Applying the mBART Multilingual Model to Optimize Translation Software's Intelligent Error Correction Capabilities in English Multi-domain Terminology Conversion

Pei Wang[1,2] , Chenxi Li[3]
[1]The College of Foreign Languages, Yanbian University, 133002, China
[2]School of Foreign Languages, University of Sanya, Sanya 572000, Hainan, China
[3]The Kyoto College of Graduate Studies for Informatics, Kyoto 606-8225, Japan
Corresponding author: Pei Wang (e-mail: 15104504767@163.com)

ABSTRACT Existing translation systems generally lack dynamic disambiguation capabilities across domains, leading to frequent mistranslations in semantically ambiguous or interdisciplinary scenarios, while their terminology retranslation and error correction mechanisms remain limited. To address these issues, this study proposes an intelligent terminology correction framework based on the mBART multilingual model. By integrating multilingual context modeling, multihead attention mechanisms, and a terminology correction candidate list, the proposed approach dynamically identifies domain-specific meanings and generates accurate correction suggestions. Fine-tuning on parallel corpora from medicine, law, engineering, and electrical engineering enables effective intelligent error correction across multiple professional domains. Such capability is particularly valuable for technical communication involving electromagnetic engineering and related disciplines, where precise terminology translation supports reliable knowledge dissemination and interdisciplinary collaboration. Experimental results demonstrate that the proposed method significantly improves translation quality. Across 15 domains, the average BLEU score reaches 57.9, compared with 38.4 for the baseline model, while the error rate for ambiguous terms decreases to 24.3%. In the electrical engineering domain, terminology recognition accuracy reaches 0.94, and semantic fidelity in engineering applications reaches 96.1%, effectively enhancing context awareness and translation consistency for multi-domain terminology.

INDEX TERMS multilingual context modeling, terminology correction, mBART pretrained model, electromagnetic engineering communication, domain-specific terminology

I. INTRODUCTION

THE acceleration of globalization has promoted interdisciplinary and multilingual information exchange. The accurate translation of professional terms is crucial for promoting knowledge sharing and technological cooperation. As the mainstream language of international academic and business exchanges, the accurate translation of cross-disciplinary professional terms has become a key goal of translation technology development [1,2]. The accuracy of terminology translation is directly related to the effectiveness of information transmission and the rigor of professional communication. However, mainstream translation software is limited by static dictionary architecture and general pre-trained models, making it difficult to achieve dynamic contextual disambiguation of terms, resulting in frequent mistranslations of terms in polysemous or cross-disciplinary contexts, and its error correction and retranslation mechanisms are limited [3,4]. In the field of terminology translation, researchers have adopted methods such as terminology enhancement training, rule constraints, and terminology knowledge base assistance to improve the accuracy of terminology recognition and translation [5,6]. On the other hand, the rise of pre-trained language models has promoted the development of deep learning-based translation systems. Researchers have combined contextual information to model terminology ambiguity, gradually improving the model's contextual understanding ability [7,8]. Bryant et al. [9] provided a methodological reference for text-level fine-grained error correction. Chen [10] verified the effectiveness of data-driven methods in foreign language teaching, but did not address the dynamic disambiguation of professional terms. Aliguliyeva et al. [11] emphasized the

impact of digital technology on language acquisition by non-native speakers, indirectly highlighting the importance of accurate understanding of professional terms. At the level of cross-modal technology, the end-to-end multilingual speech recognition framework proposed by Liang and Yan [12] provides a new approach to speech-text alignment. However, it does not solve the problem of analyzing emotional tone and rhetorical structure in literary listening. Linhares Pontes et al. [13] and El-Alami et al. [14] have demonstrated the potential of multilingual entity linking and transfer learning in cross-lingual term alignment. However, existing research has not adequately addressed the issues of semantic disambiguation and real-time error correction for multi-domain terms in dynamic contexts [15,16]. While HLSTM (Hierarchical Long Short-Term Memory)-based methods can identify the contextual intent of hate words [17,18], they are still insufficient in domain-sensitive modeling of specialized terms.

However, there are still shortcomings in model fine-tuning strategies, context recognition module design, and candidate error correction word list construction, resulting in insufficient adaptability of the model to dynamic contexts of multi-domain terms in practical applications, and the error correction accuracy still needs to be improved [19,20]. Existing research has failed to achieve close integration of Chinese-English vocabulary and terminology translation with intelligent error correction, and lacks a systematic and comprehensive solution [21,22]. At the same time, the polysemy and contextual changes of multi-domain terms are more complex and subtle, which puts forward higher requirements for traditional models. In cross-domain terminology conversion, due to the lack of communication training and in-depth understanding of semantic differences between domains, the model often makes misjudgments, affecting the overall translation effect [23,24]. In addition, existing research still has shortcomings in dealing with terminology consistency and transferability in multilingual environments, which limits the wide applicability of translation systems in international applications [25,26]. Therefore, improving the context perception ability of translation models and combining artificial intelligence technology to build a more accurate and intelligent terminology error correction mechanism has become the focus and difficulty of current research [27]. Current mainstream translation systems have significant limitations in handling specialized terminology: industry benchmark models have a high mistranslation rate in cross-domain terminology translation, especially when dealing with domain-sensitive and polysemous terms, which seriously hinders the transfer of interdisciplinary knowledge.

This paper investigates the problem of contextual ambiguity identification and correction in multi-domain English terminology translation, proposing a dynamic context-aware intelligent terminology correction framework based on the mBART multilingual model. This framework uses a pre-trained mBART model for context modeling and combines semantic similarity and attention mechanisms to design a context discrimination module that dynamically identifies polysemous terms. By combining a multilingual terminology dictionary and knowledge base, a candidate correction word list is generated, achieving semantic alignment and outputting correction suggestions. Furthermore, this paper fine-tunes the model on parallel corpora in fields such as medicine, law, and engineering to improve the accuracy of terminology translation. This research aims to improve the practicality and quality of professional domain translation software by combining dynamic context awareness and intelligent error correction technologies.

As a technology-intensive discipline, engineering employs terminology that is highly polysemous and context-dependent. Mistranslation can affect the accuracy of technical documents and even lead to production deviations or safety accidents. The terminology system in this field is cross-domain coupled, necessitating the support of a context-aware translation system. The application of this research framework in engineering not only verifies the model's robustness in high-tech manufacturing scenarios but also provides linguistic support for international standards and transnational technological collaboration in the industry.

II. CONSTRUCTION AND IMPLEMENTATION OF A MULTILINGUAL TERMINOLOGY CORRECTION MECHANISM

APPLYING MBART FOR MULTILINGUAL CONTEXT MODELING

ENCODING MULTILINGUAL CONTEXT INFORMATION

This study uses the pre-trained mBART model to perform context modeling on English sentences containing professional terminology [28,29]. The specific process first involves word segmentation of the input sentence, and then using the SentencePiece encoder to divide the text into subword units to ensure the model's generalization ability for low-frequency words and new words. The input sequence length is uniformly set to 512 subwords, and any excess is truncated to ensure consistency of the model input. Subsequently, the processed subword sequence is input to the mBART encoder, using its multi-layer Transformer structure to capture contextual information within and across sentences. The mBART model uses a stacked structure of 12 encoder layers and 12 decoder layers, with each layer containing 768-dimensional hidden units and 12 attention heads, ensuring rich semantic feature extraction capabilities.

