Improving English Interpretation Style Transfer Capabilities by Self-Supervised Semantic Enhancement Strategies

Ling Gu
Faculty of Humanity and Law, Gannan College of Science and Technology, Ganzhou 341000, Jiangxi, China
Corresponding author: Ling Gu (e-mail: alexgu2005@163.com)

ABSTRACT To enhance stylistic adaptability in English interpretation systems, this paper addresses stylistic inauthenticity, which often appears as rigid or pragmatically inappropriate output when processing domain-specific terminology and technical standards. The objective is to achieve semantically faithful and style-controllable generation without explicit style annotations. A self-supervised semantic enhancement strategy is proposed. Six fine-grained interpretation styles are first defined: Formal-Diplomatic, Academic-Expository, Business-Negotiation, Media-Interview, Casual-Conversational, and Legal-Procedural. Semantically preserved style variants are generated using the T5-large model, and a dual-filtering mechanism is applied to construct high-quality augmented data. Multi-view contrastive learning is then performed using the SimCLR framework. DeBERTa-v3-base is used to extract style-invariant semantic representations. Semantic-style decoupling is achieved by integrating gradient reversal layers with learnable style prototypes. Finally, multi-task joint training optimizes semantic fidelity and stylistic control. Experimental results show that the model outperforms baselines in semantic fidelity, with BLEU-4 of 33.9, chrF++ of 59.8, and COMET-22 of 0.714. It achieves style classification accuracy of 82.6% and a mutual information score of 0.183, validating effective decoupling and high-fidelity multi-style interpretation generation.

INDEX TERMS English interpretation, style transfer, semantic enhancement, contrastive learning, semantic fidelity

I. INTRODUCTION

WHILE neural interpretation systems have advanced semantic accuracy, their output frequently suffers from translation tone, which manifests as stiff language lacking the authenticity and pragmatic adaptability required for effective communication in specialized contexts [1,2]. The industry is selected as the research focus because it is highly representative: the documents and exchanges in this field cover technical descriptions, business negotiations, market reports, and other discourse types at the same time, posing typical challenges to style diversity, terminology consistency and pragmatic appropriateness.This defect is particularly prominent in high-demand interpretation scenarios [3,4], where diplomatic statements require tact and formality, academic speeches emphasize logical rigor, and business negotiations demand flexibility. Language style [5,6], as a core dimension of interpretation quality, directly affects the effect of information reception and the realization of communicative intentions. However, existing systems focus more on content fidelity and neglect dynamic style transfer, resulting in translations that are correct but pragmatically inappropriate [7]. Especially under the realistic conditions without explicit style annotation, how to achieve semantically unchanged and style-controllable English interpretation generation has become a key bottleneck for the implementation of intelligent interpretation technology. According to the actual needs of professional interpretation, through the analysis of the United Nations conference minutes, academic speeches, business negotiation records, media interviews, daily dialogues and legal texts, and combined with the opinions of industry experts, this study identified six most representative and differentiated interpretation styles: formal diplomacy, academic explanation, business negotiation, media interview, casual conversation, and legal procedure. These styles cover the continuum from formal writing to oral instant, from professional expression to daily communication, and have significant differences in pragmatic functions, syntactic features and vocabulary selection.

In recent years, neural interpretation models have continued to improve semantic fidelity on benchmarks such as International Workshop on Spoken Language Translation (IWSLT) [8,9] and Multilingual Speech Translation Corpus (Must-C) [10]. However, their output style is highly

dependent on the distribution of training data and is difficult to generalize to new contexts. To implement style control [11], some studies have adopted text style transfer methods, such as StyleTransformer, which guides generation through explicit style embeddings. However, these methods rely heavily on large-scale parallel data with style labels, while high-quality multi-style interpretation corpora are extremely scarce [12]. Another type of unsupervised method attempts to separate style and content through adversarial training or representation decoupling, but in highly compressed and context-dependent interpretation languages, semantic-style decoupling can easily lead to information loss or style drift [13]. Furthermore, existing work reduces stylistic variation to a formal–informal dichotomy, overlooking the multidimensionality and fine-grained differences—such as technical versus marketing or artisanal versus industrial—that are essential for accurate interpretation in industry contexts.

Self-supervised learning provides a new path to alleviate the dependency on annotations. Frameworks such as SimCSE and ConSERT learn semantically sensitive sentence representations on monolingual corpora through contrastive learning, and have been proven effective in semantic similarity tasks [14,15]. Recently, some studies have attempted to apply it to machine translation, such as SimCSE [16,17], which improves the semantic consistency of translations through sentence-level contrast. However, these methods do not explicitly model style diversity, and their enhancement strategies are difficult to generate pragmatically reasonable style variants. Interpretation languages have strong colloquialism, information compression, and context-dependent characteristics, and general text enhancement cannot capture their style evolution laws [18,19]. Although some studies have explored controllable text generation, they do not impose sufficient constraints to preserve semantic fidelity and often introduce semantic drift during style transfer [20]. In addition, existing contrastive learning [21,22] focuses more on the discriminability of the representation space and ignores the decoupling mechanism of style and semantics, which makes it difficult for the model to maintain content consistency when transferring style. In summary, previous studies have not yet established an unsupervised, multi-style semantic enhancement and transfer framework for interpretation scenarios.

To address the above limitations, this study proposes a self-supervised semantic enhancement strategy, aiming to achieve high-fidelity, multi-style English interpretation generation without explicit style annotation. Firstly, this paper defines six categories of fine-grained interpretation styles and constructs style-aware enhancement data based on public corpora, moving beyond the binary simplification of style. Secondly, a multi-view contrastive learning mechanism based on Simple Framework for Contrastive Learning of Representations (SimCLR) is designed to construct positive sample pairs through semantically preserved stylized rewriting, and learn invariant representations that are robust to surface form and sensitive to semantics. Furthermore, style prototypes and gradient reversal layers are applied to achieve deep decoupling of semantics and style, supporting flexible style control during inference. Finally, a multi-style manual evaluation set for interpretation is constructed, covering professional translators' retranslations and context-screened samples, providing a reliable benchmark for style transfer research. The self-supervised semantic enhancement strategy proposed in this paper achieves improvements in semantic fidelity (Bilingual Evaluation Understudy (BLEU-4) score of 33.9) and style classification accuracy (82.6%) without explicit style annotation. This addresses translation tone in interpretation and enhances the adaptability and controllability of intelligent interpretation systems. This work not only addresses the feasibility of unsupervised interpretation style transfer but also advances the advancement of self-supervised learning from general semantic modeling to pragmatic style perception, providing a new paradigm for the naturalness and adaptability of intelligent interpretation systems.

The core contribution of this study is to propose a unified self-supervised semantic enhancement and decoupling framework for unsupervised multi-style interpretation generation. The framework does not simply apply the existing technology, but carries out systematic integration and innovation at the architecture level: (1) a style-aware data enhancement strategy is designed to generate high-quality, fine-grained, semantic preserving style variants through structured prompt control and dual filtering, which provides a reliable basis for unsupervised comparative learning; (2) A multi-view contrastive learning mechanism for interpreting text is proposed, which adapts the SimCLR framework to the construction of positive samples with diverse styles but semantic equivalence, and is specially used to learn the sentence representation that is robust to surface style changes and sensitive to deep semantics; (3) A semantic style decoupling and migration mechanism based on learnable style prototypes and a gradient inversion layer is constructed, which realizes the style invariance of semantic coding and the flexibility and controllability of style injection, and supports zero-shot style migration. These components are optimized under the unified multi-task collaborative training goal, which jointly solves the contradiction that semantic fidelity and style control are difficult to achieve simultaneously without supervision in previous work, and provides a new methodological paradigm for the adaptive generation of style in professional interpretation.

