Construction of Chinese-Japanese-English Translation Database Based on Cross-Lingual Transformer-Based Models

Xiaodong Wang[1], Dianlong Wu[2]
[1]School of International Studies, Changchun Humanities and Sciences College, Changchun 130117, Jilin, China
[2]School of Foreign Languages, Changchun Normal University, Changchun 130032, Jilin, China

Corresponding author: Dianlong Wu (e-mail: wudianlong@ccsfu.edu.cn)

ABSTRACT Accurate semantic alignment and comprehensive multilingual coverage remain major challenges in constructing Chinese–Japanese–English translation databases for technical knowledge sharing and engineering information exchange. This study proposes a cross-lingual translation database construction framework based on Transformerbased neural networks and hierarchical contrastive representation learning. Using InfoXLM as the shared semantic encoder, dependency-syntax perturbation and cross-lingual lexical substitution are employed to generate codeswitching hard negative samples, while a three-way triplet contrastive optimization strategy jointly constrains Chinese, Japanese, and English semantic spaces to improve representation consistency. The fine-tuned model is further integrated with semantic similarity filtering and bidirectional consistency verification to construct a large-scale trilingual translation database from multilingual candidate corpora. Experimental results demonstrate an alignment precision of 94.7%, a recall of 92.3%, an F1-score of 93.5%, and a translation coverage of 89.8%, significantly outperforming conventional multilingual alignment approaches. The proposed framework provides an efficient solution for multilingual engineering knowledge organization, technical document retrieval, and intelligent information processing, offering valuable support for cross-language electromagnetic engineering documentation, antenna technology resources, and multilingual communication systems requiring robust semantic alignment and signal-aware information management.

INDEX TERMS cross-lingual Transformer,semantic alignment, trilingual translation database,electromagnetic engineering information processing,multilingual technical documentation

I. INTRODUCTION

WITH the continued growth in international engineering education and lifelong learning needs, cross-language technical documents, multilingual MOOCs (Massive Open Online Courses), and transnational technical training have created an urgent demand for high-quality Chinese-Japanese-English translation databases. However, Chinese, Japanese, and English belong to different language families and have significant typological differences in word order, morphological markers, and lexical expressions, resulting in shortcomings in the semantic alignment accuracy and trilingual translation coverage of existing translation databases [1-2]. Specifically, parallel corpora built based on statistical machine translation or artificial neural network translation typically only cover bilingual pairs, and high-quality trilingual aligned sentence pairs are extremely scarce due to annotation costs [3-4]. On the other hand, zero-shot alignment based on multilingual pre-trained models is prone to semantic drift on long-distance language pairs due to the lack of explicit modeling of syntactic variations in Chinese, Japanese, and English, making it difficult to guarantee terminology consistency and translation fluency in engineering education scenarios [5-6].

This paper addresses specific problems at three levels. The first is the sentence-level semantic equivalence determination among Chinese, Japanese, and English. Because the SOV (Subject Object-Verb) word order in Japanese contrasts with the SVO (Subject-Verb-Object) word order in Chinese and English, and Japanese heavily utilizes particles to mark case relationships, while Chinese relies on word order and English on morphological changes, there is no simple lexical alignment mapping among the three languages. Therefore, models need to learn deep semantic representations that go beyond surface structures [7-8]. The second is the large-scale automated construction of trilingual translation databases. Manually annotating trilingual parallel sentence pairs is extremely costly, especially in engineering and technical fields (such as mechanical repair manuals,

electrical drawings, and programming interface documents), where there are numerous technical terms and consistent translation styles are required. Relying on manual methods makes it difficult to obtain sufficient training data in the short term; automated semantic alignment and filtering methods are necessary [9-10]. The third is the cross-linguistic data support for continuing engineering education scenarios. In multinational corporate training, international technical certification exams, and multilingual courses, non-native language engineers often need to consult technical materials in Chinese, Japanese, and English simultaneously. For example, when a Chinese engineer is reading Japanese equipment manual, quickly aligning it with the corresponding international standard English expression can significantly reduce comprehension errors [11-12]. The importance of this is that Chinese, Japanese, and English cover the three major engineering languages in the world's major economies. Building a high-quality trilingual translation database is not only a fundamental task in cross-language natural language processing, but also a key infrastructure for reducing language barriers in international engineering education and improving the efficiency of lifelong learning [13-14].

To address the cross-lingual semantic alignment problem, existing research has proposed various solutions, but these still have shortcomings in the context of Chinese, Japanese, and English. The LASER (Language- Agnostic SEntence Representations) series of models employs a multilingual BiLSTM (Bi-directional Long Short-Term Memory) encoder and a shared byte-pair encoding vocabulary, achieving sentence vector alignment for over a hundred languages through zero-shot transfer learning. However, it is insufficient in extracting features from agglutinative languages [15-16]. Another approach is based on multilingual variant fine-tuning of BERT (Bidirectional Encoder Representations from Transformers), such as mBERT (multilingual BERT), XLM-RoBERTa (Cross-lingual Language Model-Robustly optimized BERT approach), and InfoXLM (Information-theoretic Cross-lingual Language Model), specifically designed for cross-lingual contrastive learning. Researchers typically train bilingual sentence pair classifiers using parallel sentence pairs in Chinese-Japanese, Chinese-English, and Japanese-English languages, or use triple loss for relaxed alignment of bilingual boundaries [17-18]. However, these methods share two common problems: first, they are limited to bilingual sentence pairs and lack a triple joint constraint mechanism, resulting in the semantic spaces of Chinese-Japanese and Chinese-English languages not naturally aligning into the same cluster; second, they do not explicitly model word order variations caused by differences in syntactic structure, leading to poor robustness of the model to long sentences with large dependency tree depths and natural representations of modifier shifts [19]. Furthermore, most existing methods rely on simple positive examples from existing parallel corpora and lack proactive strategies for generating hard negative examples (sentence pairs that are semantically similar but not mutually translated), making the discrimination boundaries of the model ambiguous when distinguishing near-synonymous interference items, and easily misclassifying sentences with similar terms but different semantics as mutually translated pairs. The present work addresses semantic equivalence confusion caused by typological differences in word order and lexical expression across Chinese, Japanese, and English; factual commonsense contradictions in trilingual translations are not targeted by the proposed method. The aforementioned limitations indicate a need for a framework capable of explicitly modeling syntactic variations and jointly optimizing trilingual constraints. Artificial neural networks, with their distributed representations that bridge word order and lexical differences, end-to-end joint learning that avoids error accumulation, and deep nonlinear transformations that address syntactic variations, provide an ideal computational framework for Chinese-English-Japanese trilingual alignment.

The aim of this paper is to construct a high-quality trilingual translation database based on artificial neural networks (ANNs) for Chinese, Japanese, and English, thereby overcoming the semantic alignment bottleneck caused by typological differences in languages. To this end, a cross-lingual alignment method combining hierarchical triplet contrastive learning and syntactic perturbation enhancement modules is proposed. The specific steps are as follows: First, a pre-trained InfoXLM model based on the Transformer architecture is used as the semantic encoder, inputting a pre-cleaned set of trilingual translation anchor points. Second, dependency syntactic perturbation and cross-lingual lexical substitution are performed on Chinese, Japanese, and English sentences respectively, generating code-switching-hard negative example samples, forming positive and negative example triplets in the trilingual semantic space. Third, triplet contrastive losses in three directions (Chinese-Japanese, Chinese-English, and Japanese-English) are introduced at the top layer of the model for joint fine-tuning, forcing equivalent sentence vectors to be closer together and non-corresponding samples to be further apart. Finally, the fine-tuned model is used to perform vector similarity filtering and bidirectional consistency judgment on massive web crawled candidate sentence pairs, constructing the trilingual translation database. Experimental results show that the method achieves a precision of 94.7%, a recall of 92.3%, and an F1 score of 93.5% for sentence-level alignment on the test set, achieves a coverage of 89.8 % for real translation pairs in the candidate data, and also demonstrates good application performance in continuing engineering education scenarios.

II. CROSS-LINGUISTIC SEMANTIC ALIGNMENT METHODS

A. OVERALL ARCHITECTURE AND MULTILINGUAL ENCODER BASE

The cross-lingual semantic alignment model constructed in this paper consists of four parts: a trilingual data input layer, a syntactic perturbation enhancement module, a shared

multilingual encoder, and a three-way triplet contrastive loss function. The input is a pre-cleaned and aligned set of parallel sentence pairs in Chinese, Japanese, and English, denoted as $D = \{(s_i^{zh}, s_i^{ja}, s_i^{en})\}_{i=1}^{N}$, where s_i^{zh}, s_i^{ja}, and s_i^{en} represent the Chinese, Japanese, and English sentences in the i-th trilingual translation pair, respectively, and N is the total number of parallel sentence pairs. The data is first perturbed by dependency syntax and subjected to cross-lingual lexical substitution to generate extended code-switching hard-to-bearing example samples. These samples are then fed into a shared-weight multilingual encoder to extract sentence vector representations. Finally, hierarchical triplet contrastive constraints are applied to the embedding space to reshape the semantic distribution.