The contextual semantic vector output by the encoder contains rich contextual information and can effectively express the semantic relationship between terms at different positions in a sentence and between sentences [30]. To further improve the model's ability to capture the context of multi-domain terms, this study performs weighted aggregation on the hidden state output by the mBART encoder and uses a multi-head attention mechanism to perform a weighted summation of the vector of the target term's position and its contextual neighboring word vectors. The calculation formulas are as follows:

$$h_{\text{context}} = \sum_{i=1}^{n} \alpha_i h_i \tag{1}$$

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{n} \exp(e_j)} \tag{2}$$

$$e_i = q^\top W_a h_i \tag{3}$$

h_i represents the hidden vector of the i-th word; q is the query vector; W_a is the trainable parameter matrix; α_i is the normalized attention weight; and n is the context window size. This method enables the model to dynamically focus on contextual components with high semantic relevance to the target term, enhancing its contextual discrimination capabilities.

To accurately model the semantic representation of a term's location, this paper does not employ mean pooling or CLS tagging, but instead designs a weighted attention aggregation mechanism based on a context window. Specifically, a local window is formed by taking five words before and after the target term's position in the input sequence. Then, using the hidden states output by the mBART encoder, a learnable multi-head attention module performs a weighted sum of all word vectors within this window. This attention mechanism uses the target word vector as the query and the word vectors within the window as the keys and values, thus dynamically focusing on the semantically most relevant neighboring context. This aggregation method outperforms static pooling and can capture the local context of terms more precisely.

AUXILIARY MECHANISM FOR DETERMINING TERM MEANING

The proposed "domain-aware multi-head attention mechanism" achieves domain-adaptive adjustment of attention weights by fusing domain label embeddings with word context vectors. Specifically, each input sentence is assigned a learnable domain embedding vector (such as "medical", "legal", etc.) before encoding, and this vector is concatenated with word embeddings and input into the mBART encoder. During the attention calculation stage, the domain embedding is used to generate a domain bias, which is added to the standard attention score [31,32]. This mechanism uses semantic similarity calculation as the basis for discrimination, specifically using the cosine similarity indicator:

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\|\|v_2\|} \tag{4}$$

v_1 represents the contextual semantic vector of the target term in the current sentence, and v_2 is the semantic vector representation of the candidate domain term. By calculating semantic similarity, the model can provide matching signals between the current term and candidate semantics in various domains, thereby supporting subsequent domain-aware discrimination and reducing the risk of misidentification of polysemous words. To enhance the discrimination effect, the design incorporates an attention mechanism, assigning different weights to different semantic feature dimensions during the semantic matching process.

MODEL TRAINING AND PARAMETER SETTING

The mBART model is initialized using large-scale multilingual pre-trained weights and then fine-tuned on parallel corpora covering multiple domains, including medicine and law. During fine-tuning, the Adam optimizer is used with an initial learning rate of 3×10^{-5}, a batch size of 64, and 10 training epochs, employing a learning rate scheduling strategy combining linear warm-up and cosine annealing. The loss function is primarily based on cross-entropy, with an auxiliary loss based on semantic similarity results applied to enhance term semantic matching. Early stopping is used during fine-tuning to avoid overfitting and ensure the model's generalization performance in multi-domain term context discrimination tasks.

Figure 1 visualizes the attention mechanism and model layer activation patterns in term context modeling. In Figure 1(a), the horizontal axis represents the token position in the sentence where the term is located, and the vertical axis represents the number of layers in the mBART encoder. Colors represent the attention weights assigned to each position by each layer. The matrix shows that shallow layers primarily focus on positions adjacent to the term, while deeper layers exhibit broader context coverage. This demonstrates mBART's layer-by-layer contextual awareness in term modeling. Figure 1(b) shows the activation responses at each layer for different domains, with the horizontal axis representing the encoder layer number and the vertical axis representing the activation intensity. The medical domain exhibits a significantly higher response, reflecting its greater sensitivity to local semantics, while the legal domain exhibits significantly increased activation in layers 9 to 12, reflecting its reliance on complex semantic abstraction. These features validate the adaptability of our method for multi-domain terminology context discrimination and error correction, particularly in terms of dynamic context alignment and hierarchical feature extraction, providing structural support for terminology error correction.

BUILDING THE TERM CONTEXT IDENTIFICATION SUBMODULE

SEMANTIC SIMILARITY CALCULATION AND IDENTIFICATION PROCESS

This study addresses the polysemy problem of multi-domain terms in different contexts and designs a term context discrimination submodule based on semantic similarity [33,34]. This module takes as input the semantic vectors of the target term and its context extracted by the mBART encoder. The module first matches the term semantic vectors against a set of predefined domain term vectors. The term vectors are obtained by intercepting the corresponding position vectors from the hidden state of the last encoder layer of the model.

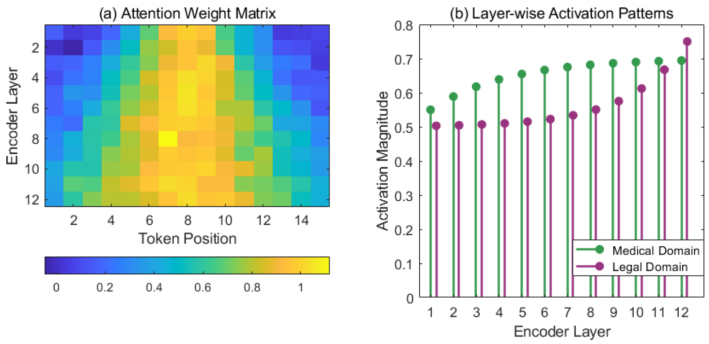

FIGURE 1. Attention distribution and layer-level activation features during multi-domain term context modeling. Figure 1(a) Term context attention weight matrix; Figure 1(b) Encoder layer activation patterns for different domains.

The domain term vectors are generated and cached from a domain-specific corpus using the same encoder, ensuring that the two are compared within a unified semantic space.

The module compares all candidate term vectors and selects the candidate words with the highest similarity as latent semantic labels. By default, the top-3 labels are retained to cover the possibility of ambiguity and prevent misjudgment. To balance recall and precision, this paper employs a grid search on the cosine similarity threshold (range 0.5–0.8, step size 0.05) on the validation set. Experiments show that when the threshold is set to 0.62, the term discrimination F1 score reaches its peak (0.89), and performs stably across the medical, legal, and engineering domains. Thresholds below this threshold introduce too many low-relevance candidates, leading to misclassification; thresholds above this threshold result in over-filtering and reduced recall. Therefore, 0.62 is ultimately adopted as the dynamic discrimination threshold, allowing a fallback to the top-3 label mechanism when confidence is insufficient, ensuring robustness.

ATTENTION MECHANISM ENHANCES SEMANTIC DISCRIMINATION

To improve the accuracy and flexibility of semantic discrimination, the module design applies a multi-head attention mechanism to deeply model the relationship between the target term and the candidate semantic label. Specifically, the target term context vector and the candidate label vector are first mapped to the query, key, and value space, and the attention weight is calculated as shown in Formula 5:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \qquad (5)$$

Q, K, and V are the query, key, and value matrices, respectively. $d_k = 64$ is the scaling factor for the key vector dimensions. The multi-head mechanism uses 8 heads to capture information from different subspaces of semantic features, enhancing the model's ability to perceive fine-grained semantic differences in complex contexts. Attention weights are optimized through backpropagation, driving the model to focus on vector dimensions with stronger semantic relevance, thereby improving semantic discrimination results. Combined with semantic similarity calculations, the attention output serves as the weighted input for the final discrimination, dynamically adjusting semantic label assignments to ensure accurate identification and proper distinction of terminology ambiguity.