II. METHODS

A. OVERALL ARCHITECTURE OVERVIEW

To achieve unsupervised multi-style English interpretation generation, this study constructs an end-to-end semantic-style decoupling framework. Its core is to obtain style-invariant semantic representations through Self-Supervised Semantic Enhancement with Contrastive Learning (SSSECL), and to achieve controllable transfer by combining

learnable style prototypes. The overall architecture is shown in Figure 1.

Figure 1 shows the proposed multi-style English interpretation architecture, which primarily consists of a source language encoder, a semantic-style decoupling module, a style prototype library, a style-conditional decoder, and a self-supervised contrastive learning branch.

The selected six styles have high frequency and strong discrimination in typical scenes such as industry reports, exhibition exchanges, contract negotiations, and media releases. Their definitions are based on actual corpus analysis and verification by experts in the field to ensure the mutual exclusion of style categories in pragmatics and the adequacy of task coverage.

The source language encoder, using a six-layer Transformer architecture, encodes the input Chinese source sentence into a contextual representation, providing source language information for subsequent decoding. The semantic-style decoupling module is a core component, consisting of a semantic encoder and a style encoder. The semantic encoder is built upon a frozen DeBERTa-v3-base model (Decoding-enhanced Bidirectional Encoder Representations from Transformers with Disentangled Attention) and extracts style-invariant semantic vectors from the English interpretation. The style encoder, using a lightweight Convolutional Neural Network (CNN) and an attention mechanism, outputs a vector representing a specific language style.

The style prototype library consists of six learnable vectors, each corresponding to six predefined interpretation styles. During inference, the user specifies a target style index, and the system retrieves the corresponding style prototype vector.

The style-conditional decoder also uses a six-layer Transformer architecture. Its initial hidden state is obtained by concatenating the semantic vector and the target style vector and performing linear projection, thereby simultaneously incorporating semantic content and style information into the generation process. A self-supervised contrastive learning branch serves as an auxiliary training module. By contrastively learning different style variants of the same semantics (based on the SimCLR framework), it forces the semantic encoder to learn content representations that are robust to surface style variations, further enhancing the system's semantic fidelity.

The language style features are shown in Table 1.

B. SELF-SUPERVISED SEMANTIC ENHANCEMENT CONTRASTIVE LEARNING

1) Style-Aware Data Augmentation Strategy

To construct semantically consistent but stylistically diverse positive sample pairs, this study proposes a style-aware data augmentation strategy. Its core is to derive semantically equivalent variants covering six interpretation styles from a single sentence through controllable text generation and strict quality filtering.

This strategy consists of three steps: style-controlled generation, dual quality filtering, and positive sample pair construction.

In order to preliminarily evaluate the basic quality of T5-large-generated style variants, this paper conducted manual verification before automatic filtering. 200 basic sentences were randomly selected from the training corpus, and each basic sentence generated 5 variants of other styles, a total of 1000 generated samples. Employ two research assistants with linguistic background to make independent annotation according to the following standards: (1) semantic retention: use a three-level scale (3 = completely retain the core semantics and facts, 2 = basically retain but slightly increase or decrease or reorganize the information, 1 = significant semantic change or distortion); (2) Style accuracy: judge whether the generated sentence conforms to the typical pragmatic and syntactic characteristics of the target style (yes/no). Krippendorff's α was 0.71 for semantic retention and 0.69 for style accuracy, indicating acceptable annotation consistency.

Given a base English interpretation sentence $y \in Y$, a style-conditioned generative model is employed to produce semantically preserved variants in five other styles. We adopt a fine-tuned T5-large model (Text-to-Text Transfer Transformer) for generation.

The input follows a structured prompt format:

"[TARGET_STYLE: s_k] Rewrite the following sentence while preserving its meaning: y", where

$$s_k \in S = \{\text{Formal-Diplomatic}, \text{Academic-Expository}, \ldots, \text{Legal-Procedural}\}$$

represents one of the six predefined styles, and $s_k \neq s_{\text{orig}}(y)$.

The model outputs a stylized rewriting denoted as $y^{(k)}$. Five such variants are generated, forming the augmentation set $\widetilde{Y}_y^+ = \{y^{(k)} | k \neq orig\}$. Semantic fidelity constraint: the semantic similarity between the original sentence y and the variant $y^{(k)}$ is calculated:

$$Sim_{sem}(y, y^{(k)}) = \frac{e(y)^\top e(y^{(k)})}{\|e(y)\| \cdot \|e(y^{(k)})\|} \tag{1}$$

In Formula (1), $e(\cdot)$ denotes the sentence vector produced by the Sentence-BERT encoder. Only samples that satisfy $Sim_{sem}(y, y^{(k)}) \geq 0.82$ are retained to ensure high semantic consistency.

To ensure the semantic fidelity of the style variants generated by T5-large model, this study adopts a multi-level semantic control strategy. First, during the generation phase, structured prompts are used in the following format: TARGET_STYLE: [style] Rewrite the following sentence while preserving its meaning: "[sentence]". This instruction emphasizes preserving the original meaning and guides the model toward interpretation rather than unrestricted generation. At the same time, the decoding temperature parameter is set to 0.3 to reduce the randomness in the generation

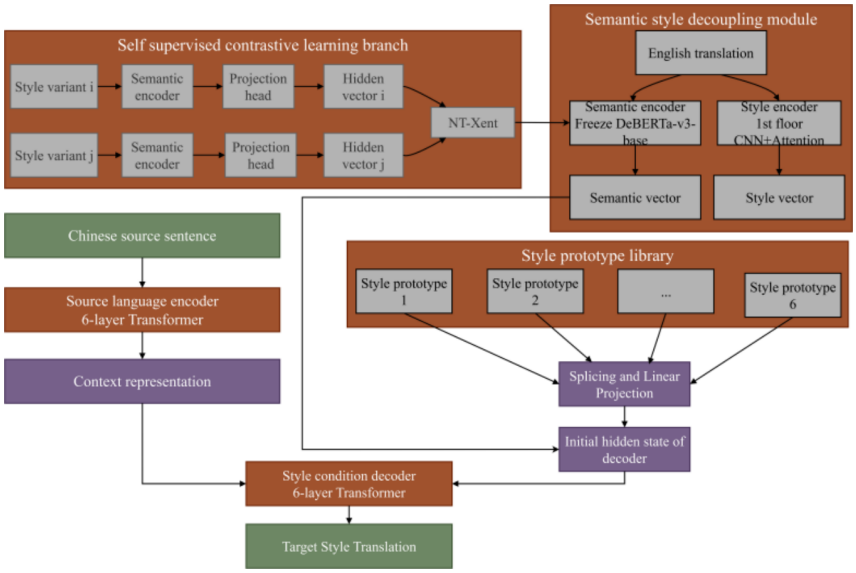

FIGURE 1. Overall architecture of this paper

TABLE 1. Language style features

Style type	Pragmatic function	Syntax features
Formal-Diplomatic	Maintain national stance and avoid direct conflict	Passive voice, nominalization structure
Academic-Expository	Elaborate on theories and logical arguments	Long compound sentence, conditional clause
Business-Negotiation	Expressing intention and leaving room for maneuver	Conditional sentences and modal verbs
Media-Interview	Attract attention and guide public opinion	Short sentences, abbreviated sentences
Casual-Conversational	Establish friendly and informal communication	Free word order and rhetorical questions
Legal-Procedural	Clarify responsibilities and ensure compliance with procedures	Programmatic structure, long modifiers

process and encourage rewriting with higher certainty. Secondly, this paper recognizes the inherent risk of semantic deviation in the generation model, so it does not directly assume the semantic fidelity of its output, but regards it as a pool of candidate variants, and then strictly filters it through a dual filtering mechanism: semantic similarity filtering ($Sim_{sem} \geq 0.82$) is based on the deep semantic coding of Sentence-BERT, which removes samples of semantic deviation from the vector level; Style purity filtering ensures that variant styles meet expectations and avoid semantic distortion caused by style confusion. This design ensures that the enhanced data used to construct positive sample pairs has high semantic consistency, and provides a reliable basis for the follow-up comparative learning.

