Improving Named Entity Recognition using Bilingual Constraints and Word Alignment

An T Dao\textsuperscript{1}, Thinh H Truong\textsuperscript{1}, Long Nguyen, Dien Dinh

Faculty of Information Technology, University of Science, Hochiminh city, Vietnam

\textsuperscript{1} These authors contributed equally to this work.

E-mail: dtan@apcs.vn, ththinh@apcs.vn, long hb nguyen@gmail.com, ddien@fit.hcmus.edu.vn

Abstract. Named entities carry essential meanings and information in natural language. Therefore, Named Entity Recognition has many applications in different Natural Language Processing tasks such as Information Retrieval, Information Extraction, Machine Translation, and Question Answering. State-of-the-art Named Entity Recognition systems are based on supervised machine learning algorithms which require huge amounts of training data. The main problem, however, is that constructing named entity annotated corpora is an expensive, labor-intensive, and time-consuming task. Therefore, in this paper, we propose an approach to improve monolingual Named Entity Recognition systems by exploiting an existing unannotated English-Chinese bilingual corpus. The system jointly recognizes named entities in both English and Chinese sentences through the use of bilingual constraints. Experimental results show an improvement in Named Entity Recognition of both Chinese and English compared to the strong baseline StanfordNER. In particular, Chinese Named Entity Recognition improves significantly by 20.81% in term of F\textsubscript{1}-score. As for the English language, Named Entity Recognition F\textsubscript{1}-score increases slightly from 75.75% to 76.08%. When comparing to the state-of-the-art system in improving Named Entity Recognition based on bilingual resources, we manage to outperform in Chinese Named Entity Recognition task by 5.99% and achieve comparable results for the English side. Our proposed method can also be generalized to apply to resource-limited languages.

1. Introduction

With the development of the Internet, the amount of data is growing exponentially [1]. Consequently, the demand for managing and interpreting data has become urgent [2]. In order to interpret information effectively, names of objects like persons, organizations, locations, and numeric expressions need to be recognized. The task of identifying references of these entities is called Named Entity Recognition (NER) which plays an important role in Information Extraction [3]. Named entities are also fundamental to other tasks such as Machine Translation [4], Question Answering [5], Information Retrieval [6], Co-reference Resolution [7], and so on.

Most state-of-the-art NER systems are based on supervised machine learning algorithms which typically require large amounts of high-quality annotated training data in order to produce accurate results [8, 9]. Because of the performance limit, most NER systems make these common types of errors [10] (example in Figure 1):

- **NE Boundary Errors:** Only some parts of an NE are recognized. For example, the entity “提姆·库克(Tim Cook)” in Chinese sentence is only recognized as “提姆(Tim)” having type PERSON (PER).
• NE Type Errors: NEs are tagged with incorrect types. For example, “Apple Inc.” is tagged with label PERSON (PER) although the correct label is ORGANIZATION (ORG)

• Unrecognized NE Errors: NEs are not recognized at all. For instance, the word “美国 (United State)” in the Chinese sentence is not recognized as an entity.

Nevertheless, annotating NE manually would take a lot of time and human resources. This problem restricts the availability of annotated corpus, especially for resource-limited language.

Instead of annotating data manually, another way to improve NER is to make use of information from a resourceful language like English. Different languages contain complementary clues which help in disambiguating each other. For example, in Figure 1 the word “库克 (Cook)” in the Chinese sentence is not recognized as an entity. However, it is aligned with the entity [Cook, PER] in English, providing information that it might also be an entity of type PERSON.

Thanks to the use of cross-lingual corpora from Machine Translation task which has been extensively studied and improved, previous works showed improvements of NER performance. In 2012, Li et al. [11] proposed a novel model based on linear chain Conditional Random Field which projects features between English and Chinese through word alignment. The information is transferred on feature-level instead of label-level. The model combined both monolingual and bilingual features and performed decoding on two languages simultaneously to help improve the tagging process. However, this model relied on a bilingual corpus in which NEs are manually annotated. A method that formulates the problem of exploring complementary cues about entities on an unannotated parallel corpus between English and Chinese was described by Che et al. [12]. This inspiring idea was based on encoding bilingual features into objective functions which are solved by Integer Linear Programming technique [13]. Nonetheless, this work required initial Named Entity Recognition systems trained on corpora with BIO annotation scheme which are not widely available. According to these two works, using additional information from different languages could help correct the errors made by monolingual taggers. In particular, the high performance of English NER can help improve Chinese NER through word alignment.