TRAINING AND OPTIMIZATION STRATEGIES

The term context discrimination submodule and the mBART main model are jointly trained, and the loss function consists of a standard cross entropy loss and a semantic matching loss. The semantic matching loss is defined by maximizing the similarity between the correct label and the input semantic vector, as shown in Formula 6:

$$L_{\text{sim}} = 1 - \cos(v_{\text{pred}}, v_{\text{true}}) \qquad (6)$$

v_{pred} is the semantic vector of the term predicted by the model, and v_{true} is the ground-truth term vector for the corresponding domain. During training, the AdamW optimizer is used with an initial learning rate of 3×10^{-5}, a batch size of 64, and 10 training epochs. Gradient accumulation techniques are used to improve the stability of small-batch training, ensuring stable convergence of the model under limited sample conditions.

Figure 2 shows the performance of the term context discrimination module during the attention mechanism configuration and training convergence process. The horizontal axis of Figure 2(a) represents the number of attention heads, the left vertical axis represents the model's accuracy in the term discrimination task, and the right vertical axis represents the precision. When the number of attention heads is increased to 5 or 6, the model's accuracy and precision reach a high level, exceeding 0.9, indicating that a reasonable number of multi-head attention heads helps the model more accurately capture the semantic characteristics of terms in context. Figure 2(b) shows the loss convergence trend during training, with the horizontal axis representing the number of training epochs and the vertical axis representing the loss value (logarithmic scale). Both the training loss and validation loss decrease to below 0.15 after 50 epochs , indicating that the model has good stability and generalization ability after applying the multilingual attention mechanism, verifying its effectiveness in multi-domain term context recognition. The 50-epoch training curve shown in Figure 2(b) is an

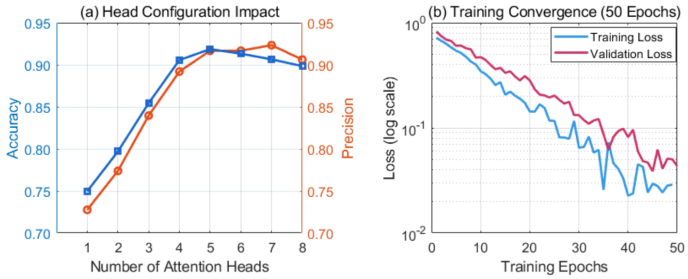

FIGURE 2. Visualization of the Contextual Discrimination Module. Figure 2(a) The Impact of Attention Head Configuration on Performance; Figure 2(b) Model Training Convergence Process

extended experiment conducted to fully observe the model's convergence trend. The actual fine-tuning process uses an early stopping strategy, and a stable convergence state is usually achieved within 10 epochs, consistent with the main experimental settings.

BUILDING A CANDIDATE CORRECTION VOCABULARY MULTI-SOURCE TERMINOLOGY RESOURCE INTEGRATION AND VOCABULARY CONSTRUCTION PROCESS

This study integrates WordNet (a terminology dictionary) and a large-scale multilingual parallel translation corpus to construct a multilingual candidate vocabulary for target terms to support semantic alignment and error correction generation. The source distribution of the terminology candidate vocabulary and the domain distribution of the parallel corpus are two independent dimensions: the former reflects the construction composition of vocabulary resources, and the latter reflects the domain matching of model training data. First, synonyms, hyponyms, and related terms of the target term are extracted from WordNet to form a basic semantic expansion set. This step leverages WordNet's hierarchical structure and semantic relationship graph to ensure that the terms in the vocabulary cover a wide range of semantic variations, meeting multi-domain coverage requirements. For professional terminology dictionaries, using automated crawling and text mining techniques to obtain authoritative terminology sets in fields such as medicine, law, and engineering to ensure the professionalism and accuracy of the vocabulary [35,36]. Entries in the terminology dictionary are cleaned and standardized to remove ambiguous terms and rare usages to ensure the high quality of the vocabulary.

Secondly, bidirectional term alignment is performed using multilingual parallel translation corpora (including English-French, English-German, and English-Chinese language pairs). A term alignment vocabulary is generated by automatically identifying the corresponding relationships between terms in the source and target languages using word alignment algorithms (such as FastAlign). The term alignment process incorporates a context window, taking into account the distribution of co-occurring words within the sentence to reduce mismatches. The alignment results are filtered using statistical frequency statistics to eliminate low-frequency term pairs and ensure the reliability of the vocabulary. Finally, the three resources are integrated to construct a comprehensive candidate correction vocabulary that includes the original form of the term, synonyms, and multilingual equivalents.

Figure 3 illustrates the key features of source contribution and semantic space distribution during the terminology vocabulary construction process. Figure 3(a) shows the contribution ratio of each source to the terminology vocabulary. Medical Database (MDB) is the primary contributor, accounting for 24%, followed by Legal Database and Engineering Database, accounting for 22% and 21%, respectively. WordNet and bilingual parallel corpora contribute relatively little, indicating that this method prioritizes specialized domain resources during the construction process. Figure 3(b) shows the distribution of terms in the semantic space after dimensionality reduction. The X-axis and Y-axis represent the two principal components after dimensionality reduction by PCA (Principal Component Analysis). The visualization shows that terms from different fields form relatively obvious clustering areas. Medical terms are concentrated near (0.5, 0.7), engineering terms are close to (1.2, 0.3), and legal terms are distributed in the range of (0.8, 1.5). The distribution of the three semantic groups is relatively separated, reflecting the model's ability to effectively distinguish the semantic boundaries of different professions. This also indirectly verifies that the vocabulary construction method has good domain discrimination and semantic representation effects.It is important to note that the source distribution of the terminology candidate vocabulary (Figure 3a) and the domain distribution of the parallel corpora (Figure 3b) are two independent dimensions: the former reflects the construction and composition of the vocabulary resources, while the latter reflects the domain coverage of the model training data. This dual-dimensional design ensures that the candidate vocabulary not only has multi-source semantic coverage capabilities but also matches the domain characteristics of the training corpus. To avoid introducing semantically broad or domain-irrelevant candidate terms from general semantic resources (such as WordNet), this paper introduces a domain-relevance filtering mechanism. Specifically, after extracting synonyms or hyponyms from WordNet, the system calculates the cosine similarity between each candidate word and the current sentence's

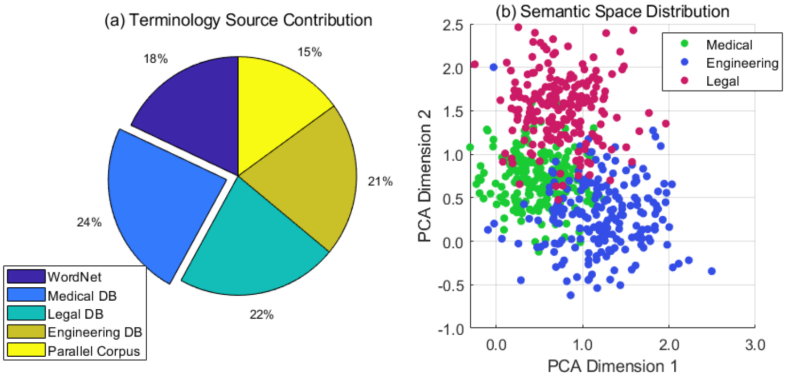

FIGURE 3. Dynamic graph of terminology. Figure 3(a) Term source contribution graph; Figure 3(b) Semantic space distribution graph

context embedding, and retains only terms with a similarity higher than a dynamic threshold (set to 0.65, determined through validation set optimization). Furthermore, if a candidate word does not appear in a target domain-specific dictionary or parallel corpus alignment table, its priority is significantly reduced. This dual filtering strategy effectively suppresses semantic drift and generalization errors, ensuring that the error correction candidate set retains both semantic diversity and domain specificity.

