number of training iterations for model convergence.
Contrastive Learning	Temperature Parameter	0.07	Controls the sharpness of similarity distribution in contrastive objectives.
Model Scale	BART Baseline Parameters	406M	Total trainable parameters of the pretrained BART Large model.
Model Scale	CASR-F Parameters	421M	Parameter size after introducing dynamic mask and semantic anchors based on BART baseline.

model training are shown in Table 2.

Table 2 shows the design orientation of training configuration between optimizing stability and semantic restoration performance. The weight ratio exhibits a coordinated effect on discourse level consistency and word level accuracy in multiple rounds of validation. Dropout and weight decay jointly maintain the parameter balance of the model, enabling dynamic masking, anchor attention, and joint loss to converge stably under the same optimization framework, thereby achieving deep contextual semantic restoration.

IV. RESULT ANALYSIS

A. OVERALL TRANSLATION QUALITY ASSESSMENT: COMPARATIVE ANALYSIS WITH BASELINE MODEL

To verify the overall effectiveness of the method at a macro level, this paper compares CASF-F with the original anchor free BART baseline and Transformer baseline models using commonly used translation quality indicators. BLEU-4 and METEOR are used as characterization indicators to measure vocabulary matching, phrase reproduction, and local fluency. Experiments have shown that based on the same corpus and hyperparameter settings, CASR-F significantly outperforms baseline in BLEU-4 and METEOR metrics across different sentence length intervals. In scenarios with more than 30 words, CASR-F has a minimum BLEU-4 value of only 34.5 and a minimum METEOR value of only 29.3. When the BART baseline is greater than 30 words, BLEU-4 reaches its highest value of 31.43, and its highest METEOR value does not exceed the lowest value of CASR-F. Compared with Transformer, CASR-F also maintains a robust advantage, further demonstrating the contribution of the designed dependency guided mask and gate anchor attention to macro translation quality Under finer grained sentence length conditions, this article divides sentences into several length intervals and compares the BLEU-4 and METEOR score trends of three types of models in each length interval. The experimental results showed that as the sentence length increased, the score of the traditional baseline model decreased significantly, while the score of CASR-F fluctuated more smoothly and remained at a higher level overall. This indicates that dynamic noise masks and gated semantic anchors help maintain more stable vocabulary alignment and phrase reproduction ability in long sentences or complex dependency structures. The comparison of curves in Figure 2 shows that CASR-F decays more smoothly and scores higher, indicating its context driven optimization effect in vocabulary matching and word order generation. The dynamic masking mechanism and anchor attention complement each other in improving surface fluency. The effectiveness of the proposed model in terms of overall translation quality has been verified, laying a solid foundation for subsequent discourse level and semantic level analysis.

