# Refining Idioms Semantics Comprehension via Contrastive Learning and Cross-Attention

**Mingmin Wu**[1,2,3,4,*], **Guixin Su**[1,2,3,4,*], **Yongcheng Zhang**[1], **Zhongqiang Huang**[1,2,3,4], **Ying Sha**[1,2,3,4,†]
[1]College of Informatics, Huazhong Agricultural University, Wuhan, China
[2]Key Laboratory of Smart Farming for Agricultural Animals, Wuhan, China
[3]Hubei Engineering Technology Research Center of Agricultural Big Data, Wuhan, China
[4]Engineering Research Center of Intelligent Technology for Agriculture, Ministry of Education
{wmm_nlp, cometsue, zhyc, hzq}@webmail.hzau.edu.cn, shaying@mail.hzau.edu.cn

**Abstract**

Chinese idioms on social media demand a nuanced understanding for correct usage. The Chinese idiom cloze test poses a unique challenge for machine reading comprehension due to the figurative meanings of idioms deviating from their literal interpretations, resulting in a semantic bias in models' comprehension of idioms. Furthermore, given that the figurative meanings of many idioms are similar, their use as suboptimal options can interfere with optimal selection. Despite achieving some success in the Chinese idiom cloze test, existing methods based on deep learning still struggle to comprehensively grasp idiom semantics due to the aforementioned issues. To tackle these challenges, we introduce a **R**efining **I**dioms **S**emantics **C**omprehension **F**ramework (**RISCF**) to capture the comprehensive idioms semantics. Specifically, we propose a semantic sense contrastive learning module to enhance the representation of idiom semantics, diminishing the semantic bias between figurative and literal meanings of idioms. Meanwhile, we propose an interference-resistant cross-attention module to attenuate the interference of suboptimal options, which considers the interaction between the candidate idioms and the blank space in the context. Experimental results on the benchmark datasets demonstrate the effectiveness of our RISCF model, significantly outperforming state-of-the-art methods.

**Keywords:** Machine Reading Comprehension, Semantic Sense Contrastive Learning, Interference-Resistant Cross-Attention

## 1. Introduction

Chinese machine reading comprehension is a challenging task that requires models to have a profound understanding of natural language. The cloze task (Kobayashi, 2002) is a special form of machine reading comprehension that requires models to choose appropriate options from candidate words or sentences based on the context. Due to the simple and intuitive form of the cloze task, it has been widely used as a classic method to evaluate the language comprehension ability of models (Fotos, 1991). At present, Chinese idiom reading comprehension has not received extensive attention in the field of computational linguistics. However, the model's understanding ability of idioms has a significant impact on the performance of various natural language processing tasks, such as computer-assisted essays and machine translation (Shao et al., 2017; Liu et al., 2019).

Most Chinese idioms are derived from stories in ancient literature from Chinese history, and often convey the moral behind these tales. However, the semantics of each Chinese idiom cannot be literally understood through the composition of its characters. For example, Table 1 illustrates the literal meaning of the idiom "雨后春笋" as **'Bamboo shoots after a spring rain'**, while its figurative meaning signifies **'Many new things are rapidly emerging'**. This implies a semantic bias between literal and figurative meanings of idioms, which brings great challenges to the model of learning the semantic representation of idioms.

| **Idiom** | 雨后春笋 |
|---|---|
| **Literal Meaning** | Bamboo shoots after a spring rain. |
| **Figurative Meaning** | Many new things are rapidly emerging. |

Table 1: An example shows the semantic bias between the literal and figurative meanings in idioms.

In addition, since many idioms are near-synonyms and their figurative meanings are similar, those suboptimal options will interfere with the model to select the best option when they are simultaneously used as candidate options for a passage to fill in the blank space. Table 2 shows one such case where the candidate idioms "本末倒置", "得不偿失", "背道而驰" and "因小失大" collectively express the meaning of outcomes deviating from anticipated results in a specific action, resulting in unfavorable consequences. As a result, other idioms as the suboptimal options interfere with the selection of the best idiom answer "本末倒置".

[*]Equal contribution.
[†]Corresponding author.



*LREC-COLING 2024*, pages 13785–13795
20-25 May, 2024. 

| | | | |
|---|---|---|---|
| **Passage** | 安装监控补光灯的目的在于识别车辆信息，但这种手段也对交通安全构成了隐患，那么无疑是__。<br>The purpose of installing monitoring supplementary lights is to identify vehicle information,<br>but this approach also poses a hidden danger to traffic safety, so it is undoubtedly____. | | |
| **Candidate Options** | Option A | 本末倒置 | Reverse the priorities of things. |
| | Option B | 得不偿失 | The loss outweighs the gain. |
| | Option C | 背道而驰 | Run in the opposite direction. |
| | Option D | 因小失大 | Losing big benefits for small ones. |
| **Best Option** | Option A | 本末倒置 | Reverse the priorities of things. |

Table 2: An example of Chinese idiom cloze test. The candidate options have similar figurative meanings.

To measure the ability of understanding Chinese idioms, the Chinese idiom cloze test was proposed: given a passage and a set of idiom candidates, the model needs to capture the comprehensive idiom semantics and attenuate the interference of other idioms in the candidate idioms to select the most appropriate option. Table 2 shows an example of the Chinese idiom cloze test. Nowadays, the Chinese idiom cloze test has been explored using deep learning models, such as the Attentive Reader model (Luong et al., 2015) and the Stanford Attention Reader model (Hermann et al., 2015) . These models use BiLSTM (Hochreiter and Schmidhuber, 1997) to encode given sequences and obtain hidden states to select the most suitable idiom. Pre-trained language models based on BiLSTM have limitations in transmitting text information over long distances. The text sequences in the Chinese idiom cloze test datasets are relatively long, so these models perform poorly.