VOCABULARY UPDATE AND DYNAMIC MAINTENANCE MECHANISM

To ensure the timeliness and professional adaptability of the candidate's terminology vocabulary, the system has designed an automated dynamic vocabulary maintenance mechanism. This mechanism extracts potential new terms from authoritative domain terminology databases and the latest bilingual translation corpora on a monthly basis, performs preliminary screening using pre-trained models, and then incorporates them into the candidate set. To ensure the high quality and practical effectiveness of the vocabulary content, the system incorporates a joint weighting strategy based on TF-IDF (Term Frequency-Inverse Document Frequency) and word frequency trend analysis to rank new terms by importance. It also regularly removes redundant terms whose frequency of use continues to decline or whose semantics have drifted across multiple contexts. Furthermore, the system integrates fuzzy matching and subword decomposition algorithms to automatically identify and merge stem variations, spelling variations, common misspellings, and professional abbreviations, improving the vocabulary's coverage of diverse terminology in real-world applications.

ERROR CORRECTION CANDIDATE GENERATION AND SEMANTIC ALIGNMENT APPLICATION

The constructed candidate terminology word list not only serves the terminology recognition task but also plays a core role in the intelligent translation error correction module. After detecting potential terminology usage errors in the translation results, the system first uses the word list to perform semantic neighbor retrieval to generate a set of candidate error correction words. To improve the accuracy of recommendations, a semantic alignment mechanism is used in combination with contextual embedding representation to calculate the semantic similarity between the original word and the candidate word in a specific context. At the same time, the word frequency statistics and terminology credibility scores are referred to, and a comprehensive ranking is performed to generate the optimal error correction suggestions. The final output candidate can be used as a model retranslation input, and also supports manual review and interactive correction, effectively improving the consistency and professional accuracy of terminology translation.

INTEGRATING CONTEXTUAL INFORMATION FOR TERMINOLOGY RETRANSLATION CORRECTION

SEMANTIC EMBEDDING EXTRACTION AND AUXILIARY INPUT CONSTRUCTION

This study extracts the contextual semantic embedding of the sentence in which the term is located and uses it as auxiliary input to guide the mBART model to optimize term translation and error correction. The specific operation is to first use the pre-trained mBART encoder to encode the input sentence and obtain the context representation vector of each word in the sentence. For the target term, select its corresponding hidden state vector in the encoding layer. At the same time, collect the context word vectors with a window size of ± 5 words around the term to form the term context semantic embedding. The window range is determined by experimental tuning to take into account both context richness and noise suppression. Subsequently, the semantic embedding is fused with the encoding result of the source language sentence to form a multi-level input feature. The fusion method adopts a weighted splicing strategy, that is, Formula 7:

$$E_{\text{fused}} = \alpha E_{\text{term}} + (1 - \alpha) E_{\text{sent}} \tag{7}$$

E_{term} represents the term context embedding, and E_{sent} represents the full sentence encoding vector. The weight parameter α is initially set to 0.6, and the optimal value is determined through fine-tuning. This fused feature is fed

into the decoder layer and serves as an important condition for generating the translation.

MBART TRANSLATION GENERATION AND ERROR CORRECTION GUIDANCE MECHANISM

During the decoding phase, the mBART model dynamically adjusts the output of term translations based on auxiliary inputs fused with semantic embeddings. The decoder employs a multi-head self-attention mechanism, combined with contextual information from the encoder output, and uses a cross-attention module to capture the relationships between terms and other words in the sentence. To enhance error correction capabilities, a term error correction loss function is designed, combining a cross-entropy loss with a term accuracy weight.

$$L = L_{\mathrm{CE}} + \beta \cdot L_{\mathrm{term}} \tag{8}$$

In formula (8), L_{CE} is the standard word-level cross-entropy loss, L_{term} is the error correction accuracy loss of the term translation, and the weight β is set to 0.8 to enhance the accuracy of term translation. To further enhance error correction, the model dynamically matches candidate terms during decoding, incorporating the previously constructed terminology candidate correction word list.

Figure 4 systematically illustrates the core mechanism of terminology retranslation and error correction.

FINE-TUNING THE MBART MODEL FOR TERMINOLOGY TRANSLATION TASKS

This study constructed a multi-domain parallel corpus containing a total of 152,840 English-Chinese sentence pairs. The medical field accounts for 41.2% (62,980 pairs), covering five sub-fields including cardiovascular, neurology, and oncology; the legal field accounts for 32.7% (49,980 pairs), including contract law and intellectual property law; and the engineering field accounts for 26.1% (39,880 pairs), covering electrical, mechanical, and civil engineering. All corpora underwent terminology alignment and manual verification to ensure consistent terminology translation. The data was divided into training, validation, and test sets in an 8:1:1 ratio.

Fine-tuning is performed on a pre-trained mBART-large model with approximately 610M parameters. The Adam optimizer is used during training, with an initial learning rate of 3×10^{-5}, linear warmup, and a cosine annealing schedule. The batch size is set to 64, and the maximum input sequence length is limited to 512 tokens. Ten epochs are trained, and the early stopping strategy is determined based on the term translation accuracy on the validation set. The loss function is a weighted loss combining a cross-entropy loss and a term correction regularization term. The weight coefficient adjusted the impact of the term correction component on the overall training. Model training is performed in parallel on four NVIDIA A100 GPUs to ensure training efficiency and maximize resource utilization.

During fine-tuning, a context adaptation mechanism is specifically designed to assist the model in identifying the domain category information of the input sentence by applying domain label embeddings. Domain labels and word embeddings are fused at the encoding level, enabling the model to capture domain-specific differences in terminology usage.

Figure 5 illustrates the key optimization strategies and performance improvements of the term translation model during training. Figure 5(a) compares the downward trend of model training loss under four different learning rates, with epochs as the horizontal axis and training loss as the vertical axis. The model with a learning rate of 5×10^{-5} converges faster, with the loss dropping to approximately 0.1 after the sixth epoch, indicating that 5×10^{-5} is a more optimal learning rate. Figure 5(b) shows the evolution of the True Positive Rate (TPR) for term recognition during training, with epochs as the horizontal axis and TPR values as the vertical axis. The TPR for medical term recognition improved significantly over 10 epochs of training, while that for engineering term recognition is slightly lower overall, reflecting that the model is more susceptible to learning medical term recognition within the specific corpus. Although the comparative experiment shown in Figure 5(a) only runs for 10 epochs to highlight the difference in convergence trend, the main model training continues until the early stopping condition is triggered (up to 20 epochs) after verifying that the loss has stabilized. The final selected learning rate (3×10^{-5}) has achieved effective convergence within 10 rounds, and the validation metrics continue to improve, proving that the training budget is sufficient to cover the complexity of this task. The 5×10^{-5} learning rate comparison experiment shown in Figure 5(a) is an exploratory experiment in the parameter tuning phase. The main experiment used a 3×10^{-5} learning rate and achieved effective convergence within 10 epochs. The differences in convergence speed under different learning rate settings reflect the impact of hyperparameters on training dynamics, but do not affect the final model performance.