The specific criteria and indicators of the dual filtering mechanism are set as follows: the cosine similarity calculated based on Sentence-BERT is used as the indicator for semantic fidelity filtering. The threshold is strictly set to 0.82, which is determined based on ablation experiments on reserved data: below this threshold, the proportion of semantic deviation found by manual evaluation increases significantly. The style-purity filter employs a BERT-base classifier fine-tuned on six categories of style-annotation data and uses the softmax probability of the corresponding target category as the confidence score. The threshold value is set to 0.90 to ensure that variants are clearly classified as target styles and to exclude samples with vague or mixed styles.

Style purity constraint: the confidence that the variant belongs to the target style s_k is calculated using the 6-way style classifier C_{style}:

$$p_k = C_{style}(y^{(k)})[k] \tag{2}$$

The sample is retained only when $p_k \geq 0.90$ to exclude instances of style mixing or generation failure.

For each base sentence, a high-quality multi-style variant subset $\widetilde{Y}_y^+ \subseteq Y_y^+$ can be obtained. Finally, positive sample pairs are constructed by any pairwise combination in $\{y\} \cup \widetilde{Y}_y^+$:

$$P_y = \left\{ (y^{(i)}, y^{(j)}) \mid y^{(i)}, y^{(j)} \in \{y\} \cup \widetilde{Y}_y^+, i \neq j \right\} \tag{3}$$

2) Contrastive Learning Based on SimCLR

To learn sentence representations that are robust to language style and sensitive to semantics, this study adopts SimCLR [23,24] as the basic architecture for self-supervised contrastive learning and adapts it to the characteristics of interpreted text. Given a batch of N base English interpretation sentences, $2N$ enhanced views are constructed. These views are fed into a shared-weight semantic encoder to extract representations. The semantic encoder uses DeBERTa-v3-base [25,26]. For the input sentence, the encoder outputs a

token-level representation, which is then average-pooled to obtain a sentence-level semantic vector:

$$h = MeanPool(DeBERTa(y)) \in \mathbb{R}^{768} \tag{4}$$

In Formula (4), the semantic vector h is regarded as a style-invariant semantic core representation. To improve the discriminative ability of contrastive learning, a nonlinear projection head $g(\cdot)$ is applied to map the semantic vector to a space dedicated to contrastive learning:

$$g(h) = MLP(h) = W_2 \cdot \sigma(Dropout(W_1 h + b_1)) + b_2 \in \mathbb{R}^{256} \tag{5}$$

The projection head participates only in the loss calculation during pre-training and is discarded during fine-tuning. The final semantic representation still uses h. The contrast loss adopts the Normalized Temperature-scaled Cross-Entropy Loss (NT-Xent) [27,28], and the formula is:

$$L_{cont} = -\frac{1}{|P|} \sum_{(i,j)\in P} \log \frac{\exp(sim(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \exp(sim(z_i, z_k)/\tau)} \tag{6}$$

In Formula (6), P is the set of all positive sample pairs in the batch; $z_i = g(h_i)$ is the projection vector of the i-th view; $\tau = 0.05$ is the temperature hyperparameter, which controls the sharpness of the distribution. A small value enhances the discrimination between positive and negative samples.

C. SEMANTIC-STYLE DECOUPLING AND TRANSFER MECHANISM

1) Style-Invariant Representation Learning Based on Gradient Reversal

To achieve the decoupled generation of semantic content and language style, this paper designs a dual-path decoupling architecture, which models style-invariant semantic representation and pluggable style representation, respectively, and strengthens the decoupling effect through gradient reversal training. The semantic representation is extracted by a frozen semantic encoder E_{sem} (pre-trained DeBERTa-v3-base). Given an English reference translation y, its semantic vector is defined as Formula (7):

$$h_{sem} = MeanPool(E_{sem}(y)) \in \mathbb{R}^{768} \tag{7}$$

The parameters of this vector are frozen during the fine-tuning phase to ensure that it only encodes deep semantic information that is independent of style.

In the style training phase, the style vector is extracted from the reference translation through a lightweight style encoder E_{style} (1-layer CNN + multi-head attention):

$$s = E_{style}(y) \in \mathbb{R}^{256} \tag{8}$$

To prevent the semantic encoder E_{sem} from leaking style information, a gradient reversal layer (GRL) [29,30] is applied:

$$L_{GRL} = -\lambda_{grl} E_y \left[\log P_\phi(\hat{k} = k \mid h_{sem})\right] \tag{9}$$

In Formula (9), P_ϕ is the auxiliary style classifier (parameter ϕ), and k is the true style label of y. λ_{grl} is the adversarial weight. GRL is identical during forward propagation and multiplies the gradient by $-\lambda_{grl}$ during backpropagation, forcing h_{sem} to be indistinguishable from the style.

2) Learnable Style Prototype

The inference stage does not rely on the reference translation, but directly uses the learnable style prototype vector s_k, each s_k corresponding to a predefined interpretation style. These prototypes are used as model parameters and are jointly optimized during training through the style classification loss:

$$L_{style\text{-}cls} = E_y \left[CrossEntropy(k, Classifier(E_{style}(y)))\right] \tag{10}$$

The style conditional decoder is a 6-layer Transformer [31,32], whose initial hidden state h_0 is generated by linear projection after splicing the semantic and style vectors:

$$h_0 = W \begin{bmatrix} h_{sem} \\ s_k \end{bmatrix} + b, \quad W \in \mathbb{R}^{d\times(768+256)}, \quad b \in \mathbb{R}^d \tag{11}$$

In Formula (11), d is the decoder hidden dimension. This design ensures that the generation process is both constrained by semantic content and guided by the target style.

The style control is explicitly specified by the user during inference. The target style index $k \in \{1, \cdots, 6\}$, and the system automatically injects the corresponding prototype s_k, achieving zero-shot style transfer without retraining or fine-tuning. This mechanism supports fine-grained and interpretable dynamic adaptation of language style while maintaining semantic fidelity.

D. MULTI-TASK JOINT TRAINING AND INFERENCE

1) Training Objective Function

To collaboratively optimize semantic fidelity, style controllability, and semantic-style decoupling, this study adopts a multi-task joint training strategy. The total loss function is defined as the weighted sum of four sublosses:

$$L = L_{NLL} + \lambda_1 L_{cont} + \lambda_2 L_{style\text{-}cls} + \lambda_3 L_{GRL} \tag{12}$$

In Formula (12), L_{NLL} is the translation loss using the standard negative log-likelihood loss to supervise the sequence generation from the source language to the target language.

$$L_{NLL} = -\sum_{t=1}^{T} \log P(y_t \mid y_{<t}, h_{sem}, s_k, H_x; \theta) \tag{13}$$

In Formula (13), y_t is the reference English translation token; H_x is the source language encoder output; θ is a trainable parameter; h_{sem} does not participate in the gradient update due to freezing.