The encoder base uses the InfoXLM model. The InfoXLM model introduces a cross-lingual contrastive learning pre-training task based on the XLM-RoBERTa architecture. By increasing the representation distance between parallel sentence pairs and random negative examples within the same language, it initially achieves alignment of multilingual semantic spaces [20-21]. This paper retains its 24-layer Transformer encoding structure, with a hidden layer dimension of 1024 and 16 attention heads. After loading the pre-trained weights, the model performs subword segmentation on the input sentence, generating a context representation sequence $\{h_1, h_2, \ldots, h_T\}$, and takes the output vector corresponding to the [CLS] position as the sentence vector representation after linear projection and layer normalization:

$$v = \text{LayerNorm}\left(W_p h_{[\text{CLS}]} + b_p\right) \tag{1}$$

In Formula (1), $h_{[\text{CLS}]}$ is the last hidden state at the [CLS] position, and W_p and b_p are the projection matrix and bias term, respectively. Layer normalization is used to stabilize the numerical distribution of sentence vectors during training.

After encoding, the three sentences yield Chinese sentence vectors v^{zh}, Japanese sentence vectors v^{ja}, and English sentence vectors v^{en}, respectively. To ensure that the three language representations reside in the same comparable vector space, this paper freezes the parameters of the bottom 16 Transformer layers of InfoXLM, fine-tuning only the top 8 layers and the sentence vector projection layer [22-23]. This strategy preserves the general cross-linguistic knowledge learned during pre-training while allowing the model to adapt to the specific goal of Chinese-Japanese-English alignment, avoiding alignment bias degradation caused by full parameter fine-tuning [24-25].

At the semantic similarity measurement level, cosine similarity is used as the criterion for cross-linguistic vector distance. For any two sentence vectors v_1 and v_2, the cosine similarity is defined as:

$$\text{sim}(v_1, v_2) = \frac{v_1^T v_2}{\|v_1\|\|v_2\|} \tag{2}$$

This metric is unaffected by the vector magnitude scale and can be directly used as the similarity score for positive and negative pairs in the contrastive loss function for margin calculation. Subsequent hierarchical triplet contrastive loss can construct three-way convergence and divergence constraints based on this similarity.

The InfoXLM network employs comprises a 24-layer Transformer encoder. In balancing depth and breadth, the deep network constructs hierarchical semantic features through layer-by-layer abstraction, while the shallow layers capture parts of speech and local phrase patterns, the mid-layers learn syntactic structures, and the deep layers extract cross-linguistic semantic equivalence. The breadth provides sufficient representational capacity for the lexical-semantic mapping of Chinese, Japanese, and English, avoiding the loss of language- specific features due to information bottlenecks. Regarding the source of network nonlinearity, the GELU (Gaussian Error Linear Unit) activation function in each feedforward network employs a probabilistic gating mechanism, enabling the network to selectively transmit semantically relevant information while suppressing surface features related to language identification. Finally, in terms of training strategy, freezing the bottom 16 layers is equivalent to preserving the cross-linguistic general knowledge obtained during pre-training, while fine-tuning only the top 8 layers is considered as training a lightweight "trilingual alignment classifier" on the existing feature space. The overall architecture of the trilingual semantic alignment model is shown in Figure 1.

Figure 1 illustrates the overall architecture of the trilingual semantic alignment model, which mainly consists of four parts: the input layer receives parallel sentence pairs in Chinese, Japanese, and English; the data augmentation module generates hard negative example samples by using dependency syntax perturbation and cross-language lexical substitution; the shared multilingual encoder (InfoXLM) fine-tunes the top layer while freezing the bottom layer to extract sentence vectors; finally, the loss region calculates the loss through three-way triplet contrast loss (covering Chinese-Japanese, Chinese-English, and Japanese-English directions), forcing equivalent sentence vectors to be pulled closer to each other and non-corresponding samples to be pushed further away, thereby achieving trilingual semantic alignment.

Regarding the adaptability of alignment between Chinese, Japanese, and English, there are differences in word order, morphological markers, and dependency distance among the three languages. The deep Transformer, through layer-by-layer abstraction, can capture parts of speech and local phrase patterns at the low level, learn syntactic structures at the mid-level, and extract cross-linguistic semantic equivalence at the high level. Its 16 attention heads provide sufficient diversity in alignment patterns; some heads focus on aligning the subject-object order in Chinese and English, while others learn the rearrangement mapping from Japanese subject-object-verb word order to Chinese word

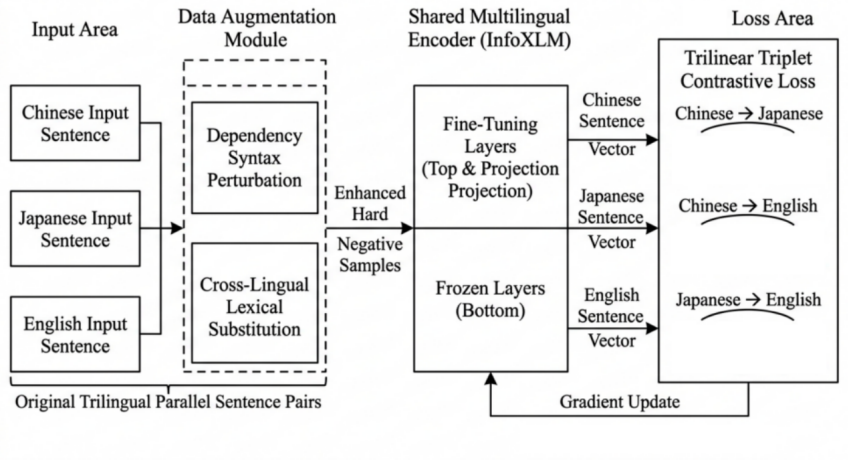

FIGURE 1. Overall architecture of the trilingual semantic alignment model

order.

The overall training process is executed iteratively in batches. Each batch samples trilingual positive example anchors from group B of D. These are then processed by a syntactic perturbation module to generate enhanced variants and cross-lingual negative examples for each group. These are fed into the encoder to obtain three sets of sentence vectors. The three-way triplet loss is then calculated, and the gradient is backpropagated. The optimizer uses AdamW with an initial learning rate of 3×10^{-5}, coupled with linear warm-up and cosine decay scheduling. On the validation set, the model selects the optimal checkpoint based on the trilingual alignment recall rate, which is used for subsequent large-scale candidate sentence pair vector extraction and similarity filtering for database construction.

B. CODE-SWITCHING ENHANCEMENT BASED ON DEPENDENCY SYNTAX PERTURBATION

To address the sensitivity of alignment models to surface word order and lexical variations caused by differences in syntactic structures between Chinese, Japanese, and English languages, this paper introduces a code-switching enhancement method based on dependency syntactic perturbations during the training data construction phase. This method applies structure-aware perturbations and cross-lingual lexical substitutions to each sentence in the original trilingual translation anchor set, generating syntactic variants and lexical interference samples. This expands the set of hard examples, forcing the model to ignore surface differences and focus on deep semantic equivalence during training.

First, the dependency parsing parsers for Chinese, Japanese, and English sentences are called separately for each language. For Chinese, a Stanza dual-affine dependency parser trained based on the CTB (Chinese Treebank) annotation specification is used; for Japanese, a Camphr dependency parser trained based on the UD (Universal Dependencies) Japanese GSD treebank is used; and for English, a Stanza parser trained based on the UD English EWT treebank is used [26-27]. After parsing, each sentence is represented as a dependency tree structure $T = (V, E)$, where $V = \{w_1, w_2, \ldots, w_n\}$ is the set of word nodes; $E \subseteq V \times V \times R$ is the set of directed edges with relation labels; R is the set of dependency relation types. For each word node w_i, its part-of-speech tag p_i, headword index h_i, and dependency relation label r_i are recorded.

