Enhancing Bilingual Lexicon Induction via Bi-directional Translation Pair Retrieving

Qiuyu Ding, Hailong Cao*, **Tiejun Zhao**

Harbin Institute of Technology

qiuyuding@stu.hit.edu.cn, caohailong@hit.edu.cn, tjzhao@hit.edu.cn

Abstract

Most Bilingual Lexicon Induction (BLI) methods retrieve word translation pairs by finding the closest target word for a given source word based on cross-lingual word embeddings (WEs). However, we find that solely retrieving translation from the source-to-target perspective leads to some false positive translation pairs, which significantly harm the precision of BLI. To address this problem, we propose a novel and effective method to improve translation pair retrieval in cross-lingual WEs. Specifically, we consider both source-side and target-side perspectives throughout the retrieval process to alleviate false positive word pairings that emanate from a single perspective. On a benchmark dataset of BLI, our proposed method achieves competitive performance compared to existing state-of-the-art (SOTA) methods. It demonstrates effectiveness and robustness across six experimental languages, including similar language pairs and distant language pairs, under both supervised and unsupervised settings.

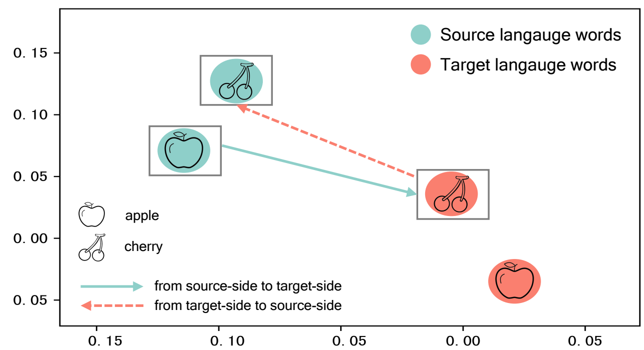


Figure 1: The illustration of asymmetry in retrieving translations, traversing from the source-side language to the target-side language and vice versa in cross-lingual WEs. The blue points are words in the source language, and the pink points are words in the target language.

Introduction

Bilingual Lexicon Induction (BLI) or word translation aims to find word translation pairs across large monolingual corpora of two languages, and has wide applications in cross-lingual NLP tasks, such as cross-lingual text classification (Klementiev, Titov, and Bhattarai 2012; Mogadala and Rettinger 2016), unsupervised machine translation (Artetxe, Labaka, and Agirre 2018b; Yang et al. 2018; Sun et al. 2019; Duan et al. 2020), and cross-lingual named entity recognition (Mayhew, Tsai, and Roth 2017; Xie et al. 2018).

Most BLI models map monolingual word embeddings (WEs) into a shared WE space (Lample et al. 2018; Artetxe, Labaka, and Agirre 2018a; Ruder, Vulić, and Søgaard 2019; Li et al. 2022). When searching for the translation of a source word, the closest target word is retrieved in the cross-lingual WE space and regarded as a translation result.

However, we find an interesting phenomenon where a source word has its nearest neighbor in the target language, but the nearest neighbor of that target word may not be the original source word. As illustrated in Figure 1, the cross-lingual nearest neighbor of the source word "*apple*" is "*cherry*", but the cross-lingual nearest neighbor of the target word "*cherry*" is the source word "*cherry*". This indicates that the translation pairs obtained from the two perspectives exhibit inequality. Clearly, the translation pair {*apple, cherry*} obtained from the source-side perspective, which is commonly used, is not a reliable translation pair. One of the reasons for this false translation issue is that the BLI system's retrieval predominantly focuses on a single perspective: source-to-target, which we refer to as the "source-side perspective", raising the problem of "**source-side bias**" (The formal definition can be found in Sec. Methodology).

Although the previous study (Lample et al. 2018) notes the importance of retrieval of translation pairs, they did not resolve the problem of false positive word pairs arising from source-side bias. In this paper, we investigate the impact of false positive translation pairs caused by source-side bias on BLI's accuracy. Then we propose a novel method of word translation retrieval that incorporates a bi-directional perspective to alleviate this problem. Specifically, when retrieving word translation pairs for a given source word, we first retrieve its cross-lingual nearest neighbor set from the source-side perspective. Then, for each word in this set, a reverse nearest neighbor set retrieval is conducted from the target-side perspective. The final translation result is collaboratively determined by the set of translations obtained from

*Corresponding author

both perspectives.

To evaluate the effectiveness of our method, we perform a comprehensive set of BLI experiments on the standard BLI benchmark, the experimental language pairs including six language pairs, three similar language pairs, and three distant language pairs, in several setups. Besides, experiments were conducted in two scenarios: supervised BLI and unsupervised BLI. Our method gains a significant improvement on almost all language pairs than several state-of-the-art (SOTA) BLI systems.

Overall, this paper makes the following contributions:

- We find an interesting "source-side bias" problem in BLI, which lead to false positive translation pairs and significantly harms the precision of BLI.
- We propose a novel method that effectively addresses the "source-side bias" problem. Our method introduces a novel retrieval strategy that integrates multiple perspectives, generating higher-quality translation pairs.
- Extensive experiments demonstrate that our method significantly improves the precision of both supervised and unsupervised BLI on six experimental language pairs, and analytical experiments further verify the robustness of our proposed method.