III. METHODS FOR EVALUATING THE ACCURACY OF ENGLISH TERMINOLOGY CORRECTION

This study selects a multilingual parallel corpus covering multiple professional fields, such as medicine, law, and engineering, as the experimental dataset to ensure that the terminology coverage is extensive and has complex contextual characteristics. The corpus includes Cardiovascular Medicine, Neurology, Orthopedic Surgery, Oncology, Pediatrics, Contract Law, Intellectual Property Law, Criminal Law, Civil Law, International Law, Mechanical Engineering, Electrical Engineering, Civil Engineering, Chemical Engineering, and Software Engineering. The corpus sources include public professional translation data sets and self-constructed domain terminology comparison libraries, all of which have undergone strict manual verification to ensure the quality of annotation. Each piece of data contains the

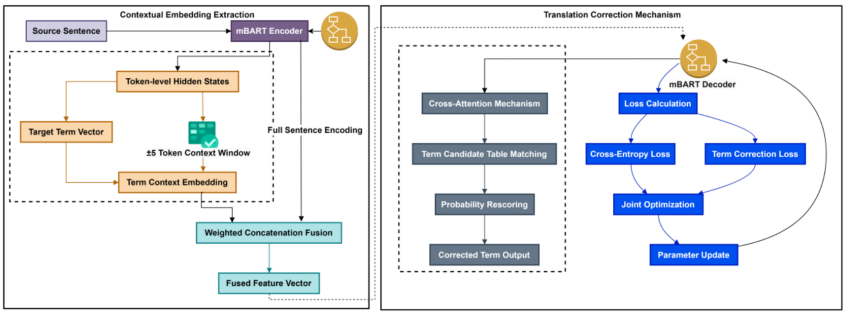

FIGURE 4. The core mechanism of terminology retranslation and error correction

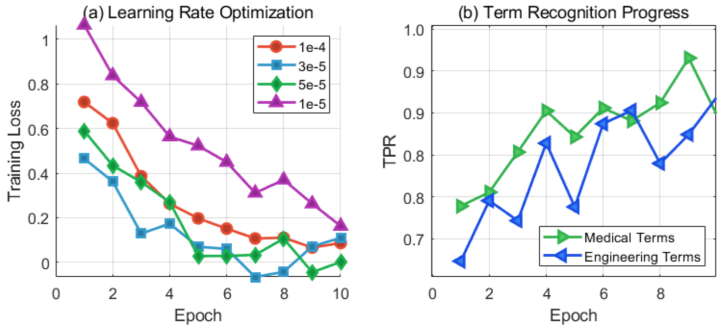

FIGURE 5. Visualization of the Model Training Optimization Process. Figure 5(a) Comparison of Learning Rate Optimization Results; Figure 5(b) Trends in Term Recognition Rate Improvement for Different Domains

source language sentence, the corresponding translation, and the accurately annotated term position and standard translation method, which facilitates the objective evaluation of term recognition and error correction effects. The data set is divided into a training set, a validation set, and a test set. The training set contains rich and balanced term types. The validation set is used for model fine-tuning and parameter selection, and the test set is specifically used for comprehensive performance evaluation.

Using BLEU and TER to Evaluate Term Translation Quality

BLEU (Bilingual Evaluation Understudy) and TER (Translation Error Rate) are standard metrics for measuring machine translation quality. BLEU measures the accuracy and fluency of a translation by counting the overlap between n-grams in the translation and the reference translation. TER measures the translation error rate by calculating the number of editing operations (such as insertions, deletions, and substitutions) required to return the translation to the reference translation. This paper calculates BLEU and TER before and after terminology correction, comparing the changes in these metrics to objectively reflect the degree of improvement in terminology translation. The evaluation data uses a manually annotated set of professional terminology translations to ensure the reliability and relevance of the results.

This paper sets up two baseline models for fair comparison: (1) Vanilla mBART: the original mBART-large model (unfine-tuned) is used directly for translation; (2) Domain-Finetuned mBART: mBART fine-tuned on the same multi-domain corpus but without integrating the context discrimination module and the error correction vocabulary. All comparative experiments are run on the same test set to ensure consistent evaluation.

Figure 6 illustrates the improved performance of the model in the terminology translation task. Figure 6(a) compares BLEU scores across different domains, with the x-axis representing 15 professional domains and the y-axis representing the BLEU score, a measure of translation quality. Clearly, the proposed method significantly outperforms the baseline method in all domains. In the software domain, the BLEU score improves from 47.3 to 67.4, and in IP (Intellectual Property) Law, it improves from 42.3 to 63.1. This demonstrates that the proposed method is particularly effective in processing terminology in highly technical domains, with an average BLEU score of 57.9 across 15 domains, exceeding the baseline model's 38.4. Figure 6(b) shows the reduction in error rates for different types of terminology, with the x-axis representing the five categories of terminology and the y-axis representing the percentage of error. Comparisons show that the proposed method reduces the error rate for ambiguous terms from 58.4% to 24.3%, and for medical terms from 42.3% to 16.5%, demonstrating its significant advantage in handling ambiguous and cross-domain terminology. This shows that the proposed terminology identification and translation optimization strategy can effectively reduce errors and improve translation quality in multiple professional contexts.

The manual evaluation employed a rigorous blind review process: 500 sentences containing ambiguous terms were

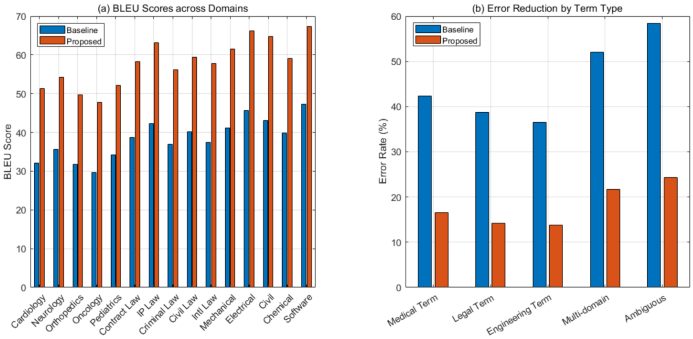

FIGURE 6. Translation Quality Assessment Comparison. Figure 6(a) Comparison of BLEU Scores for Various Fields; Figure 6(b) Changes in Term Type Error Rate

randomly selected from the test set (stratified sampling by domain, ensuring approximately one-third each from medicine, law, and engineering). After removing model identifiers, these sentences were distributed to three bilingual reviewers with master's degrees or higher and relevant professional backgrounds. The scoring criteria included semantic accuracy (1–5 points) and professional appropriateness (1–5 points), with the average of the two scores used as the final score. The three reviewers received pre-evaluation training including 20 sample sentences for calibration. After the evaluation, Krippendorff's α was calculated to be 0.82 (semantic dimension) and 0.79 (professional dimension), indicating good inter-rater consistency.

All BLEU scores were calculated using the sacreBLEU toolkit (v2.3.1) with the default configuration: n-gram order of 4, additive smoothing, and raw word segmentation based on spaces (without additional preprocessing), and case insensitive. This setting ensures consistency with mainstream machine translation evaluation standards and improves the reproducibility of results.