Recently, researchers proposed Chinese idiom cloze test methods based on the BERT model (Devlin et al., 2018), which performs better in Chinese reading comprehension tasks. In order to learn the actual meaning of the idiom, the CM model (Wang et al., 2020) and the SKER model (Long et al., 2020) use the BERT model to encode the interpretation sequence of the idiom, and use the representation of the interpretation sequence or the near-synonym representation of the idiom as the idiom representation. Introducing the interpretation of idioms as external knowledge can alleviate the semantic bias between the figurative and literal meanings of idioms to a certain extent. However, by directly integrating idioms and idiom interpretations, it is difficult to determine whether the model learns the representation of the interpretation sequence or the representation of the idioms themselves. Moreover, using the interpretation of near-synonyms to represent the idioms each other does not take into account that the idioms with similar figurative meanings as the suboptimal options will interfere with the selection of the best option.

In order to diminish the semantic bias between figurative and literal meanings of idioms and attenuate the interference of suboptimal options on the best selection, we introduce a Refining Idioms Semantics Comprehension Framework (RISCF) for the Chinese idiom cloze test. Specifically, we use a semantic sense contrastive learning module to select the correct idiom representation as a positive example for the blank space of the given passage, and negative examples for the other idiom representations. The semantic sense contrastive learning module diminishes the semantic bias between the figurative and literal meanings of idioms, and enables the model to capture comprehensive figurative features that can distinguish idioms in the semantic feature space. Building upon this, we utilize an interference-resistant cross-attention module that enables multiple idioms in the set to assist the model in understanding the difference between the best option and the suboptimal option. The interference-resistant cross-attention module learns the relationship between different idioms in the candidate set and the blank space of the given passage, attenuating the interference of suboptimal options on the best selection. Experimental results on the benchmark datasets demonstrate the effectiveness and robustness of the RISCF model, especially for datasets constructed with interference options that have similar meanings, which outperform state-of-the-art methods significantly.

Our main contributions are as follows:

- We introduce a novel RISCF model for the Chinese idiom cloze test, which uses contrastive learning and cross-attention mechanism to refine the semantic representation of idioms and capture the comprehensive idioms features.
- We propose a semantic sense contrastive learning module, which enables the model to enhance the representation of idiom semantics, diminishing the semantic bias between the figurative and literal meanings of idioms.
- We propose an interference-resistant cross-attention module, which considers the interaction between the candidate idioms and the blank space in the context, attenuating the interference of suboptimal options on the best selection.

## 2. Related Work

The cloze-style reading comprehension is an important form of assessing machine reading ability (Mihaylov and Frank, 2018; Wang et al., 2020; Long et al., 2020; Foolad and Kiani, 2023; Yue et al., 2023). The Chinese idiom cloze test is more challenging because there is a semantic bias between the figurative and literal meanings of idioms (Wei et al., 2022; Sha et al., 2023). Furthermore, when many near-synonymous idioms are used as candidate options, those suboptimal options will interfere with the selection of the best option due to their similar figurative meanings (Mirxanova, 2022; Rustamova, 2023). To evaluate the ability of the Chinese cloze test with idioms, Zheng et al. created a large-scale Chinese idiom cloze test dataset. Based on the ChID dataset (Zheng et al., 2019), early works proposed methods such as the Attentive Reader model (Luong et al., 2015) and the Stanford Attention Reader model (Hermann et al., 2015) based on BiLSTM (Xu et al., 2019) and attention mechanisms.

Due to the limitations of BiLSTM (Li et al., 2020) and the stronger feature extraction capabilities of the Pre-trained language model based on Transformer (Vaswani et al., 2017), most of the research currently utilizes the BERT model (Lan et al., 2019) as the backbone. Tan et al. (2021) successively proposed a BERT-based dual-embedding model and a two-stage model. The dual-embedding model matches the embedding of each candidate idiom with the representations corresponding to the blanks in the context and the hidden representations of all the tokens respectively (Tan and Jiang, 2020). The two-stage model uses the ChID dataset to re-pretrain and fine-tune the BERT model (Tan et al., 2021). However, the basic BERT model cannot correctly understand the actual meanings of idioms, ignoring the consideration of the semantic bias between the figurative and literal meanings of idioms.

Subsequent studies aimed to improve idiom representation by incorporating idiom interpretation and near-synonymous relationships as external knowledge (Wang et al., 2020; Long et al., 2020). Wang et al. introduced a CM model with an Attribute Attention mechanism to balance the importance of idiom characters and their interpretations. However, simply concatenating these representations doesn't distinguish whether the model learns the interpretation sequence or the actual idiom meanings. Long et al. proposed the SKER model, which builds a near-synonym graph based on cosine similarity between near-synonym dictionary annotations. Nevertheless, this approach overlooks the situation where near-synonymous idioms, with similar figurative meanings, could interfere when used as candidates for filling in the blanks in a passage.

In this paper, we propose a semantic sense contrastive learning module and an interference-resistant cross-attention module to simultaneously address the issues of semantic bias and interference from suboptimal options. The comprehensive idiom semantics will be refined to facilitate the model to choose the appropriate option to fill in the blank space by using contrastive learning (Gao et al., 2021; Yeh et al., 2022) and cross-attention (Chen et al., 2021; Hertz et al., 2022) without introducing external knowledge.

## 3. RISCF Model

In this section, we first introduce the problem formulation of the Chinese idiom cloze test, and then elaborate on the details of the RISCF model. The overview of the RISCF model is shown in Figure 1. Generally, we propose two innovative modules: the semantic sense contrastive learning module and the interference-resistant cross-attention module, to respectively address the aforementioned issues regarding the semantic bias between the figurative and literal meanings of idioms and the interference of suboptimal options on the best selection.