Background

In this section, we provide a brief introduction to the background and fundamentals, including the BLI task and similarity evaluation measure.

Preliminaries and Task Formulation

In BLI, we assume there are two languages L_1 and L_2, and the monolingual WEs of them are X and Y. The vocabularies of them are V_X and V_Y, and the dim d of X and Y usually is 300. BLI aims to learn a mapping matrix W which can map X and Y into a shared WE space, and ensure the maximum source words are close to their translations in language L_2. The mapped WE space is XW_x and YW_y respectively, and they are in the same cross-lingual WE space Z.

In the process of learning the mapping matrix, a seed lexicon D_0 is commonly employed to provide anchor information, aiding the initialization of W. D_0 consists of ground-truth translation pairs, consisting of source words and their corresponding target words. During the self-learning of the BLI model, W can be optimized, and this self-learning process involves using certain fictitious translation pairs with high confidence, treating them as new anchor points to ensure their proximity. After the mapping, the cross-lingual word embedding space Z is obtained. Subsequently, for a given source language word w_s, we search for its translation in Z, by computing the similarity between all the words in the target language vocabulary and selecting the one with the highest similarity as the translation result.

Similarity Evaluation

In the cross-lingual WE space, translation pairs are retrieved using word similarity measures. Two approaches are employed commonly: Cosine and Cross-Domain Similarity Local Scaling (CSLS) (Lample et al. 2018).

Cosine. For a source-side word w_s and a target-side word w_t, the word vectors of them are x and y, and the cosine similarity is calculated as follows:

$$cos(x, y) = \frac{x \cdot y}{\|x\| \|y\|} \tag{1}$$

A higher cosine value indicates that the vector representations of two words have highly similar semantics.

CSLS.

Previous research (Lample et al. 2018) has identified the presence of a hubness problem in cross-lingual WE spaces, which means a single word becomes the nearest neighbor of multiple other words. In such cases, relying solely on cosine similarity to measure word similarity can fall into local optimization. The CSLS method (Lample et al. 2018) is proposed to alleviate the hubness problem, it incorporates a local scaling factor to optimize the comparison of word vectors. The CSLS similarity is calculated as follows:

$$CSLS(x, y) = 2cos(x, y) - \tau(x) - \tau(y) \tag{2}$$

where $\tau(x)$ and $\tau(y)$ are the average cosine similarity of $\tau(x)$ (or $\tau(y)$) to its k nearest neighbors, respectively, and the value of k is often set to 10.

Methodology

In this section, we present a novel approach to enhance BLI by improving the retrieval of word translation pairs. We first introduce the "source-side bias" phenomenon that we have identified in the word translation retrieval process of BLI. Inspired by this observation, we propose an effective framework to alleviate the influence of the "source-side bias", thereby improving the overall performance of BLI.

Source-side Bias

When translating a word from a source language to a target language, the typical approach involves searching for the target word that has the highest word similarity within the cross-lingual WEs. The current retrieval process for translation follows the perspective of $source \rightarrow target$, in this way, the inverse of the nearest neighbor identified from the source-side perspective might not always hold true for a source word (Artetxe, Labaka, and Agirre 2018b) in a high-dimensional space. This does not satisfy the view in linguistics: word translation is bidirectional. In particular, when a source word w_s is translated as word w_t in the target language, and conversely, w_t can be translated back to w_s in the vast majority of cases, as Eq. (3), where Trans(a,b) means the translation of a is b:

$$\text{Trans}(w_s, w_t) \Leftrightarrow \text{Trans}(w_t, w_s) \tag{3}$$

While this symmetry is not satisfied in the current word translation retrieval method, which only supports retrieval from $source \rightarrow target$, and conversely, when retrieval is performed from $target \rightarrow source$, the translation result of w_t may not be this source-side word w_s. This observation gives rise to a noteworthy phenomenon, and we call it "***source-side bias***".

As shown in Figure 2, for a given word w, when we retrieve its translation in the target language, it falls into the

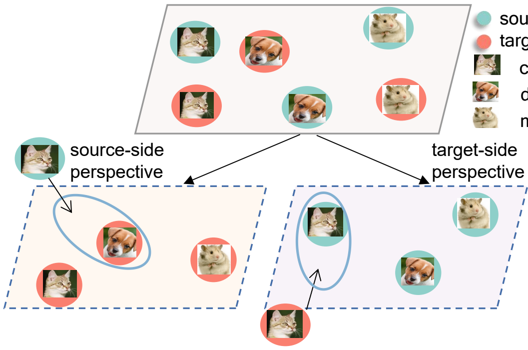


Figure 2: An illustration of word translation retrieval from source-side perspective and target-side perspective in cross-lingual WEs. The translation pairs circled by blue curves are retrieved from the two perspectives, separately.

source-side bias. Due to w belonging to the source language, its perspective is source-side which follows its language, and only includes target-side words $T_w = \{w_t^1, w_t^2, \dots\}$, then the word with the highest similarity is selected from target-side words as translation. However, existing methods do not consider whether the translation word selected from the target-side perspective w_t is the nearest neighbor of the source word.

It is evident that source-side bias can lead to asymmetric translation pairs. For a word w_s, its translation found in the target-side, denoted as w_t, may not necessarily have w_s as its translation when searching for it on the source-side. As shown in Figure 2, for a word *cat*, when searching from a different language, its retrieval results are different. It is challenging to determine which perspective provides the most accurate translation without any supervision. In cases of ambiguity, where $T(w_s) = w_t$ but $S(w_t) \neq w_s$, it will be doubted whether the found translation is the most correct one. We refer to this phenomenon as ***bidirectional confusion*** (See Figure 2).

