



Translation Discrimination Model of Scientific Text Ambiguous Words based on multi-granularity attention Mechanism

Juan Tan^{1,*}

¹ School of Languages and Media, Anhui University of Finance and Economics, Bengbu 233030, Anhui, China

SUMMARY: *In order to solve the problems of ambiguous word translation boundary, insufficient utilization of context constraints, and unstable differentiation of candidate terms in scientific and technological texts, this paper proposes a translation discrimination model based on multi-granularity attention mechanism, which integrates word-level, phrase-level, and sentence-level semantic information into a unified discrimination framework, and combines candidate matching and result correction strategies to improve the accuracy of translation selection. Experiments are carried out based on 42 600 bilingual sentence pairs and 18 240 ambiguous word examples in scientific texts. The results show that the Accuracy, Macro-F1 and MRR of the model reach 91.8%, 90.9% and 0.946, respectively, and the average reasoning time is 20.2 ms. The overall performance is better than that of the comparison methods, indicating that the model has good discrimination ability and application value in complex scientific and technological contexts.*

KEYWORDS: *science and technology text; Ambiguous word translation; Multi-granularity attention mechanism; Models for Translation Discrimination*

1 Introduction

Under the background of the continuous acceleration of global scientific and technological communication and the increasingly frequent cross-lingual transmission of professional knowledge, scientific text translation is no longer a general problem of word substitution and syntactic conversion, but gradually evolves into a complex computational task oriented to term recognition, context modeling and semantic discrimination. Compared with general texts, scientific texts have the characteristics of dense terms, nested sentence patterns, strong domain dependence and subtle semantic boundaries. The same word often corresponds to different translations in different subject scenarios, different collocation structures and even different discussion levels. If the model only relies on local neighboring words or static word vectors to complete the judgment, it is easy to cause problems such as word sense misselection, professional semantic deviation and context cohesion imbalance, which will affect the accuracy and professionalism of the whole translation. This phenomenon is more obvious in fields with rapid terminology update such as artificial intelligence, medicine, biological engineering, and material science, which also makes the translation of ambiguous words in scientific and technological texts a key link in the quality improvement of machine translation.

From the existing research, ambiguity resolution, context-aware translation and

*tjaufe1109@163.com

<https://doi.org/10.65102/is2026799>

professional text machine translation have formed a rich accumulation of technologies. Some studies use explicit word sense information to improve the semantic bias in neural machine translation, some studies use context-aware mechanism to enhance the model's ability to capture long-distance dependence and discourse information, and some studies have begun to pay attention to the translation stability of terms in scientific texts, medical texts and other professional corpora. These results show that the improvement of translation discrimination effect does not simply depend on a larger parameter scale, but is closely related to the organization of semantic cues, the extraction level of context information, and the attention allocation strategy. However, there are still two shortcomings in the existing methods. First, the semantic modeling of ambiguous words mostly remains in a single granularity, and the intra-sentence lexical association can be identified, but the document-level topic constraints, domain semantic bias and collocation structure differences are often weakened. Second, in the science and technology text scene, the model still does not fully utilize the term neighborhood, syntactic role and cross-sentence semantic cues, resulting in unstable boundary recognition between similar meanings.

Based on this, this paper focuses on the problem of translation discrimination of ambiguous words in scientific and technological texts, and constructs a translation discrimination model under the multi-granularity attention mechanism. The word level, phrase level and sentence level semantic information are integrated into a unified representation framework, which retains the fine-grained word sense differences and strengthens the model's ability to recognize the domain context and context evolution relationship. The model consists of ambiguous word feature representation, multi-granularity semantic extraction, attention weighted discrimination and output decision, trying to establish a more stable translation selection mechanism in complex scientific and technological contexts. This paper mainly focuses on three issues: (1) Whether multi-granularity semantic joint modeling can effectively improve the accuracy of translation discrimination of scientific and technological text ambiguous words; (2) Whether the attention mechanism can enhance the ability of the model to distinguish confusable meanings and cross-domain context changes; (3) Whether the model has good generalization and discrimination stability under different ambiguity types and different scientific and technological corpus conditions. Through subsequent experiments and comparative analysis, this paper attempts to provide a more targeted modeling path for fine-grained semantic discrimination in scientific text machine translation.

2 Related Research

With the deepening of scientific text translation research, ambiguity translation discrimination has become an important link affecting the quality of machine translation. Traditional neural machine translation systems have long relied on large-scale parallel corpora and end-to-end parameter learning, and have achieved high performance in general text translation. However, when dealing with technical texts with dense terms, obvious syntactic compression and strong context dependence, they are still prone to problems such as semantic fragmentation, insufficient context constraints and incorrect translation selection. Especially in the scenarios of cross-disciplinary term transfer, nested expression of long sentences, and high frequency of homonyms, although the existing models can complete surface alignment, it is difficult to stably depict the true orientation of ambiguous words in different semantic levels. Campolungo et al. (2022) constructed the DiBiMT benchmark around the bias of word sense disambiguation in machine translation, and pointed out that translation models systematically deviated in the processing of polysemous words, which provided direct evidence for

subsequent research on ambiguous word translation [1]. Mehrparvar and Pezzelle (2024) further investigated the processing ability of multilingual large-scale models for language ambiguity detection and translation, and found that although the model has certain discrimination potential in complex contexts, it is still unstable in fine-grained sense boundary identification [4].

On the one hand, existing research has begun to improve the translation discrimination ability of ambiguous words by two ways: explicit word sense information and context enhancement. Campolungo et al. (2022) proposed to use explicit word sense information to reduce the disambiguation bias in neural machine translation, indicating that external sense knowledge has a positive effect on translation selection [2]. Bejgu et al. (2024) promote ambiguity discrimination in open environment from the perspective of word sense linking, make word sense identification no longer limited to closed vocabulary, and provide new ideas for translation and sense mapping in complex contexts [5]. Proietti et al. (2024) analyzed the homonymous word resolution ability of pre-trained language models and pointed out that the model can capture part of the context cues, but it is still easy to confuse when facing words with similar semantics and obvious domain differences [7]. Sumanathilaka et al. (2024) quantitatively evaluated the performance of word sense disambiguation of multiple types of large language models, and the results show that the growth of model scale cannot naturally replace refined context modeling, and more targeted structure design is still needed for accurate discrimination of ambiguous words [8].

On the other hand, research on context-aware machine translation and professional text translation provides important support for the identification of ambiguous words in scientific and technological texts. Vu et al. (2024) introduced the source anaphora interpretation mechanism into context-aware machine translation, which improved the model's ability to grasp the cross-sentence semantic relationship [9]. Appicharla et al. (2024) combined multi-task learning to construct a context-aware translation method, and proved that the explicit introduction of context information was helpful to improve the consistency of translation in complex contexts [12]. Honda et al. (2023) verified the role of context-aware mechanism in promoting fine-grained semantic selection in English-Japanese business dialogue translation [13]. Castilho and Knowles (2025) conducted a systematic review of "context" research in neural machine translation, and pointed out that current methods generally have problems such as single context modeling granularity and scattered evaluation criteria [14]. In the scenario of scientific and academic texts, Yue et al. (2024) focused on the translation process of term abbreviations, emphasizing that the correct recognition of professional terms depended on the support of task-based context [15]. Roussis et al. (2024) pointed out from the perspective of scientific text machine translation that term function, inter-sentence cohesion and argument structure in scientific texts will jointly affect translation quality [16]. These studies show that the processing of ambiguous words in scientific texts can no longer rely on local clues at the level of single sentence, but needs more detailed multi-level semantic organization mechanism. To more clearly present the relationship between existing studies and the work in this paper, Table 1 provides an inductive comparison of related directions.