![Figure 1](image_url)  
**Figure 1.** Example of NE errors in a parallel English-Chinese sentence pair.

To fix the mentioned Named Entity Recognition errors, our initial intuition is to make use of additional information in a bilingual corpus. These observations motivate us to make the following contributions:

• We explore a projecting method that takes advantage of an unannotated English-Chinese bilingual corpus aligned at sentence level to boost the performance of monolingual NER systems in this language pair (Section 2.3). This simple but effective method does not require any further annotations such as named entities annotation or gazetteers.

• We introduce an additional bilingual feature which penalize words tagged with none-type (Section 2.7) to improve the proposed method.

The experimental results (Section 3) agree with our intuition, in which the proposed method could improve the strong English and Chinese NER baselines. The related work is discussed in Section 4. Finally, Section 5 is the conclusion and future work.
2. Method

In this section, we describe in details the proposed system where NEs are jointly tagged in both English and Chinese sentences. Our main idea is to combine bilingual information obtained from both monolingual taggers and word alignment results to project NE tags. In this work, we focus on recognizing the four most common types of NE, which are geopolitical entities (GPE), location names (LOC), organization names (ORG), and person names (PER).

2.1. Overview

This section describes the general idea of our proposed system. Following the work of Che et al. (2013) [12], we re-use the Stanford Word Segmenter with LDC standard [14] to segment Chinese text into words. Given a pair of parallel sentences in English and Chinese, we first use StanfordNER toolkit [15] to recognize NEs in these sentences. Then, we project the NE tags from English to Chinese and vice-versa. In order to project NEs, for each word in the target sentence, we get the Constraint-based Monolingual NER scores \( P_{\text{target}} \) (Section 2.2). These scores show how the target sentence contributes in defining the tags of its own words, which describes its “self-context” information. This is when most monolingual Named Entity Recognition systems get their results.

Nonetheless, to take advantages of bilingual text, we also get the “cross-lingual context” of each word in the target sentence from the source sentence. In other words, we want to use additional information from the aligned words in the source sentence (\( P_{\text{source}} \)) to help annotate the corresponding target words with the correct labels. As a result, by combining these source and target context information, we calculate Bilingual Type Reassignment Score \( P_{\text{bi}} \) (Section 2.4) for each target word. However, not all words in the source sentence should contribute to finding the correct label of the target word. The main problem here is that we do not know which word in the source sentence should affect recognizing the type of the target word. Therefore, we need to find the alignment set \( A \) which contains suitable source words contributing to the probability of assigning type for the corresponding target word. This set \( A \) is constructed using two separate strategies: TopK (Section 2.3.1) and Threshold (Section 2.3.2). Finally, based on this Bilingual Type Reassignment Score, we assign the appropriate type for each word (Section 2.5).

From the example in Figure 1, suppose that we want to find the appropriate tag for the word “美
国(United States)” in the Chinese sentence. The whole process of our system goes through these steps which are illustrated in Figure 2:

- First, we calculate the Constraint-based Chinese Monolingual NER score \( P_{\text{mono}}(T_R \mid \text{中国(United States)}) \) for \( T_R \in \{\text{GPE, LOC, ORG, PER, O} \} \) respectively.
- Next, we find the alignment set \( A \) of that word “美国”. Assume that \( A = \{\text{the, United, States}\} \).
- We calculate the scores of the alignment set \( A \), which denote the contribution of source sentence in defining the tag of target word “美国”. In detail, the probability that word
美 has tag $T_R P(T_R \mid 美国(United States))$ is calculated based on $P(T_R \mid the)$, $P(T_R \mid United)$ and $P(T_R \mid States)$ in the English sentence.

- Finally, we calculate the Bilingual Type Reassignment Score $P_{bi}(T_R \mid 美国(United States))$ and tag it with type $T_R - LOC$ which has the highest score.