This phenomenon it is widely in multiple languages (Artetxe, Labaka, and Agirre 2018b; Radovanovic, Nanopoulos, and Ivanovic 2010). And it is also verified by our subsequent experiments.

Translation Retrieving with Target-side Fusion

Based on the previously mentioned issues of source-side bias and bidirectional confusion, we propose the following methods to address these problems and enhance the translation process in BLI. We denote the translation retrieval of words from source language to target language as the **source-side perspective**, and conversely, the translation direction from target to source is referred to as the **target-side perspective**.

To retrieve the translation of a word, we initiate the process from its source-side perspective. We identify several target-side words with high similarity, then for each target-side word, we explore its nearest neighbors in the source language. The final translation depends on both source-side results and target-side results.

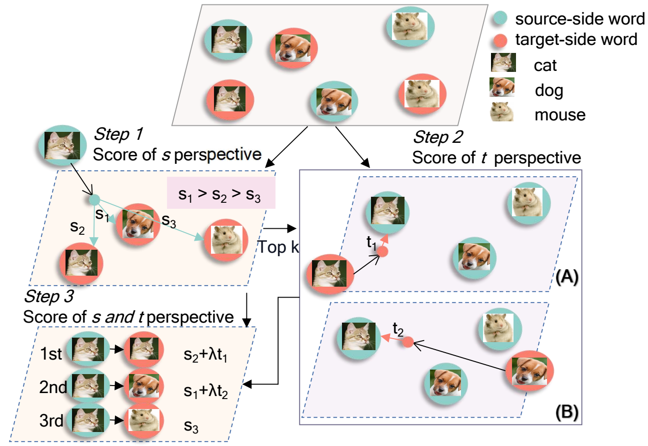


Figure 3: An illustration of proposed bi-directional perspective word translation retrieval method. It combines translation of both source-side perspective and target-side perspective. First, given a source word, its multiple target translations are retrieved from the source-side perspective. Then corresponding source translations are retrieved for these target words. The candidate translations of the two perspectives are combined for the final translation result.

Specifically, as depicted in Figure 3, we aim to integrate the target-side perspective into word translation retrieval to alleviate the source-side bias. The initial step involves word translation retrieval using the word similarity evaluation method. For a given source word w_s^i, we curate a set, denoted as $S_{\text{per}}(w_s^i) = \{w_{t_1}, w_{t_2}, \dots, w_{t_n}\}$, comprising the top n target words with the highest word similarity to it (S_{per} represents the source-side perspective). Then, for each word in $S_{\text{per}}(w_s^i)$, we retrieve their translations in the source-side language, and similarly, for each target word, we retrieve their top m translations, where T_{per} represents the target-side perspective:

$$T_{per}(w_t^i) = \{w_{s_1}, w_{s_2}, \dots, w_{t_m}\} \tag{4}$$

Subsequently, using the retrieved translations from the target-side perspective $T_{\text{per}}(w_t^i)$, we assign a target-side score to each word w_t^i. For this purpose, two methods are proposed to calculate the target-side score: Back Retrieving of Top Rank (BRTR) and Back Retrieving of Score Rank (BRSR). Where BRSR is an advanced version of the BRTR.

Back Retrieving of Top Rank (BRTR). Our proposed BRTR method aims to re-rank the translation candidate words by fusing the rank of the target-side perspective. We calculate the target-side score based on the rank of the source word w_s^i among the target words. If the source-side word w_s^i appears in $T_{\text{per}}(w_t^i)$, we simply obtain its rank as $R(w_s^i)$ and assign the target-side score as $1/R(w_s^i)$. The calculation formula is as Eq. (5), where $Score_T$ represent the score calculated in target-side perspective, and source-side perspective score $Score_S$ is calculated in the same.

$$Score_T = \frac{1}{R(w_s^i)} \tag{5}$$

Back Retrieving of Similar Rank (BRSR). In order to fine-tune the ranking of translation candidate words more precisely, we propose another method called BRSR that incorporates word similarity into the ranking process. BRSR not only calculates the target-side score based on the rank of the source words but also incorporates the similarities. For a given target word w_t^i, if it is present in the set $T(w_s^i)$ and ranks within the top n translations, we compute the similarity between w_t^i and w_s^i as the target-side score. As shown in Eq. (6) (S_{per} is the same) :

$$Score_T = \text{Sim}(w_t^i, w_s^i) \tag{6}$$

The Final Score. The final score of word translation retrieval is composed of both source-side score and target-side score. In order to control the weight of the impact of the two scores, we introduced a weighting factor to control the target-side score. The final score S for a source word w_s^i and one of its candidate target-side words w_t^i is computed as follows:

$$S = (1-\lambda)Score_S(w_s^i, w_t^i) + \lambda Score_T(w_t^i, w_s^i) \tag{7}$$

The similarity $S(w_s^i, w_t^i)$ between source word w_s^i and target word w_t^i can be decomposed into a weighted sum of the source language similarity $S_{\text{per}}(w_s^i, w_t^i)$ and the target language similarity $T_{\text{per}}(w_s^i, w_t^i)$, where λ is a tunable hyperparameter. We aim to find the target word w_t^i that maximizes S with respect to w_s^i, and regard it as the translation result. And the selection of λ is influenced by the closeness of similarity between the two languages, and a comprehensive experiment is presented in Sec. Effect of Hyper-parameters.