Table 1: Comparison of representative works related to the translation of ambiguous words in scientific and technological texts

Research Direction	Representative Literature	Core Approach	Main Characteristics	Existing Limitations	Implications for This Study
Bias Analysis of Word Sense Disambiguation in Machine Translation	Campolungo et al. (2022) [1]	Constructing evaluation benchmarks for ambiguous-word bias	Able to reveal weaknesses in model disambiguation	Focuses more on evaluation, with insufficient design of discrimination mechanisms	A dedicated discrimination model for ambiguous words is needed
Translation Enhancement with Explicit Word Sense Information	Campolungo et al. (2022) [2]	Introducing word sense information into NMT	Can alleviate some translation bias	Limited adaptability to complex scientific contexts	Domain-specific context should be incorporated for further extension
Word Sense Linking in Open Environments	Bejgu et al. (2024) [5]	Performing linking discrimination in an open sense space	Suitable for complex semantic mapping	Limited specificity for translation scenarios	Its idea of sense association can be adopted
Homograph Processing in Pre-trained Language Models	Proietti et al. (2024) [7]	Analyzing the disambiguation ability of PLMs for homographs	Reflects the potential of contextual modeling	Insufficient capture of fine-grained domain differences	Multi-granularity semantic representation should be strengthened
Evaluation of Ambiguity Handling in Large Language Models	Sumanathilaka et al. (2024) [8]	Conducting quantitative tests on multiple LLMs	Shows the upper limit of large-model discrimination ability	Lacks dedicated structural constraints	Parameter scale cannot replace mechanism design
Context-Aware Machine Translation	Vu et al. (2024) [9]; Appicharla et al. (2024) [12]	Integrating cross-sentence information and auxiliary tasks	Helps improve contextual consistency	Insufficient fine-grained lexical discrimination	Local and global contexts should be combined
Scientific Text and Terminology Translation Research	Yue et al. (2024) [15]; Roussis et al. (2024) [16]	Conducting translation analysis for terminology and scientific discourse	Closely aligned with scientific text scenarios	Limited dedicated modeling of ambiguous words	The characteristics of scientific text scenarios should be highlighted

It can be seen from Table 1 that the existing research has promoted related issues from the perspectives of word sense disambiguation, context enhancement, pre-trained model analysis and scientific text translation, but there is still a lack of system modeling of "ambiguous word discrimination in scientific text" as an independent task. Existing methods either focus on external word sense injection or discourse context utilization, and there is still a lack of research that can truly integrate lexical level, phrase level, sentence level and local discourse level information at the same time. Based on this, this paper intends to construct a translation discrimination model of scientific and technological text ambiguous words under the multi-granularity attention mechanism, which organizes semantic clues at different levels within a unified framework, and enhances the model's comprehensive recognition ability for

professional context, meaning boundaries and translation differences, which lays a foundation for subsequent model design and experimental analysis.

3 Construction of the translation discrimination model of scientific text ambiguous words based on multi-granularity attention mechanism

3.1 Feature representation of ambiguous words and multi-granularity semantic information extraction in scientific and technological texts

In the process of constructing a translation discrimination model for ambiguous words in scientific texts, feature representation and semantic information extraction are the basic links that determine the accuracy of subsequent translation selection. Different from general texts, ambiguous words in scientific texts are often constrained by term attributes, local collocations, syntactic roles and discourse topics. If the representation is only based on static word embedding or fixed window context, the model can capture the surface adjacency relationship, but it is difficult to distinguish the real semantic orientation of the same word in different scientific and technological scenarios, resulting in professional translation deviation and blurred discrimination boundaries. Based on this, this paper constructs a feature extraction method around "lexical representation-multi-granularity modeling-semantic fusion", and incorporates the word-level, phrase-level and sentence-level information of the sentence in which the ambiguous word is contained into a unified computing framework to enhance the model's ability to recognize subtle differences in scientific and technological context (as shown in Figure 1).

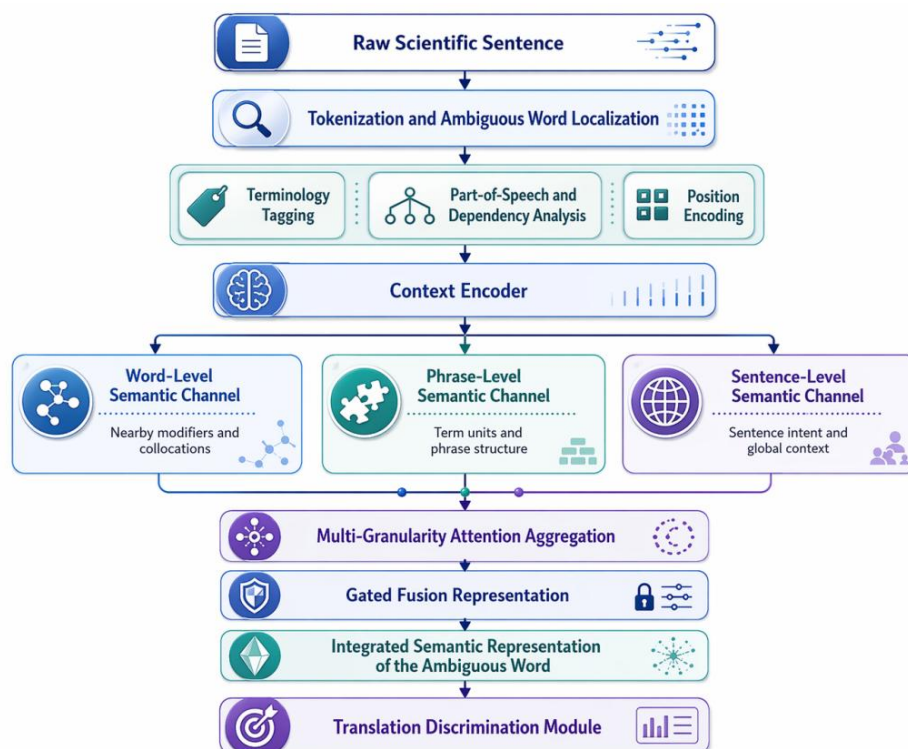


Figure 1: Feature representation of ambiguous words and multi-granularity semantic information extraction framework in scientific and technological texts

Specifically, the feature representation of ambiguous words in scientific texts includes four steps: text encoding, term tagging, structural complement and context mapping. For input sentence $S = \{w_1, w_2, \dots, w_n\}$ after word segmentation, word embeddings, position embeddings, term category embeddings and dependency role embeddings are extracted and concatenated to form the initial representation:

$$x_i = [e_i^{\text{tok}} \parallel e_i^{\text{pos}} \parallel e_i^{\text{term}} \parallel e_i^{\text{dep}}] \quad (1)$$

where, e_i^{tok} represents the lexical semantic vector, e_i^{pos} represents the positional feature, e_i^{term} identifies whether the word is inside the term unit, and e_i^{dep} reflects its syntactic dependency role. All input vectors are mapped by the context encoder to obtain a sequence of context representations: input, term category embeddings and dependency role embeddings, which are concatenated to form an initial representation:

$$H = \text{Encoder}(X) = \{h_1, h_2, \dots, h_n\} \quad (2)$$

Among them, h_i is no longer an isolated word representation, but a semantic state fused with local context, which provides a basic representation for subsequent discrimination of ambiguous words.

In the multi-granularity semantic extraction stage, we construct word-granularity neighborhood $\Omega_i^{(w)}$, phrase-granularity neighborhood $\Omega_i^{(p)}$ and sentence-granularity neighborhood $\Omega_i^{(s)}$, respectively, with the ambiguous word w as the center. Word granularity is used to capture neighboring modifiers, collocations and functional determiners. Phrase granularity highlights noun phrases, verb-object structure and term combination units. The sentence granularity is used to represent the goal of the whole sentence and the semantic direction of science and technology. For any granularity g , the attention weight is defined as follows.

$$\alpha_{ij}^{(g)} = \frac{\exp\left((W_q^{(g)} h_i)^\top (W_k^{(g)} h_j) / \sqrt{d}\right)}{\sum_{m \in \Omega_i^{(g)}} \exp\left((W_q^{(g)} h_i)^\top (W_k^{(g)} h_m) / \sqrt{d}\right)} \quad (3)$$

Accordingly, the semantic aggregation result under this granularity can be obtained as follows.

$$c_i^{(g)} = \sum_{j \in \Omega_i^{(g)}} \alpha_{ij}^{(g)} W_v^{(g)} h_j \quad (4)$$

This process enables the model to redistribute the focus of attention in different semantic ranges and avoid premature judgment of ambiguous words based on only a single context cue.

Considering that the sense discrimination in scientific texts is often not determined by a single granularity, we further use the gated fusion mechanism to generate the comprehensive representation of ambiguous words:

$$z_i = \lambda_i^{(w)} c_i^{(w)} + \lambda_i^{(p)} c_i^{(p)} + \lambda_i^{(s)} c_i^{(s)} \quad (5)$$

Here, $\lambda_i^{(w)}$, $\lambda_i^{(p)}$ and $\lambda_i^{(s)}$ are the normalized weights, which measure the contribution of different granularity semantics to the discrimination of the current ambiguous word. The

fused z_i can preserve local collocation constraints, structural unit information and overall sentential semantic orientation at the same time, which is more suitable for fine-grained translation sense discrimination in scientific texts.