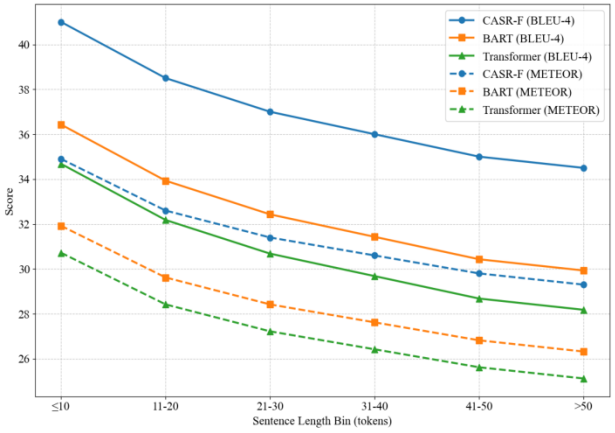

FIGURE 2. Curve of BLEU-4 and METEOR scores under different sentence lengths

B. TEXT LEVEL SEMANTIC COHERENCE ANALYSIS

The semantic coherence of a translation at the discourse level largely depends on the ability to maintain the predicate argument structure during the translation process. This article selects the semantic role annotation F1 value as the evaluation index, compares the CASR-F model output with the standard translation for semantic role comparison, and systematically evaluates the performance of dynamic masking mechanism and gated semantic anchor attention mechanism in deep semantic restoration from two dimensions: structural consistency and semantic coherence. Experimental data shows that the CASR-F model outperforms the BART baseline model in F1 values for the main semantic role types. On the core argument A0, the F1 value of CASR-F reached 83.9%, an increase of 3.8% from the baseline. It reaches 82.1% on the objective argument A1, an increase of 4.3%; The performance is particularly outstanding on the additional argument AM, with the F1 value increasing from 73.5% of the baseline to 80.8%, an increase of 7.3%. In the encoder section, the dynamic masking mechanism effectively enhances the recognition ability of core predicates and their related arguments. In the decoder section, the gated semantic anchor attention mechanism dynamically adjusts the attention intensity to the discourse structure during cross sentence generation, effectively maintaining the integrity of the predicate argument structure. In addition, by

introducing inter sentence contrastive learning constraints in the joint training objectives, the model can maintain the continuity of adjacent sentence distributions in the semantic space, thereby achieving more natural semantic coherence in logical transitions and event progression processes. The consistency improvement of CASR-F in various semantic roles, especially in AM class arguments, validates the strong semantic coherence and structural preservation ability of the model at the discourse level. Figure 3 shows the comparison of F1 values for different models in semantic role annotation tasks.

From Figure 3, it can be seen that the CASR-F model has a more significant improvement in consistency among the main semantic role categories, especially in the processing scenarios of cross sentence arguments and implicit predicates, where the error distribution is relatively concentrated. This phenomenon indicates that the model has strong semantic coherence and structural preservation ability at the discourse level.

C. LONG RANGE CONTEXT DEPENDENT MODELING CAPABILITY

In college English translation, the ability to maintain long-distance semantic dependency relationships directly affects the coherence of the translation. The CASR-F model enhances the perception of implicit semantics in the previous text through the synergistic effect of dynamic masking and gated semantic anchors. The experiment used a combination of referential resolution accuracy and attention weight distribution to examine its semantic memory and logical coherence ability within the discourse scope, in order to verify the role of gated attention mechanism in information transmission and referential resolution. Experimental data shows that in decoding step 8, the attention weights of CASR-F remain at 0.52, 0.66, and 0.72 for remote positions 5-7, significantly higher than the BART baseline of 0.3, 0.48, and 0.55 at the same position. This indicates that the joint training and anchor mechanism effectively expands the semantic memory capacity of the model and enhances its ability to model long-range semantic dependencies. Figure 4 compares the attention distribution characteristics of two types of models through heat maps.

In Figure 4, it can be seen that CASR-F is more concentrated at the discourse connections, indicating that the model maintains a high level of attention to inter sentence connections and semantic transitions. This phenomenon is consistent with the design goals of dynamic masking strategy and semantic anchor mechanism, indicating that the two have synergistically improved the cross semantic coherence modeling ability to a certain extent, thereby enhancing the overall coherence of the translation.

D. SEMANTIC RESTORATION CONSISTENCY ANALYSIS UNDER DIFFERENT TEXT TYPES

Different genres of discourse present diverse constraints on semantic modeling: expository texts emphasize logical progression relationships, argumentative texts focus on the coherence of argumentation chains, while narrative texts emphasize the continuity of event development and referential consistency. To verify the cross genre adaptability of the CASR-F model, this paper compared BLEU-4, METEOR, semantic role annotation F1 values, and referential resolution accuracy on three genre corpora, and combined the strength of cross sentence semantic constraints and attention distribution patterns to analyze the generalization performance of the model in depth. It should be noted that during the model training process, cross genre adaptability verification is achieved through implicit alignment of discourse structural features. In the encoding process, the dynamic masking mechanism dynamically adjusts the noise application strategy based on the dependency path. In the decoding stage, the gated semantic anchor adjusts its participation in the attention mechanism through a semantic memory function, ensuring that the evidence in the argumentative essay is logically coherent with the events in the narrative during the generation process. Figure 5 shows the comparison of various indicators between two models in three genres.