The core of our projection method is Bilingual Type Reassignment Score, which describes the probability that a word $w_T$ has type $T_R$ in the target sentence is introduced in equation (1).

$$P_{bi}(T_R \mid w_T) = \lambda P_{target}(T_R \mid w_T) + (1 - \lambda) P_{source}(T_R \mid A)$$  (1)

where the interpolation parameter $\lambda$ (mono_weight), which indicates the weight of the target word monolingual NER score, will be tuned on the development set. This weight shows the contribution of “self-context” in defining the tag of $w_T$. $P_{target}(T_R \mid w_T)$ is the Constraint-based Monolingual NER probability that a target word $w_T$ has type $T_R$. $P_{source}(T_R \mid A)$ is the probability that a target word $w_T$ has type $T_R$ given the alignment set $A$.

### 2.2. Constraint-based Monolingual NER

The Constraint-based Monolingual NER probability is obtained from StanfordNER toolkit. StanfordNER implements a linear chain Conditional Random Field (CRF) model [16], which is a popular approach in sequence-tagging tasks. At first, using StanfordNER, we get the conditional monolingual probability for each class as in equation (2).

$$P_{CRF}(y \mid x) = \frac{1}{Z(x)} \prod_i M_i(y_i, y_{i-1} \mid x)$$  (2)

where $x$ is the input sequence, $y$ is the output sequence, $Z(x)$ is the partition function, $M_i$ is the clique potential for edge clique $i$.

In particular, we calculate $P_{target}(T_R \mid w_T)$ where $T_R \in \{GPE, LOC, ORG, PER, O\}$ which is the probability that the target word $w_T$ has type $T_R$ as in equation (3).

$$P_{target}(T_R \mid w_T) = P_{CRF}(y_i \mid x_i)$$  (3)

Where $x_i$ word is recognized with tag $y_i$.

Previous work showed that word clustering could improve CRF tagger performance [12]. Therefore, we also use word clustering for both English NER and Chinese NER. In particular, we re-use the word cluster results obtained from the work of Che et. al. (2013) [12] to train monolingual CRF taggers.

### 2.3. Projecting Method

In this section, our main projection mechanism is described. Theoretically, only word alignment is needed to project named entities from source language to target language. Thus, we can just simply assume two words aligned with each other would have the same type. Nonetheless, in practice, word alignment systems can have errors and require large bilingual corpora to generate reliable results. Specifically, word alignment for named entities is less reliable than other words since NEs are usually proper names and one proper name is not frequently repeated in the corpus (i.e., NEs are sparsely distributed) [10]. Therefore, instead of relying solely on the word alignment results, we consider the alignment probability to build the Alignment Set $A$. Corresponding to one target word, this set $A$ contains source words that contribute to the probability of assigning the NE type of that word. The larger the alignment probability of the
target word with the source word is, the more contribution the source word make to the tagging process of that target word. The source score for each word \( w_T \) is calculated as in equation (4)

\[
P_{source}(T_R \mid A) = \frac{\sum_{w_S \in A}[1 + P_{mono}(T_R \mid w_S)]P_{align}(w_T \mid w_S) - 1}{|A|} \tag{4}
\]

where \( P_{align}(w_T \mid w_S) \) is the word alignment probability that the target word \( w_T \) is aligned with the source word \( w_S \) which can be obtained from IBM models [17]. Similar to \( P_{target} \), we can estimate \( P_{mono}(T_R \mid w_S) \) by \( P_{mono}(T_R \mid w_S) = P_{CRF}(T_R \mid w_S) \).

For example in Figure 1, assume that due to errors in word alignment, the alignment set \( A \) of “苹果(Apple)” contains [“Apple”, “ORG”], [“Tim”, “PER”], [“Cook”, “PER”]. Nevertheless, the correct alignment “苹果(Apple)” - “Apple” should have far higher alignment probability compared to “苹果(Apple)” - “Tim” and “苹果(Apple)” - “Cook”. Therefore, the word “Apple” with tag “ORG” contribute more in choosing the final tag for “苹果(Apple)” than the other two words. Thus, we are able to correctly tag “苹果(Apple)” as an entity of type “ORG”.

In order to find the alignment set \( A \), we proposed two ranking strategies based on beam search: TopK and Threshold. Beam search has been proven to be effective in many Natural Language Processing tasks [18, 19].