EVALUATING TERMINOLOGY PERCEPTION ABILITY USING TERM RECOGNITION PRECISION (TPR)

The term recall (TPR) for term recognition is defined here as the term-level recall rate, which is the proportion of terms correctly identified by the model among all manually labeled terms. This metric evaluates the model's sensitivity to technical terms at the term level, without involving aggregation at the sentence or document level.

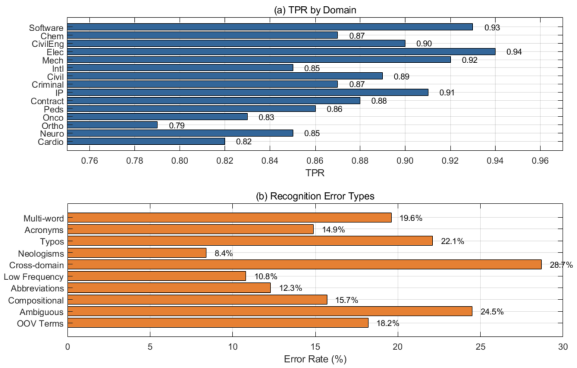

FIGURE 7. Terminology Recognition Accuracy Analysis. Figure 7(a) Comparison of Terminology Recognition Accuracy by Field; Figure 7(b) Distribution of Recognition Error Types

Figure 7 illustrates the performance of term recognition after optimization based on the mBART multilingual model, analyzing it from two dimensions: domain distribution and error type. In Figure 7(a), the vertical axis represents different professional domains, and the horizontal axis represents the TPR for term recognition. The results show that the model performs well across various domains, with TPRs generally exceeding 0.80 and reaching a maximum of 0.94 (in the electrical domain), demonstrating the system's strong generalization capabilities in multi-domain term recognition. Figure 7(b) shows the incidence rate of various recognition errors, with the vertical axis representing different error types and the horizontal axis representing the corresponding error rate (%). As can be seen, the error rates for cross-domain terms and ambiguous terms are relatively high, at 28.7% and 24.5%, respectively. This indicates that term polysemy and contextual drift remain key challenges in term translation, and also provides a target direction for subsequent vocabulary maintenance and semantic alignment optimization.

The error types in Figure 7(b) were manually labeled by two annotators with linguistics backgrounds based on a predefined classification system (including five categories: "cross-domain misclassification," "unresolved polysemous ambiguity," "mismatched word forms," "missing knowledge base," and "ignoring context"). Consistency training was

performed before labeling, and Cohen's κ = 0.85 was calculated after labeling. Disagreements were arbitrated by a third senior annotator. The final error distribution was based on the consensus labeling results.

TERM CORRECTION ACCURACY

The terminology correction accuracy rate is specifically used to measure the model's ability to correct mistranslated terms. This indicator calculates the proportion of terms corrected by the model that fully match the standard translation. The evaluation process involves screening out sentences containing incorrect terms from the original translation, applying the model to correct them, and comparing them with the standard translation to determine whether the correction is accurate. The terminology correction accuracy rate reflects the effectiveness of the model in improving translation quality in actual application scenarios, with special attention paid to the correction performance of 10 complex contexts and polysemous terms to ensure the practical value of the system.

TABLE 1. Terminology type retention and system response time in different professional fields

Domain/Term Type	Simple Terms (%)	Ambiguous Terms (%)	Cross-domain Terms (%)	Average Response Time (ms)	Confidence Interval (95%)
Cardiovascular Medicine	94.3	84.7	81.3	42	[93.5, 95.1]
Neurology	93.1	83.5	79.6	45	[92.2, 94.0]
Orthopedics	92.8	85.2	82.1	44	[91.8, 93.8]
Oncology	91.5	82.9	78.5	47	[90.4, 92.6]
Contract Law	95.2	88.6	86.3	39	[94.3, 96.1]
Intellectual Property Law	96.1	90.1	88.9	37	[95.3, 96.9]
Criminal Law	94.7	87.2	84.5	41	[93.8, 95.6]
Mechanical Engineering	94.3	87.8	85.2	38	[93.4, 95.2]
Electrical Engineering	95.6	90.3	88.7	36	[94.7, 96.5]
Software Engineering	97.8	94.2	92.6	33	[97.0, 98.6]
Cross-domain Average	94.5	87.5	84.8	40.2	-

Table 1 shows the retention of terminology types and system response time in different professional fields, reflecting the model's performance in cross-domain terminology processing. Software engineering has the highest terminology correction retention rate, reaching 97.8% for simple terms. Retention rates for ambiguous terms and cross-domain terms are also as high as 94.2% and 92.6%, respectively, demonstrating relatively accurate and stable terminology processing in this field. While the retention rates for simple terminology in medical fields such as cardiovascular and neurology are slightly lower than those in software engineering, at 94.3% and 93.1%, respectively, the retention rate for ambiguous terms remains above 80%, demonstrating the model's adaptability in complex semantic environments. Furthermore, the terminology retention rates for contract law and intellectual property law in the legal field both exceeded 90%, with short response times (39ms and 37ms, respectively), demonstrating the model's high efficiency and accuracy in the legal field. The overall cross-domain average terminology retention rate is 94.5% for simple terms, 87.5% for ambiguous terms, and 84.8% for cross-domain terms, with an average response time of approximately 40.2ms. This demonstrates the model's strong terminology recognition and retention capabilities, effectively supporting professional translation needs across multiple fields and ensuring the rigor and integrity of translated texts.

All response times were measured on an NVIDIA A100 GPU (40GB VRAM) with FP16 precision and a batch size of 1. End-to-end response time includes term detection, candidate retrieval, semantic reordering, and retranslation generation, without quantization compression. The system employs caching optimization, and the candidate vocabulary is preloaded into GPU memory to minimize I/O overhead.

TERM PRESERVATION RATE

The terminology retention rate assesses whether the model accurately preserves the core meaning and form of source terms during the translation process, avoiding oversimplification or incorrect substitution. The model's terminology fidelity is determined by measuring the proportion of correctly retained or appropriately translated terms in the translation. The evaluation dataset includes multiple professional fields to verify the model's consistency in terminology across these fields.

TABLE 2. Model quality dimension evaluation results in the three fields of medicine, law, and engineering

Quality Dimension	Medical Domain (%)	Legal Domain (%)	Engineering Domain (%)	Statistical Significance
Semantic Fidelity	93.5	95.2	96.1	$p < 0.001$
Formal Consistency	88.7	92.1	93.4	$p = 0.002$
Professional Expression	85.9	90.8	93.2	$p < 0.001$
Domain Adaptability	91.3	94	95	$p = 0.001$
Overall Preservation Rate	89.9	93	94.4	$p < 0.001$
95% Confidence Interval	[88.5, 91.3]	[91.8, 94.2]	[93.3, 95.5]	-

Table 2 presents the model quality assessment results for the three domains of medicine, law, and engineering, covering five dimensions: semantic fidelity, formal consistency, professional expression, domain adaptability, and overall retention. Overall, the engineering domain performs best across all indicators, with semantic fidelity reaching 96.1%, formal consistency reaching 93.4%, and overall retention reaching 94.4%, demonstrating that terminology and expressions in this domain are well preserved and conveyed. The legal domain ranks second, with a semantic fidelity of 95.2% and an overall retention of 93.0%. Medicine performs slightly lower, with a semantic fidelity of 93.5% and an overall retention of 89.9%. Furthermore, all indicators are statistically significant (p-values less than 0.005), indicating statistically significant differences between the domains. The confidence intervals for medicine (88.5, 91.3) and engineering (93.3, 95.5) did not overlap, confirming significant differences between the domains. This shows that there are differences in the model's ability to maintain terminology and text quality in different professional fields, and the terminology processing effect in the engineering field is more prominent, reflecting the model's adaptability and accuracy in identifying and communicating professional terminology.