From Figure 5, it can be clearly observed that CASR-F outperforms the baseline model significantly in all genres. The model shows stable performance in explanatory and argumentative texts, with a particularly prominent improvement in narrative texts, verifying the enhanced role of the gate anchor mechanism in event coherence and reference tracking.

E. CONTRIBUTION ANALYSIS OF ABLATION EXPERIMENT TO CORE COMPONENTS

To evaluate the independent and synergistic effects of dynamic masking, semantic anchor, and joint training in the CASR-F framework, this paper designed ablation experiments. The experiment observes the degradation pattern of model performance and the trend of semantic restoration ability by sequentially removing or replacing each core module while keeping other parameters unchanged. The experiment selected BLEU-4, METEOR, semantic role annotation F1 value, and Coreference Resolution Accuracy (Coref-Acc) as core evaluation indicators to comprehensively measure the performance of the model from multiple dimensions such as vocabulary matching, syntactic structure, and semantic coherence. When using random masks instead of dependency syntax guided masking strategies, there is a significant decline in vocabulary matching metrics, which confirms the enhancing effect of masking mechanisms on semantic path representation. Removing the semantic anchor component significantly reduces the accuracy of referential resolution, indicating that this module has critical value in maintaining the continuity of entities and arguments within the text. After canceling the joint consistency training, the decrease in F1 value of semantic role annotation confirms the important contribution of sentence level contrastive constraints in maintaining semantic structural

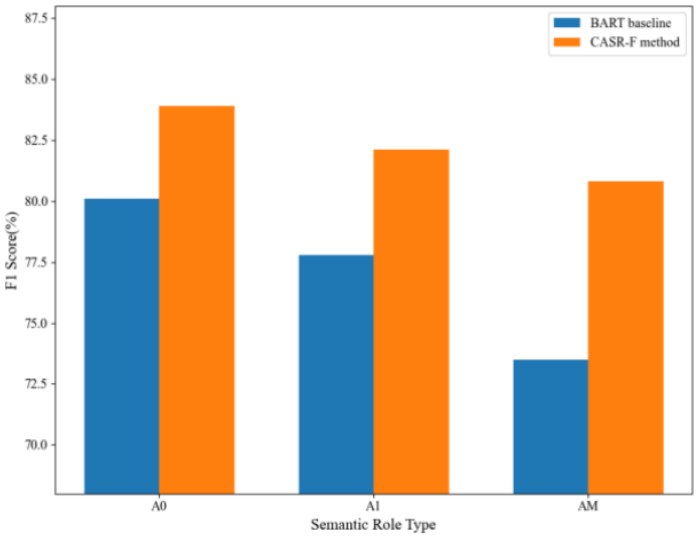

FIGURE 3. Comparison of F1 values of different models on semantic role types

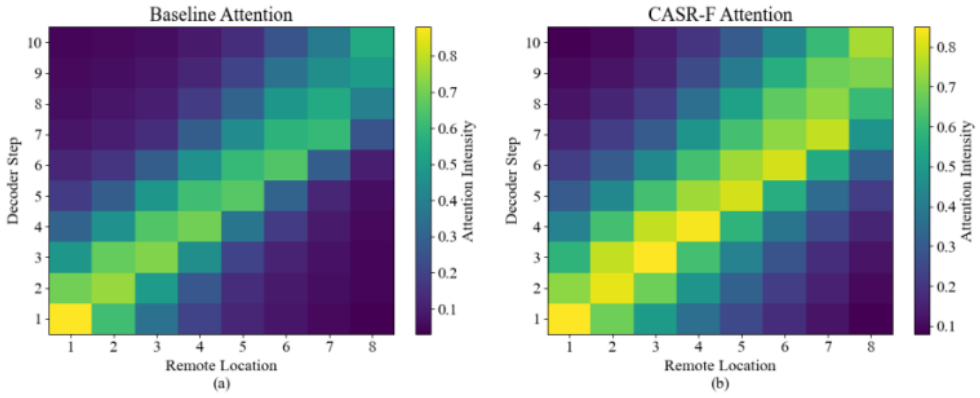

FIGURE 4. Heat distribution map of cross sentence attention: (a) BART baseline model; (b) CASR-F model

integrity. Compared to the complete model, when both dynamic masks and gate anchors are removed simultaneously, all indicators of the model show significant degradation: The BLEU-4 and METEOR scores decreased by 5.8 and 4.9 respectively, and the accuracy of referential resolution decreased by about 10.2 percentage points, verifying the complementary characteristics of the two modules in local vocabulary reconstruction and global semantic aggregation. For the convenience of unified display, the F1 value has been increased by 100 times. Figure 6 shows the performance comparison of various evaluation indicators under different ablation configurations.