The perturbation operation uses the predicate verb node in the dependency tree as the root, extracting the core argument nodes such as the subject and object directly governed by it, forming an indivisible syntactic skeleton set K [28-29]. For modifiers outside the skeleton—including adjective modifiers, adverbial modifiers, prepositional phrase adjuncts, and tense particles—positional substitution is performed according to the word order typology of the target language. The Chinese substitution rule is: the relative position of the attributive and the headword remains fronted, and the adverbial is randomly shifted within the range after the subject and before the predicate; the Japanese substitution rule is: modifiers dependent on the predicate are randomly swapped in the sequence to the left of the core predicate, but the position of the predicate verb at the end of the sentence remains fixed; the English substitution rule is: adverbial modifiers are rearranged with uniform probability in the three positions of the beginning, middle, and end of the sentence. The substitution probability is set to 0.4, and the substitution amplitude is randomly selected $k \in \{1, 2, 3\}$ displacement steps per word node.

At the lexical level, a random subset of real word nodes is selected from each perturbed sentence for cross-language replacement, where the number of selected nodes does not exceed 25% of the total word count in that sentence and selection is restricted to three open word classes: nouns, verbs, and adjectives. Let $C(w_i)$ be the candidate set of equivalent translation words for word w_i in the other two languages. If $C(w_i)$ is not empty, a replacement word is extracted from it using a uniform distribution and embedded in the original position.

The perturbed and replaced sentence and the original sentence form a positive sample pair, while any sentence

from the other three language groups in the same batch is considered a negative sample. Let the original Chinese sentence be x^{zh}, and the perturbated and enhanced variant be $\tilde{x}^{zh}$. The sentence vectors corresponding to the two are brought closer together during contrastive learning to enhance the model's invariance to syntactic variations.

$$L_{\text{aug}} = -\log \frac{\exp\left(\text{sim}(v^{zh}, \tilde{v}^{zh})/\tau\right)}{\sum_{j=1}^{B} \exp\left(\text{sim}(v^{zh}, v_{\text{neg},j})/\tau\right) + \sum_{k=1}^{B} \exp\left(\text{sim}(v^{zh}, \tilde{v}_k^{zh})/\tau\right)} \tag{3}$$

In Formula (3), $\tilde{v}^{zh}$ is the sentence vector of the perturbation variant; τ is the temperature coefficient; B is the batch size; $v_{\text{neg},j}$ represents the sentence vectors of other Chinese sentences in the batch, excluding the current positive one. Japanese and English sentences are calculated using the same method to enhance contrast loss within their respective languages. If the cosine similarity between the substituted sentence and the original sentence calculated by InfoXLM is less than 0.6, the variant is discarded. From the perspective of training artificial neural networks, syntactic perturbation and code-switching enhancement are essentially constructing task-related invariant transformations. Traditional ANNs (Artificial Neural Networks) may fluctuate with small changes in the input distribution. However, by exposing the network to a large number of syntactic variants and lexical interference samples under the same semantics, the network's representation function $f_\theta(\cdot)$ is forced to satisfy: $|f_\theta(x) - f_\theta(\tilde{x})| \ll |x - \tilde{x}|$, where $\tilde{x}$ is the perturbation variant and x is the original input sentence. This property is called semantic equivariance, and it is the core mechanism for neural networks to generalize from finite samples to infinite natural language variants. This paper constructs hard negative example triples and further increases the curvature of the decision boundary in the region near positive examples, forcing the network to form a smoother semantic manifold in the embedding space.

C. HIERARCHICAL TRIPLE CONTRASTIVE LEARNING MECHANISM

After syntactic perturbation and code-enhancing processing, each training batch obtains an expanded set of trilingual samples. This paper designs a hierarchical triplet contrastive learning mechanism. By constructing semantic constraints in three directions, it aggregates equivalent vectors of Chinese, Japanese, and English into compact semantic clusters in the embedding space, while simultaneously increasing the margin distance between non-corresponding language samples [30-31].

For the i-th trilingual parallel anchor point within a batch, the InfoXLM encoder outputs the Chinese sentence vector v_i^{zh}, the Japanese sentence vector v_i^{ja}, and the English sentence vector v_i^{en}, respectively. Taking Chinese as the source language perspective, a triplet $(v_i^{zh}, v_i^{ja}, v_j^{ja})$ is constructed, where v_i^{ja} is the positive example Japanese vector, and v_j^{ja} is a randomly selected non-corresponding Japanese vector within the batch ($j \neq i$). The loss function for this directional triplet is defined as:

$$L_{zh\to ja} = \frac{1}{B}\sum_{i=1}^{B} \max\left(0, \text{sim}(v_i^{zh}, v_j^{ja}) - \text{sim}(v_i^{zh}, v_i^{ja}) + \alpha\right) \tag{4}$$

In Formula (4), B is the batch size; $\text{sim}(\cdot,\cdot)$ is the cosine similarity function; α is a preset interval hyperparameter, controlling the minimum boundary of the difference between positive and negative example similarities. When the positive example similarity exceeds the negative example similarity by less than α, a non-zero loss occurs, and gradient backpropagation forces v_i^{zh} and v_i^{ja} closer together, while moving them further away from v_j^{ja}.

Similarly, the Chinese-to-English triplet $(v_i^{zh}, v_i^{en}, v_k^{en})$ and the Japanese-to-English triplet $(v_i^{ja}, v_i^{en}, v_m^{en})$ are constructed, where v_k^{en} and v_m^{en} are non-corresponding English vectors sampled from the batch ($k \neq i$, $m \neq i$). The corresponding loss terms $L_{zh\to en}$ and $L_{ja\to en}$ are calculated using the same function form. To balance the constraint strength in the three directions and avoid the gradient of a particular language pair dominating the training process, the three-directional losses are combined in a weighted summation manner as the final optimization objective:

$$L_{\text{triplet}} = \lambda_1 L_{zh\to ja} + \lambda_2 L_{zh\to en} + \lambda_3 L_{ja\to en} \tag{5}$$

In Formula (5), λ_1, λ_2, and λ_3 are the weight coefficients of the loss in each direction, satisfying $\lambda_1 + \lambda_2 + \lambda_3 = 1$. Based on the trilingual alignment recall rate on the validation set, this paper sets $\lambda_1 = 0.35$, $\lambda_2 = 0.35$, and $\lambda_3 = 0.30$ to reflect the relative balance of training data volume between Chinese and Japanese, and Chinese and English languages, as well as the indirectness of semantic transfer between Japanese and English.

During training, each batch first calculates the three-language sentence vectors through forward propagation, and then constructs negative example index matrices in the three directions respectively. Negative example sampling employs an in-batch online hard sample mining strategy: for each anchor point, the top $K = 3$ non-corresponding samples with the highest cosine similarity (excluding their corresponding positive samples) in the batch are selected as hard negative examples, and the loss is calculated for each as a mean. This strategy accelerates the convergence of semantic boundaries without introducing additional computational overhead. Furthermore, to maintain semantic consistency between the enhanced samples and the original samples, an in-language enhancement contrastive loss term L_{aug} is added to the total loss, which has been defined in the previous section. The final joint loss function is $L = L_{\text{triplet}} + \gamma L_{\text{aug}}$, where γ is set to 0.2 to control the contribution ratio of the enhanced signal.

Gradient backpropagation only updates the parameters of the top 8 Transformer layers of InfoXLM and the sentence

vector projection matrix, while the parameters of the bottom 16 layers remain frozen [32-33]. The optimizer is trained iteratively for 15 rounds with a batch size of $B = 64$ and an initial learning rate of 3×10^{-5}, and the interval hyper-parameter α is set to 0.3. After training, the model's three-way alignment loss on the validation set tends to stabilize, and the equivalent sentence vectors of Chinese, Japanese, and English form a tight cross-lingual semantic cluster in the embedding space. From the loss landscape analysis of artificial neural networks, the hierarchical triplet contrastive loss creates a multi-objective constrained optimization problem. Traditional bilingual sentence pair contrastive learning only establishes local alignment between the two semantic spaces, and its loss landscape contains multiple equivalent minimum basins. The network may converge to a suboptimal state where Chinese-English alignment is good but Japanese alignment is poor. Introducing a joint constraint in three directions is equivalent to imposing a geometric consistency condition in the parameter space, i.e., the three similarity functions must simultaneously satisfy the triangle inequality. This constraint restricts the feasible region to the intersection of the three loss basins, reducing the probability of the network getting trapped in bad local optima. The dimensionality reduction distribution of t-SNE (t-distributed Stochastic Neighbor Embedding) is shown in Figure 2.

Figure 2 on the left shows the t-SNE dimensionality reduction distribution of Chinese, Japanese, and English samples in the embedding space before fine-tuning. Colored by language, the three colored point groups are interspersed and have no clear boundaries, indicating semantic mixing. The right side shows the distribution after fine-tuning, colored by semantic clusters, forming five separate and compact clusters: Cluster 1 is for mechanical repair sentences; Cluster 2 is for electrical engineering sentences; Cluster 3 is for programming interface sentences; Cluster 4 is for everyday conversation sentences; Cluster 5 is for legal and regulatory sentences. Within each cluster, the samples of the three languages are closely clustered, demonstrating that the model successfully brings the translated sentence pairs closer together and pushes different semantics apart, achieving cross-language alignment with semantics as the primary organizing principle.