The 95% confidence intervals in Tables 1 and 2 are calculated based on 1000 bootstrapping samplings. Each time, the same number of samples were sampled with replacement from the corresponding domain test set, and the 2.5% and 97.5% quantiles were used as the upper and lower confidence limits after calculating the indicators. This method is applicable to non-normally distributed indicators and is robust to small samples. The significance tests for the quality dimensions of each domain in Table 2 were performed using one-way ANOVA, with post-hoc multiple comparisons corrected using Tukey HSD. All p-values were calculated based on the F-distribution, and the significance threshold was set to α=0.05. The data satisfy homogeneity of variance (Levene test p>0.05) and approximate normality (Shapiro-Wilk test p>0.01), meeting the ANOVA prerequisites.

CONTEXTUAL CONSISTENCY SCORE

The terminology contextual consistency score measures the contextual applicability and semantic coherence of the model's translated terms. Through both manual review and automated methods based on semantic similarity, the translated terms are evaluated for their consistency with the overall context of the sentence and domain expertise. Scoring criteria include accurate semantic communication, consistency with the context, and avoidance of ambiguity.

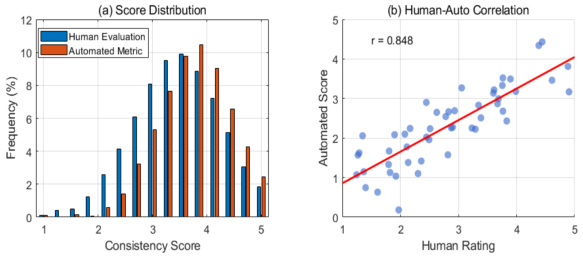

FIGURE 8. Contextual consistency assessment. Figure 8(a) Consistency score distribution; Figure 8(b) Correlation between manual and automatic scoring

Figure 8(a) shows the consistency score on the horizontal axis, ranging from 1 to 5, and the frequency percentage of the corresponding score on the vertical axis. The frequency distribution of manual and automatic evaluations is compared, showing that both have a higher frequency in the 3.0-4.0 score range, indicating that the automatic indicators are generally similar to the manual scoring distribution; Figure 8(b) shows the scatter distribution and linear fit of human and automatic scores, with human scores on the horizontal axis and automatic scores on the vertical axis. The Shapiro-Wilk test confirms that the data approximately follow a normal distribution ($p > 0.05$), therefore the Pearson correlation coefficient was chosen instead of the Spearman correlation coefficient. The correlation is significant at the $\alpha = 0.01$ level ($p < 0.001$).

IV. CONCLUSION

This paper presents an intelligent error correction system for English multi-domain terminology translation, based on the pre-trained mBART multilingual model. The system uses multilingual context modeling to extract semantic features and, combined with a semantic discrimination module, dynamically identifies the meanings of terms in different contexts. It constructs a multilingual error correction vocabulary to support error correction and utilizes contextual information to guide accurate retranslations. The model is fine-tuned on parallel corpora from multiple fields, including medicine, law, and engineering, improving its ability to translate multi-domain terminology. Experimental results show a significant improvement, with a BLEU score of 57.9 across 15 domains, compared to the baseline of 38.4. The ambiguous term error rate is reduced to 24.3%, and term recognition accuracy reaches 0.94 in the electrical field, highly correlated with human scores (r=0.848). This research addresses the contextual awareness challenge in multi-domain terminology translation and demonstrates the potential of intelligent translation software in interdisciplinary communication. Future work can optimize the contextual mechanism, expand language pairs, and improve the model's practicality.

AUTHOR CONTRIBUTIONS

Conceptualization – Pei Wang and Chenxi Li; methodology – Pei Wang and Chenxi Li; investigation – Pei Wang and Chenxi Li; writing-original draft preparation – Pei Wang and Chenxi Li. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

Funded by the 2025 Hainan Provincial Philosophy and Social Sciences Planning Project, "Research on Strategies for Cross-Cultural Empathetic Communication of Qiong Merchants' Stories from the Perspective of the Maritime Silk Road" (Project No. HNSK(YB)25-68).

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] D. Maier, C. Baden, D. Stoltenberg, M. De Vries-Kedem, and A. Waldherr, "Machine translation vs," multilingual dictionaries assessing two strategies for the topic modeling of multilingual text collections. Communication methods and measures, vol. 16, no. 1, pp. 19-38, 2022, doi: 10.1080/19312458.2021.1955845.

[2] R. Litschko, I. Vulić, S. P. Ponzetto, and G. Glavaš, "On cross-lingual retrieval with multilingual text encoders," Information Retrieval Journal, vol. 25, no. 2, pp. 149-183, 2022, doi: 10.1007/s10791-022-09406-x.

[3] H. Rathnayake, J. Sumanapala, R. Rukshani, and S. Ranathunga, "Adapter-based fine-tuning of pre-trained multilingual language models for code-mixed and code-switched text classification," Knowledge and Information Systems, vol. 64, no. 7, pp. 1937-1966, 2022, doi: 10.1007/s10115-022-01698-1.

[4] A. Fan, S. Bhosale, H. Schwenk, Z. Ma, A. El-Kishky, and S. Goyal, "Beyond english-centric multilingual machine translation," Journal of Machine Learning Research, vol. 22, no. 107, pp. 1-48, 2021.

[5] A. L. Zorrilla and M. I. Torres, "A multilingual neural coaching model with enhanced long-term dialogue structure," ACM Transactions on Interactive Intelligent Systems (TiiS), vol. 12, no. 2, pp. 1-47, 2022, doi: 10.1145/3487066.

[6] B. Heinisch, "Large language models for terminology work: A question of the right prompt?," Journal for Language Technology and Computational Linguistics, vol. 38, no. 2, pp. 13-30, 2025, doi: 10.21248/jlcl.38.2025.280.

[7] Mamta and A. Ekbal, "Transformer based multilingual joint learning framework for code-mixed and english sentiment analysis," Journal of Intelligent Information Systems, vol. 62, no. 1, pp. 231-253, 2024, doi: 10.1007/s10844-023-00808-x.

[8] A. E. Pierson, D. B. Clark, and C. E. Brady, "Scientific modeling and translanguaging: A multilingual and multimodal approach to support science learning and engagement," Science Education, vol. 105, no. 4, pp. 776-813, 2021, doi: 10.1002/sce.21622.

[9] C. Bryant, Z. Yuan, M. R. Qorib, H. Cao, H. T. Ng, and T. Briscoe, "Grammatical error correction: A survey of the state of the art," Computational Linguistics, vol. 49, no. 3, pp. 643-701, 2023, doi: 10.1162/coli_a_00478.

[10] Y. Chen, "Analyzing the design of intelligent English translation and teaching model in colleges using data mining," Soft Computing, vol. 27, no. 19, pp. 14497-14513, 2023, doi: 10.1007/s00500-023-09096-7.

[11] L. H. Aliguliyeva, A. R. Khalilova, E. R. Sadiqova, S. S. Suleymanova, and T. Y. Maharramzade, "The impact of globalization and digitalization on the Azerbaijani language," Universidad y Sociedad, vol. 17, no. 4, Art. no. e5310-e5310, 2025.