Experimental Setting

In this section, we present the setup of our experiment on the task of BLI. We test our method on BLI task on several similar and distant language pairs, include both sides of English-France (EN-FR), English-Italian (EN-IT), English-Spanish (EN-ES), English-Chinese (EN-ZH), English-Japanese (EN-JA), English-Thai (EN-TH).

Datasets

We follow the standard BLI setup from prior work (Lample et al. 2018; Artetxe, Labaka, and Agirre 2018b; Joulin et al. 2018; Mohiuddin, Bari, and Joty 2020; Peng, Lin, and Stevenson 2021; Li et al. 2022). We use the widely used MUSE dataset (Lample et al. 2018), which consists of 300-dim embeddings pre-trained with FastText (Bojanowski et al. 2017) which is trained on the monolingual corpora of full Wikipedias for each language, and the vocabularies are trimmed to the 200k most frequent words. We also employ the test sets released by (Lample et al. 2018) that are widely used in BLI evaluations. These test sets contain 1500 ground-truth translation pairs for each language, and they serve as standard benchmarking test sets for BLI models. Again following prior work, 5k translation pairs are used as seed lexicon D_0 (Lample et al. 2018).

Evaluation Metrics

In this paper, we employ two widely used similarity evaluation metrics: Cosine and the standard CSLS (Lample et al. 2018). Besides, P@1, 5, and 10 (Precision @1, 5, and 10) are employed to measure the precision of BLI models.

Training Setup and Hyperparameters

We select best hyperparameters by searching a combination of λ, n, m with the following range: λ: $\{0.05, 0.1, \ldots, 1.0\}$ with 0.05 step size; n, m: $\{3, 4, \ldots, 20\}$ with 1 step size. All experiments are performed on a single Nvidia RTX A6000.

Baseline Models

- **MUSE** (Lample et al. 2018) is a classical BLI system based on adversarial training to map monolingual WEs.
- **VECMAP** (Artetxe, Labaka, and Agirre 2018b) is a robust BLI system based on self-learning, widely adopted in many BLI systems.
- **RCSLS** (Joulin et al. 2018) introduces a novel relaxed CSLS loss and learns a non-orthogonal mapping between monolingual WEs.
- **LNMAP** (Mohiuddin, Bari, and Joty 2020) is a non-linear map-based BLI system that utilizes non-linear autoencoders to map WEs.
- **L1-refinement** (Peng, Lin, and Stevenson 2021) is a BLI system that adds a post-processing step to improve the cross-lingual WEs, which is based on the Manhattan norm (aka. ℓ_1 norm)
- **CLBLI** (Li et al. 2022) is one of the most SOTA BLI systems which applies contrastive learning into BLI, and the positive examples and hard negative examples contribute for optimize the mapping matrices.

For all baseline methods, we confirm the parameters based on their suggested optimal settings. Moreover, we ensure that all BLI systems utilize the static WEs offered by (Lample et al. 2018). When dealing with systems that have the potential to leverage dynamic word representations from various language models (e.g., CLBLI), we do not integrate these dynamic word representations. Because the advantages of static WEs and dynamic word representations bring in enhancing BLI are orthogonal (Zhang et al. 2021).

Results and Discussion

In this section, we empirically demonstrate the effectiveness of our proposed method. First, the main findings are reported. Then, we analyze the effect of the hyperparameters used in our method. Thirdly, CSLS and our proposed method are compared in particular. Last, we demonstrate the results obtained using other evaluation methods.

Main Results

We first compare the overall performance of our proposed method (BRTR and BRSR) with the baseline models, as shown in Table 1. On average, BRSR (BRTR) significantly outperforms 1.04 (0.62) P@1 baseline models.

For similar language pairs (EN-ES, EN-FR, EN-IT), despite the baseline systems already achieving commendable