2.3.1. TopK Beam Search In order to find the most appropriate source words, we need to rank all possible alignments according to some heuristics. Previous work showed the importance of word translation in Named Entity Alignment task [10]. Therefore, we use the word alignment probability between a target word and a source word obtained from IBM models [17] as the heuristic to compare. We rank all the alignments in descending order and get top K source words which have the highest alignment probability with the considered target word to obtain the alignment set \( A \). Heuristically, the alignment set \( A \) is considered as the most appropriate source words set describing the “cross-lingual context”. To find the optimal value for K, a development set is used.

2.3.2. Threshold Beam Search Similar to TopK approach, we also sort the possible alignments set in descending order based on word alignment probability. We get all source words \( w_S \) with word alignment probabilities greater than or equal to some threshold value \( \theta \) as in equation (5).

\[
A = \{w_S \mid P_{align}(w_T \mid w_S) \geq \theta\} \tag{5}
\]

where \( P_{align}(w_T \mid w_S) \) is the word alignment probability and the threshold \( \theta \) would be chosen using a development set.

2.4. Bilingual Type Reassignment Score

Finally, we combine all the mentioned scores by using interpolation as in equation (6) to calculate the Bilingual Score, which describes the probability that a word \( w_T \) has type \( T_R \) in the target sentence.

\[
P_{bd}(T_R \mid w_T) = \lambda P_{target}(T_R \mid w_T) + (1 - \lambda) P_{source}(T_R \mid A) = \lambda P_{target}(T_R \mid w_T) + (1 - \lambda) \frac{\sum_{w_S \in A}[1 + P_{mono}(T_R \mid w_S)]P_{align}(w_T \mid w_S) - 1}{|A|} \tag{6}
\]

where the interpolation parameter \( \lambda \) will be tuned on the development set.
2.5. Type Assignment
For each word, we select the highest Bilingual Type Reassignment Score and assign it with the corresponding label, as in equation (7).

\[ T_R(w_T) = \arg\max_{T_R \in \{GPE, PER, LOC, ORG, O\}} P_{bi}(T_R | w_T) \] (7)

2.6. Distortion Feature
To further refine the Bilingual Type Reassignment Score, we apply the distortion feature. The intuition for this feature is that the positions of words in sentences are also clues to align target words and source words between languages. If a word appears at the \(i^{th}\) position in English sentence, the aligned Chinese word may locate close to the \(i^{th}\) position in Chinese sentence. The larger the difference in position is, the less possibility they are aligned together. Assume that a target word \(w_T\) appears at position \(pos_T\) in the target sentence \(S_T\), and its aligned word \(w_S\) appears at position \(pos_S\) in the source sentence \(S_S\). Following the work of Nguyen et al [10], the distortion score for \(w_T\) and \(w_S\) is calculated as in equation (8)

\[ Dist(w_T, w_S) = 1 - f(\frac{pos_T}{\text{len}(S_T)} - \frac{pos_S}{\text{len}(S_S)}) \] (8)

where \(f(x)\) is the distance function. In this work, we try \(f\) with modulus function. Thus, the source score is redefined as in equation (9)

\[ P_{\text{source}}(T_R | A) = \frac{\sum_{w_S \in A} [1 + P_{\text{mono}}(T_R | w_S)] P_{\text{align}}(w_T | w_S) \cdot Dist(w_T, w_S) - 1}{|A|} \] (9)

2.7. O-type Penalty
Typically, the \(O\) label (not entity) appears far more frequently than other labels in a corpus. This leads to the unbalanced class problem of training data in which the number of instances for each class is not distributed equally. Therefore, we only tag a word with label \(O\) when the score corresponding to class \(O\) is considerably larger than the score of other classes, as in equation (10). We introduce the O-type penalty, which is the amount of difference in score between \(O\) and other classes needed to tag the word with type \(O\). To find the optimal O-type penalty \(\theta_O\), we use a development set.

\[ \forall T_R' \in \{PER, LOC, ORG, GPE\}, P(O | w) > P(T_R' | w) \Leftrightarrow P(O | w) - P(T_R' | w) \geq \theta_O \] (10)

3. Experiment
3.1. Experimental Settings
3.1.1. Data
We conducted experiments with the data used in the work of Che et al. (2013) [12]. In particular, this dataset is extracted from named entity annotated corpus of OntoNotes 4.0 and contains 4 classes: GPE, LOC, ORG, PER.