The weight configuration is determined by grid search, and the loss weights are shown in Figure 2.

The loss weights satisfy $\lambda_1+\lambda_2+\lambda_3=1$. Grid search results indicate that setting $\lambda_1=0.5$, $\lambda_2=0.3$, and $\lambda_3=0.2$ achieves the optimal balance between semantic fidelity, style controllability, and decoupling strength. A high λ_1 ensures that contrastive learning effectively shapes semantically invariant representations; a moderate λ_2 strengthens style discrimination; a moderate λ_3 prevents semantic information loss caused by adversarial training, overall improving the model's overall performance in unsupervised style transfer.

2) Style Control at the Inference Stage

The style vector visualization is shown in Figure 3.

Figure 3 shows the t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization of the style vectors. The six style categories form clear, separate clusters. Within-category samples are closely clustered, indicating high consistency in expression within the same style. The inter-category distance is significantly greater than the intra-category distance, reflecting that different pragmatic functions are effectively distinguished in the vector space. This result validates the effectiveness of the collaborative learning between the style prototype library and the encoder, providing an interpretable representation foundation for subsequent precise style control.

III. EXPERIMENTAL SETUP

A. DATASET CONSTRUCTION AND PREPROCESSING

This phase aims to construct a large-scale, high-quality monolingual English corpus for style-aware data augmentation in self-supervised contrastive learning. The corpus is organized according to six predefined interpretation styles to ensure the typicality of the stylistic register and the discriminability of linguistic features. The specific data sources and scale are shown in Table 2.

100 Chinese source sentences are randomly selected from the IWSLT test set. The sample covers four thematic areas: science and technology (40%), society (30%), culture (20%), and politics (10%), reflecting the diversity of real-world interpretation scenarios. Sentence length is strictly controlled to 10–25 words, consistent with the typical span of information units in professional interpretation.

Multi-style reference translation generation:

Formal-Diplomatic and Legal-Procedural: two individuals with United Nations language service qualifications are hired to stylize all 100 source sentences according to the standards of their respective professional domains. This results in 200 high-quality reference translations, ensuring the formality, precision, and formality required for diplomatic documents or legal proceedings.

Academic-Expository, Media-Interview, Business-Negotiation, and Casual-Conversational styles were manually screened from the original IWSLT and MuST-C test sets based on speaker identity, context, and discourse function:

Academic: presentations delivered by scientists or professors (e.g., on topics such as AI and biomedicine);

Media: question and answering (Q&A) clips from press conferences;

Business: presentations by entrepreneurs or investors at business roadshows;

Casual: natural conversations in roundtable discussions or informal interviews.

All 600 reference translations are reviewed by a native-speaker reviewer (with a Master's degree in Translation and at least five years of professional editing experience) for grammatical correctness, pragmatic appropriateness, and stylistic consistency. Examples exhibiting mixed styles, ambiguous semantics, or cultural incompatibility are removed, ultimately retaining 600 sentences (100 from each style category) to form a balanced, high-quality evaluation benchmark. This study was approved by the Ethics Committee of Gannan College of Science and Technology, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

Test set statistics are shown in Table 3.

B. BASELINE MODELS

To fully validate the effectiveness of the proposed method in unsupervised multi-style English interpretation generation, it is compared with representative baseline models. All baselines are replicated under the same data conditions to ensure fair comparison. The models used for the comparison are shown in Table 4.

TABLE 4. Comparison models

Model	Type	Need a style label?
Transformer	Standard interpretation	No
mBART (Multilingual Bidirectional and Auto-Regressive Transformers)	Pre-trained interpretation	No
StyleTransformer	Explicit style	Yes
ByT5 (Byte-level T5)	Byte-level pre-training	No
T5	Text-to-text pre-training	No
SimCSE	Contrastive learning	No
This paper model	Self-supervised style transfer	No

C. EVALUATION METRICS

Semantic fidelity measures the semantic consistency between the translation and the reference translation. The evaluation metrics used include BLEU-4 [33,34], chrF++, Translation Edit Rate (TER), Crosslingual Optimized Metric for Evaluation of Translation (COMET-22) [35], and BERTScore (F1).

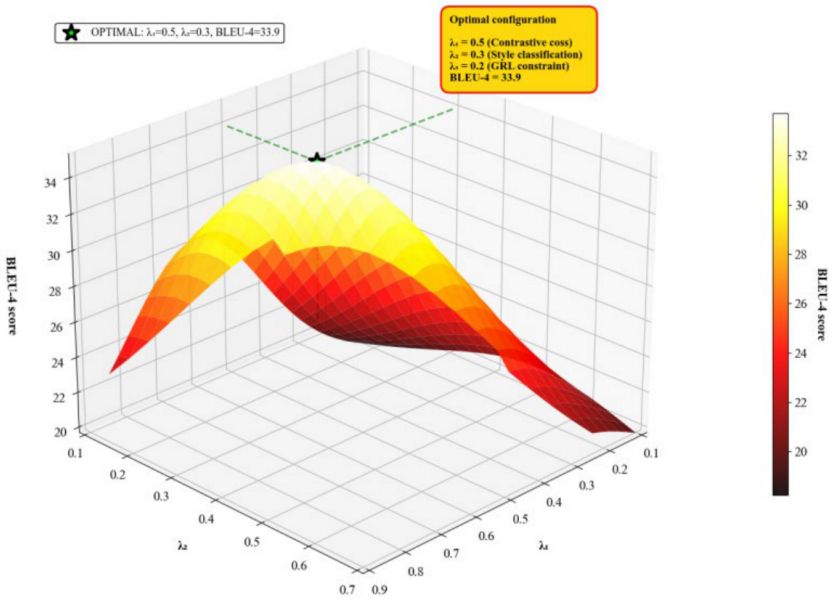

FIGURE 2. Loss weight

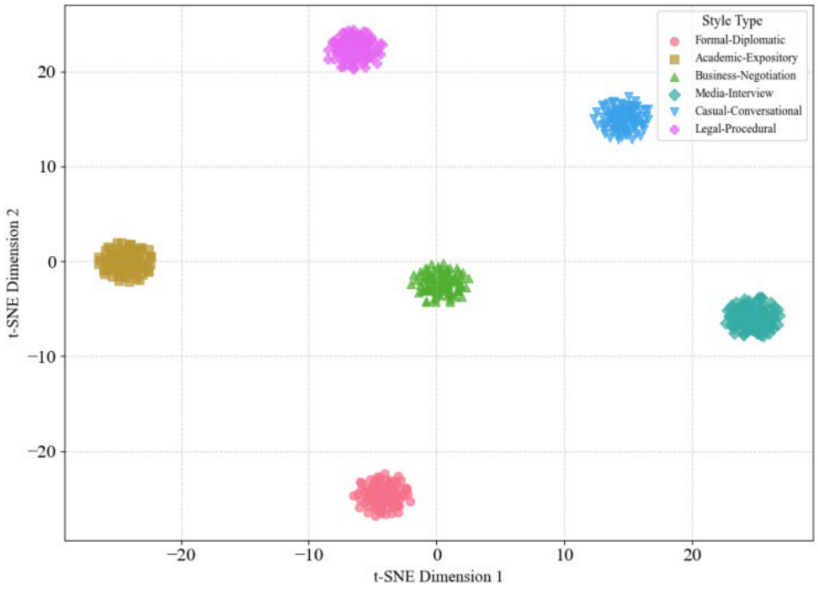

FIGURE 3. Style vector visualization

In the evaluation of style matching, the automatic evaluation uses the accuracy of the 6-way style classifier, which is a BERT-base architecture. For the translation $\widehat{y}$ generated by the model, the proportion of it correctly classified as the target style k is calculated:

$$StyleAcc = \frac{1}{|D_{test}|} \sum_{\widehat{y} \in D_{test}} I\left[\arg\max_{k'} C_{style}(\widehat{y})[k'] = k_{target}\right] \tag{14}$$

The manual evaluation is conducted by three native English speakers (all with a master's degree in translation or linguistics) who give the generated results a five-dimensional Likert rating (1–5 points), including semantic fidelity, style adaptability, language naturalness, fluency, and adequacy. The rating process uses a blind evaluation, which shows that the evaluation results are reliable.