At the implementation level, this paper integrates term dictionary matching, dependency parsing and context encoding in scientific and technological texts into a unified preprocessing link, and uses a multi-granularity attention module to complete semantic extraction and representation update. The ambiguous word representation vector output by the model will be directly fed into the subsequent translation discrimination module to complete the classification and ranking of candidate translations. In this way, the discriminative evidence originally scattered in different semantic levels of scientific texts is organized into a computable and comparable unified representation, which not only enhances the adaptability of the model to the professional context, but also provides a clear input basis for the subsequent structure design and experimental analysis.

3.2 Model design of ambiguity translation Discrimination based on multi-granularity attention mechanism

In the task of translation of ambiguous words in scientific texts, most of the traditional translation models implicitly deal with ambiguous words in the end-to-end decoding process. Although the models can rely on large-scale corpora to learn general context correspondences, faced with the constraints of term reuse, cross-domain sense borrowing and long-distance semantics that occur frequently in scientific texts, the model is able to learn the general context correspondences. There is often no explicit discrimination mechanism for candidate translations of ambiguous words, which is prone to local semantic matching and overall imbalance of translation meaning. Based on this, this paper proposes an ambiguity translation discrimination model based on multi-granularity attention mechanism, which integrates the candidate translation representation, context-aware representation and category decision process into the same framework, so as to improve the stability and interpretability of the translation selection of ambiguous words in scientific texts.

From the perspective of the core discrimination process, the model is composed of a candidate translation encoding module, a context discrimination module, a decision output module and a result correction module. The candidate encoding module is responsible for vectorizing multiple candidate translation terms corresponding to ambiguous words, and mapping dictionary definitions, term attributes and historical co-occurrence features into a set of candidate vectors $Q_i = \{q_i^1, q_i^2, \dots, q_i^m\}$. The context discrimination module receives the multi-granularity fusion representation z_i generated in Section 3.1, and further introduces the sentence-level topic vector g_i and the term environment vector u_i to form the joint discrimination state for translation selection:

$$r_i = \tanh(W_r[z_i \parallel g_i \parallel u_i] + b_r) \quad (6)$$

where, r_i represents the comprehensive discriminative representation of ambiguous words in the current technological context, and W_r and b_r are learnable parameters. This representation does not stay in local neighborhood aggregation, but compresses the word surrounding cues, sentential semantic constraints and term environment into a unified discriminant space.

In the decision output stage, the model matches the comprehensive discriminant representation r_i with the candidate translation vector q_i^k to calculate the discrimination score of each candidate translation:

$$s_i^k = r_i^T W_s q_i^k + b_s \quad (7)$$

Furthermore, the selection probability of a candidate translation y_i^k can be expressed as follows.

$$P(y_i^k | x_i) = \frac{\exp(s_i^k)}{\sum_{j=1}^m \exp(s_i^j)} \quad (8)$$

where, m is the number of translation candidates. The model finally selects the translation with the largest probability as the discrimination result of the current ambiguous word. Different from the implicit processing of ambiguous words by the translation decoder directly, the design separates the "translation judgment" from the general sequence generation, so that the discrimination process has more explicit target constraints, and is more suitable for the application scenarios with narrow semantic boundaries and high mistranslation costs in scientific and technological texts.

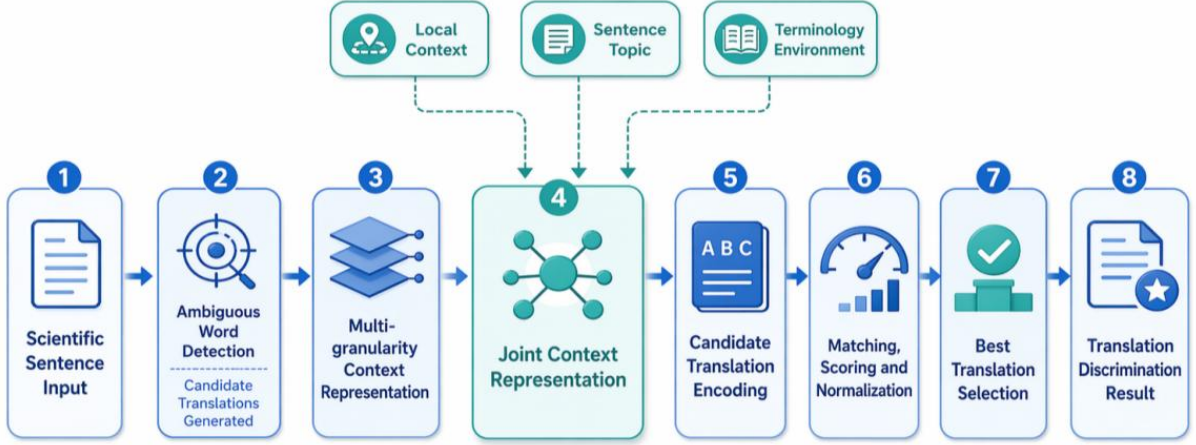


Figure 2: Model design for translation discrimination of ambiguous words based on multi-granularity attention mechanism

At the implementation level, this paper integrates term dictionary matching, dependency parsing and context encoding in scientific and technological texts into a unified preprocessing link, and uses a multi-granularity attention module to complete semantic extraction and representation update. The overall process is shown in Figure 2. During training, the cross-entropy loss is used to optimize the distribution of candidate translations, and the class weight is combined to alleviate the imbalance problem of different sense samples. The output of the model can not only directly serve the independent ambiguous word translation discrimination task, but also be used as the pre-discrimination module in the subsequent scientific text machine translation system to provide more stable candidate constraints for the decoding stage. Through this design, the translation sense selection of ambiguous words in scientific texts changes from "implicit generation" to "explicit discrimination", which helps to enhance the robustness and discrimination accuracy of the model in complex professional contexts.

3.3 Structural composition of the translation discrimination model for ambiguous words in scientific texts

In the task of translation discrimination of ambiguous words in scientific and technological texts, the performance of the model depends not only on the discrimination ability of a single attention unit, but also on the organization mode and information transmission mechanism between each functional module. If the structure of single-channel context encoding and direct classification output is still adopted, although the model can obtain certain semantic representation ability, it is often difficult to maintain the consistency and stability of translation selection in the scientific and technological context with dense terms, syntactic nesting and obvious cross-sentence constraints. Based on this, this paper constructs a hierarchical structural framework composed of input representation layer, multi-granularity semantic aggregation layer, candidate translation discrimination layer and result correction layer, so that the translation discrimination of ambiguous words forms a closed-loop link from the entry of the text to the output of the result. The overall structure is shown in Figure 3.

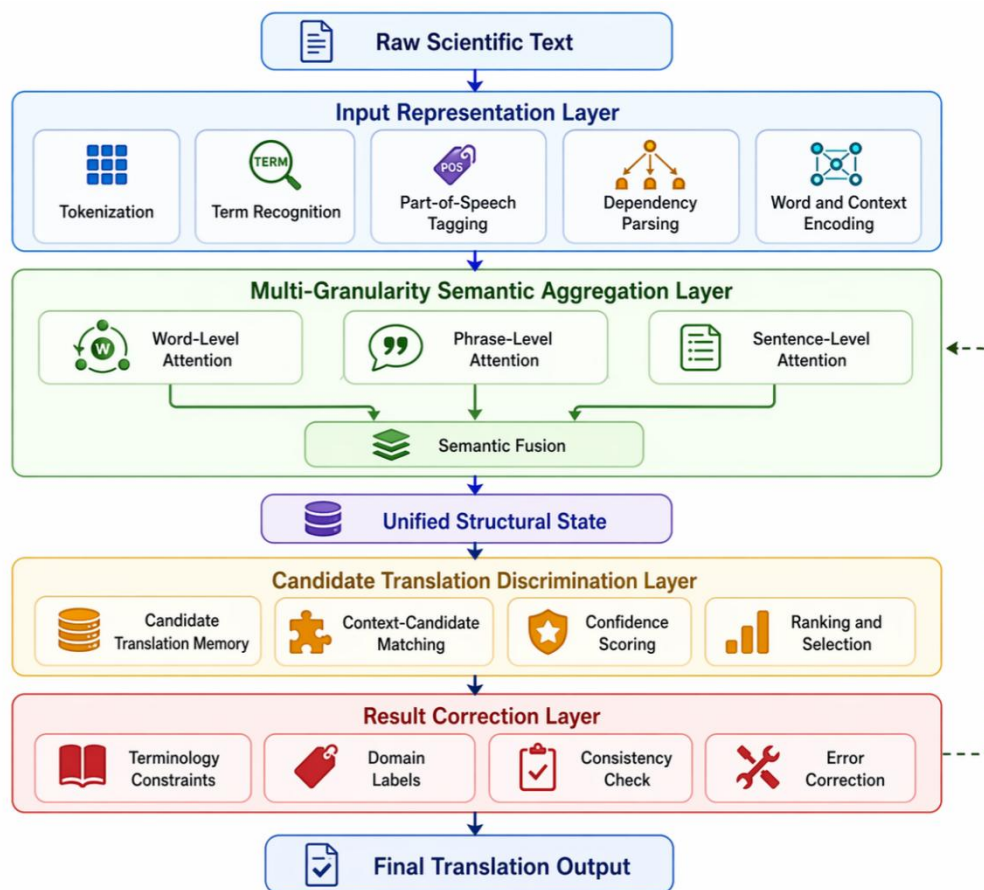


Figure 3: Structural composition of the discriminative model for translation of ambiguous words in scientific and technological texts