Model	EN-FR		EN-ES		EN-IT		EN-ZH		EN-JA		EN-TH		AVG
	→	←	→	←	→	←	→	←	→	←	→	←	
MUSE	82.00	81.87	81.67	83.33	67.45	60.22	25.00	16.80	0.00	0.00	20.87	14.99	44.52
MUSE-BRTR	81.93	**82.00**	81.87	83.47	67.52	60.10	32.60	**26.27**	**1.92**	0.07	22.67	15.26	46.31
MUSE-BRSR	**82.20**	81.87	**82.20**	**83.80**	**67.90**	**60.65**	**33.60**	25.73	**1.92**	**0.76**	**22.84**	**15.60**	**46.59**
VECMAP	81.60	84.13	81.27	85.47	67.90	62.75	43.33	42.20	52.84	41.49	25.13	18.97	57.26
VECMAP-BRTR	81.60	84.47	81.60	85.87	67.90	62.75	43.07	42.67	53.27	**42.25**	25.33	19.67	57.54
VECMAP-BRSR	**81.80**	**84.93**	**81.93**	**86.20**	**68.65**	**62.93**	**45.13**	**42.67**	**53.53**	42.04	**25.60**	**19.92**	**57.94**
RCSLS	83.00	84.93	84.20	86.53	67.49	60.90	47.67	47.40	38.52	38.67	22.40	15.40	56.43
RCSLSE-BRTR	83.27	**85.13**	84.40	**86.67**	66.84	61.70	47.40	47.67	39.63	39.77	23.80	17.69	57.00
RCSLS-BRSR	**83.40**	85.07	**84.73**	**86.67**	**67.30**	**62.20**	**48.60**	**48.20**	**40.99**	**40.20**	**24.00**	**17.89**	**57.44**
LNMAP	80.07	80.87	80.13	79.93	65.25	59.42	33.87	40.07	39.96	42.87	19.07	26.87	54.03
LNMAP-BRTR	80.27	80.87	80.40	80.20	65.70	59.70	34.00	40.40	40.27	**43.25**	19.40	27.00	54.29
LNMAP-BRSR	**80.60**	**80.93**	**80.53**	**80.70**	**66.48**	**60.20**	**34.20**	**40.67**	**41.67**	42.87	**19.53**	**27.97**	**54.70**
L1	81.67	84.20	81.27	85.40	67.98	62.87	43.13	42.13	52.84	41.56	25.07	19.04	57.26
L1-BRTR	81.87	84.40	81.40	85.67	68.20	62.99	44.20	**42.60**	52.93	42.25	25.33	19.18	57.59
L1-BRSR	**81.93**	**84.67**	**82.07**	**86.27**	**68.40**	**63.06**	**45.20**	**42.60**	**53.26**	**42.80**	**25.80**	**20.40**	**58.04**
CLBLI	82.54	83.47	84.40	85.63	67.02	63.06	51.37	48.40	49.17	41.80	34.59	21.54	59.42
CLBLI-BRTR	82.64	**84.27**	84.67	85.63	67.52	**64.00**	51.37	48.67	50.77	42.67	34.59	**22.53**	59.94
CLBLI-BRSR	**83.31**	84.00	**84.73**	**86.20**	**67.90**	63.79	**51.87**	**49.60**	**53.02**	**43.20**	**35.25**	22.25	**60.43**

Table 1: The P@1 accuracy of BLI systems with $|D_0| = 5k$ on similar language pairs and distance language pairs. Bold numbers indicate the best scores. The results are statistically significant (paired t-test, p-value < 0.01).

Model	EN-FR		EN-ES		EN-IT		EN-ZH		EN-JA		EN-TH		AVG
	→	←	→	←	→	←	→	←	→	←	→	←	
MUSE	82.73	81.53	82.07	83.33	67.15	60.34	30.67	31.00	0.00	4.89	0.00	0.00	43.64
MUSE-BRSR	**83.13**	**82.33**	**82.80**	**83.80**	**68.96**	**60.84**	**32.40**	**31.27**	**0.14**	**6.80**	**0.07**	0.00	**44.38**
VECMAP	82.33	83.20	82.20	85.13	69.34	62.56	0.00	0.00	48.18	0.28	0.00	0.00	42.77
VECMAP-BRSR	**83.00**	**84.80**	**82.80**	**85.60**	**70.40**	**63.00**	**1.52**	**2.20**	**49.00**	**2.40**	0.00	0.00	**43.73**

Table 2: The accuracy of unsupervised BLI systems on similar language pairs and distance language pairs. The word similarity measure is CSLS. The results are statistically significant (paired t-test, p-value < 0.01).

results, our method still brings enhancements. For instance, in the EN-ES direction based on MUSE system, the accuracy is improved from 81.6 P@1 to 82.20 P@1. Even against the most SOTA baseline system CLBLI, our method still yields significant improvements (from 82.54 P@1 to 83.31 P@1 on the EN-FR language pair). For distant language pairs (EN-ZH, EN-JA, EN-TH), a more significant improvement bring by BRSR and BRTR. On the EN-ZH language pair, our method achieves an 8.6 P@1 improvement over MUSE, and a 2.07 P@1 improvement over L1. These findings substantiate the efficacy and robustness of our proposed methods, which attain convergence to a favorable solution across similar and distant language pairs.

Table 2 shows the unsupervised BLI results. Clearly, our method not only enhances the performance of supervised BLI but also delivers a substantial improvement in the realm of unsupervised BLI.

Effect of Hyper-parameters

To assess the impact of the weight parameter λ on the target-side score, we conduct an analysis on several language pairs using the VECMAP framework. We explore a range of λ values from 0.05 to 1.0, with a step size of 0.05. The corresponding plots are shown in Figure 4, illustrating the direct influence of the target-side score on word translation accuracy. First, we observe that the trend remains consistent across all language pairs. As the value of λ increases, the accuracy gradually decreases. Conversely, the best performance is achieved for most language pairs when λ is chosen from the interval $\{0.05, 0.1, 0.15, 0.2\}$. These findings highlight the significance of the target-side score incorporation and emphasize the importance of selecting an appropriate value for λ to optimize the word translation accuracy.

Although the consistent trends in P@1 accuracy across different language pairs with varying λ values, we observe that the value of λ depends on the quality of cross-lingual WEs and the similarity between the language pairs. For closely related language pairs and well-performing cross-lingual WEs, smaller λ values tend to yield the best P@1 results. On the other hand, for distant language pairs and lower-quality cross-lingual WEs, appropriately increasing the value of λ leads to the optimal P@1 performance. The quality of cross-lingual WEs can be reflected in the accuracy of word translation.