The semantic-style decoupling degree is used to quantify the independence of semantic representation from style, and the mutual information (MI) is calculated:

$$MI(K; h_{sem}) = \sum_{k \in K} \int p(k,h) \log \frac{p(k,h)}{p(k)p(h)}\, dh \tag{15}$$

In Formula (15), K is the style label. The lower the MI value, the less dependent the semantic representation is on

TABLE 2. Data sources and scale

Style type	Data source	Corpus description	Number of sentences
Formal-Diplomatic	United Nations Parallel Corpus v1.0	Minutes of United Nations General Assembly and Security Council meetings, including diplomatic statements, draft resolutions, etc	50,000
Legal-Procedural	Europarl v10	European Parliament debate text, covering treaty interpretation, rules of procedure, and legislative procedures	30,000
Academic-Expository	ACL Anthology Reference Corpus	Conference abstracts and speeches in the field of computational linguistics, with rigorous logic and dense terminology	25,000
Media-Interview	News Commentary v17 dataset from the Workshop on Machine Translation (WMT)	News commentary articles, simulating the perspective of a journalist, with concise language and distinct viewpoints	20,000
Casual-Conversational	OpenSubtitles 2018	Subtitles for movies and TV dramas, highly colloquial	40,000
Business-Negotiation	Technology, Entertainment, Design (TED) Talks Transcripts (from IWSLT & MuST-C)	Business-themed TED talks, focusing on collaboration, strategy, and interest expression	25,000

Note: ACL = ACL Anthology Reference Corpus. HyperText Markup Language (HTML) tags, non-English sentences (filtered using the langdetect library), and sentences shorter than 5 or longer than 100 words are removed. Sentence-level deduplication is performed using MinHash and locality-sensitive hashing, with a Jaccard similarity threshold of 0.95 to eliminate duplicates within and across corpora. 1,000 sentences are randomly selected from each corpus type and independently annotated by two graduate students with a linguistics background to determine whether they met predefined stylistic features. Krippendorff's $\alpha = 0.74$ is calculated, indicating good inter-annotator agreement.

TABLE 3. Test set statistics

Style type	Number of sentences	Average sentence length	TTR	Typical language domain
Formal-Diplomatic	100	22.3	0.68	United Nations resolutions, diplomatic statements
Legal-Procedural	100	24.1	0.65	Interpretation of treaties and rules of procedure
Academic-Expository	100	19.8	0.71	Academic argumentation and technical interpretation
Media-Interview	100	14.2	0.75	News Q&A and opinion guidance
Business-Negotiation	100	16.7	0.70	Business roadshows and cooperation negotiations
Casual-Conversational	100	12.5	0.78	Daily conversations and informal communication
Total	600	18.3	0.71	–

Note: Type-Token Ratio (TTR) = number of word types / total number of words, used to measure lexical richness; higher values indicate greater vocabulary diversity.

style, and the stronger the decoupling effect.

D. IMPLEMENTATION DETAILS

This study conducts model training and evaluation in a unified hardware and software environment. All models are implemented based on the PyTorch 2.0 deep learning framework and integrate the HuggingFace Transformers 4.30 library to load pre-trained language models. The training process is performed on 8 NVIDIA A100 80 GB processors, and mixed precision training is used to improve computational efficiency. In terms of optimization strategy, the AdamW optimizer is used with a weight decay coefficient of 0.01. The learning rate adopts a linear warm-up strategy: it increases linearly from 0 to a peak of 2×10^{-5} in the first 1000 steps and then remains constant. To prevent gradient explosion, gradient clipping is applied with a threshold of 1.0. Model regularization is achieved through Dropout (rate = 0.1) and applied to all fully connected layers and attention modules.

The training batch size is dynamically adjusted according to the task stage: a large batch size (512) is used in the contrastive learning pre-training stage to enhance the diversity of negative samples; a batch size of 256 is used in the end-to-end fine-tuning stage to balance memory usage and convergence stability.

IV. RESULTS

A. MAIN EXPERIMENT

1) Semantic Fidelity Performance

Semantic fidelity performance evaluation aims to verify whether the model accurately preserves the core semantics of the source language during style transfer and is a basic indicator for measuring the reliability of interpretation. The semantic fidelity performance comparison results are shown in Figure 4.

The proposed model achieves state-of-the-art performance across all semantic fidelity metrics, reaching 33.9, 59.8, 43.7, 0.714, and 0.842 in BLEU-4, chrF++, TER, COMET-22, and BERTScore (F1), respectively, outperforming strong baselines including SimCSE. ByT5 performs the worst due to fuzzy word segmentation and semantic boundaries in Chinese-to-English translation, caused by byte-level modeling. Although StyleTransformer applies style control, it relies on style labels and does not decouple semantics, which sacrifices content fidelity. T5 and mBART achieve good basic performance through large-scale pre-training, but lack the ability to model the diversity of interpretation styles. In contrast, the proposed framework constructs style-invariant semantic representations through self-supervised contrastive learning and performs style decoupling at the decoding stage. By preventing style interference in semantic encoding, it enhances cross-style semantic consistency and improves both semantic fidelity and style transfer perfor-

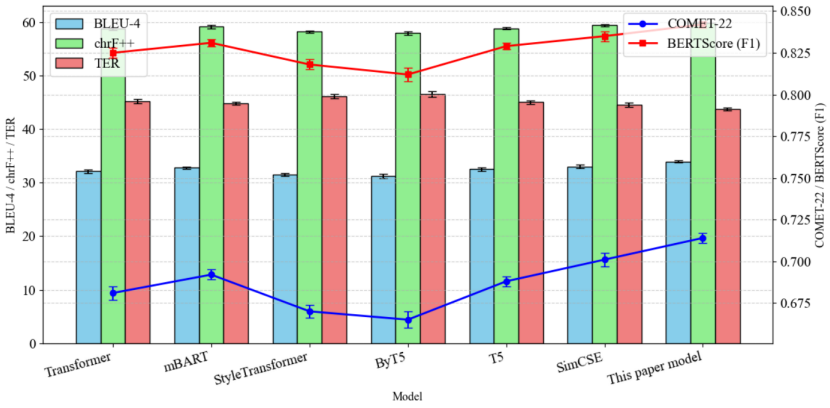

FIGURE 4. Semantic fidelity performance analysis results

mance without requiring style labels.

2) Style Transfer Accuracy

Style matching evaluation aims to measure the appropriateness and naturalness of the model-generated translation in the target pragmatic context and is a core dimension for verifying the effectiveness of style transfer. A comparison of style adaptation results is shown in Figure 5.