The input representation layer is responsible for receiving the original scientific sentences, completing word segmentation, term recognition, part-of-speech tagging and dependency extraction, and mapping the ambiguous words and their contexts into a unified vector space. The multi-granularity semantic aggregation layer extracts semantic clues at three levels: word level, phrase level and sentence level, and strengthens the information related to the target word discrimination through attention allocation. The candidate translation discrimination

layer matches the aggregated context representation with the candidate translation library, and outputs the confidence score of each translation. The result correction layer combines term constraints, domain labels and intra-sentence consistency rules to re-filter the initial discrimination results to reduce misjudgments caused by local optima. Each layer is not simply connected in series, but a linkage structure is formed by the output of the former layer constraining the judgment of the later layer, and the feedback of the later layer correcting the attention of the former layer.

In order to describe the information convergence process in this structure, let the base representation of ambiguous words in the input layer be h_i , the multi-granularity semantic aggregation result be z_i , and the semantic memory matrix of candidate translation items be M_i . Then the structure fusion state can be expressed as follows.

$$m_i = \text{LayerNorm}(W_m[h_i \parallel z_i \parallel M_i] + b_m) \quad (9)$$

where, m_i represents the unified structure state before entering the discriminative layer, W_m and b_m are learnable parameters. Based on this state, the model evaluates the set of candidate translations $Y_i = \{y_i^1, y_i^2, \dots, y_i^K\}$ to score and output the final discrimination result:

$$\hat{y}_i = \arg \max_{k \in \{1, \dots, K\}} \text{Softmax}(W_o m_i + b_o)_k \quad (10)$$

The formula shows that the final output of the model does not depend on a single context fragment, but a global choice made after considering the semantic aggregation results and candidate translation memory in a unified structural state. In the running process, the input representation layer completes the text parsing, the multi-granularity semantic aggregation layer generates the context evidence for discrimination, the candidate translation discrimination layer ranks the translation items, and the result correction layer verifies the consistency of professional terms and the semantic coherence within the sentence.

3.4 Process of translation discrimination of ambiguous words in scientific and technological texts

In the task of translation of ambiguous words in scientific texts, the model not only needs to complete the semantic discrimination of candidate translation items, but also needs to maintain the stability and coherence of the discrimination process in complex contexts. If the processing of ambiguous words is completely left to the single-step classifier, although the computing link is short, in the scientific text with dense terms, obvious syntactic compression and overlapping context constraints, the problem of excessive local evidence and overall semantic imbalance is often prone to occur. Therefore, based on the above feature representation, model design and structure organization, this paper further constructs the translation discrimination process of ambiguous words for scientific and technological texts. The input parsing, semantic extraction, candidate scoring, result correction and final output are connected as continuous computing links, so that the discrimination process turns from static recognition to dynamic screening. The overall process is shown in Figure 4.

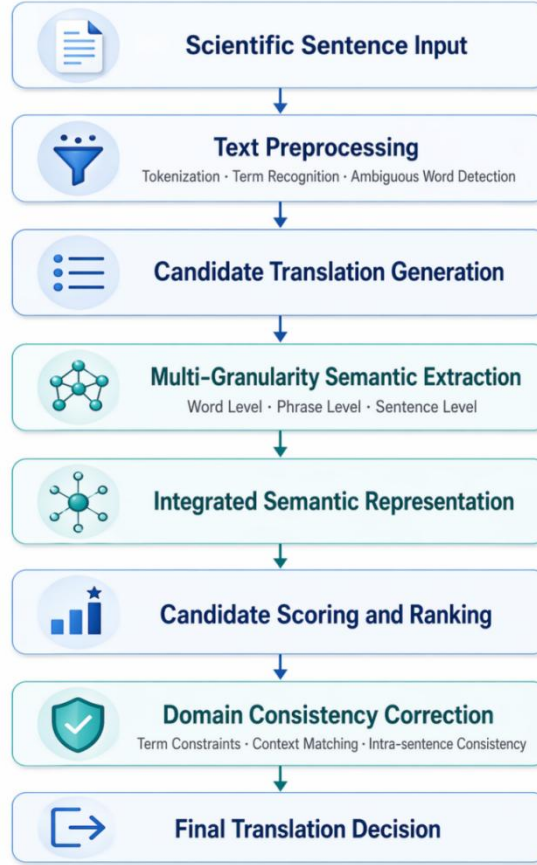


Figure 4: Process of translation discrimination of ambiguous words in scientific and technological texts

The process mainly includes four steps. First, the input processing stage receives the original scientific sentence, completes word segmentation, term recognition, ambiguous word location and candidate translation generation, and maps the internal structure information into a unified representation. Secondly, the multi-granularity semantic extraction stage constructs word-level, phrase-level and sentence-level semantic evidence around the target ambiguous word, and obtains the comprehensive semantic vector under the current context. Thirdly, the candidate translation discrimination stage scores and sorts the candidate translations to generate the initial discrimination results. Fourth, in the result correction stage, high similarity candidates are re-checked according to term constraints, domain labels and intra-sentence semantic consistency, thereby reducing the risk of misjudgment. Through this link, the meaning selection in scientific texts no longer relies on a single point of judgment, but gradually converges under the common constraints of multiple evidences.

In order to formally describe the discrimination and correction mechanism in this process, let the initial set of candidate translations of ambiguous word x_i be $C_i = \{c_i^1, c_i^2, \dots, c_i^K\}$, the model obtains the original score \tilde{s}_i^k of the candidate translation based on the comprehensive semantic representation m_i . Further introducing the domain consistency constraint r_i^k , the corrected discriminant score can be expressed as follows.

$$\tilde{s}_i^k = \eta s_i^k + (1 - \eta) r_i^k \quad (11)$$

where $\eta \in [0,1]$ is the trade-off coefficient, and r_i^k is used to describe the consistency between the candidate translation and the current context of the science and technology field.

Based on the corrected scores, the final discrimination result is defined as follows.

$$y_i^* = \arg \max_{k \in \{1, \dots, K\}} \tilde{s}_i^k \quad (12)$$

This formula shows that the final output is not simply to select the candidate with the highest original probability, but to make a more robust translation selection under the joint effect of model scoring and domain verification. In the specific operation, after the input processing module completes the text parsing, the multi-granularity attention module outputs the context representation of the ambiguous word, the candidate discrimination module generates the preliminary ranking results, and the correction module combines the term rules and the semantic relationship in the sentence to complete the screening, and finally obtains the translation output.

4 Experimental Results

4.1 The dataset

In order to test the adaptability and stability of the model constructed in this paper in the task of translation discrimination of ambiguous words in scientific and technological texts, the experiment uses the data organization method of "public corpus + manual annotation corpus". The public part mainly selects parallel sentence pairs with high term density in English and Chinese scientific and technological texts, and the self-built part focuses on four fields of artificial intelligence, biomedicine, material engineering and information communication, and supplements the candidate translation items of ambiguous words and context labels. After cleaning, 42 600 bilingual sentence pairs of scientific texts were finally formed, including 18 240 ambiguous word instances and 54 730 candidate translation records. In order to ensure the pertinence of the discrimination task, three kinds of high-frequency ambiguity phenomena, such as homograph, synonym competition and term escape, are retained in the corpus. All the data were divided into training set, validation set and test set according to 8:1:1, corresponding to 34 080, 4 260 and 4 260 groups of samples.