Retrieval Method	EN-ES $\rightarrow$	EN-ES $\leftarrow$	EN-ZH $\rightarrow$	EN-ZH $\leftarrow$
MUSE k=15				
CSLS	81.47	83.30	25.13	16.47
CSLS-BRTR	**81.73**	**83.63**	**33.27**	**24.73**
CLBLI k=15				
CSLS	84.27	85.37	51.13	48.20
CSLS-BRTR	**84.63**	**86.20**	**51.79**	**49.52**
MUSE k=20				
CSLS	81.67	83.47	25.13	16.60
CSLS-BRTR	**81.87**	**83.8**	**33.60**	**25.73**
CLBLI k=20				
CSLS	84.37	85.47	51.37	48.20
CSLS-BRTR	**84.73**	**86.20**	**51.87**	**49.60**

Table 3: Comparison of performance with restricted k value of our proposed method BRTR and CSLS.

Comparison of Local Scaling Between BRTR and CSLS

Since the proposed method also involves the scale selection of candidate items (hyperparameter m and n), similar to the parameter k in CSLS, it is essential to differentiate the improvements brought by our method from the scaling of the k value in CSLS. So we impose restrictions on the k value used in the baseline system, ensuring that it had the same value as our proposed method during the retrieval of translation pairs. The experimental results are summarized in Table 3, which highlights the impact of this constraint on the performance of the baseline system. It can be seen that the k value has only a very weak effect on CSLS. This demonstrates the effectiveness of our proposed method, as the enhancements are not due to the manipulation of the k value in CSLS but are instead a result of the innovative aspects introduced by our approach.

Other Evaluations

To validate the effectiveness of our proposed approach, we use additional evaluation measures to perform a comprehensive comparative analysis. Besides CSLS, we introduce the cosine similarity measure with P@1, P@5, and P@10 as evaluation metrics to facilitate a more thorough comparison.

Cosine Similarity. As for cosine is one of the most classical methods to calculate word similarities, and is popular used in BLI for word translation retrieving, we use it to assess the performance of our proposed method. As depicted in Table 4, our method outperforms all baseline models in terms of precision.

Precision@5 and Precision@10. Additionally, to examine the impact of our method on the entire set of candidate items, we calculated the Precision@5 and Precision@10 scores using both cosine and CSLS similarity measures.

Model	EN-ES $\rightarrow$	EN-ES $\leftarrow$	EN-JA $\rightarrow$	EN-JA $\leftarrow$	AVG
MUSE	79.00	80.00	0.00	0.07	39.77
MUSE-BRSR	**79.80**	**80.20**	**2.40**	**0.41**	**40.70**
VECMAP	79.47	82.53	48.05	34.32	61.09
VECMAP-BRSR	**79.60**	**83.00**	**49.76**	**35.70**	**62.02**
RCSLS	81.20	84.33	21.86	35.98	55.84
RCSLS-BRSR	**81.87**	**85.02**	**22.57**	**36.96**	**56.61**
LNMAP	78.67	78.20	36.20	34.32	56.85
LNMAP-BRSR	**79.60**	**79.20**	**37.47**	**35.47**	**57.94**
L1	79.67	82.53	48.39	34.32	61.23
L1-BRSR	**80.07**	**83.47**	**49.75**	**35.97**	**62.32**
CLBLI	82.85	84.60	34.31	34.87	59.16
CLBLI-BRSR	**83.39**	**85.20**	**39.34**	**36.52**	**61.11**

Table 4: The accuracy of BLI systems using Cosine word similarity measure.

The results are presented in Table 5, demonstrating that our method achieves substantial improvements on both similarity measures (Cosine and CSLS) across all baseline systems.

Related Work

BLI is an important task of machine learning, involves two main components: WEs mapping and word translation retrieval. So we provide a condensed summary of the most relevant research.

BLI methods. The BLI methods can be divided into three main types, linear mapping-based BLI methods, non-linear mapping-based BLI methods, and statistical-based BLI methods. For linear mapping-based BLI methods, it is based on linear mapping to first map two monolingual WEs into a shared WE space, then the mapping matrix is fine-tuned by self-learning.

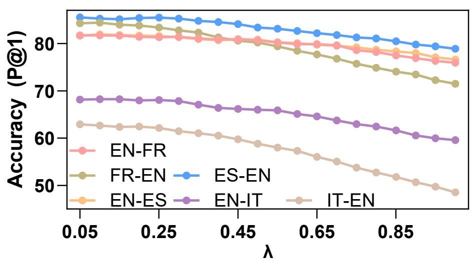
For the **linear mapping-based** BLI methods, (Artetxe, Labaka, and Agirre 2018b) employs adversarial training, (Artetxe, Labaka, and Agirre 2018a) leverages word distributions in different languages for similarity, (Joulin et al. 2018) utilizes a relaxed CSLS (Artetxe, Labaka, and Agirre 2018b) loss, BLISS (Patra et al. 2019) relaxes the isometric assumption, (Jawanpuria et al. 2018) uses classification on smooth Riemannian manifolds, and (Li et al. 2022) optimizes the mapping matrix through contrastive learning.

For the **non-linear mapping-based** BLI methods, recent notable contributions include (Glavaš and Vulić 2020) employs instance-based approach that incorporates non-linear maps, (Mohiuddin, Bari, and Joty 2020) utilizes non-linear auto-encoders to learn a non-linear mapping in the derived space, (Peng, Lin, and Stevenson 2021) introduces a post-processing step to enhance cross-lingual WEs.