III. TRILINGUAL TRANSLATION DATABASE BASED ON SEMANTIC SIMILARITY FILTERING AND ENGINEERED CONSTRUCTION

A. SIMILARITY FILTERING ALGORITHM

After model fine-tuning, the parameters are frozen and deployed in the semantic filtering and alignment determination process for large-scale candidate trilingual parallel sentence pairs. The candidate data source is a set of sentence pair triples generated by cross-language retrieval coarse matching of Chinese, Japanese, and English monolingual texts crawled from multilingual news websites, open-source subtitle libraries, and technical document forums, denoted as ξ. The sentences in each candidate triple already have potential mutual translation relationships in the coarse matching stage, but there are a large number of incorrect pairings introduced by polysemy, non-literal correspondence, and crawling noise.

First, the fine-tuned InfoXLM encoder extracts sentence vectors for each sentence in ξ. Taking the Chinese sentence c_i^{zh} as an example, after word segmentation and sub-word encoding, it is input into the model, and the normalized projection [CLS] vector is taken as the sentence representation u_i^{zh}. The same method is used to obtain u_i^{ja} and u_i^{en} for Japanese and English sentences. All vectors are pre-computed and persistently stored in a memory-mapped file to avoid redundant encoding overhead.

Subsequently, cross-linguistic semantic similarity is calculated for each candidate triplet. Using Chinese as the source, the cosine similarity with the candidate Japanese and English translations is calculated separately [34-35]:

$$\begin{aligned} s_i^{zh\leftrightarrow ja} &= \frac{u_i^{zh} \cdot u_i^{ja}}{\|u_i^{zh}\|\|u_i^{ja}\|}, \\ s_i^{zh\leftrightarrow en} &= \frac{u_i^{zh} \cdot u_i^{en}}{\|u_i^{zh}\|\|u_i^{en}\|} \end{aligned} \tag{6}$$

In Formula (6), u_i^{zh}, u_i^{ja}, and u_i^{en} are the Chinese, Japanese, and English sentence vectors of the i-th candidate sentence pair, respectively, and $\|\cdot\|$ represents the L2 norm of the vector. Both similarity scores are in the range [-1, 1].

To determine whether the candidate triples constitute a valid translation pair, consistency filtering is implemented. The similarity between the Chinese and Japanese sentence vectors, $s_i^{zh\leftrightarrow ja}$, is not lower than the threshold θ, and the similarity between the Chinese and English sentence vectors, $s_i^{zh\leftrightarrow en}$, is also not lower than θ. The similarity between the Japanese and English sentence vectors, $s_i^{ja\leftrightarrow en}$, is not lower than θ. The final admission criteria are:

$$\begin{aligned} &(s_i^{zh\leftrightarrow ja} \geq \theta) \wedge (s_i^{ja\leftrightarrow en} \geq \theta) \\ &\wedge (s_i^{zh\leftrightarrow en} \geq \theta) \end{aligned} \tag{7}$$

In Formula (7), $\wedge$ represents the logical AND operation, and the threshold θ is determined to be 0.82 through parameter tuning on the validation set. Only when the directional similarity of the three language pairs simultaneously meets the threshold constraint is the triple accepted as a high-quality translation pair; otherwise, it is filtered out. This bidirectional consistency constraint guarantees the semantic translatability between any two languages within the database.

The filtered valid triples are written to the database in a structured format. Each record is assigned a globally unique identifier, which is constructed by concatenating the language prefix with the SHA-256 (Secure Hash Algorithm 256-bit) hash value. Physical storage adopts the Apache Parquet columnar format, storing the original text, sentence vector index, and metadata fields separately by language. The compression algorithm used is Snappy. To support

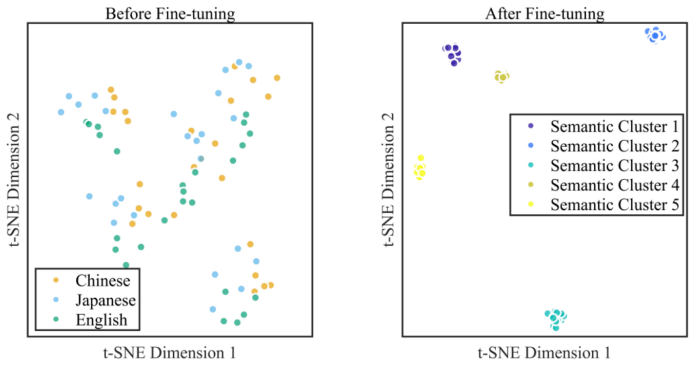

FIGURE 2. Evolution of the t-SNE dimensionality reduction distribution of trilingual semantic clusters in the embedding space

efficient cross-language retrieval, an inverted index is built for the Chinese column, and a full-text search index is built for the Japanese and English columns. The indexing engine is based on Lucene. The distribution of cross-language semantic similarity scores for candidate sentences is shown in Figure 3.

Figure 3 shows the cosine similarity density distribution of candidate sentence pairs in the trilingual crosslinguistic semantic similarity filtering stage. All three density curves show obvious unimodal characteristics. The density peak in the Chinese-English direction is slightly higher than the other two groups, while the peak height in the Japanese-English direction is slightly lower and the tail is extended slightly longer, indicating that the similarity dispersion of Japanese and English sentences in the vector space is relatively large.

B. ENGINEERED STORAGE AND INDEXING

The process of using semantic similarity to solve the alignment issue requires various engineering procedures before developing a practical database, including data cleaning, deduplication, quality assurance and the design of a storage facility. Cleaning of candidate triple sets commences with the original name of each candidate triple being cleaned first: standardizing the NFKC (Normalization Form KC) encoding, removing any control characters, verifying if the language and frame of reference is represented correctly and removing samples with incorrect identifiers via fastText language detection; eliminating misaligned samples having unusual ratios of length; removing duplicate sentences in all languages using MinHash and Locality Sensitive Hashing techniques. Redundant copies from multiple sources are resolved through graph clustering: candidate triples are treated as nodes, with edges being established if the Chinese characters match and the Japanese and English languages have similarities ≥ 0.85; then a connected component retains the candidate triple with the highest overall quality score, with redundant candidates going into the candidate pool. There are three levels of quality control: the Level 1 - every sample subjected to automatic constraints fails and are passed onto a status awaiting an expert review; Level 2 - professional experts evaluate a sample of 5 000 using their domain specific expertise to attribute semantic equivalence, terminology consistency and fluency, thereby establishing thresholds for future assessments using $< 2\%$ error rates; Level 3 - an open community user feedback mechanism is established to facilitate innovation through iterative model improvement.

The indexing layer employs three heterogeneous indexes: the inverted index supports keyword retrieval in Chinese, Japanese, and English; the vector index uses an IVF_PQ (Inverted File with Product Quantization) structure; and the primary key index is based on a combination of SHA-256 prefix and auto-incrementing integers. Physical storage uses a hybrid architecture: raw text and metadata are stored in Apache Parquet columnar format, partitioned by domain, with Snappy compression algorithm used, totaling approximately 6.2 GB of storage; sentence vectors are stored independently in index files, while a NumPy backup of the original floating-point numbers is retained; the inverted index is managed by Lucene and stored in a local directory; dynamic fields from user feedback and metadata are stored in SQLite tables. The database provides three query interfaces: vector retrieval, keyword retrieval, and batch export. All interfaces are encapsulated through a REST (Representational State Transfer) style API (Application Programming Interface), supporting API key rate limiting.

IV. EXPERIMENTAL SETUP AND DATA PREPARATION

A. EXPERIMENTAL ENVIRONMENT AND HYPERPARAMETER CONFIGURATION

The hardware environment consists of a single machine with dual NVIDIA A100 80GB GPUs (Graphics Processing Units), an AMD EPYC 7742 64-core CPU (Central Processing Unit), and 512GB of RAM. The software environment uses PyTorch 2.0.1 and Transformers 4.31.0 as the training framework, with CUDA (Compute Unified Device Architecture) version 11.8. InfoXLM pre-trained weights are loaded from the HuggingFace model library. Stanza 1.5.0 and Camphr 0.7.2 parsers handle Chinese-English and Japanese dependency structures, respectively.