In the experimental implementation, sentence pair alignment, duplicate sample elimination, term standardization and sub-word segmentation are completed in a unified way, and the position of ambiguous words, candidate translation categories and domain labels are synchronously annotated for all texts. Considering the strong context constraints in scientific texts, in addition to retaining source sentences and target sentences, the dataset also records term boundaries, syntactic dependencies, domain categories and ambiguity types to support multi-granularity semantic modeling and subsequent ablation analysis. Table 2 shows the structural composition of the data set in this study and the experimental use. Overall, the dataset not only covers the common professional contexts in scientific texts, but also retains the fine-grained label information required for ambiguity discrimination, which can provide a more stable data basis for model training, comparative evaluation and error analysis.

Table2: Data set structure and experimental purposes

Data Type	Quantity	Main Fields	Purpose Description
Bilingual Scientific Text Sentence Pairs	42,600 pairs	ID, source sentence, target sentence, sentence length, domain label	Used for model training and overall performance evaluation
Annotated Ambiguous Word Instances	18,240 entries	Ambiguous word position, part of speech, ambiguity type, standard translation	Used for supervised learning in the translation discrimination task
Candidate Translation Records	54,730 entries	Candidate translation, term attribute, semantic label, frequency	Used for candidate ranking and discrimination output
Contextual Auxiliary Information	42,600 entries	Term boundary, dependency relation, context window, discourse position	Used for multi-granularity semantic extraction and error analysis

4.2 Data preprocessing

In the task of translation discrimination of scientific and technical ambiguous words, the original corpus often contains problems such as inconsistent terminology, alignment deviation of sentence pairs, redundant annotation of candidate translations, and uneven distribution of domain labels. Without effective data preprocessing, the model is susceptible to noise samples and abnormal features during the training process, which will lead to semantic boundary blurring, aggravated class bias and unstable convergence. Based on this, this paper constructs a multi-stage preprocessing mechanism of "corpus cleaning - ambiguous word alignment - candidate normalization - feature normalization" to improve the consistency, computability and reproducibility of the input data. The overall process is shown in Figure 5.

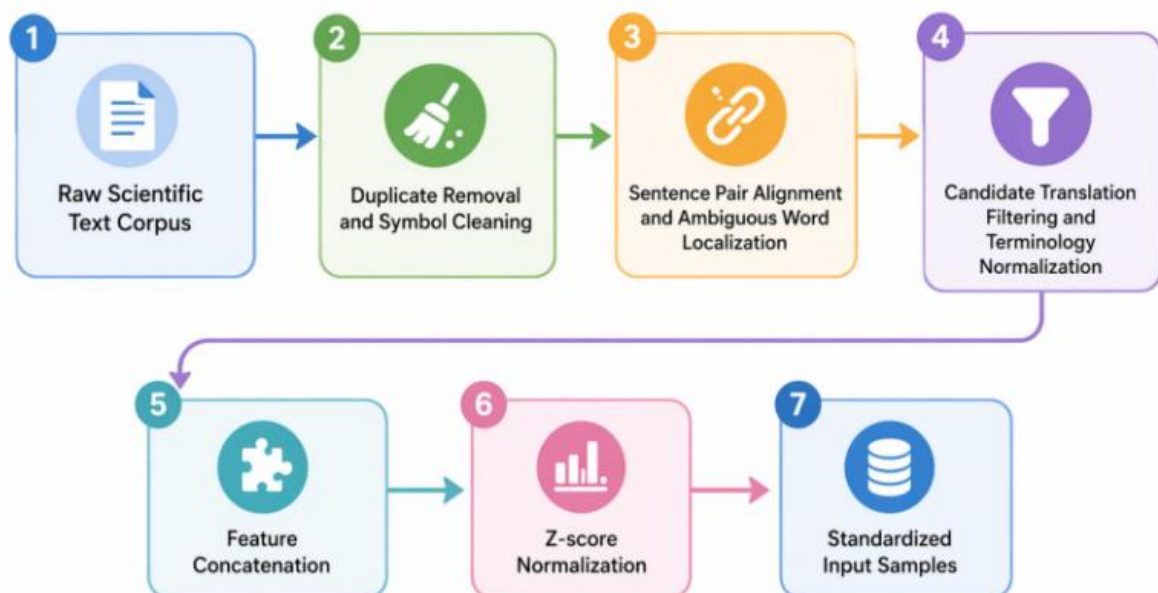


Figure 5: Data preprocessing process

At the corpus level, the repeated samples are eliminated, abnormal symbols are filtered and sentence length is truncated to remove the missing source language or target language, the mismatch of sentence pairs and the garble code of terms. Then, combined with the glossary

and manual verification results, the position of ambiguous words in the sentence is relocated, and the case, hyphen, abbreviation and plural forms in different translations are unified. Aiming at the problem that the set of candidate translation items is too large and there are many low-frequency interference items, this paper retains the semantically valid candidate items that appear no less than 5 times. At the same time, the translation items that are synonymous but not applicable to the current domain context are separately marked as interference categories, so as to enhance the difficulty of discrimination task.

In the feature curation stage, we combine the lemma semantics, ambiguity position, term boundary and domain label into a unified input vector. Let the original representation of the i th lemma be v_i , ambiguity position markers p_i , term boundary features t_i , and domain label embeddings d_i , then the preprocessed input representation is defined as:

$$\tilde{x}_i = W_p[v_i \parallel p_i \parallel t_i \parallel d_i] + b_p \quad (13)$$

Here, W_p and b_p are mapping parameters, and \tilde{x}_i represents the normalized input features before entering the model. In order to reduce the influence of feature scale differences of different dimensions on the training process, this paper further uses the Z-score method for standardization processing:

$$\hat{x}_{ij} = \frac{\tilde{x}_{ij} - \mu_j}{\sigma_j + \varepsilon} \quad (14)$$

where, μ_j and σ_j represent the mean and standard deviation of the JTH dimension feature respectively, and ε is a minimal constant set to prevent the denominator from being zero. After this process, the features from different sources are compressed to a relatively consistent numerical interval, which is conducive to the model's stable learning of multi-granularity semantic information. Through the above processing, the original scientific text corpus is transformed into standardized data input with clear structure, unified annotation and comparable scale, which provides a reliable basis for subsequent evaluation index calculation, model training and ablation experiments.

4.3 Evaluation Metrics

In order to comprehensively test the effectiveness of the translation discrimination model of scientific and technological text ambiguous words under the multi-granularity attention mechanism, this paper establishes an evaluation system from four dimensions of discrimination accuracy, class balance, candidate ranking ability and reasoning efficiency. Considering that this task is not general machine translation generation, but the discrimination and selection of candidate translation items, the evaluation index is no longer BLEU and other whole sentence generation indicators. Accuracy, Macro-F1, MRR and Latency are selected as the core measures. Macro-F1 is used to reflect the comprehensive recognition level of various ambiguous words under the condition of unbalanced class distribution. MRR is used to describe the ranking position of correct translations in the candidate list.

For the class balance evaluation, this paper uses the macro average F1 value as the main index, which is defined as:

$$\text{Macro-F1} = \frac{1}{C} \sum_{c=1}^C \frac{2P_c R_c}{P_c + R_c} \quad (15)$$

where C represents the total number of ambiguous categories, P_c and R_c represent the

precision and recall of class c , respectively. This index can avoid excessive dominance of high-frequency categories on the overall results, and is more suitable for evaluating the discrimination stability of different ambiguity types in scientific texts. For candidate ranking evaluation, the average reciprocal ranking is used in this paper, which is expressed as:

$$\text{MRR} = \frac{1}{N} \sum_{i=1}^N \frac{1}{\text{rank}_i} \quad (16)$$

where N is the number of test samples and rank_i represents the ranking position of the correct translation item in the candidate list in the i th sample. The higher the MRR value, the better the model ranks the correct translation. In terms of efficiency evaluation, the average inference duration is defined as:

$$T_{\text{avg}} = \frac{1}{N} \sum_{i=1}^N (t_i^{\text{out}} - t_i^{\text{in}}) \quad (17)$$

where, t_i^{in} and t_i^{out} represent the timestamp of the i th sample entering the model and the output result, respectively. This index can be used to measure the responsiveness of the model while maintaining the discrimination accuracy. All results in the experiment were run repeatedly with three random seeds and reported in the form of "mean \pm standard deviation" to improve the robustness of the interpretation of the results. Through the combination of the above indicators, the performance of the model in the task of translation discrimination of ambiguous words in scientific and technological texts can be more completely presented, and a unified evaluation basis for subsequent ablation experiments and comparative analysis can be provided. In order to further analyze the discrimination certainty and error distribution of the model under different contexts and different ambiguity types, Top-1 discrimination accuracy, average confidence difference and false positive rate are added as auxiliary statistical indicators.