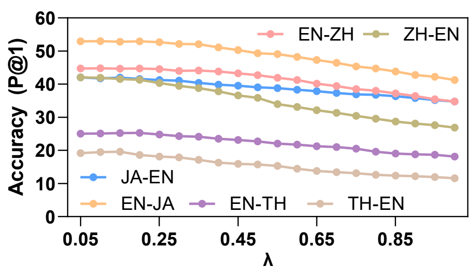
For **statistic-based** BLI methods, it automatically obtains the correspondence between vocabulary in two languages through statistical analysis of large-scale corpora. (Haghighi et al. 2008) involves using a generative model based on typical correlation analysis to extract translations of terms. (Vulić, De Smet, and Moens 2011) proposes a bilingual la-

Model	EN-IT (Cosine)			EN-IT (CSLS)			EN-ZH (Cosine)			EN-ZH (CSLS)		
	P@1	P@5	P@10	P@1	P@5	P@10	P@1	P@5	P@10	P@1	P@5	P@10
MUSE	59.88	75.70	80.02	67.45	81.76	85.39	22.60	43.33	51.87	25.00	45.60	52.87
MUSE-BRSR	**60.78**	**77.06**	**80.62**	**67.90**	**82.20**	**85.77**	**29.00**	**52.40**	**60.47**	**33.60**	**55.47**	**63.33**
VECMAP	63.74	79.18	82.44	67.90	83.42	86.30	39.60	60.73	66.60	43.33	64.53	70.33
VECMAP-BRSR	**63.82**	**79.41**	**82.67**	**68.65**	**83.77**	**86.87**	**41.40**	**61.67**	**66.87**	**45.13**	**65.13**	**70.80**
RCSLS	60.26	78.73	83.72	67.49	82.82	86.45	44.13	65.33	69.73	47.67	67.80	72.67
RCSLS-BRSR	**62.82**	**79.49**	**84.25**	**67.30**	**83.50**	**86.83**	**44.80**	**65.73**	**71.73**	**48.60**	**68.33**	**73.00**
LNMAP	60.20	76.20	81.45	65.25	79.71	83.50	31.80	48.20	53.73	33.87	51.60	57.00
LNMAP-BRSR	**61.40**	**77.67**	**82.27**	**66.48**	**81.45**	**84.00**	**32.60**	**49.27**	**54.20**	**34.20**	**52.00**	**57.67**
L1	63.96	79.56	82.43	67.98	83.42	86.30	39.47	60.87	66.60	43.13	64.47	70.33
L1-BRSR	**64.57**	**80.73**	**83.27**	**68.40**	**83.60**	**86.83**	**41.60**	**61.80**	**67.73**	**45.20**	**64.87**	**70.67**
CLBLI	63.89	80.70	84.56	67.02	82.89	86.75	44.73	67.17	73.84	51.37	73.67	79.62
CLBLI-BRSR	**64.47**	**81.80**	**84.86**	**67.90**	**83.42**	**87.42**	**47.26**	**68.13**	**75.17**	**51.87**	**74.00**	**79.90**

Table 5: Precision@1, 5, 10 of BLI systems.



(a) Similar language pairs.



(b) Distance language pairs.

Figure 4: Effect of hyper-parameter λ on proposed BRTR, on several similar language pairs and distant language pairs.

tent Dirichlet Assignment model for finding translations of terms in comparable corpora without using any linguistic resources. (E and Zhou 2022) suggests using a Markovian semantic model to characterize the meaning of words.

Word translations' retrieval. After obtaining cross-lingual WEs, another important step is the retrieval of word translations, which usually involves computing the similarities between word vectors. Currently, there are three commonly used methods for calculating word vector similarity: Cosine Similarity, CSLS, and Euclidean Distance. **Cosine Similarity** (Johnson, Douze, and Jégou 2019) is a widely used method for measuring the similarity between two vectors. For two given word vectors, their cosine similarity can be computed by taking the dot product of the vectors and dividing it by the product of their norms. **CSLS** (Cross-Domain Similarity Local Scaling) (Lample et al. 2018) incorporates a local scaling factor to optimize the comparison of word vectors. It enhances the quality of similarity measurements by considering the neighborhood of each word.

The BLI methods above mainly focus on the mapping of monolingual WEs. Different from them, we notice the importance of translation pair retrieval in BLI and leverage the bi-direction perspective into translation retrieval to address the source-side bias and achieve better accuracy.

Conclusion and Future Work

In this study, we propose a novel and effective approach for word translation retrieval to enhance BLI. Our method addresses the issue of source-side bias in the retrieval of translation pairs by incorporating the target-side perspective. Through extensive experiments on a benchmark dataset, we demonstrate significant improvements in BLI performance across six language pairs, including both similar and distant language pairs. Theoretical and experimental analyses confirm the effectiveness of our method in mitigating the source-side bias. Notably, our approach not only enhances the precision of P@1 but also improves the overall word translation accuracy. Besides, our method does not require additional data or prior knowledge and can be applied to any language pair. Overall, our work presents a simple yet powerful solution for word translation retrieval in BLI, addressing the limitations of source-side bias and offering promising results across various language pairs.