The data in Figure 6 indicates that the complete model performs the best in all four indicators. Among them, removing both dynamic masks and semantic anchors has the most significant impact; In the case of removing a single module, removing the dynamic mask has the greatest impact on BLEU-4 and METEOR, removing the semantic anchor has the greatest impact on the accuracy of referential resolution, and canceling joint training destroys the overall balance of the model. This indicates that the three core components form a hierarchical complementary mechanism within the BART autoencoder framework, jointly promoting the deep semantic restoration capability of context.

V. CONCLUSION AND PROSPECT

On the basis of the BART architecture, this study introduces a dynamic noise mask guided by dependency syntax to enhance the encoder's ability to capture the core semantic path, and designs a gated semantic anchor attention mech-

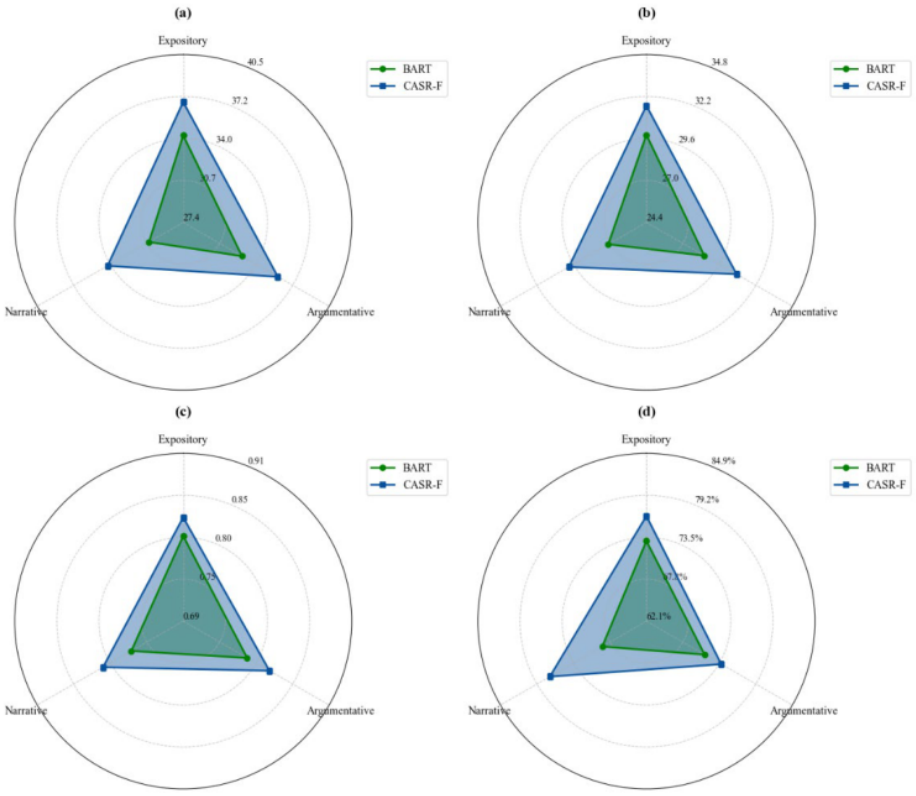

FIGURE 5. Performance Comparison of CASR-F and BART Models under Different Text Types: (a) BLEU-4 comparison; (b) METEOR comparison; (c) Comparison of SRL-F1;(d) refers to the comparison of digestion accuracy