During the fine-tuning phase, the batch size is set to

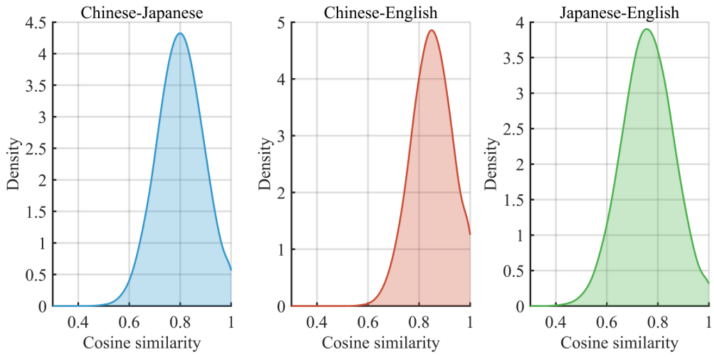

FIGURE 3. Distribution of cross-linguistic semantic similarity scores for candidate sentences

64, and the maximum sequence length is truncated to 128 tokens. The optimizer used is AdamW, with a weight decay coefficient of 0.01 and an initial learning rate of 3×10^{-5}. After a linear warm-up of the first 800 steps, the learning rate decays to 3×10^{-7} using a cosine curve. The triplet loss interval hyperparameter α is set to 0.3, and the temperature coefficient τ is set to 0.07. The three-way loss weights λ_1, λ_2, and λ_3 are set to 0.35, 0.35, and 0.30, respectively, and the augmentation loss coefficient γ is set to 0.2. The training consists of 15 epochs. At the end of each epoch, the alignment recall is evaluated on the validation set, and the optimal checkpoint is retained for subsequent database construction.

B. DATASET CONSTRUCTION AND PARTITIONING STRATEGIES

The initial trilingual anchor data is taken from the TED-Multilingual-Parallel-Corpus multilingual caption collection, which contains 389,764 parallel sentence pairs in Chinese, Japanese, and English (https://github.com/ajinkyakulkarni14/TED-Multilingual-Parallel-Corpus), and 1,620 sentence pairs in Chinese-Japanese, Chinese-English, and Japanese-English from WikiMatrix (https://github.com/facebookresearch/LASER/tree/main/tasks/WikiMatrix), totaling 391,384 pairs. The TED caption collection and WikiMatrix are general-domain resources covering science lectures and general knowledge entries. To incorporate domain-specific technical content, the training anchor set is augmented with the JPO Patent Corpus (http://lotus.kuee.kyoto-u.ac.jp/WAT/patent/jpc2018.html). This corpus consists of a Chinese-Japanese and English-Japanese patent description corpus of 1,000,000 parallel sentences with four sections: Chemistry, Electricity, Mechanical Engineering, and Physics, each accounting for 25% of the data, based on the International Patent Classification. The Mechanical Engineering section of the JPO Patent Corpus directly addresses the technical documentation scenarios referenced in the introduction, including maintenance procedures and equipment descriptions. The initial trilingual anchor set of 391,384 pairs is augmented with the Chinese-Japanese and English-Japanese portions of the JPO Patent Corpus, bringing the total number of trilingual anchor points to 2,391,384 across all three languages before candidate data processing. Candidate data consists of monolingual texts scraped from public caption libraries, which undergo coarse matching via cross-language retrieval to generate 12,370,000 candidate triples. The dataset is randomly divided into a training set of 1,913,107 pairs, a validation set of 239,138 pairs, and a test set of 239,139 pairs in an 8:1:1 ratio.

C. BASELINE MODEL

This paper selects two cross-lingual alignment methods as comparative baselines. LASER: This is LASER zerosample alignment, which uses the Facebook LASER2 multilingual encoder to extract sentence vectors and directly calculates cosine similarity for alignment determination, without task-specific fine-tuning. mBERT: This is mBERT fine-tuned alignment, which, based on bert-base-multilingual-cased, uses three sets of parallel sentence pairs from the training set (Chinese-Japanese, Chinese-English, and Japanese-English) to fine-tune the bilingual sentence pair classification heads respectively. The three-way alignment result is obtained by merging the average of the three classification scores.

V. RESULTS

A. EVALUATION OF TRILINGUAL ALIGNMENT PRECISION, RECALL, AND F1 SCORE

This section evaluates the trilingual alignment performance of the proposed method on the test set. The test set is completely isolated from the training set, and each sample pair is manually verified to ensure a strict mutual translation relationship. During evaluation, the fine-tuned InfoXLM encoder is frozen; sentence vectors are extracted for each Chinese, Japanese, and English sentence in the test set; and the cosine similarity of each candidate triplet in the Chinese-Japanese, Chinese-English, and Japanese-English directions is calculated. Based on the three-way consistency threshold θ=0.82, a triplet is considered a valid mutual translation pair only if the similarity in all three directions is not lower than this threshold. Based on the comparison between the judgment results and the true labels, statistical precision, recall, and F1 score are used as the core evaluation metrics. The comparison groups include the proposed method, LASER, and mBERT. Table 1 lists the precision, recall, and F1 score

TABLE 1. Performance comparison of the proposed method and baseline methods on the trilingual alignment task

Method	Language Direction	Precision (%)	Recall (%)	F1 Score (%)
Proposed Method	Trilingual joint	94.7	92.3	93.5
LASER	Trilingual joint	71.4	68.9	70.1
mBERT	Trilingual joint	82.6	79.3	80.9

of the three methods on the trilingual joint alignment task.

As shown in Table 1, with a threshold of θ= 0.82 , the proposed method achieves a precision of 94.7%, a recall of 92.3%, and an F1 score of 0.935 on the trilingual joint alignment task. The performance of the LASER and mBERT groups is lower than that of the proposed method in all metrics. This result indicates that the model successfully maps equivalent semantics of Chinese, Japanese, and English to adjacent regions in the embedding space, and that cross-lingual semantic similarity calculation is sufficiently robust to syntactic variations and lexical interference. To examine the performance of the method under different levels of strictness, the precision and recall corresponding to each threshold are recorded using the similarity threshold θas a variable parameter, and precision-recall curves are plotted. Figure 4 shows the precision-recall curves of the proposed method under different similarity thresholds.

Figure 4 shows that when the threshold is 0.78, the recall rate increases to 96%, but the precision rate drops to 86.6%, indicating that low-quality candidate pairs are mistakenly included. When the threshold is 0.86, the precision rate reaches 96%, while the recall rate is 88%, indicating that overly strict constraints filter out some semantically correct but slightly low vector similarity real translation pairs, resulting in a waste of data resources.

B. ABLATION EXPERIMENTS

To quantify the independent contribution of each key component in the proposed method to the trilingual alignment performance, five ablation variants are set up for comparison with the full model: removing the syntactic perturbation enhancement module (w/o Aug), removing the Chinese-Japanese constraint (w/o zhja), removing the Chinese-English constraint (w/o zh-en), removing the Japanese-English constraint (w/o ja-en), and replacing the hierarchical triplet loss with an independent bilingual sentence pair contrast loss (w/ Bi-Encoder). All variants are independently trained to convergence under the same training set, validation set, and hyperparameter configuration. The evaluation metric is the trilingual joint alignment F1 score. Figure 5 shows the comparison of the F1 scores of the full model and the five ablation variants over ten rounds of experiments.

According to the ablation experiment results in Figure 5, the mean F1 score of the complete model in ten rounds of experiments is 93.48%, with a minimum of 92.07% and a maximum of 95.44%. After removing the syntactic perturbation enhancement module, the mean F1 score drops to 86.69%; after replacing the hierarchical triplet loss with the independent bilingual sentence pair contrast loss, the mean F1 score drops to 83.54%, the lowest among all variants. When removing the Chinese-Japanese, Chinese-English, or Japanese-English single constraints respectively, the mean F1 scores are 89.14%, 88.66%, and 90.63%, respectively, all lower than the complete model. From the model mechanism perspective, the syntactic perturbation enhancement, by generating code-switching hard negative examples, forces the model to learn deeper semantic representations beyond the surface word order. After its removal, the model's robustness to natural word order variants decreases. Among the three-way constraints, the Chinese-English direction contributes relatively more, which is related to English's mediating role in connecting the semantic spaces of Chinese and Japanese; although the Japanese-English direction contributes slightly less, it still plays a moderating role in the overall tightness of the trilingual semantic clusters. Independent bilingual encoders, lacking a unified joint embedding basis, exhibit the lowest performance due to increased false positive rates in three-way consistency filtering. In summary, the overall design of the proposed method demonstrates a clear synergistic gain effect in cross-lingual semantic alignment. Comprehensive ablation results show that the syntactic perturbation enhancement module, the three-way integrity of the hierarchical triplet loss, and the joint training strategy are the core pillars of the proposed high-performance method; the absence of any one component leads to a degradation in the quality of trilingual alignment.