### 3.1. Problem Formulation

The Chinese idiom cloze test can be formally defined as follows: given a passage $P = \{w_1, w_2, \cdots, w_{\mathrm{b}}, \cdots, w_n\}$ with $n$ Chinese characters, the goal of the task is to select the most appropriate idiom from a set of candidate idioms which is denoted as $I = \{i_1, i_2, \cdots, i_{\mathrm{k}}\}$ to fill the blank space $w_b$ in the passage. An example illustrating the definition of the Chinese idiom cloze test is shown in Table 2. Given the passage "The purpose of installing monitoring supplementary lights is to identify vehicle information, but this approach also poses a hidden danger to traffic safety, so it is undoubtedly____." The best option to fits the blank space for this example is option A "本末倒置(Reverse the priorities of things)".

### 3.2. Input and Encoding Layer

BERT (Devlin et al., 2018) has demonstrated its effectiveness in various tasks. We utilize BERT as the passage encoder to extract the contextual hidden states of idioms. For $t$-th input passage $P_t = \{w_1, \cdots, w_{\mathrm{b}}, \cdots, w_{\mathrm{n}}\}$ with $n$ Chinese characters, the [CLS] and [SEP] tokens are first added to the start and end of the passage $P_t$ respectively. Then, the blank space $w_b$ that needs to be filled is marked as a [MASK] token to represent the idiom. Therefore, the sequence of passage input to the BERT model for encoding can be represented as

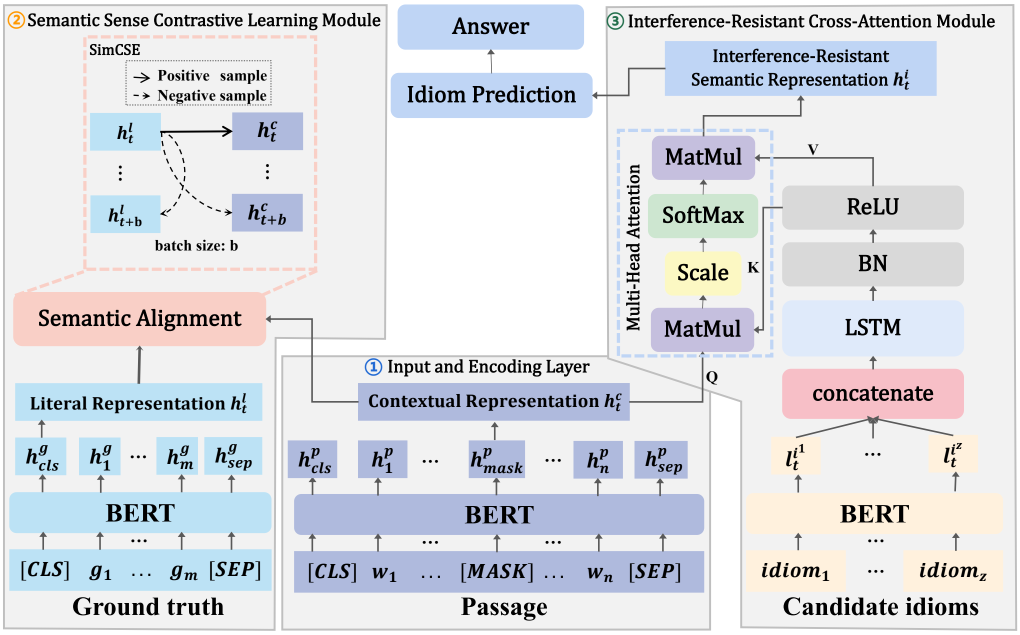



Figure 1: The architecture of our RISCF model. The serial number 1 is the input and encoding layer, the serial number 2 is the semantic sense contrastive learning module, and the serial number 3 is the interference-resistant cross-attention module.

$P_t = \{[\mathrm{CLS}], w_1, \cdots, [\mathrm{MASK}], \cdots, w_n, [\mathrm{SEP}]\}$. Taking Table 2 as an illustration, the passage can be rephrased as "安装监控补光灯的目的在于识别车辆信息，但这种手段也对交通安全构成了隐患，那么无疑是[MASK]". Next, the Chinese whole word mask BERT model (Cui et al., 2021) obtains the hidden states $H_t^P$ of the $t$-th passage:

$$H_t^P = \mathrm{BERT}\left(\{[\mathrm{CLS}]w_1, \cdots [\mathrm{MASK}] \cdots w_n[\mathrm{SEP}]\}\right) \quad (1)$$

where $H_t^P = \{h_{cls}^p, h_1^p, \cdots, h_{mask}^p, \cdots, h_n^p, h_{sep}^p\} \in \mathbb{R}^{(n+2)\times d}$, and $d$ is the hidden dimension. The hidden state $h_{mask}^p$ of the blank space marked as the [MASK] token can be directly used as the $t$-th contextual representation $h_t^c$ of the idiom to predict the idiom. However, it is not optimal because the basic contextual representation $h_t^c$ cannot comprehensively represent the semantics of idioms, ignoring the issue of semantic bias between the figurative and literal meanings of idioms.

## 3.3. Semantic Sense Contrastive Learning Module

Due to the semantic bias between the figurative and literal meanings of idioms, the semantics of each Chinese idiom cannot be literally understood through the composition of its characters. In response to this issue, we propose a semantic sense contrastive learning module to enhance the semantic representation of idioms, diminishing the semantic bias between figurative and literal meanings of idioms. Specifically, since the demonstrated effectiveness of the SimCSE model (Gao et al., 2021) in contrastive learning, we utilize the SimCSE model to perform contrastive learning by constructing the positive and negative samples for the corresponding literal and contextual representations of idioms.