The framework can still be improved in several aspects. First, considering the great power that Large Language Models have shown on NLP tasks, the monolingual WEs and cross-lingual WEs can be explored for improvement by incorporating word representations from Large Language Models. Besides, there is room to improve the training of BLI models by introducing anchor information from Large Language Models.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 62376076 and the Key R&D Program of Yunnan under Grant 202203AA080004. The authors would like to thank the anonymous reviewers for their helpful comments.

References

Artetxe, M.; Labaka, G.; and Agirre, E. 2018a. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 789–798.

Artetxe, M.; Labaka, G.; and Agirre, E. 2018b. Unsupervised Statistical Machine Translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 3632–3642. Brussels, Belgium.

Bojanowski, P.; Grave, E.; Joulin, A.; and Mikolov, T. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5: 135–146.

Duan, X.; Ji, B.; Jia, H.; Tan, M.; Zhang, M.; Chen, B.; Luo, W.; and Zhang, Y. 2020. Bilingual Dictionary Based Neural Machine Translation without Using Parallel Sentences. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 1570–1579.

E, W.; and Zhou, Y. 2022. A Mathematical Model for Universal Semantics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3): 1124–1132.

Glavaš, G.; and Vulić, I. 2020. Non-Linear Instance-Based Cross-Lingual Mapping for Non-Isomorphic Embedding Spaces. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7548–7555. Online: Association for Computational Linguistics.

Haghighi, A.; Liang, P.; Berg-Kirkpatrick, T.; and Klein, D. 2008. Learning bilingual lexicons from monolingual corpora. In *Proceedings of Annual Meeting of the Association for Computational Linguistics*, 771–779.

Jawanpuria, P.; Balgovind, A.; Kunchukuttan, A.; and Mishra, B. 2018. Learning Multilingual Word Embeddings in Latent Metric Space: A Geometric Approach. arXiv:1808.08773.

Johnson, J.; Douze, M.; and Jégou, H. 2019. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3): 535–547.

Joulin, A.; Bojanowski, P.; Mikolov, T.; Jégou, H.; and Grave, É. 2018. Loss in Translation: Learning Bilingual Word Mapping with a Retrieval Criterion. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2979–2984.

Klementiev, A.; Titov, I.; and Bhattarai, B. 2012. Inducing crosslingual distributed representations of words. In *Proceedings of International Conference on Computational Linguistics*, 1459–1474.

Lample, G.; Conneau, A.; Ranzato, M.; Denoyer, L.; and Jégou, H. 2018. Word translation without parallel data. In *International Conference on Learning Representations*.

Li, Y.; Liu, F.; Collier, N.; Korhonen, A.; and Vulić, I. 2022. Improving Word Translation via Two-Stage Contrastive Learning. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 4353–4374.

Mayhew, S.; Tsai, C.-T.; and Roth, D. 2017. Cheap translation for cross-lingual named entity recognition. In *Proceedings of the conference on empirical methods in natural language processing*, 2536–2545.

Mogadala, A.; and Rettinger, A. 2016. Bilingual word embeddings from parallel and non-parallel corpora for cross-language text classification. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 692–702.

Mohiuddin, M. T.; Bari, M. S.; and Joty, S. 2020. LNMap: Departures from Isomorphic Assumption in Bilingual Lexicon Induction Through Non-Linear Mapping in Latent Space. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2712–2723.

Patra, B.; Moniz, J. R. A.; Garg, S.; Gormley, M. R.; and Neubig, G. 2019. Bilingual Lexicon Induction with Semi-supervision in Non-Isometric Embedding Spaces. In *Annual Meeting of the Association for Computational Linguistics*.

Peng, X.; Lin, C.; and Stevenson, M. 2021. Cross-Lingual Word Embedding Refinement by ℓ_1 Norm Optimisation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2690–2701. Online: Association for Computational Linguistics.

Radovanovic, M.; Nanopoulos, A.; and Ivanovic, M. 2010. Hubs in space: Popular nearest neighbors in high-dimensional data. *Journal of Machine Learning Research*, 11(sept): 2487–2531.

Ruder, S.; Vulić, I.; and Søgaard, A. 2019. A survey of cross-lingual word embedding models. *Journal of Artificial Intelligence Research*, 65: 569–631.

Sun, H.; Wang, R.; Chen, K.; Utiyama, M.; Sumita, E.; and Zhao, T. 2019. Unsupervised bilingual word embedding agreement for unsupervised neural machine translation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 1235–1245.

Vulić, I.; De Smet, W.; and Moens, M. F. 2011. Identifying word translations from comparable corpora using latent topic models. In *Proceedings of the annual meeting of the Association for Computational Linguistics: Human language technologies*, 479–484.

Xie, J.; Yang, Z.; Neubig, G.; Smith, N. A.; and Carbonell, J. 2018. Neural cross-lingual named entity recognition with minimal resources. *arXiv preprint arXiv:1808.09861*.

Yang, Z.; Chen, W.; Wang, F.; and Xu, B. 2018. Unsupervised Neural Machine Translation with Weight Sharing. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 46–55.

Zhang, J.; Ji, B.; Xiao, N.; Duan, X.; Zhang, M.; Shi, Y.; and Luo, W. 2021. Combining Static Word Embeddings and Contextual Representations for Bilingual Lexicon Induction. In *Findings of the Association for Computational Linguistics*, 2943–2955.