C. SEMANTIC CLUSTER COHESION AND SEPARABILITY

This section evaluates the effect of hierarchical triplet contrastive learning mechanism on reshaping the semantic distribution of three languages from the perspective of embedded spatial geometry. Five thousand pairs of trilingual translation sentences are randomly selected from the test set, and sentence vectors are extracted using InfoXLM encoders before and after fine-tuning, resulting in two sets of semantic representation sequences (before and after fine-tuning). The evaluation metrics used are the silhouette coefficient and the Davies-Bouldin index, which measure the cohesion and separability of semantic clusters, respectively. Figure 6 shows the comparison between the semantic cluster silhouette coefficient and the Davies-Bouldin index before and after fine-tuning.

Figure 6 shows the comparison results of semantic cluster cohesion and separability before and after fine-tuning. In Figure 6(a), the mean silhouette coefficient increases from 0.416 before fine-tuning to 0.790 after fine-tuning; in Figure 6(b), the mean Davies-Bouldin index decreases from 1.867 to 0.928. The changes in these two indicators indicate that after processing by the method presented in this paper, the trilingual samples with the same semantics are more tightly clustered in the embedding space, and the bound-

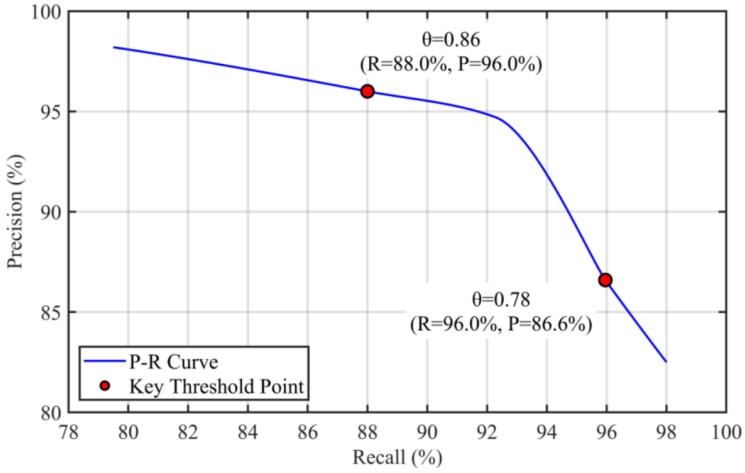

FIGURE 4. Precision-Recall curves for the trilingual alignment task (based on the method in this paper)

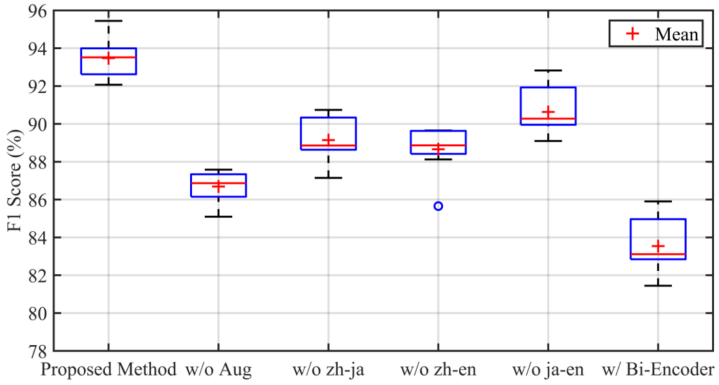

FIGURE 5. Comparison of F1 values in ablation experiments (10 rounds of experiments)

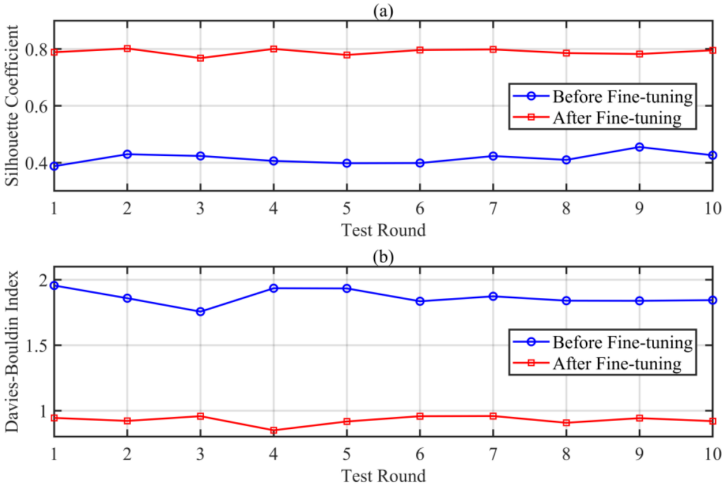

FIGURE 6. Comparison of semantic cluster cohesion and separability before and after fine-tuning: (a) Comparison of semantic cluster contour coefficients before and after fine-tuning; (b) Comparison of semantic cluster Davies-Bouldin index before and after fine-tuning

aries between different semantic clusters are clearer. This change stems from the combined effect of hierarchical triplet contrastive learning and syntactic awareness. Before fine-tuning, the pre-trained model exhibits a mixed distribution of language-dominated trilingual representations due to significant differences between Chinese, Japanese, and English

family systems. During fine-tuning, the three-way triplet loss forces equivalent sentence vectors to move closer together and non-corresponding samples to move further apart. The syntactic perturbation module further enhances the model's robustness to syntactic variations, thereby dissolving linguistic typological differences under semantic equivalence, ultimately forming a cross-lingual alignment space with semantic content as the primary organizing principle.

D. ROBUSTNESS UNDER DIFFERENT TEXT LENGTHS AND SYNTACTIC COMPLEXITIES

This section evaluates the alignment performance stability of the proposed method under different text lengths and syntactic complexities. The test set is divided into five groups based on the number of

Chinese sentence characters: short sentences (4-20 characters), medium-short sentences (21-40 characters), medium-long sentences (41-60 characters), long sentences (61-80 characters), and very long sentences (>80 characters). It is also divided into three groups based on the maximum depth of the dependency tree: low complexity (1-5 characters), medium complexity (6-10 characters), and high complexity ($\geq$11 characters). The evaluation metric is the F1 score of the trilingual joint alignment. Figure 7 compares the F1 scores of the proposed method, mBERT fine-tuned alignment, and LASER zero-sample alignment under different syntactic complexities.

Figure 7 compares the F1 scores of the proposed method, mBERT fine-tuned alignment, and LASER zero-shot alignment under different syntactic complexities. The proposed method achieves an F1 score of 95.03% at low complexity (dependency tree depth 1-5), 93.49% at medium complexity (depth 6-10), and 91.63% at high complexity (depth $\geq$ 11), with a difference of 3.40 percentage points from low to high. The F1 scores of mBERT fine-tuned alignment are 83.91%, 80.19%, and 74.36% for the three groups, with a difference of 9.55 percentage points. The corresponding values for LASER zero-shot alignment are 73.32%, 65.42%, and 57.35%, with a difference of 15.97 percentage points. The data indicates that the proposed method maintains a good F1 score under high complexity conditions, and the performance degradation with increasing syntactic complexity is smaller than that of the two comparative methods. This advantage is primarily attributed to the dependency syntactic perturbation and code-switching enhancement module introduced in this paper. This module actively generates syntactic variant samples during training, such as modifier position substitutions and core argument rearrangements, giving the model stronger semantic invariance to changes in deep dependency structures. Simultaneously, the hierarchical triple contrastive loss forces equivalent sentence vectors from Chinese, Japanese, and English to form compact semantic clusters in the embedding space. Even with highly complex sentences, the model can stably extract deep semantic representations rather than relying on surface word order, thus effectively suppressing the alignment performance degradation caused by increased syntactic complexity.

E. DATABASE CONSTRUCTION QUALITY ASSESSMENT

Based on a semantic similarity filtering process, 3.8 million high-quality trilingual translation pairs are selected from 12.37 million candidate triplet pairs to form the final database. This section conducts an independent quality assessment of the final database of 3.8 million trilingual translation pairs. 10,000 trilingual translation records are randomly selected from the database, covering four text types: news, science and technology, law, and everyday conversation. The fluency of the translations is assessed using the BLEU-4 (Bilingual Evaluation Understudy-4) index. The mean, standard deviation, score distribution interval, and sample proportion of the BLEU-4 scores for the Chinese-Japanese, Chinese-English, and Japanese-English language pairs are calculated respectively. Table 2 shows the BLEU-4 evaluation results for the three language pairs.