For each ground truth idiom with $m$ (usually 4) characters, we set the [CLS] and [SEP] tokens to represent the beginning and end of the ground truth idiom, and utilize the BERT model to obtain the literal representation of the correct idiom in the blank space. For example, the correct answer in Table 2 can be rewritten as "$\{[\mathrm{CLS}], 本, 末, 倒, 置, [\mathrm{SEP}]\}$". Next, the ground truth idiom for the $t$-th passage is denoted as $G_t = \{[\mathrm{CLS}], g_1, \cdots, g_m, [\mathrm{SEP}]\}$ encoding by the Chinese whole word mask BERT model to obtain the hidden states $H_t^G$:

$$H_t^G = \mathrm{BERT}\left(\{[\mathrm{CLS}], g_1, \cdots, g_m, [\mathrm{SEP}]\}\right) \quad (2)$$

where $H_t^G = \{h_{cls}^g, h_1^g, \cdots h_m^g, h_{sep}^g\} \in \mathbb{R}^{(m+2)\times d}$, and $d$ is the hidden dimension. We take the hidden state $h_{cls}^g$ of the [CLS] token as the literal representation $h_t^l$ of the idiom.

Subsequently, building upon the idea of utilizing the SimCSE model for constructing positive and negative samples in contrastive learning, we adopt a strategy where the contextual representation $h_t^c$ and the literal representation $h_t^l$ serve as positive sample pairs, while the contextual and literal representations of other idioms within the same batch act as negative samples for $h_t^c$ and $h_t^l$. The semantic sense contrastive learning module aligns the figurative and literal meanings of idioms in the semantic space, diminishing the semantic bias and enhancing the semantic representation of idioms.

The training objective for ($h_t^c$, $h_t^l$) with a batch of $N$ paired examples is:

$$\mathcal{L}_{cl} = -\log \frac{e^{sim\left(h_t^c, h_t^l\right)/\tau}}{\sum_{t'=1}^{N} e^{sim\left(h_t^c, h_{t'}^l\right)/\tau}} \quad (3)$$

where $sim(h_t^c, h_t^l)$ represents the cosine similarity $\frac{(h_t^c)^T h_t^l}{\|h_t^c\| \cdot \|h_t^l\|}$ between the contextual and literal representation of the idioms, and $\tau$ is a temperature hyperparameter. The training loss $\mathcal{L}_{cl}$ of semantic sense contrastive learning will serve as an additional loss value to guide the optimization of the parameters, facilitating learning the semantic meaning of idioms.

## 3.4. Interference-Resistant Cross-Attention Module

In the aforementioned semantic sense contrastive learning is utilized to effectively diminish the semantic bias between the figurative and literal meanings of idioms. However, since many idioms are near-synonyms and their figurative meanings are similar, those suboptimal options will interfere with the model to select the best choice when they are simultaneously used as candidate options for a passage to fill in the blank space. To address this issue, we propose an interference-resistant cross-attention module to capture the semantic relationship between the representation of candidate idioms and the contextual representation of the blank space, and attenuate the interference of candidate options on the best option.

Firstly, for each idiom in the candidate idiom set $I_t$ corresponding to the $t$-th passage, we set the [CLS] and [SEP] tokens to represent the beginning and end of each candidate idiom, and then utilize the BERT model to encode each candidate idiom separately. Taking the $k$-th idiom in the candidate idiom set $I_t$ corresponding to the $t$-th passage as an example, assuming that the idiom is composed of $m$ (usually 4) characters, the idiom can be denoted as $I_t^k = \{[\mathrm{CLS}], a_1, \cdots, a_m, [\mathrm{SEP}]\}$, encoding by the Chinese whole word mask BERT to obtain the corresponding hidden state $H_t^{I^k}$:

$$H_t^{I^k} = \mathrm{BERT}\left(\{[\mathrm{CLS}], a_1, \cdots, a_{\mathrm{m}}, [\mathrm{SEP}]\}\right) \quad (4)$$

where $H_t^{I^k} = \{h_{cls}^{i^k}, h_1^{i^k}, \cdots h_{\mathrm{m}}^{i^k}, h_{sep}^{i^k}\} \in \mathbb{R}^{(m+2)\times d}$, and $d$ is the hidden dimension. We take the hidden state $h_{cls}^{i^k}$ of the [CLS] token as the literal representation $l_t^{i^k}$ of the $k$-th idiom in the candidate set corresponding to the $t$-th passage. Next, we concatenate all the literal representations of idioms in the candidate idiom set (assuming a total of $z$ idioms) and utilize an LSTM layer to obtain the interactive representation $L_t^i$ among the candidate idioms:

$$L_t^i = \mathrm{LSTM}\left(\mathrm{Concat}\left([l_t^{i^1}, \ldots, l_t^{i^k}, \ldots, l_t^{i^z}]\right)\right) \quad (5)$$

where $L_t^i \in \mathbb{R}^{z \times d_l}$, and $d_l$ is the hidden dimension of the LSTM. As LSTM is more capable of extracting key features over subwords like Chinese idioms (Ma et al., 2020; Ács et al., 2021), the interactive representation $L_t^i$ among the candidate idioms captures the crucial semantic of each idiom. To obtain a more stable semantic representation distribution and facilitate model training and convergence, we further process the interactive representation $L_t^i$ to obtain the robust idiom representation $h_t^r$:

$$h_t^r = \mathrm{ReLU}\left(\mathrm{BN}\left(L_t^i\right)\right) \quad (6)$$

where the BN function is the batch-normalization operation and ReLU is the activation function. In order to enable the model to attenuate the interference of suboptimal options on the best option, particularly when multiple near-synonymous idioms are simultaneously used as candidate options, we utilize an interference-resistant cross-attention to calculate the correlation between the contextual representation $h_t^c$ and the robust idiom representation $h_t^r$, which yields the interference-resistant semantic representation $h_t^i$ of the idioms:

$$h_t^i = \mathrm{Attention}\left(h_t^c, h_t^r, h_t^r\right) = \mathrm{Softmax}\left(\frac{h_t^c (h_t^r)^T}{\sqrt{d_k}}\right) h_t^r \quad (7)$$

where $\sqrt{d_k}$ is the size of the first dimension of the input of query and key. Specifically, $\sqrt{d_k}$ is the feature dims on each head of the multi-head. The interference-resistant semantic representation of idioms $h_t^i$ combines the semantic information of the context and the crucial interactive information among idioms, and attenuates the interference of near-synonymous information from the suboptimal options.