4.4 Ablation experiments

In order to further verify the practical role of the multi-granularity attention mechanism and the result correction strategy in the translation discrimination of ambiguous words in scientific and technological texts, this paper designs multiple sets of ablation experiments to gradually remove the key components of the model while maintaining the consistency of data set division, training rounds, learning rate and batch size. And observe its influence on the discrimination accuracy, macro average F1 value, average reciprocal ranking and reasoning time. A total of five groups of models were set up in the experiment: (1) Word-level attention was removed, and only phrase-level and sentence-level semantic modeling were retained; (2) removing phrase-level attention and retaining only word-level and sentence-level semantic modeling; (3) Removing sentence-level attention and retaining only word-level and phrase-level semantic modeling; (4) The result correction module is removed, and the discriminant results are output only according to the initial candidate ranking. (5) complete model, that is, retaining multi-granularity attention and result correction mechanism at the same time. The models of each group were repeated for 5 times, and the average result was taken as the final statistical value.

Table 3: Comparison of the results of ablation experiments

Model Setting	Accuracy (%)	Macro-F1 (%)	MRR	Latency (ms)
Without Word-Level Attention	88.6	87.5	0.913	19.4
Without Phrase-Level Attention	89.1	88.0	0.901	19.1
Without Sentence-Level Attention	87.9	86.8	0.895	18.7
Without Result Correction Module	90.3	89.2	0.928	17.9
Full Model	91.8	90.9	0.946	20.2

Table 3 shows the experimental results with different ablation Settings. From the overall trend, the complete model achieves the best performance in the three effectiveness indicators of Accuracy, Macro-F1 and MRR, but its average inference time is slightly higher than that of each ablation model. Specifically, after removing the word-level attention, the Accuracy decreases from 91.8% to 88.6%, and the Macro-F1 decreases from 90.9% to 87.5%, indicating that the local collocation information has an important impact on the instant discrimination of scientific text ambiguous words. After removing the phrase-level attention, the MRR of the model decreases significantly. From the perspective of overall reduction, removing sentence-level attention has a more prominent impact on model performance. After removing sentence-level attention, the overall performance decreases the most, with Accuracy reduced to 87.9% and Macro-F1 reduced to 86.8%, which indicates that the discrimination of ambiguous words in scientific texts is not completely determined by the local neighborhood, and the semantic direction of the whole sentence has a strong constraint on the translation selection. After removing the result correction module, the accuracy and ranking ability of the model are still better than the first three groups of ablation models, but the performance is still significantly lower than that of the full model, which indicates that term consistency verification and domain matching correction have practical value for reducing misjudgment.

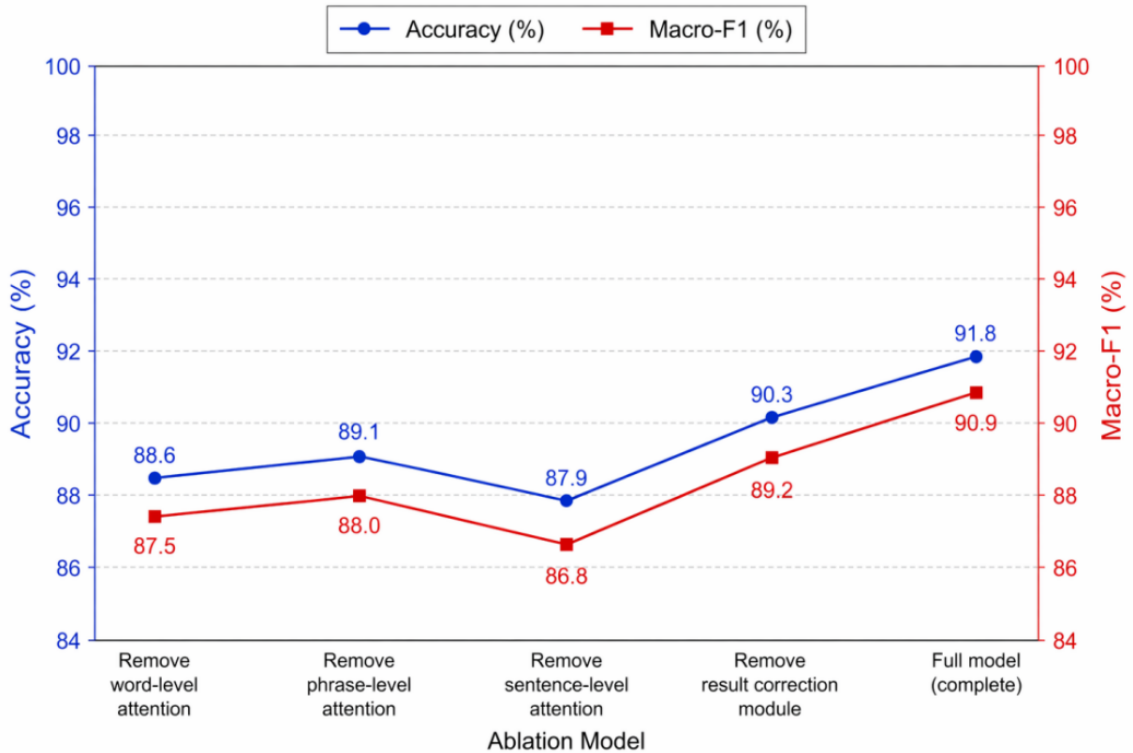


Figure 6: Comparison of different ablation models on core metrics

Figure 6 further compares the differences between the models in the two core indicators of Accuracy and Macro-F1. It can be seen that the full model maintains the highest level in both metrics, while the model with missing sentence-level attention shows the most significant decrease. This shows that the multi-granularity attention mechanism is not simply composed of several independent local modules, but forms a complementary relationship between different semantic levels. In general, the ablation experimental results verify the rationality of the model design, and also show that a more complete discriminant link between fine-grained semantic extraction and result correction should be established to obtain a stable performance of translation discrimination of ambiguous words in scientific texts, rather than relying on single-scale contextual features.

5 Discussion

5.1 Comparison with existing translation discrimination methods for ambiguous words

In order to test the relative advantages of the proposed model in the task of ambiguous word translation discrimination in scientific and technological texts, this paper selects three representative baseline methods: BiLSTM-Attention, BERT discrimination model and Context-aware Transformer. And it is retrained under the same data set partition, training rounds, parameter update strategy and hardware environment to ensure that the comparison results are comparable. The comparison results show that although the existing methods can complete the candidate translation screening in the general context, when the term reuse, local collocation compression and sentence-level semantic traction simultaneously occur in the scientific text, the model often has problems such as insufficient candidate ranking forward, confusion of near meaning and unbalanced recognition between categories. In contrast, the multi-granularity attention model proposed in this paper performs better on a number of core indicators.

From the overall results, the Accuracy of BiLSTM-Attention is 85.4%, Macro-F1 is 84.1%, and MRR is 0.881. The BERT discrimination model reached 88.7%, 87.6% and 0.914, respectively. The Context-aware Transformer is further improved to 90.2%, 89.0% and 0.928. The Accuracy, Macro-F1 and MRR of the proposed model reach 91.8%, 90.9% and 0.946, respectively, which are higher than those of the above baseline methods. It shows that it is still difficult to fully describe the multi-layer semantic constraints of ambiguous words in scientific texts by only relying on single-layer context coding or general context enhancement. The multi-granularity attention mechanism can absorb word-level adjacency information, phrase-level structure information and sentence-level semantic orientation at the same time, so that the translation selection no longer stays in local matching, but turns to more stable comprehensive discrimination. In terms of inference efficiency, the average inference time of the proposed model is 20.2 ms, which is slightly higher than 16.8 ms of BiLSTM-Attention and 18.9 ms of BERT model, but it is still in the acceptable range.