Table 2 shows that the mean BLEU-4 score for the Chinese-Japanese translation is 32.4, and for the Chinese-English translation it is 35.8. Approximately 67% of the samples for the Chinese-Japanese translation has BLEU-4 scores between 28 and 38, while approximately 72% of the samples for the Chinese-English translation has scores between 32 and 42. The slightly higher scores for the Chinese-English translation are consistent with the prior knowledge that the Chinese and English language structures are relatively similar and that there is less information loss during literal translation. For the Japanese-English translation, the mean BLEU-4 score is 30.2, slightly lower than the 32.4 for the Chinese-Japanese translation and the 35.8 for the Chinese- English translation. Its score distribution range is [25, 36], with approximately 63% of the samples falling within this range, a lower proportion than the Chinese-Japanese and Chinese-English translations, indicating that semantic alignment between Japanese and English presents a certain challenge in terms of surface matching in the translation. This is mainly due to syntactic differences, with inconsistent word order and morphological changes during literal translation.

In addition to measuring the quality of translations, this study presents supplementary measures to assess the database from an engineering perspective. Five indicators are evaluated: query response time, storage efficiency, data integrity, database coverage, and manual verification pass rate. Coverage is defined as the ratio of the number of trilingual pairs accepted into the final database to the estimated total number of valid translation triples present in the raw candidate set. The estimated total number of valid triples is derived from a manually annotated random sample of 10,000 candidate triples. Three professional annotators judge each candidate triple as a valid mutual translation or not. In the sampled 10,000 candidate triples, 3,420 triples are confirmed as valid, giving an estimated valid proportion of 34.2% for the entire candidate set of 12.37 million triples, which yields 4.23 million estimated valid triples.

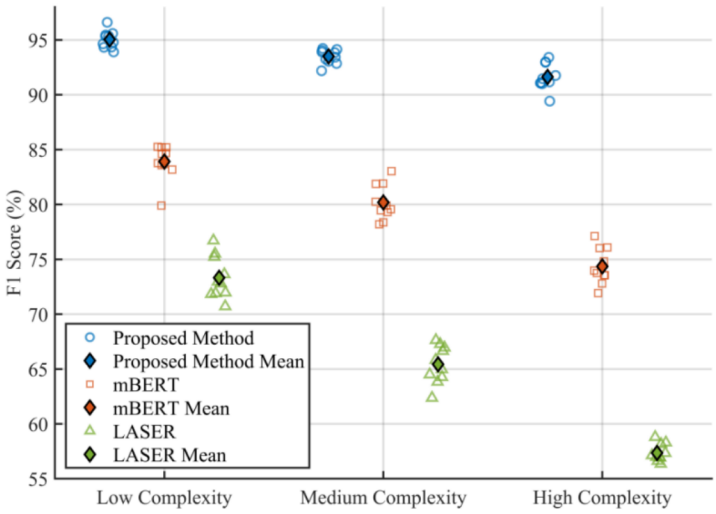

FIGURE 7. Comparison of alignment F1 scores under different syntactic complexities (10 rounds of experiments)

TABLE 2. Fluency assessment of translations from the trilingual translation database (BLEU-4)

Language Pair	BLEU-4 Mean	BLEU-4 Std Dev	BLEU-4 Score Distribution Interval	Proportion of Samples within Interval
Chinese-Japanese	32.4	5.1	[28, 38]	67%
Chinese-English	35.8	5.3	[32, 42]	72%
Japanese-English	30.2	5.6	[25, 36]	63%

The final database contains 3.8 million trilingual pairs, and the coverage is computed as 3.8 million divided by 4.23 million, resulting in 89.8%. The results of the core quality evaluation of the database are shown in Table 3.

Table 3 shows that the database satisfies all five dimensions of the criteria satisfactorily. The manual verification ratio of 96.3% demonstrates good reliability, and the average latency for vector retrieval is 38 ms. This indicates it can support real-time interactions. The total data storage size of 6.2 GB indicates an acceptable level of efficiency in storage; 99.8% of the data maintains integrity following cleaning/encoding, which indicates that the cleaning and encoding process is effective. The database has an overall coverage rate of 89.8%, demonstrating that the cleaning and sorting method can identify and retain the majority of valid trilingual translation pairs from the source candidate data. As such, it has demonstrated usability as measured by accuracy, efficiency, scalability, and completeness.

F. EVALUATION OF APPLICATION EFFECTIVENESS IN CONTINUING ENGINEERING EDUCATION SCENARIOS

This section verifies the practical support capability of the constructed database for continuing engineering education scenarios. The participants consist of 30 Chinese engineers and 20 Japanese engineers. Each participant completes two tasks: Task A is "Source-Translation Matching" (given a Chinese equipment maintenance manual, the most accurate translation is selected from three candidate Japanese translations), and Task B is "Cross-Language Terminology Location" (given English technical terms, the corresponding expressions in Japanese circuit diagram descriptions is highlighted). The comparison group includes the proposed method and two alternative cross-lingual embedding approaches LASER zero-shot alignment and mBERT fine-tuned alignment both described in the Baseline Model section. Evaluation metrics include: matching accuracy of Task A, average location time of Task B, the proportion of participants who agree on "reducing the burden of understanding complex syntactic structures," and the semantic mapping accuracy of the English technical terms in Task B by the Japanese engineers. Table 4 shows the evaluation results of these metrics.

Table 4 shows that the proposed method achieves a Task A matching accuracy of 91.3%, higher than the 65.8% of LASER and the 78.4% of mBERT. In Task B, the average localization time for participants using the proposed method is 28.4 seconds, compared to 61.2 seconds for LASER and 44.7 seconds for mBERT. Postexperience interviews indicate that 87% of engineers using the proposed method report that the trilingual alignment examples reduce the comprehension burden of complex syntactic structures, whereas only 31% of LASER users and 56% of mBERT users express the same agreement. Japanese engineers achieve an 89.5% semantic mapping accuracy for English technical terms with the proposed method, outperforming the 59.6% accuracy of LASER and the 72.3% accuracy of mBERT.

TABLE 3. Database core quality assessment results

Evaluation Dimension	Metric	Measured Value
Alignment Accuracy	Human Validation Pass Rate (Semantic Equivalence)	96.3%
Query Performance	Average Latency of Vector Retrieval	38 ms
Storage Footprint	Total Storage Usage (including indexes)	6.2 GB
Coverage	Estimated Coverage of Valid Triples in Candidate Set	89.8%
Data Integrity	Proportion of Records with Complete Fields and Valid Encoding	99.8%

TABLE 4. Evaluation results of application effectiveness in continuing engineering education scenarios

Evaluation Metric	Proposed Method	LASER	mBERT
Task A: Source-translation matching accuracy	91.3%	65.8%	78.4%
Task B: Cross-lingual term localization average time	28.4 seconds	61.2 seconds	44.7 seconds
Proportion of participants who agreed that the system "significantly reduces the comprehension burden of complex syntactic structures"	87%	31%	56%
Japanese engineers' semantic mapping accuracy for English technical terms	89.5%	59.6%	72.3%

VI. CONCLUSION

This paper proposes a hierarchical triplet contrastive learning and syntactic perturbation enhancement method based on artificial neural networks to construct a Chinese-Japanese-English trilingual translation database. Using InfoXLM as the semantic encoder, code-switching hard negative examples are generated through dependency syntactic perturbation and cross-linguistic lexical substitution. A three-way triplet contrastive loss is introduced for joint fine-tuning, forcing equivalent sentence vectors to aggregate into semantic clusters that separate non-corresponding samples in the embedding space. The fine-tuning model performs vector similarity filtering and bidirectional consistency judgment on candidate sentence pairs, constructing a database containing 3.8 million trilingual translation sentence pairs. On the test set, the sentence-level alignment precision is 94.7%; recall is 92.3%; F1 score is 93.5%. The database translation fluency scores (BLEU-4) are 32.4, 35.8, and 30.2 in the Chinese-Japanese, Chinese-English, and Japanese-English directions, respectively, with a human verification pass rate of 96.3%. This database provides data support for multilingual technical document retrieval in cross-linguistic information processing and engineering education scenarios.

AUTHOR CONTRIBUTIONS

Conceptualization – X.W.; methodology – D.W.; formal analysis –D.W.; investigation –X.W.; resources –X.W.; writing-original draft preparation – X.W. and D.W.; All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

REFERENCES

[1] C. Mi and S. Xie, "Language relatedness evaluation for multilingual neural machine translation," Neurocomputing, vol. 570, no. 1, Art. no. 127115, 2024, doi: 10.1016/j.neucom.2023.127115 .

[2] J. Luo, C. Cherry, and G. Foster, "To diverge or not to diverge: A morphosyntactic perspective on machine translation vs human translation," Transactions of the Association for Computational Linguistics, vol. 12, no. 1, pp. 355-371, 2024, doi: 10.1162/tacl_a_00645.

[3] S. Ranathunga, E. S. A. Lee, M. Prifti Skenduli, et al., "Neural machine translation for low-resource languages: A survey," ACM Computing Surveys, vol. 55, no. 11, pp. 1-37, 2023, doi: 10.1145/3567592.