## 3.5. Idiom Prediction and Loss Function

Finally, we feed the interference-resistant semantic representation $h_t^i$ into a linear layer, followed by a softmax function to produce the probability distribution of candidate idioms given the passage $P_t$, which scores each candidate idiom $I_t^c (c \in \{1, \cdots, z\})$ in the candidate idiom set $I_t$ and conduct idiom prediction. The process mentioned above can be formulated as:

$$\Pr\left(I_t^c \mid P_t\right) = \mathrm{softmax}\left(w \cdot h_t^i + b\right) \quad (8)$$

where $w$ and $b$ are the learnable weight and bias.

The training objective of the idiom prediction is to minimize the cross-entropy loss $\mathcal{L}_{pr}$ between the ground truth and predictions:

$$\mathcal{L}_{pr} = -\sum_{c=1}^{z} \mathrm{y}_g \log \Pr\left(I_t^c \mid P_t\right) \tag{9}$$

where $z$ represents the number of idioms in the candidate set, and $y_g$ represents the one-hot label distribution of the ground truth. The overall training objective is to minimize the sum of the loss $\mathcal{L}_{pr}$ of idiom prediction and the loss $\mathcal{L}_{cl}$ of semantic sense contrastive learning. The total loss function $\mathcal{L}$ can be formulated as:

$$\mathcal{L} = \mathcal{L}_{pr} + \mathcal{L}_{cl} \tag{10}$$

Idiom prediction and contrastive learning are trained simultaneously, and we minimize the total training loss $\mathcal{L}$ to fine-tune all parameters.

# 4. Experimental results

## 4.1. Data and Experimental Setup

Next, we will introduce several datasets we use, evaluation metrics and experimental settings.

### 4.1.1. Datasets

**ChID.** (Zheng et al., 2019) ChID is the first large-scale and general Chinese idiom cloze test dataset, covering various data types from the internet, including news, novels, and prose. Among them, news and novels belong to the in-domain data, while prose falls into the out-of-domain data. The in-domain data consists of a training set, a validation set (Dev), and a test set (Test), while the out-of-domain data includes an additional test set called Out. Furthermore, there is an extra test set named Sim. Sim and Test have identical passage content, but the construction of candidate idiom sets differs. In Sim, the candidate idioms share similarities with the correct answers. The presence of similar idioms in the candidate set results in more scattered model attention and increases interference choices, making Sim more challenging than Test in terms of comprehension difficulty. The statistics of the ChID dataset are shown in Table 3.

**ChID-Competition.**[1] ChID-Competition is a dataset for the online Chinese idiom comprehension competition. It is a modified version of ChID, with a key difference being that each data entry in ChID-Competition contains a list of paragraphs with multiple blanks, all sharing the same set of candidate idioms. Each idiom can only be used once. In ChID-Competition, the true answers and distractor options are semantically similar, and the model needs to distinguish their differences to make the correct selection. ChID-Competition is divided into four subsets: Train, Dev, Test, and Out. The specific statistics of the ChID-Competition dataset are shown in Table 3.

**CCT.** (Jiang et al., 2018) While the ChID dataset is commonly used for idiom reading comprehension, the CCT dataset is a more challenging idiom cloze dataset with more idioms. The CCT dataset collects a total of 108,987 sentences, 7,395 different idioms, and the definitions of these idioms. The training and test sets contain 108,432 and 555 sentences, and 7,071 and 508 idioms, respectively.

### 4.1.2. Evaluation Metrics

The evaluation metric used is accuracy, which calculates the percentage of correct answers predicted by the model in the total test data set.

### 4.1.3. Experimental Settings

We set the maximum input sequence length to 128 and utilized an NVIDIA GeForce RTX 3080Ti GPU for model execution. The training was performed in the PyTorch 1.4.0 (Paszke et al., 2019) and Transformers 3.1.0 (Wolf et al., 2020) environment, with a total of 5 epochs. The initial learning rate was 5e-5, and we adopted the warm-up linear schedule strategy with 1000 warm-up steps. The optimization was carried out using the AdamW optimizer (Loshchilov and Hutter, 2018).

## 4.2. Results and Discussion

### 4.2.1. Methods to Compare

**Language Model (LM)** (Zhou et al., 2016) employs a BiLSTM model to encode the input sequence and compare its hidden state with the representation of each candidate idiom for idiom selection.

**Attentive Reader (AR)** (Luong et al., 2015) enhances the BiLSTM model using an attention mechanism to focus on important information during the encoding process.

**Standard Attentive Reader (SAR)** (Hermann et al., 2015) is an improved version of AR that utilizes a bilinear function as a matching function to calculate attention weights.

**BERT-WWM** (Cui et al., 2021) is an upgraded version of the BERT model that utilizes whole word masking, masking the entire word rather than just individual tokens.

**Synonym Knowledge Enhanced Reader (SKER)** (Long et al., 2020) constructs a near-synonym graph using cosine similarity of idiom embeddings and encodes the graph using a graph attention network to replace the original idiom representation.

**BERT-based two-stage model (BTSM)** (Tan et al., 2021) is pre-trained on a large Chinese corpus and fine-tuned for idiom prediction.