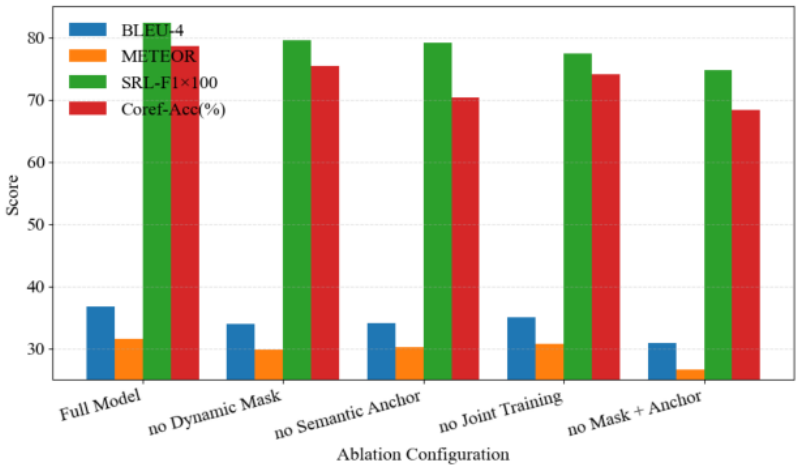

FIGURE 6. Impact of Core Component Ablation on Translation Quality and Semantic Consistency

anism to establish cross-sentence semantic associations, improving the model's robustness for handling the complex technical vocabulary and long-range semantic dependencies often found in engineering standards. By jointly optimizing sequence reconstruction and achieving cross sentence semantic consistency goals, the synchronous improvement of vocabulary accuracy and discourse coherence has been achieved. The experimental results show that the CASR-F model performs better than the BART baseline in college English translation tasks: BLEU-4 reached 36.72, METEOR

was 31.45, and the accuracy of referential resolution was improved to 78.6%. The F1 value exceeded 80% on both the core and additional arguments of semantic annotation. These results demonstrate that the proposed method improves translation fluency, semantic structure recovery, and long-distance dependency modeling performance in college English translation scenarios, providing effective validation for text-level machine translation research. This study offers a methodological reference for translation quality assessment and automatic writing tasks in educational settings, but its application value in professional and technical text processing still needs further testing in more targeted task settings.

AUTHOR CONTRIBUTIONS

Pandeng Li designed, collected and analyzed the data, and drafted the manuscript. Pandeng Li conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Pandeng Li participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This work was supported by the Henan Province Philosophy and Social Science Planning Project entitled "Study on Contemporary Classical Translators' Cultural Orientation" (Grant No. 2022BYY003). 2026 Higher Education Teaching Reform Research and Practice Project of Henan University of Urban Construction(Grant No.2026JG012)

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] B. Klimova, M. Pikhart, A. D. Benites, C. Lehr, and C. Sanchez-Stockhammer, "Neural machine translation in foreign language teaching and learning: A systematic review," Education and Information Technologies, vol. 28, no. 1, pp. 663-682, 2023, doi: 10.1007/s10639-022-11194-2.

[2] J. Son and B. Kim, "Translation performance from the user's perspective of large language models and neural machine translation systems," Information, vol. 14, no. 10, p. 574, 2023, doi: 10.3390/info14100574.

[3] Y. Xiao, L. Wu, J. Guo, J. Li, M. Zhang, and T. Qin, "A survey on non-autoregressive generation for neural machine translation and beyond," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 10, pp. 11407-11427, 2023, doi: 10.1109/TPAMI.2023.3277122.

[4] X. Lin, M. Afzaal, and H. S. Aldayel, "Syntactic complexity in legal translated texts and the use of plain English: a corpus-based study," Humanities and Social Sciences Communications, vol. 10, no. 1, pp. 1-9, 2023, doi: 10.1057/s41599-022-01485-x.

[5] N. M. Guerreiro, R. Rei, D. van Stigt, L. Coheur, P. Colombo, and A. F. T. Martins, "Xcomet: Transparent machine translation evaluation through fine-grained error detection," Transactions of the Association for Computational Linguistics, vol. 12, pp. 979-995, 2024, doi: 10.1162/tacl_a_00683.