[12] S. Liang and W. Q. Yan, "A hybrid CTC+ Attention model based on end-to-end framework for multilingual speech recognition," Multimedia Tools and Applications, vol. 81, no. 28, pp. 41295-41308, 2022, doi: 10.1007/s11042-022-12136-3.

[13] E. Linhares Pontes, L. A. Cabrera-Diego, J. G. Moreno, E. Boros, A. Hamdi, and A. Doucet, "MELHISSA: a multilingual entity linking architecture for historical press articles," International journal on digital libraries, vol. 23, no. 2, pp. 133-160, 2022, doi: 10.1007/s00799-021-00319-6.

[14] F. El-Alami, S. O. El Alaoui, and N. E. Nahnahi, "A multilingual offensive language detection method based on transfer learning from transformer fine-tuning model," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 8, pp. 6048-6056, 2022, doi: 10.1016/j.jksuci.2021.07.013.

[15] H. Al-Khalifa, K. Al-Khalefah, and H. Haroon, "Error analysis of pre-trained language models (PLMs) in English-to-Arabic machine translation," Human-Centric Intelligent Systems, vol. 4, no. 2, pp. 206-219, 2024, doi: 10.1007/s44230-024-00061-7.

[16] S. Khurana, A. Laurent, and J. Glass, "Samu-xlsr: Semantically-aligned multimodal utterance-level cross-lingual speech representation," IEEE Journal of Selected Topics in Signal Processing, vol. 16, no. 6, pp. 1493-1504, 2022, doi: 10.1109/JSTSP.2022.3192714.

[17] S. Shekhar, H. Garg, R. Agrawal, S. Shivani, and B. Sharma, "Hatred and trolling detection transliteration framework using hierarchical LSTM in code-mixed social media text," Complex & Intelligent Systems, vol. 9, no. 3, pp. 2813-2826, 2023, doi: 10.1007/s40747-021-00487-7.

[18] J. Acs, E. Hamerlik, R. Schwartz, N. A. Smith, and A. Kornai, "Morphosyntactic probing of multilingual BERT models," Natural Language Engineering, vol. 30, no. 4, pp. 753-792, 2024, doi: 10.1017/S1351324923000190.

[19] J. Bjerva, "The role of typological feature prediction in NLP and linguistics," Computational Linguistics, vol. 50, no. 2, pp. 781-794, 2023, doi: 10.1162/coli_a_00498.

[20] G. Manias, A. Mavrogiorgou, A. Kiourtis, C. Symvoulidis, and D. Kyriazis, "Multilingual text categorization and sentiment analysis: a comparative analysis of the utilization of multilingual approaches for classifying twitter data," Neural Computing and Applications, vol. 35, no. 29, pp. 21415-21431, 2023, doi: 10.1007/s00521-023-08629-3.

[21] M. F. Arslan and M. Ahmed, "Building a Multilingual Vocabulary Database: A Comprehensive Study of 24 Global and Local Languages," Journal of Asian Development Studies, vol. 14, no. 2, pp. 1327-1342, 2025, doi: 10.62345/jads.2025.14.2.103.

[22] A. Fernando, S. Ranathunga, D. Sachintha, L. Piyarathna, and C. Rajitha, "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.

[23] I. Banerjee, J. M. Lambert, B. A. Copeland, J. L. Paranczak, K. M. Bailey, and C. M. Standish, "Extending functional communication training to multiple language contexts in bilingual learners with challenging behavior," Journal of Applied Behavior Analysis, vol. 55, no. 1, pp. 80-100, 2022, doi: 10.1002/jaba.883.

[24] W. Duan, N. Yamashita, Y. Shirai, and S. R. Fussell, "Bridging fluency disparity between native and nonnative speakers in multilingual multiparty collaboration using a clarification agent," Proceedings of the ACM on Human-Computer Interaction, vol. 5, no. CSCW2, pp. 1-31, 2021, doi: 10.1145/3479579.

[25] S. Bala Das, D. Panda, T. Kumar Mishra, B. Kr Patra, and A. Ekbal, "Multilingual neural machine translation for indic to indic languages," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 23, no. 5, pp. 1-32, 2024, doi: 10.1145/3652026.

[26] M. R. Kanfoud and A. Bouramoul, "SentiCode: A new paradigm for one-time training and global prediction in multilingual sentiment analysis," Journal of Intelligent Information Systems, vol. 59, no. 2, pp. 501-522, 2022, doi: 10.1007/s10844-022-00714-8.

[27] B. Kotiyal, H. Pathak, and N. Singh, "Debunking multi-lingual social media posts using deep learning," International journal of information technology, vol. 15, no. 5, pp. 2569-2581, 2023, doi: 10.1007/s41870-023-01288-6.

[28] V. Kuperman, N. Siegelman, S. Schroeder, C. Acartürk, S. Alexeeva, and S. Amenta, "Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus," Studies in Second Language Acquisition, vol. 45, no. 1, pp. 3-37, 2023, doi: 10.1017/S0272263121000954.

[29] M. Saqlain, "Evaluating the readability of English instructional materials in Pakistani Universities: A deep learning and statistical approach," Education science and management, vol. 1, no. 2, pp. 101-110, 2023, doi: 10.56578/esm010204.

[30] I. Cushing, A. Georgiou, and P. Karatsareas, "Where two worlds meet: language policing in mainstream and complementary schools in England," International Journal of Bilingual Education and Bilingualism, vol. 27, no. 9, pp. 1182-1198, 2024, doi: 10.1080/13670050.2021.1933894.

[31] D. Van Thin, H. Quoc Ngo, D. Ngoc Hao, and N. Luu-Thuy Nguyen, "Exploring zero-shot and joint training cross-lingual strategies for aspect-based sentiment analysis based on contextualized multilingual language models," Journal of Information and Telecommunication, vol. 7, no. 2, pp. 121-143, 2023, doi: 10.1080/24751839.2023.2173843.

[32] S. Chauhan, S. Saxena, and P. Daniel, "Fully unsupervised word translation from cross-lingual word embeddings especially for healthcare professionals," International Journal of System Assurance Engineering and Management, vol. 13, no. Suppl 1, pp. 28-37, 2022, doi: 10.1007/s13198-021-01182-z.

[33] Y. Chen, H. T. Hsu, and H. Y. Tai, "The relative effects of corrective feedback and language proficiency on the development of L2 pragmalinguistic competence: the case of request downgraders," International Review of Applied Linguistics in Language Teaching, vol. 63, no. 1, pp. 341-366, 2025, doi: 10.1515/iral-2023-0036.

[34] S. Ghadiri, Z. Tajeddin, and M. Alemi, "Teacher Corrective Feedback on Learners' Pragmatic Failure: Types of Feedback in Online Pragmatics Instruction," Journal of English Language Teaching and Learning, vol. 16, no. 33, pp. 172-193, 2024.

[35] M. Memari and F. Hafez, "The Effect of Corrective Feedback on Iranian English as a Foreign Language Learners' Interlanguage Pragmatics Development," Journal of English Language Teaching and Learning, vol. 17, no. 35, pp. 267-282, 2025.

[36] G. Moro, N. Piscaglia, L. Ragazzi, and P. Italiani, "Multi-language transfer learning for low-resource legal case summarization," Artificial Intelligence and Law, vol. 32, no. 4, pp. 1111-1139, 2024, doi: 10.1007/s10506-023-09373-8.