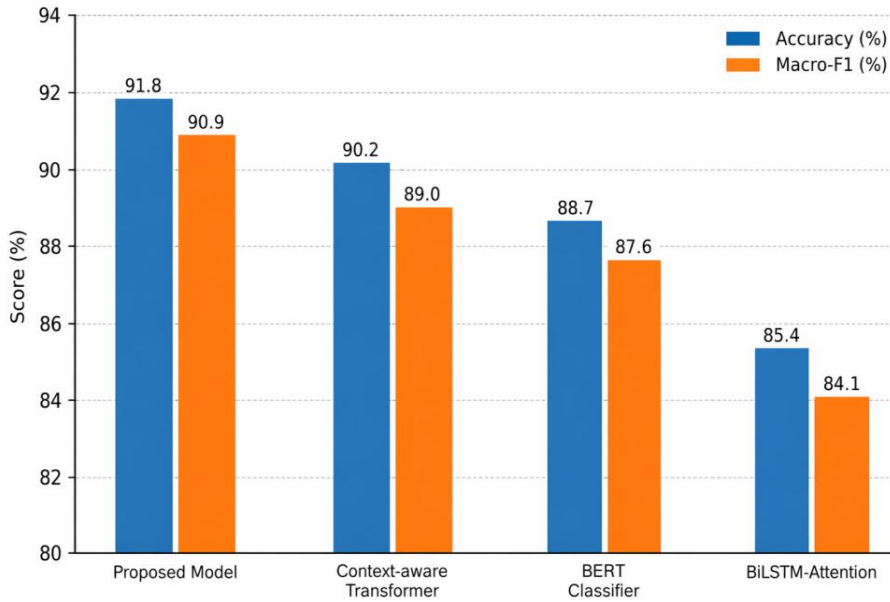


Figure 7: Performance comparison of different translation discrimination methods for ambiguous words

Figure 7 shows the comparison results of different methods on the two indicators Accuracy and Macro-F1. It can be seen that the proposed model maintains the highest level in both indicators, and the advantages are not concentrated on a single indicator, but reflected in both overall accuracy and class balance. This indicates that the proposed method does not gain superficial improvement by favoring high-frequency senses, but maintains a relatively stable discrimination ability under different ambiguity types. In general, compared with the existing ambiguous word translation discrimination methods, the multi-granularity attention mechanism shows stronger semantic integration ability and more robust translation discrimination ability in the scientific text scene.

5.2 Analysis of the role of multi-granularity attention mechanism

Translation discrimination of ambiguous words in scientific texts is not always carried out in a single and stable context. In the face of complex situations such as dense distribution of terms, nested expression of long sentences, cross-domain semantic transfer and competition of synonym candidates, traditional single-layer context modeling methods are often prone to the problems of too strong local features and insufficient overall semantic constraints. In order to test the actual role of multi-granularity attention mechanism in complex context, this paper sets up four typical test scenarios, namely "term intensive context", "long sentence nested context", "cross-domain transfer context" and "semantic competition context", and uses Accuracy, Macro-F1 and MRR as evaluation indicators to test the complete model. The results are presented in Table 4.

Table 4: Performance of multi-granularity attention mechanism in different complex contexts

Test Scenario	Accuracy (%)	Macro-F1 (%)	MRR
Terminology-Intensive Context	92.4	91.7	0.951
Long Sentence Nested Context	90.8	89.6	0.938
Cross-Domain Transfer Context	89.9	88.8	0.927
Synonym Competition Context	89.1	87.9	0.921

The experimental results show that the multi-granularity attention mechanism maintains a relatively stable discrimination ability in different complex scenes. In the "terminological intensive context", the accuracy of the model reaches 92.4%, and Macro-F1 reaches 91.7%, indicating that word-level and phrase-level attention can better capture the term adjacency and fixed collocation structure, thereby avoiding the wrong selection of professional translation items. In the "long sentence nested context", the model Accuracy is 90.8%, which is slightly lower than the former, but the MRR still remains at 0.938, indicating that sentence-level attention has a significant compensation effect on cross-phrase dependence and overall sentential semantic orientation. In the "cross-domain transfer context", the accuracy of the model is 89.9%, and the Macro-F1 is 88.8%, which indicates that when the same ambiguous word crosses different scientific and technological subfields, the multi-granularity mechanism can still complete the semantic discrimination through different levels of semantic clues. As for the "near-meaning competition context", the model performance is relatively more challenging due to the high similarity of multiple candidate translations on the surface semantics, but the Accuracy still reaches 89.1%, indicating that the mechanism still has a certain ability to distinguish subtle semantic differences.

5.3 Analysis of the discriminant effect of the model in the context of scientific and technological texts

In order to further investigate the actual discrimination performance of the model in different scientific text contexts, this paper divides the test set into four categories according to the discourse function: term definition context, method process context, result explanation context and comparison argument context. The Top-1 discrimination accuracy, average confidence difference and false positive rate of each sample are calculated. Here, the average confidence difference refers to the average gap between the score of the correct candidate and the score of the second-best candidate, which can reflect the certainty of the model under the condition of candidate translation competition. Experimental results show that the multi-granularity attention mechanism shows good adaptability in different scientific and technological contexts, but there are still obvious differences in the discrimination difficulty of different contexts, and the relevant results are shown in Figure 8.

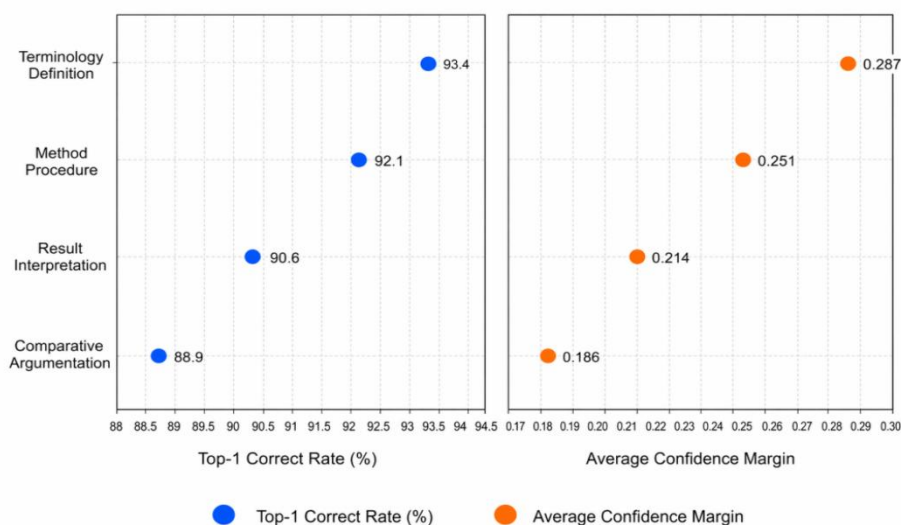


Figure 8: Model discrimination effect in different contexts of scientific and technological texts

In terms of specific performance, the model in the context of term definition has the best performance, with the Top-1 accuracy of 93.4%, the average confidence difference of 0.287, and the false positive rate of 6.6%. This result shows that the model can make full use of phrase boundary and sentential semantic orientation to complete translation item locking when the ambiguous word is in the defining sentence pattern or concept description structure. In the context of method process, the accuracy is 92.1%, and the average confidence difference is 0.251. The overall performance is still relatively stable, which is related to the strong procedural and fixed collocation of scientific and technological method description. In contrast, in the context of result interpretation, the accuracy drops to 90.6%, and the average confidence difference is 0.214, which indicates that when ambiguous words are embedded in conclusion interpretation, phenomenon attribution and trend description, semantic constraints will show stronger implicacy. The discrimination of comparative argument context was the most difficult, with the accuracy of 88.9%, the average confidence difference reduced to 0.186, and the false positive rate increased to 11.1%. The main reason for this phenomenon is that comparative sentences often introduce multiple reference objects and evaluation dimensions at the same time, and the translation sense selection of ambiguous words is not only affected by local collocation, but also by argument direction and discourse focus.

5.4 Differential Analysis of translation discrimination under different ambiguity types

The ambiguity phenomenon in scientific texts is not a single form. Different types of ambiguity have obvious differences in semantic source, context dependence strength and translation competition mode, which also determines the difficulty of model discrimination is not consistent. In order to further investigate the adaptability of the proposed model under different ambiguity types, this paper divided the test samples into four categories: homograph type, term escape type, synonym competition type and word class conversion type, and made statistics from the four dimensions of Accuracy, Macro-F1, MRR and misjudgment rate, and the results are shown in Table 5. On the whole, the model maintains a high level in all types of ambiguity tasks, but there is still a clear performance gradient between different types.