[6] Y. Li, J. Li, J. Jiang, S. Tao, H. Yang, and M. Zhang, "P-transformer: Towards better document-to-document neural machine translation," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 3859-3870, 2023, doi: 10.1109/TASLP.2023.3313445.

[7] R. He, C. Palominos, H. Zhang, M. F. Alonso-Sanchez, L. Palaniyappan, and W. Hinzen, "Navigating the semantic space: Unraveling the structure of meaning in psychosis using different computational language models," Psychiatry Research, vol. 333, p. 115752, 2024, doi: 10.1016/j.psychres.2024.115752.

[8] Y. A. Mohamed, A. Khanan, M. Bashir, A. H. H. M. Mohamed, M. A. E. Adiel, and M. Elsadig, "The impact of artificial intelligence on language translation: A review," IEEE Access, vol. 12, pp. 25553-25579, 2024, doi: 10.1109/ACCESS.2024.3366802.

[9] W. Xu, S. Agrawal, E. Briakou, M. J. Martindale, and M. Carpuat, "Understanding and detecting hallucinations in neural machine translation via model introspection," Transactions of the Association for Computational Linguistics, vol. 11, pp. 546-564, 2023, doi: 10.1162/tacl_a_00563.

[10] Y. Wang, "Artificial intelligence technologies in college English translation teaching," Journal of Psycholinguistic Research, vol. 52, no. 5, pp. 1525-1544, 2023, doi: 10.1007/s10936-023-09960-5.

[11] X. Zhao, Y. Deng, M. Yang, L. Wang, R. Zhang, H. Cheng, et al., "A comprehensive survey on relation extraction: Recent advances and new frontiers," ACM Computing Surveys, vol. 56, no. 11, pp. 1-39, 2024, doi: 10.1145/3674501.

[12] X. Liu, J. He, M. Liu, Z. Yin, L. Yin, and W. Zheng, "A scenario-generic neural machine translation data augmentation method," Electronics, vol. 12, no. 10, p. 2320, 2023, doi: 10.3390/electronics12102320.

[13] S. Chauhan and P. Daniel, "A comprehensive survey on various fully automatic machine translation evaluation metrics," Neural Processing Letters, vol. 55, no. 9, pp. 12663-12717, 2023, doi: 10.1007/s11063-022-10835-4.

[14] T. K. Lee, "Artificial intelligence and posthumanist translation: ChatGPT versus the translator," Applied Linguistics Review, vol. 15, no. 6, pp. 2351-2372, 2024, doi: 10.1515/applirev-2023-0122.

[15] M. H. A. Abdullah, N. Aziz, S. J. Abdulkadir, H. S. S. Alhassan Alhussian, and N. Talpur, "Systematic literature review of information extraction from textual data: Recent methods, applications, trends, and challenges," IEEE Access, vol. 11, pp. 10535-10562, 2023, doi: 10.1109/ACCESS.2023.3240898.

[16] H. Wang, X. Li, Z. Ren, M. Wang, and C. Ma, "Multimodal sentiment analysis representations learning via contrastive learning with condense attention fusion," Sensors, vol. 23, no. 5, p. 2679, 2023, doi: 10.3390/s23052679.

[17] S. Castilho and R. Knowles, "A survey of context in neural machine translation and its evaluation," Natural Language Processing, vol. 31, no. 4, pp. 986-1016, 2025, doi: 10.1017/nlp.2024.7.

[18] L. Kang, S. He, M. Wang, F. Long, and J. Su, "Bilingual attention based neural machine translation," Applied Intelligence, vol. 53, no. 4, pp. 4302-4315, 2023, doi: 10.1007/s10489-022-03563-8.

[19] S. Sharma, M. Diwakar, P. Singh, V. Singh, S. Kadry, and J. Kim, "Machine translation systems based on classical-statistical-deep-learning approaches," Electronics, vol. 12, no. 7, pp. 1716-1722, 2023, doi: 10.3390/electronics12071716.

[20] H. Mercan, Y. Akgun, and M. C. Odacioglu, "The evolution of machine translation: A review study," International Journal of Language and Translation Studies, vol. 4, no. 1, pp. 104-116, 2024. Available from: https://izlik.org/JA28TD85GN.