**Correcting the Misuse (CM)** (Wang et al., 2020)

[1] https://github.com/chujiezheng/ChID-Dataset

| Dataset | ChID | | | | ChID-competition | | | |
|---|---|---|---|---|---|---|---|---|
| | In-domain | | | Out-domain | In-domain | | | Out-domain |
| | Train | Dev | Test/Sim | Out | Train | Dev | Test/Sim | Out |
| **Passages** | 520,711 | 20,000 | 20,000 | 20,096 | 84,709 | 3218 | 3231 | 3754 |
| **Total blanks** | 648,920 | 24,822 | 24,948 | 30,023 | 577,157 | 23,011 | 23,209 | 27,704 |
| **Distinct idioms** | 3848 | 3458 | 3502 | 3626 | 3848 | 3414 | 3434 | 3599 |

Table 3: Statistics of the ChID and the ChID-Competition datasets.

| Model | Dev | Test | Sim | Out |
|---|---|---|---|---|
| **Human** | – | 87.1 | 82.2 | 86.2 |
| **LM** | 71.8 | 71.5 | 65.6 | 61.5 |
| **AR** | 72.7 | 72.4 | 66.2 | 62.9 |
| **SAR** | 71.7 | 71.5 | 64.9 | 61.7 |
| **BERT-wwm** | 75.4 | 75.7 | 70.2 | 66.1 |
| **SKER** | 76.0 | 76.3 | 68.8 | 68.3 |
| **BTSM** | 81.9 | 81.8 | 74.1 | 72.0 |
| **CM** | 83.0 | 83.1 | 76.1 | 77.6 |
| **BERT-IDM** | – | 83.2 | – | 67.5 |
| **RMFNet** | 86.6 | 86.8 | 80.9 | 84.8 |
| **GPT-3.5** | – | 38.6 | 31.7 | 28.4 |
| **RISCF-CL** | 79.2 | 79.4 | 72.9 | 71.7 |
| **RISCF** | **93.8** | **93.7** | **97.1** | **90.3** |

Table 4: Experimental results of accuracy on the dataset ChID.

| ChID-Competition | | | | CCT | |
|---|---|---|---|---|---|
| **Model** | **Dev** | **Test** | **Out** | **Model** | **Test** |
| **AR** | 65.4 | 65.6 | 55.6 | **BiLSTM** | 89.5 |
| **BERT** | 72.7 | 72.4 | 64.7 | **BTSM** | 93.7 |
| **BTSM** | 92.4 | 92.0 | 90.2 | **RMFNet** | 93.9 |
| **GPT-3.5** | – | 22.9 | 23.1 | **GPT-3.5** | 72.3 |
| **RISCF** | **94.5** | **94.3** | **92.2** | **RISCF** | **95.6** |

Table 5: Experiment results of accuracy on ChID-Competition and CCT.

introduces idiom interpretations and employs an attribute attention mechanism to balance the weights of different attributes among different representations of idioms.

**BERT-IDM** (Dai et al., 2023) utilizes a two-stage semantic expansion method that leverages semantic knowledge during the pre-training stage and extracts idiom interpretation information during the fine-tuning stage.

**Retrospective Multi-granularity Fusion Network (RMFNet)** (Yue et al., 2023) equipped with the multi-granularity passage fusion module to enhance the passage representation and the retrospective reading to concentrate on critical Chinese idioms.

**GPT-3.5-Turbo (GPT-3.5)** (OpenAI, 2023) is a large language model developed by OpenAI, which demonstrates the superior performance in various NLP tasks with its powerful ability in language generation and comprehension.

**RISCF-CL** only uses the semantic sense contrastive learning module in our RISCF model.

#### 4.2.2. Results and Analysis

As shown in Table 4 and Table 5, we summarize the results of the comparison experiments with the baseline model on each benchmark dataset. The results of all baseline models on different datasets are obtained from the original papers. Overall, our RISCF model significantly outperforms the baseline models on various benchmark datasets, even surpassing the performance of humans (directly derived from ChID). Furthermore, the RISCF model exhibits powerful performance on the Sim dataset, which includes numerous near-synonymous idiom options.

To summarize, BERT-based models outperform BiLSTM-based models on the Chinese idiom cloze test due to BERT's superior feature extraction capabilities. Previous BERT-based models incorporate external knowledge like idiom interpretations and near-synonym sets, aiding idiom representation and understanding to some extent. However, these methods fall short on the Sim dataset, which contains more near-synonymous idioms. This is because when near-synonymous idioms are used as options, the models struggle to select the best choice due to the similarity in figurative meanings. We also evaluated the GPT-3.5 model on the Chinese idiom cloze test datasets. Surprisingly, as shown in Table 4 and Table 5, GPT-3.5 faced challenges in effectively comprehending the semantics of Chinese idioms. Our GPT-3.5 prompt template for the idiom cloze task is presented in Table 6.

In Table 4 and Table 5, we present the experimental results for our proposed models: RISCF-CL and RISCF. The RISCF-CL model, which doesn't rely on external knowledge like idiom interpretations, shows the inferior performance compared to methods using such knowledge. However, it excels in handling newly emerging Chinese idioms without standardized definitions frequently found on the Internet. The RISCF-CL model leverages contrastive learning to align the figurative and literal meanings of idioms, reducing semantic bias and enhancing idiom representation. This is crucial in addressing the core challenge of Chinese idiom cloze tests, where semantic bias between

| Role | Prompt Template |
| --- | --- |
| system | 你需要完成中文成语完形填空测试。<br>You need to complete the Chinese idiom cloze. |
| user | '{passage}'里的#idiom#可以换成哪个成语？候选成语：'{options}'，请你给出正确选择:<br>Which idiom could replace'#idiom#' in '{passage}'? Candidate idioms: '{options}'. please give the correct option: |

Table 6: The prompt template for GPT-3.5. {passage} is the context passage of the idiom, {options} is the candidate set of the idioms.