The proposed model outperforms other models in style matching, achieving a Style Accuracy of 82.6%, a 7.8 percentage point improvement over the best baseline, SimCSE (74.8%). Although StyleTransformer applies explicit style embedding, its reliance on supervised labels and lack of semantic decoupling limits its generalization in unlabeled testing scenarios. ByT5 and Transformer lack style modeling capabilities, resulting in output styles that closely resemble the training data distribution and struggle to adapt to formal or legal domains. T5 and mBART achieve some style diversity through pre-training, but lack fine-grained control. In contrast, the proposed model constructs style-invariant semantic representations through self-supervised contrastive learning and combines them with learnable style prototypes for unsupervised style injection. This method improves style adaptation (4.4) and linguistic naturalness (4.2) while maintaining high semantic fidelity (4.3). The high consistency between manual evaluation and automatic metrics demonstrates the model's practical value in real-world interpretation scenarios.

A comparison of Style Accuracy results across six styles is shown in Figure 6.

The proposed model outperforms the baseline across all six styles, achieving Style Accuracy scores of 80.8%–86.2%. It performs exceptionally well in high-frequency styles such as Media-Interview (86.2%) and Academic-Expository (82.1%), while also maintaining a leading position in low-frequency professional domains such as Formal-Diplomatic (83.6%) and Legal-Procedural (80.8%). This demonstrates that selfsupervised semantic enhancement effectively improves cross-style generalization. In contrast, while SimCSE performs strongly in some styles, its reliance on general semantic clustering makes it difficult to capture the stylized expressions of diplomatic or legal texts. Despite incorporating explicit style embeddings, StyleTransformer still lags significantly behind in Formal-Diplomatic (72.4%) and Legal-Procedural (67.5%), indicating limited generalization without sufficient style supervision data. T5 and mBART achieve good basic performance thanks to large-scale pre-training, but lack explicit modeling of stylistic pragmatic rules. ByT5 suffers from low performance across all metrics due to its limited expressive power at the morphological and pragmatic levels of byte-level modeling. This study constructs a style-invariant semantic representation through contrastive learning and combines it with learnable style prototypes for precise control. This method maintains semantic consistency while activating representative linguistic features across various registers, achieving robust and high-quality transfer across both high- and low-frequency styles, validating the method's universality and practicality.

3) Qualitative Analysis of Multi-Style Transfer

The style types include Formal-Diplomatic, Legal-Procedural, Academic-Expository, Media-Interview, Business-Negotiation, and Casual-Conversational. The results of multi-style transfer are shown in Figure 7.

Figure 7 shows a comparison of the six types of style generation between the proposed model and StyleTransformer under the same source sentence "the agreement has not yet taken effect", demonstrating the advantages of the proposed method in style control accuracy and pragmatic appropriateness. While StyleTransformer can generate generally fluent translations, its stylistic features remain ambiguous. For example, in the Legal-Procedural style it only produces "the agreement shall not take effect", lacking the conditional constraints and formalized structure typical of legal texts. In contrast, the proposed model, through its learnable style prototypes and decoupled architecture, precisely activates linguistic paradigms across various domains: Formal-Diplomatic uses "instrument" instead of "agreement" to align with diplomatic terminology; Legal-Procedural applies "until the stipulated conditions are fulfilled" to reflect the

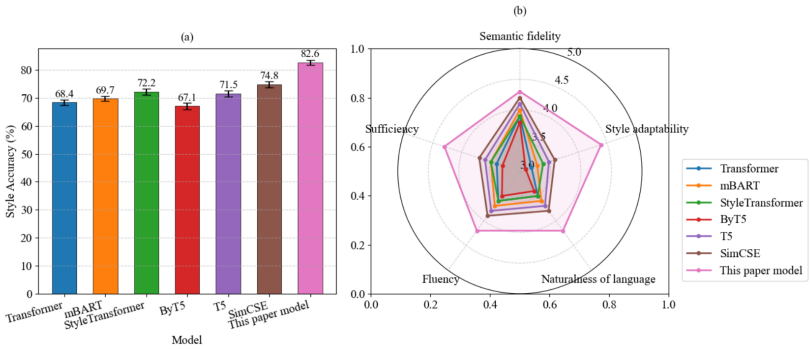

FIGURE 5. Comparison of style adaptation results: (a). Automatic evaluation; (b). Manual scoring

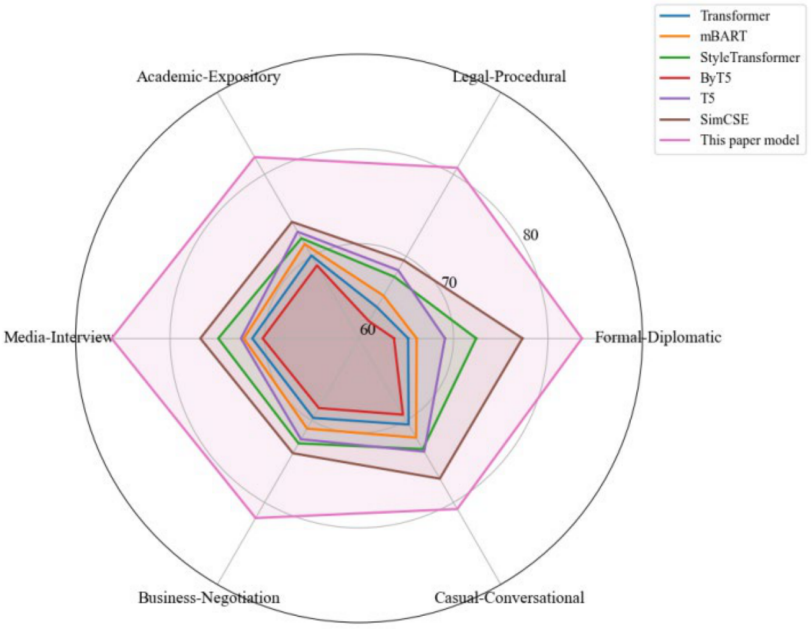

FIGURE 6. Style Accuracy radar chart for six styles

rigor and conditional constraints of legal texts; Media-Interview uses direct quotations and dashes to simulate the rhythm of real question-and-answer sessions; Casual-Conversational uses idioms such as "up and running" to enhance spoken naturalness. Crucially, all outputs maintain a distinct style while strictly adhering to the core semantics of "the protocol is not in effect", verifying the design goal of "variable style, unchanged content", and significantly outperforming baseline methods that rely on static style embedding. MI is used to quantify the dependence of semantic representations on style labels across different models. Lower MI values indicate stronger semantic–style decoupling. Results show that the proposed model (MI = 0.183) outperforms all baselines and is 37.8% lower than the next-best SimCSE (0.294), validating the effectiveness of the proposed decoupling mechanism. Although the StyleTransformer applies explicit style embeddings, its semantic encoder is still subject to style interference (MI = 0.365), reflecting the unavoidable coupling of style and content during end-to-end training. Pre-trained models such as T5 and mBART, trained on mixed-style corpora, have implicit style bias in their semantic representations, resulting in intermediate MI values. This paper achieves deep decoupling through a three-pronged mechanism: firstly, the semantic encoder pre-trained through contrastive learning is frozen to ensure that it captures only style-invariant semantics; secondly, the GRL adversarial training is applied to actively suppress the style predictability in the semantic vector; finally, the style prototype library isolates the style information into an independent vector space to avoid semantic contamination in the decoding stage. The low MI value not only explains the high fidelity of style transfer but also demonstrates, from the perspective of representation learning, that the model possesses true style generalization capabilities—that is, it can generate semantically consistent and stylistically adapted interpretation output even under unseen or weakly supervised conditions, providing solid theoretical support for unsupervised style transfer.