Table 5: Comparison of translation discrimination results under different ambiguity types

Ambiguity Type	Accuracy (%)	Macro-F1 (%)	MRR	Misclassification Rate (%)
Homograph Ambiguity	93.1	92.3	0.954	6.9
Terminological Shift Ambiguity	91.6	90.8	0.943	8.4
Synonym Competition Ambiguity	89.4	88.2	0.926	10.6
Part-of-Speech Conversion Ambiguity	88.7	87.5	0.918	11.3

From the results, the discrimination effect of homograph and heterograph samples is the best, with an Accuracy of 93.1%, Macro-F1 of 92.3%, MRR of 0.954, and a false positive rate of only 6.9%. This indicates that when the ambiguity mainly comes from the surface consistency of words and the semantic diversion is clear, the multi-granularity attention mechanism can quickly locate the correct translation with the help of local collocation and sentence-level semantics. The performance of the term escape sample is the second, with an Accuracy of 91.6%, indicating that the model has a strong ability to identify the "common words to term words" conversion that is common in the field of science and technology. The discrimination difficulty of the sample with similar meaning competition increases significantly, and the accuracy drops to 89.4%. The reason is that multiple candidate

translations are relatively close in semantic core, and it is difficult to distinguish stably by only one context feature. The results of part-of-speech conversion sample are the lowest, with an Accuracy of 88.7% and Macro-F1 of 87.5%, indicating that when ambiguous words are accompanied by functional changes and semantic transfer in sentences, the model is still affected by syntactic boundary ambiguity and structure compression.

5.5 Discussion

Comprehensive experimental results show that the proposed model outperforms the existing baseline methods in terms of accuracy, Macro-F1 and candidate ranking ability, and its advantages mainly come from the joint modeling of different semantic level information and the introduction of the result correction mechanism. The synergistic effect of word-level, phrase-level and sentence-level semantic cues enables the model to identify the true orientation of ambiguous words more stably in scientific texts, and maintain a good discrimination effect in complex contexts such as dense terms, long sentence nesting and synonym competition. However, the model in this paper still has some limitations. On the one hand, when the context of the ambiguous word is too dependent on the domain background knowledge, or there are only slight semantic differences between the translation candidates, the discriminative stability of the model will still decrease. On the other hand, although the current experimental corpus covers a variety of scientific and technological fields, there is still room for further expansion of data scale and annotation density in more subdivided professional scenarios. It can be seen that although multi-granularity semantic modeling improves the overall performance of translation discrimination of ambiguous words in scientific and technological texts, it still needs to be continuously optimized for highly confused samples and cross-domain transfer problems. The follow-up research can be promoted from two directions: first, enhance the modeling of domain knowledge constraints and term relationships to improve the adaptability of the model to complex professional contexts. Second, a more fine-grained error analysis and interpretability mechanism is introduced to improve the transparency and reliability of the model in practical scientific translation applications.

6 Conclusions

Focusing on the problems of unstable selection of translation items of ambiguous words, insufficient utilization of context constraints, and limited accuracy of candidate discrimination in scientific and technological texts, this paper constructs a translation discrimination model under multi-granularity attention mechanism. The model integrates word-level, phrase-level and sentence-level semantic information into a unified discriminative framework, and combines the result correction mechanism to enhance the model's ability to identify fine-grained semantic differences in scientific and technological contexts. Experimental results based on 42 600 bilingual sentence pairs and 18 240 ambiguous word instances in scientific and technical texts show that the Accuracy of the proposed model on the test set reaches 91.8%, Macro-F1 reaches 90.9%, MRR reaches 0.946, and the average inference time is 20.2 ms, which is better than the comparison methods. It shows that the method has a good effect in the discrimination accuracy, sorting ability and operation efficiency. The research also shows that the multi-granularity semantic joint modeling has strong adaptability to complex scenes such as dense terms, long sentence nesting and synonym competition. Nevertheless, there is still room for further optimization of the model under the conditions of high confusion sense and cross-subdivision domain transfer. In the future, the generalization

ability and practical application value of the model can be further improved by combining domain knowledge constraints, term relationship modeling and larger scientific and technological corpus.

References

- [1] Campolungo N, Martelli F, Saina F, et al. DiBiMT: A novel benchmark for measuring Word Sense Disambiguation biases in Machine Translation[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022: 4331-4352.
- [2] Campolungo N, Pasini T, Emelin D, et al. Reducing disambiguation biases in NMT by leveraging explicit word sense information[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2022: 4824-4838.
- [3] Iyer V, Barba E, Birch A, et al. Code-switching with word senses for pretraining in neural machine translation[C]//Findings of the Association for Computational Linguistics: EMNLP 2023. 2023: 12889-12901.
- [4] Mehrparvar B, Pezzelle S. Detecting and translating language ambiguity with multilingual llms[C]//Proceedings of the Fourth Workshop on Multilingual Representation Learning (MRL 2024). 2024: 310-323.
- [5] Bejgu A S, Barba E, Procopio L, et al. Word Sense Linking: Disambiguating Outside the Sandbox[C]//Findings of the Association for Computational Linguistics: ACL 2024. 2024: 14332-14347.
- [6] Ghinassi I, Tedeschi S, Marongiu P, et al. Language pivoting from parallel corpora for word sense disambiguation of historical languages: A case study on latin[C]//Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). 2024: 10073-10084.
- [7] Proietti L, Perrella S, Tedeschi S, et al. Analyzing homonymy disambiguation capabilities of pretrained language models[C]//Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). 2024: 924-938.
- [8] Sumanathilaka D K, Micallef N, Hough J. Can llms assist with ambiguity? a quantitative evaluation of various large language models on word sense disambiguation[C]//Proceedings of the First International Conference on Natural Language Processing and Artificial Intelligence for Cyber Security. 2024: 97-108.
- [9] Vu H H, Kamigaito H, Watanabe T. Context-aware machine translation with source coreference explanation[J]. Transactions of the Association for Computational Linguistics, 2024, 12: 856-874.
- [10] Małka P, Semerci Y, Scholtes J, et al. Sequence shortening for context-aware machine translation[C]//Findings of the Association for Computational Linguistics: EACL 2024.

2024: 1874-1894.

- [11] Gete H, Etchegoyhen T. Promoting Target Data in Context-aware Neural Machine Translation[C]//Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1). 2024: 9-23.
- [12] Appicharla R, Gain B, Pal S, et al. A case study on context-aware neural machine translation with multi-task learning[C]//Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1). 2024: 246-257.
- [13] Honda S, Fernandes P, Zerva C. Context-aware neural machine translation for english-japanese business scene dialogues[C]//Proceedings of Machine Translation Summit XIX, Vol. 1: Research Track. 2023: 272-285.
- [14] Castilho S, Knowles R. A survey of context in neural machine translation and its evaluation[J]. Natural Language Processing, 2025, 31(4): 986-1016.
- [15] Yue R, Ortega J, Church K. On translating technical terminology: A translation workflow for machine-translated acronyms[C]//Proceedings of the 16th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track). 2024: 48-54.
- [16] Roussis D, Sofianopoulos S, Piperidis S. Enhancing scientific discourse: Machine translation for the scientific domain[C]//Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1). 2024: 275-285.
- [17] Deilen S, Lapshinova-Koltunski E, Garrido S, et al. Evaluation of intralingual machine translation for health communication[C]//Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1). 2024: 469-479.
- [18] Deilen S, Lapshinova-Koltunski E, Garrido S H, et al. Towards AI-supported health communication in plain language: Evaluating intralingual machine translation of medical texts[C]//Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health)@ LREC-COLING 2024. 2024: 44-53.
- [19] Neves M, Yepes A J, Siu A, et al. Findings of the WMT 2022 biomedical translation shared task: Monolingual clinical case reports[C]//Proceedings of the Seventh Conference on Machine Translation (WMT). 2022: 694-723.
- [20] Neves M, Yepes A J, Névéal A, et al. Findings of the WMT 2023 biomedical translation shared task: Evaluation of ChatGPT 3.5 as a comparison system[C]//Proceedings of the Eighth Conference on Machine Translation. 2023: 43-54.
- [21] Yepes A J, Névéal A, Neves M, et al. Findings of the wmt 2017 biomedical translation shared task[C]//Proceedings of the Second Conference on Machine Translation. 2017: 234-247.
- [22] Uguet C, Bane F, Aymo M, et al. LLMs in post-translation workflows: comparing performance in post-editing and error analysis[C]//Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1). 2024: 373-386.

- [23] Macken L, De Wilde V, Tezcan A. Machine translation for open scholarly communication: Examining the relationship between translation quality and reading effort[J]. *Information*, 2024, 15(8): 427.
- [24] Hatami A, Arcan M, Buitelaar P. Enhancing translation quality by leveraging semantic diversity in multimodal machine translation[C]//*Proceedings of the 16th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*. 2024: 154-166.