[21] H. Hu, X. Wang, Y. Zhang, Q. Chen, and Q. Guan, "A comprehensive survey on contrastive learning," Neurocomputing, vol. 610, pp. 128645-128652, 2024, doi: 10.1016/j.neucom.2024.128645.

[22] Y. Hao, T. Zhang, P. Zhao, Y. Liu, V. S. Sheng, and J. Xu, "Feature-level deeper self-attention network with contrastive learning for sequential recommendation," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 10, pp. 10112-10124, 2023, doi: 10.1109/TKDE.2023.3250463.

[23] Y. Zhao, J. Zhang, and C. Zong, "Transformer: A general framework from machine translation to others," Machine Intelligence Research, vol. 20, no. 4, pp. 514-538, 2023, doi: 10.1007/s11633-022-1393-5.

[24] R. Patil, S. Boit, V. Gudivada, and J. Nandigam, "A survey of text representation and embedding techniques in nlp," IEEE Access, vol. 11, pp. 36120-36146, 2023, doi: 10.1109/ACCESS.2023.3266377.

[25] L. Yu, Y. Cheng, Z. Wang, V. Kumar, W. Macherey, and Y. Huang, "Spae: Semantic pyramid autoencoder for multimodal generation with frozen llms," Advances in Neural Information Processing Systems, vol. 36, pp. 52692-52704, 2023, doi: 10.48550/arXiv.2306.17842.

[26] L. Zhong, J. Wu, Q. Li, H. Peng, and X. Wu, "A comprehensive survey on automatic knowledge graph construction," ACM Computing Surveys, vol. 56, no. 4, pp. 1-62, 2023, doi: 10.1145/3618295.
[27] Z. Lu, R. Li, K. Lu, X. Chen, E. Hossain, and Z. Zhao, "Semantics-empowered communications: A tutorial-cum-survey," IEEE Communications Surveys & Tutorials, vol. 26, no. 1, pp. 41-79, 2023, doi: 10.1109/COMST.2023.3333342.
[28] M. Nagahisarchoghaei, N. Nur, L. Cummins, N. Nur, M. M. Karimi, and S. Nandanwar, "An empirical survey on explainable ai technologies: Recent trends, use-cases, and categories from technical and application perspectives," Electronics, vol. 12, no. 5, pp. 1092-1098, 2023, doi: 10.3390/electronics12051092.
[29] L. Xu, H. Xie, Z. Li, F. L. Wang, W. Wang, and Q. Li, "Contrastive learning models for sentence representations," ACM Transactions on Intelligent Systems and Technology, vol. 14, no. 4, pp. 1-34, 2023, doi: 10.1145/3593590.
[30] M. Apidianaki, "From word types to tokens and back: A survey of approaches to word meaning representation and interpretation," Computational Linguistics, vol. 49, no. 2, pp. 465-523, 2023, doi: 10.1162/coli_a_00474.
[31] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent neural networks: A comprehensive review of architectures, variants, and applications," Information, vol. 15, no. 9, pp. 517-524, 2024, doi: 10.3390/info15090517.
[32] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," Artificial Intelligence Review, vol. 57, no. 4, pp. 99-105, 2024, doi: 10.1007/s10462-024-10721-6.
[33] Z. Zhang, K. Chen, R. Wang, M. Utiyama, E. Sumita, and Z. Li, "Universal multimodal representation for language understanding," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 7, pp. 9169-9185, 2023, doi: 10.1109/TPAMI.2023.3234170.
[34] J. Li, T. Tang, W. X. Zhao, J. Y. Nie, and J. R. Wen, "Pre-trained language models for text generation: A survey," ACM Computing Surveys, vol. 56, no. 9, pp. 1-39, 2024, doi: 10.1145/3649449.
[35] C. Zhou, C. Qiu, L. Liang, and D. E. Acuna, "Paraphrase identification with deep learning: A review of datasets and methods," IEEE Access, vol. 13, pp. 65797-65822, 2025, doi: 10.1109/ACCESS.2025.3556899.