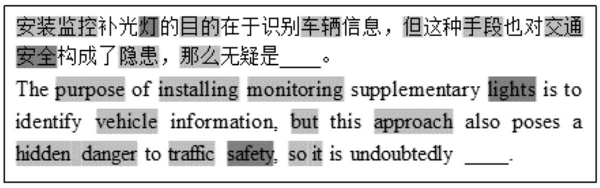
figurative and literal meanings is prominent. Only by mitigating this bias can the model reduce the interference of suboptimal options. Building upon the RISCF-CL model, the RISCF model incorporates an interference-resistant cross-attention module, which significantly improves performance. This module helps the model better distinguish between the best and suboptimal options within the context, thus minimizing the interference of suboptimal choices on the final selection.

## 4.3. Case Study

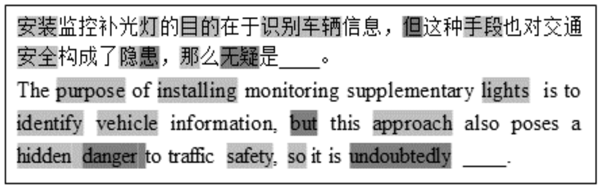
In order to intuitively illustrate the efficiency of the semantic sense contrastive learning module and interference-resistant cross-attention module in our proposed RISCF model, we conduct the attention interaction between the representation of the entire passage and the semantic representations of idioms (which were respectively learned from the BERT-wwm model, the RISCF-CL model and the RISCF model for the final idiom prediction). And then we extract the attention weights and visualize them to observe the semantic information learned from different models in the context how it supports predicting the correct idiom.

We take the example from Table 2 for visualization, and the results are shown in Figure 2. The word with darker color represents the closer semantic connection between the word in the passage and the predicted idiom. Figure 2(a), Figure 2(b) and Figure 2(c) respectively represent the visualization results of semantic information supported to predict the idiom, which respectively learned from BERT-wwm model, RISCF-CL model and RISCF model in the context.

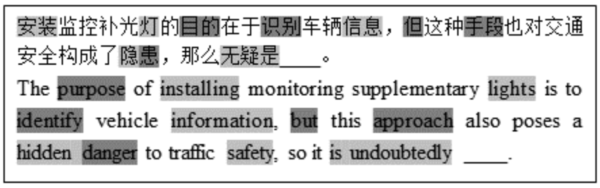
From Figure 2(a), we observe that the BERT-wwm model exhibits scattered semantic connections, capturing some irrelevant information that introduces bias in idiom comprehension. The RISCF-CL model, which builds on BERT-wwm with the semantic sense contrastive learning module, produces more focused attention. Figure 2(b) shows that this model pays specific attention to words like




(a) The visualization results of the BERT-wwm model.




(b) The visualization results of the RISCF-CL model.




(c) The visualization results of the RISCF model.

Figure 2: Visualization of the attention interaction between the representation of the predicted idiom learned from different models and the representation of the entire passage.

"但 (but)" and "隐患 (danger)," which align with the key semantic aspects expressed by the four candidate idioms in Table 2. This implies that the RISCF-CL model captures information closely related to idiom semantics in context, mitigating semantic bias between literal and figurative meanings. However, as the figurative meanings of the four candidate idioms are similar, the model struggles to capture semantic nuances that would help distinguish the best option "本末倒置(Reverse the priorities of things)."

To address this limitation, we introduce the RISCF model, an extension of RISCF-CL, which includes the interference-resistant cross-attention module. As shown in Figure 2(c), the RISCF model not only emphasizes the connection between idioms and words like "但 (but)" and "隐患 (danger)" but also recognizes the reversal of the primary and secondary relationship between "目的 (purpose)" and "手段 (approach)" linked by "但 (but)." This aligns with the meaning of "Reverse the priorities of things" expressed by the best option "本末倒置," while the suboptimal options lack this relationship. Thus, the interference-resistant cross-attention captures the semantic link between candidate idioms and the context, refining idiom semantics and reducing interference from suboptimal options.

## 5. Conclusion

In this paper, we present the Refining Idioms Semantics Comprehension Framework (RISCF) for Chinese idiom cloze tests. To narrow the semantic bias between the figurative and literal meanings of idioms, we first introduce a semantic sense contrastive learning module to align these meanings in the semantic space. Additionally, an interference-resistant cross-attention module captures semantic relationships between candidate idioms and context, refining idiom semantics and reducing interference from suboptimal options. Experimental results on benchmark datasets demonstrate the effectiveness and robustness of our RISCF model compared to strong baselines. Our case study underscores the importance of these core modules. Future work will explore model transferability for cloze tasks involving slang in different languages.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62272188), the Fundamental Research Funds for the Central Universities 2662021JC008, the Inner Mongolia Autonomous Region Major Science and Technology Project 2021ZD0046. Thank the experimental teaching center of College of informatics in Huazhong Agricultural University for providing the experimental environment and computing resources.

## Ethical Statement

We affirm that our research on Chinese idiom cloze test was conducted with the highest ethical standards. The study involved human participants and their personal information was handled with strict confidentiality. Participants provided their informed consent before participating in the study, and were informed of their right to withdraw from the study at any time. The research was approved by the Institutional Review Board (IRB) and was conducted in accordance with the guidelines provided by the IRB.

We also affirm that the data used in our research was obtained legally and ethically, and that all necessary permissions were obtained prior to collecting and using the data (Zheng et al., 2019; Jiang et al., 2018). We acknowledge that our research has limitations and potential biases, and we have made efforts to address these limitations and biases to the best of our ability.

Furthermore, we declare that there are no competing interests that may have influenced the outcome of our research or its interpretation. We have fully disclosed the funding sources of our research, and declare that our research was not influenced by any commercial or financial interests.

Finally, we acknowledge the contributions of all individuals who participated in this research, including the participants and the research team members. We express our gratitude to them for their willingness to contribute to our research, and for their trust in our ability to handle their personal information with care and confidentiality.