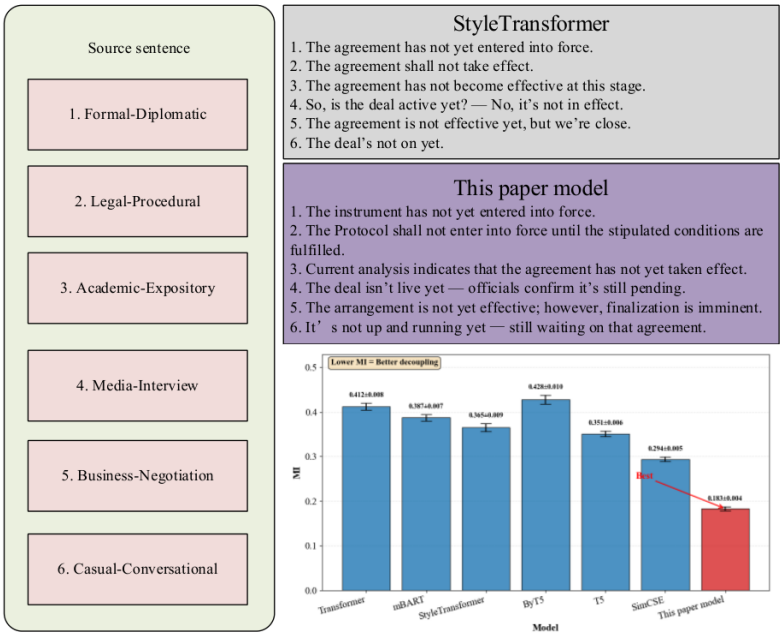

FIGURE 7. Multi-style transfer results

B. ABLATION EXPERIMENTS

Ablation experiments are used to verify the contributions of each core component in this paper to semantic fidelity, style transfer, and semantic-style decoupling, ensuring the necessity and effectiveness of the model design. The results of the ablation experiments are shown in Table 5.

TABLE 5. Ablation experiment results

Model	BLEU-4	chrF++	Style Accuracy (%)	MI
Full	33.9	59.8	82.6	0.183
w/o SS-SECL	32.1	58.2	68.4	0.297
w/o Style Prototype	33.2	59.1	74.3	0.231

Ablation experiments show that each component contributes significantly to model performance. Removing SS-SECL results in a 14.2 percentage point decrease in Style Accuracy and an increase in MI to 0.297, demonstrating that contrastive learning is central to constructing style-invariant semantic representations. The proposed method effectively decouples content and form by narrowing semantic distances through semantically preserved style variants. Removing the style prototypes results in unstable style control by relying on a real-time style encoder, validating the advantages of learnable prototypes in unsupervised transfer. Removing GRL has a minimal impact on BLEU, but Style Accuracy decreases by 5.8 percentage points, and MI increases to 0.252, demonstrating that adversarial training is essential for enforcing semantic invariance. Removing semantic freezing impairs performance by undermining the semantic robustness achieved by pre-training and causing style information to flow back into the semantic representation. In summary, the complete model achieves a balance between semantic fidelity, style controllability, and representation decoupling by combining a frozen semantic encoder, contrastive learning, style prototypes, and GRL. The absence of any one component disrupts this balance.

V. CONCLUSION

To effectively manage the distinct yet related information streams in design and manufacturing reports, this study constructs a semantic-style decoupling architecture utilizing a self-supervised semantic enhancement framework. A style-conditional generative model is used to derive semantically consistent and stylistically diverse positive sample pairs from a single corpus. Strict semantic similarity and style purity filtering are used to ensure enhanced quality. A multi-view contrastive learning mechanism based on SimCLR is used to force the semantic encoder to learn content representations that are robust to surface style variations. This method combines a frozen DeBERTa-v3-base semantic encoder with a lightweight style encoder to achieve representation decoupling. A learnable style prototype library and GRL are applied to support zero-shot style transfer during inference without the need for retraining or style annotation. The results demonstrate that the proposed method outperforms existing baselines in terms of semantic fidelity, style adaptability, and decoupling. The model achieves classification accuracy exceeding 80% across six interpretation styles, reaching a peak of 86.2% for the Media-Interview style. Mutual information is also as low as 0.183, demonstrating the high independence of semantic representations from style. This study proposes a multi-style interpretation generation framework that does not require style annotation

and reduces reliance on parallel corpora. By integrating semantic enhancement and contrastive learning, the framework decouples content and style, which improves cross-style generalization for diverse documentation tasks. This approach validates self-supervised learning for pragmatic style modeling, driving semantic perception toward stylistic awareness and offering an explainable path for natural and appropriate language generation in practical interpretation systems.

AUTHOR CONTRIBUTIONS

Ling Gu designed, collected and analyzed the data, and drafted the manuscript. Ling Gu conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Ling Gu 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.

HUMAN RESEARCH SUBJECTS

This study was approved by the Ethics Committee of Gannan College of Science and Technology, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

FUNDING

This research was funded by:

1. 2024 Jiangxi Provincial Department of Education, "Research on the Multimodal Integration and Dissemination Mechanism of Jiangxi Red Cultural Resources in the Era of Convergent Media", Project No. XW24101.

2. 2024 University-level Education Reform Project, "Research on the Application of the "One line – Three Dimensions – Six Integration" Foreign Language Teaching Model in the Cultivation of Chinese Cultural Communication Ability of College Students", Project No. XJG-2024-5.

3. 2025 College-level Education Reform Project, "Research on the Construction of 'Micro-Majors' in Local Application-Oriented Undergraduate Universities Under the OBE Concept", Project No. XJG-2025-9.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] M. De Coster, D. Shterionov, M. Van Herreweghe, and J. Dambre, "Machine translation from signed to spoken languages: State of the art and challenges," Universal Access in the Information Society, vol. 23, no. 3, pp. 1305-1331, 2024, doi: 10.1007/s10209-023-00992-1.

[2] A. S. Dhanjal and W. Singh, "An automatic machine translation system for multi-lingual speech to Indian sign language," multimedia Tools and Applications, vol. 81, no. 3, pp. 4283-4321, 2022, doi: 10.1007/s11042-021-11706-1.

[3] Y. Zhang, "A Study on the Translation of Spoken English from Speech to Text," Journal of ICT Standardization, vol. 12, no. 4, pp. 429-441, 2024, doi: 10.13052/jicts2245-800X.1244.

[4] B. R. Taira, V. Kreger, A. Orue, and L. C. Diamond, "A pragmatic assessment of Google Translate for emergency department instructions," Journal of General Internal Medicine, vol. 36, no. 11, pp. 3361-3365, 2021, doi: 10.1007/s11606-021-06666-z.

[5] T. Pu, "Accurate translation system of spoken English based on attribute features," Journal of Jilin University (Information Science Edition), vol. 42, no. 6, pp. 1155-1163, 2024.

[6] Y. Ying, "Design of an Interactive Spoken Language Machine Translation System Based on Relevance-Driven Semantic Fuzzification Analysis," 2025 IEEE 12th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), vol. 12, no. 5, pp. 2693-2873, 2025, doi: 10.1109/ITAIC64559.2025.11163367.

[7] U. Sulubacak, O. Caglayan, S. A. Grönroos, A. Rouhe, D. Elliott, L. Specia, et al., "Multimodal machine translation through visuals and speech," Machine Translation, vol. 34, no. 2, pp. 97-147, 2020, doi: 10.1007/s10590-020-09250-0.

[8] J. Zhu, R. Han, S. Zhang, and S. Chen, "Low-resource neural machine translation enhanced with compressed multilingual BERT knowledge," Journal of Computer Engineering & Applications, vol. 61, no. 8, pp. 163-163, 2025, doi: 10.3778/j.issn.1002-8331.2312-0244.