[4] F. Li, C. Chi, H. Yan, et al., "STA: An efficient data augmentation method for low-resource neural machine translation," Journal of Intelligent & Fuzzy Systems, vol. 45, no. 1, pp. 121-132, 2023, doi: 10.3233/jifs-230682.

[5] A. Gui and H. Xiao, "Multi-level multilingual semantic alignment for zero-shot cross-lingual transfer learning," Neural Networks, vol. 173, no. 1, pp. 106217-106231, 2024, doi: 10.1016/j.neunet.2024.106217.

[6] Z. Mao, C. Chu, and S. Kurohashi, "Linguistically driven multi-task pre-training for low-resource neural machine translation," Transactions on Asian and Low-Resource Language Information Processing, vol. 21, no. 4, pp. 1-29, 2022, doi: 10.1145/3491065.

[7] H. Yanaka and K. Mineshima, "Compositional evaluation on Japanese textual entailment and similarity," Transactions of the Association for Computational Linguistics, vol. 10, no. 1, pp. 1266-1284, 2022, doi: 10.1162/tacl_a_00518.

[8] J. Zhang, Y. Tian, J. Mao, et al., "WCC-JC 2.0: A web-crawled and manually aligned parallel corpus for Japanese-Chinese neural machine translation," Electronics, vol. 12, no. 5, pp. 1140-1157, 2023, doi: 10.3390/electronics12051140.

[9] A. Fernando, S. Ranathunga, D. Sachintha, et al., "Exploiting bilingual lexicons to improve multilingual embedding-based document and sentence alignment for low-resource languages," Knowledge and Information Systems, vol. 65, no. 2, pp. 571-612, 2023, doi: 10.1007/s10115-022-01761-x.

[10] S. Zhu, S. Gu, S. Li, et al., "Mining parallel sentences from internet with multi-view knowledge distillation for low-resource language pairs," Knowledge and Information Systems, vol. 66, no. 1, pp. 187-209, 2024, doi: 10.1007/s10115-023-01925-3.

[11] Z. Li, J. Ma, and F. Ren, "Exploiting Multi-Level Data Uncertainty for Japanese-Chinese Neural Machine Translation," IEICE Transactions on Information and Systems, vol. 108, no. 5, pp. 440-443, 2024, doi: 10.1587/transinf.2024edl8059.

[12] Z. Man, Y. Zhang, Y. Li, et al., "An ensemble strategy with gradient conflict for multi-domain neural machine translation," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 23, no. 2, pp. 1-22, 2024, doi: 10.1145/3638248 .

[13] J. Zhang, K. Su, H. Li, et al., "Neural machine translation for low-resource languages from a chinese-centric perspective: A survey," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 23, no. 6, pp. 1-60, 2024, doi: 10.1145/3665244.

[14] J. Zhang, K. Su, Y. Tian, et al., "WCC-EC 2.0: enhancing neural machine translation with a 1.6 M+ web-crawled English-Chinese parallel corpus," Electronics, vol. 13, no. 7, pp. 1381-1392, 2024, doi: 10.3390/electronics13071381.

[15] R. Choenni and E. Shutova, "Investigating language relationships in multilingual sentence encoders through the lens of linguistic typology,"

Computational Linguistics, vol. 48, no. 3, pp. 635-672, 2022, doi: 10.1162/coli_a_00444.
[16] Z. Mao, C. Chu, and S. Kurohashi, "Ems: Efficient and effective massively multilingual sentence embedding learning," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, no. 1, pp. 2841-2856, 2024, doi: 10.1109/TASLP.2024.3402064.
[17] V. Mathur, T. Dadu, and S. Aggarwal, "Evaluating neural networks' ability to generalize against adversarial attacks in crosslingual settings," Applied Sciences, vol. 14, no. 13, pp. 5440-5463, 2024, doi: 10.3390/app14135440.
[18] G. Tang, O. Yousuf, and Z. Jin, "Improving BERTScore for machine translation evaluation through contrastive learning," IEEE access, vol. 12, no. 1, pp. 77739-77749, 2024, doi: 10.1109/access.2024.3406993.
[19] L. Ding, L. Wang, and S. Liu, "Recurrent graph encoder for syntaxaware neural machine translation," International Journal of Machine Learning and Cybernetics, vol. 14, no. 4, pp. 1053-1062, 2023, doi: 10.1007/s13042-022-01682-9.
[20] F. Azadi, H. Faili, and M. J. Dousti, "Mismatching-aware unsupervised translation quality estimation for low-resource languages," Languages Resources and Evaluation, vol. 58, no. 4, pp. 1207-1231, 2024, doi: 10.1007/s10579-024-09727-x.
[21] B. Xing and I. W. Tsang, "HC^{2}2L: Hybrid and Cooperative Contrastive Learning for Cross-Lingual Spoken Language Understanding," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 12, pp. 8094-8105, 2024, doi: 10.1109/TPAMI.2024.3402746.
[22] A. Banerjee, B. K. Singh, V. Kumar, et al., "Research Challenges and Future Directions in Transformer-Based Neural Machine Translation," Expert Systems with Applications, vol. 1, no. 1, Art. no. 131062, 2026, doi: 10.1016/j.eswa.2025.131062.
[23] I. Sel and D. Hanbay, "Efficient adaptation: enhancing multilingual models for low-resource language translation," Mathematics, vol. 12, no. 19, pp. 3149-3156, 2024, doi: 10.3390/math12193149.
[24] G. Tolegen, A. Toleu, and R. Mussabayev, "Contrastive learning for morphological disambiguation using large language models in low-resource settings," Applied Sciences, vol. 14, no. 21, pp. 9992, 2024, doi: 10.3390/app14219992 .
[25] X. Chen, Y. Yang, and H. Hu, "Automatic evaluation of English translation based on multi-granularity interaction fusion," Neural Processing Letters, vol. 57, no. 1, pp. 19-26, 2025, doi: 10.1007/s11063-025-11716-2.
[26] V. K. Pant, R. Sharma, and S. Kundu, "An overview of stemming and lemmatization techniques," Advances in networks, intelligence and computing, vol. 1, no. 1, pp. 308-321, 2024, doi: 10.1201/9781003430421-31.
[27] J. Nivre, A. Basirat, L. Dürlich, et al., "Nucleus composition in transition-based dependency parsing," Computational Linguistics, vol. 48, no. 4, pp. 849-886, 2022, doi: 10.1162/coli_a_00450.
[28] S. Duan, H. Zhao, and D. Zhang, "Syntax-aware data augmentation for neural machine translation," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, no. 1, pp. 2988-2999, 2023, doi: 10.1109/taslp.2023.3301214.
[29] L. Li, A. Zhang, and M. X. Luo, "DSISA: A new neural machine translation combining dependency weight and neighbors," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 23, no. 2, pp. 1-16, 2024, doi: 10.1145/3638762.
[30] X. Liang, Z. Gu, Y. Xie, L. Wang, and Z. Tian, "MUSEDA: multilingual unsupervised and supervised embedding for domain adaption," Knowledge-Based Systems, vol. 273, no. 1, pp. 110560-110578, 2023, doi: 10.1016/j.knosys.2023.110560.
[31] Q. Tan, X. Song, G. Ye, et al., "An effective negative sampling approach for contrastive learning of sentence embedding," Machine Learning, vol. 112, no. 12, pp. 4837-4861, 2023, doi: 10.1007/s10994-023-06408-8.
[32] B. Kairatuly and A. Shomanov, "Balancing speed and performance with layer freezing strategies for transformer models," Scientific Journal of Astana IT University, vol. 22, no. 1, pp. 153-162, 2025, doi: 10.37943/22oxky5402.
[33] T. Hwang, H. Seo, J. Jung, et al., "Exploring Selective Layer Freezing Strategies in Transformer Fine-Tuning: NLI Classifiers with Sub-3B Parameter Models," Applied Sciences, vol. 15, no. 19, pp. 10434-10453, 2025, doi: 10.3390/app151910434.
[34] F. F. Ijebu, Y. Liu, C. Sun, et al., "Soft cosine and extended cosine adaptation for pre-trained language model semantic vector analysis," Applied Soft Computing, vol. 169, no. 1, Art. no. 112551, 2025, doi: 10.1016/j.asoc.2024.112551.
[35] F. R. Zagatti, G. Yuuji Shimizu, D. Lucredio, et al., "Investigating the Relationship Between Text Vectorization Cosine Similarity and Classification Performance," IEEE Access, vol. 13, no. 1, pp. 137348-137363, 2025, doi: 10.1109/access.2025.3595423 .