## 6. Bibliographical References

Judit Ács, Ákos Kádár, and Andras Kornai. 2021. Subword pooling makes a difference. *arXiv preprint arXiv:2102.10864*.

Chun-Fu Richard Chen, Quanfu Fan, and Rameswar Panda. 2021. Crossvit: Cross-attention multi-scale vision transformer for image classification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 357–366.

Danqi Chen, Jason Bolton, and Christopher D Manning. 2016. A thorough examination of the cnn/daily mail reading comprehension task. *arXiv preprint arXiv:1606.02858*.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3504–3514.

Yu Dai, Yuqiao Liu, Lei Yang, and Yufan Fu. 2023. An idiom reading comprehension model based on multi-granularity reasoning and paraphrase expansion. *Applied Sciences*, 13(9):5777.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Shima Foolad and Kourosh Kiani. 2023. Lukegraph: A transformer-based approach with gated relational graph attention for cloze-style reading comprehension. *arXiv preprint arXiv:2303.06675*.

Sandra S Fotos. 1991. The cloze test as an integrative measure of efl proficiency: A substitute for essays on college entrance examinations? *Language learning*, 41(3):313–336.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. Simcse: Simple contrastive learning of sentence embeddings. *arXiv preprint arXiv:2104.08821*.

Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. *Advances in neural information processing systems*, 28.

Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. 2022. Prompt-to-prompt image editing with cross attention control. *arXiv preprint arXiv:2208.01626*.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Zhiying Jiang, Boliang Zhang, Lifu Huang, and Heng Ji. 2018. Chengyu cloze test. In *BEA@ NAACL-HLT*, pages 154–158.

Miyoko Kobayashi. 2002. Cloze tests revisited: Exploring item characteristics with special attention to scoring methods. *The Modern Language Journal*, 86(4):571–586.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Peng-Hsuan Li, Tsu-Jui Fu, and Wei-Yun Ma. 2020. Why attention? analyze bilstm deficiency and its remedies in the case of ner. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8236–8244.

Yuanchao Liu, Bo Pang, and Bingquan Liu. 2019. Neural-based chinese idiom recommendation for enhancing elegance in essay writing. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5522–5526.

Siyu Long, Ran Wang, Kun Tao, Jiali Zeng, and Xin-Yu Dai. 2020. Synonym knowledge enhanced reader for chinese idiom reading comprehension. *arXiv preprint arXiv:2011.04499*.

Ilya Loshchilov and Frank Hutter. 2018. Fixing weight decay regularization in adam.(2018). In *URL https://openreview. net/forum*.

Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*.

Qianli Ma, Zhenxi Lin, Jiangyue Yan, Zipeng Chen, and Liuhong Yu. 2020. Mode-lstm: a parameter-efficient recurrent network with multi-scale for sentence classification. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6705–6715.

Todor Mihaylov and Anette Frank. 2018. Knowledgeable reader: Enhancing cloze-style reading comprehension with external commonsense knowledge. *arXiv preprint arXiv:1805.07858*.

GR Mirxanova. 2022. Stages of enhancement of synonym dictionaries. *International Journal of World Languages*, 2(1).

OpenAI. 2023. Gpt-4 technical report.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Davlatkhon Rustamova. 2023. The importance of a cognitive approach to learning synonyms in primary grades. *BARQARORLIK VA YETAKCHI TADQIQOTLAR ONLAYN ILMIY JURNALI*, 3(3):32–36.

Ying Sha, Mingmin Wu, Zhi Zeng, Xing Ge, Zhongqiang Huang, and Huan Wang. 2023. A prompt-based representation individual enhancement method for chinese idiom reading comprehension. In *International Conference on Database Systems for Advanced Applications*, pages 682–698. Springer.

Yutong Shao, Rico Sennrich, Bonnie Webber, and Federico Fancellu. 2017. Evaluating machine translation performance on chinese idioms with a blacklist method. *arXiv preprint arXiv:1711.07646*.

Minghuan Tan and Jing Jiang. 2020. A bert-based dual embedding model for chinese idiom prediction. *arXiv preprint arXiv:2011.02378*.

Minghuan Tan, Jing Jiang, and Bing Tian Dai. 2021. A bert-based two-stage model for chinese chengyu recommendation. *Transactions on Asian and Low-Resource Language Information Processing*, 20(6):1–18.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Xinyu Wang, Hongsheng Zhao, Tan Yang, and Hongbo Wang. 2020. Correcting the misuse: A method for the chinese idiom cloze test. In *Proceedings of Deep Learning Inside Out (DeeLIO):*

*The First Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*, pages 1–10.

Zhang Wei, Wang Hao, Chen Yuetong, Fan Tao, and Deng Sanhong. 2022. Identifying metaphors and association of chinese idioms with transfer learning and text augmentation. *Data Analysis and Knowledge Discovery*, 6(2/3):167–183.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pages 38–45.

Guixian Xu, Yueting Meng, Xiaoyu Qiu, Ziheng Yu, and Xu Wu. 2019. Sentiment analysis of comment texts based on bilstm. *Ieee Access*, 7:51522–51532.

Chun-Hsiao Yeh, Cheng-Yao Hong, Yen-Chi Hsu, Tyng-Luh Liu, Yubei Chen, and Yann LeCun. 2022. Decoupled contrastive learning. In *European Conference on Computer Vision*, pages 668–684. Springer.

Jianyu Yue, Yiwen Sun, Xiaojun Bi, Zheng Chen, and Yu Zhang. 2023. Retrospective multi-granularity fusion network for chinese idiom cloze-style reading comprehension. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

Chujie Zheng, Minlie Huang, and Aixin Sun. 2019. Chid: A large-scale chinese idiom dataset for cloze test. *arXiv preprint arXiv:1906.01265*.

Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. 2016. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers)*, pages 207–212.