[9] C. Ma, Y. Tian, X. Zheng, and K. Sun, "A review of neural machine translation based on knowledge distillation," Journal of Frontiers of Computer Science & Technology, vol. 18, no. 7, pp. 1725, 2024, doi: 10.3778/j.issn.1673-9418.2311027.

[10] X. Lei, "Real-time translation of English speech through speech feature extraction," Artificial Life and Robotics, vol. 29, no. 3, pp. 410-415, 2024, doi: 10.1007/s10015-024-00951-w.

[11] X. Wu, "A Corpus-Based Study on the Translation Style of Wang Rongpei: Taking the English Translation of The Book of Songs as an Example," Journal of Beijing International Studies University, vol. 45, no. 5, pp. 82, 2023, doi: 10.12002/j.bisu.478.

[12] P. Lv and D. Chen, "A Corpus-Based Comparative Study on the Translation Style of the English Translation of The Analects: Taking the Translations of Gu Hongming and Arthur Waley as Examples," Shanghai Journal of Translators, vol. 158, no. 3, pp. 61, 2021, doi: 10.3969/j.issn.1672-9358.2021.03.012.

[13] L. Wang, H. Qiu, B. Qiu, F. Meng, Q. Wu, H. Li, et al., "Trident-Cap: Image-fact-style trident semantic framework for stylized image captioning," IEEE Transactions on Circuits and Systems for Video Technology, vol. 34, no. 5, pp. 3563-3575, 2023, doi: 10.3969/j.issn.1672-9358.2021.03.012.

[14] L. Zhang, Q. Lin, F. Meng, S. Liang, J. Lu, S. Liu, et al., "Leveraging Contrastive Semantics and Language Adaptation for Robust Financial Text Classification Across Languages," Computers, vol. 14, no. 8, pp. 338, 2025, doi: 10.3390/computers14080338.

[15] C. Wang and S. Lv, "Prefix Data Augmentation for Contrastive Learning of Unsupervised Sentence Embedding," Applied Sciences, vol. 14, no. 7, pp. 2880, 2024, doi: 10.3390/app14072880.

[16] X. Liu, W. Gong, Y. Li, Y. Li, and X. Li, "A Study of Contrastive Learning Algorithms for Sentence Representation Based on Simple Data Augmentation," Applied Sciences, vol. 13, no. 18, Art. no. 10120, 2023, doi: 10.3390/app131810120.

[17] J. Hu, Y. Zhu, L. Wu, Q. Luo, F. Teng, T. Li, et al., "Text semantic matching algorithm based on the introduction of external knowledge under contrastive learning," International Journal of Machine Learning and Cybernetics, vol. 16, no. 1, pp. 741-753, 2025, doi: 10.1007/s13042-024-02285-2.

[18] R. Clouet, "Foreign languages applied to translation and interpreting as languages for specific purposes: claims and implications," RLA. Revista de Lingüística Teórica y Aplicada, vol. 59, no. 1, pp. 39-62, 2021, doi: 10.29393/RLA59-2FLRC10002.

[19] B. Zheng, S. Tyulenev, and K. Marais, "Introduction:(re-) conceptualizing translation in translation studies," Translation Studies, vol. 16, no. 2, pp. 167-177, 2023, doi: 10.1080/14781700.2023.2207577.

[20] H. Zhang, H. Song, S. Li, M. Zhou, and D. Song, "A survey of controllable text generation using transformer-based pre-trained language

models," ACM Computing Surveys, vol. 56, no. 3, pp. 1-37, 2023, doi: 10.1145/3617680.

[21] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, "A survey on contrastive self-supervised learning," Technologies, vol. 9, no. 1, pp. 2, 2020, doi: 10.3390/technologies9010002.

[22] L. Zhang, Z. Wang, S. Lian, B. Pu, Y. Liu, M. Qin, et al., "Sign language translation based on deep learning: past, present and future," Application Research of Computers/Jisuanji Yingyong Yanjiu, vol. 42, no. 8, pp. 2241, 2025.

[23] L. Xu, H. Xie, Z. Li, L. Wang F, W. Wang, Q. Li, et al., "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.

[24] Z. Yu, H. Li, and J. Feng, "Contrastive learning for unsupervised sentence embeddings using negative samples with diminished semantics," The Journal of Supercomputing, vol. 80, no. 4, pp. 5428-5445, 2024, doi: 10.1007/s11227-023-05682-6.

[25] W. Lu, D. Ming, X. Mao, J. Wang, Z. Zhao, Y. Cheng, et al., "A DeBERTa-Based Semantic Conversion Model for Spatiotemporal Questions in Natural Language," Applied Sciences, vol. 15, no. 3, pp. 1073, 2025, doi: 10.3390/app15031073.

[26] R. K. Singh, M. K. Sachan, and R. B. Patel, "Cross-domain sentiment classification using decoding-enhanced bidirectional encoder representations from transformers with disentangled attention," Concurrency and Computation: Practice and Experience, vol. 35, no. 6, pp. 1, 2023, doi: 10.1002/cpe.7589.

[27] B. Khaertdinov, S. Asteriadis, and E. Ghaleb, "Dynamic temperature scaling in contrastive self-supervised learning for sensor-based human activity recognition," IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 4, no. 4, pp. 498-507, 2022, doi: 10.1109/TBIOM.2022.3180591.

[28] D. Le, S. Truong, P. Brijesh, D. A. Adjeroh, and N. Le, "SCL-ST: Supervised contrastive learning with semantic transformations for multiple lead ECG arrhythmia classification," IEEE journal of biomedical and health informatics, vol. 27, no. 6, pp. 2818-2828, 2023, doi: 10.1109/JBHI.2023.3246241.

[29] P. Jiang and X. Cai, "Semantic Matching for Chinese Language Approach Using Refined Contextual Features and Sentence–Subject Interaction," Symmetry, vol. 17, no. 4, pp. 585, 2025, doi: 10.3390/sym17040585.

[30] P. Jiang and X. Cai, "A Symmetric Dual-Drive Text Matching Model Based on Dynamically Gated Sparse Attention Feature Distillation with a Faithful Semantic Preservation Strategy," Symmetry, vol. 17, no. 5, pp. 772-772, 2025, doi: 10.3390/sym17050772.

[31] Y. Lee, J. Son, and M. Song, "BertSRC: Transformer-based semantic relation classification," BMC Medical Informatics and Decision Making, vol. 22, no. 1, pp. 234, 2022, doi: 10.1186/s12911-022-01977-5.

[32] J. W. Lin, T. W. Su, and C. C. Chang, "Chinese Story Generation Based on Style Control of Transformer Model and Content Evaluation Method," Algorithms, vol. 18, no. 3, pp. 168, 2025, doi: 10.3390/a18030168.

[33] L. Zhou, R. Sasano, and K. Takeda, "Inference discrepancy based curriculum learning for neural machine translation," IEICE Transactions on Information and Systems, vol. 107, no. 1, pp. 135-143, 2024, doi: 10.1587/transinf.2023EDP7048.

[34] Q. Ma, "Research on English–Chinese machine translation shift based on word vector similarity," Artificial Life and Robotics, vol. 29, no. 4, pp. 585-589, 2024, doi: 10.1007/s10015-024-00964-5.

[35] Y. Dong, J. Ding, X. Jiang, G. Li, Z. Li, Z. Jin, et al., "Codescore: Evaluating code generation by learning code execution," ACM Transactions on Software Engineering and Methodology, vol. 34, no. 3, pp. 1-22, 2025, doi: 10.1145/3695991.