Training Set:
- We trained word alignment using BerkeleyAligner [20] with 8249 English-Chinese sentence pairs aligned at sentence level.
- For English NER, we trained StanfordNER using a training set of 10,000 sentences. The number of instances for each class is: GPE - 1891, LOC - 211, ORG - 1005, PER - 2044.
- For Chinese NER, we trained StanfordNER using a training set of 10,000 sentences. The number of instances for each class is: GPE - 1326, LOC - 500, ORG - 720, PER - 1440.
Test Set: The TestSet consists of 8249 English-Chinese sentence pairs. The detail distribution of each class is: English GPE - 6721, LOC - 596, ORG - 3525, PER - 3845; Chinese GPE - 6055, LOC - 932, ORG - 3387, PER - 3402.

3.1.2. Evaluation Metrics We evaluate NE recognition with 3 metrics: Recall \((R)\), Precision \((P)\), and \(F_1\)-score \((F_1)\). The same TestSet is used for all experiments.

\[
P = \frac{\text{number of correctly predict NE}}{\text{number of predict NE}}, \quad R = \frac{\text{number of correctly predict NE}}{\text{number of reference NE}}, \quad F_1 = \frac{2 \times P \times R}{P + R}
\]

3.1.3. Comparison System The StanfordNER toolkit is used as a baseline to compare with our proposed system. We also compare with the work of Che et al. (2013) \([12]\) \((\text{Che2013})\) which is considered the state-of-the-art method in improving Named Entity Recognition performance in English-Chinese.

3.2. Results

3.2.1. Monolingual NER Table 1 indicates the monolingual NER results of English and Chinese. From this table, we could see that the performance of class PER is the highest for both languages because it is the most common type in the training set. Overall, the difference in NER performance between English and Chinese is rather significant (32.92 \% in term of \(F_1\)-score). The average results \((AVG)\) of monolingual NER systems is then used as a baseline for comparison \((\text{StanfordNER})\) in further experiments.

| Class | P(\%) | R(\%) | \(F_1\)(\%) |
|-------|-------|-------|------------|
| LOC   | 57.86 | 31.13 | 46.64 / 25.57 |
| ORG   | 75.68 | 72.53 | 68.67 / 35.94 |
| PER   | 90.63 | 86.03 | 77.25 / 45.54 |
| GPE   | 88.40 | 81.72 | 75.75 / 42.83 |
| AVG   | 82.75 | 75.90 | 75.75 / 42.83 |

3.2.2. Bilingual NER Table 2 shows the results of improving NER without using Distortion feature and O-type Penalty for our system. Two beam search strategies \(\text{TopK}\) and \(\text{Threshold}\) are investigated. In both approaches, the mono weight of Chinese \(\lambda\) is 0.1125 which means that the “cross-lingual context” contributes greatly to fixing the NER errors. By projecting from English to Chinese to recover the unrecognized NEs in Chinese side, we are able to improve Recall by a large amount compared to previous methods (21.49\% and 5.71\% respectively). Therefore, there is an improvement in Chinese Named Entity Recognition from 42.83\% to 59.84\% comparing to the \(\text{StanfordNER}\) and by 2.19\% comparing to the work of Che et al. (2013). For English side, \(\text{Threshold}\) yields an \(F_1\)-score of 75.80\%, which is comparable to \(\text{StanfordNER}\) (75.75\%) but slightly lower than \(\text{Che2013}\) (77.08\%).

3.2.3. Bilingual NER with Distortion In this section, we illustrate the effectiveness of using Distortion Feature to further improve our proposed method as in Table 3. For Chinese side, the mono weight \(\lambda\) in \(\text{TopK}\) and \(\text{Threshold}\) approaches are relatively small (0.05 and 0.0625 respectively), which showed again that the “cross-lingual context” plays an important role in tagging unrecognized NEs. In details, \(F_1\)-score improves from 42.83\% (\(\text{StanfordNER}\)) and 57.65
Table 2. The Bilingual NER Results of English / Chinese

| Method    | P(%)     | R(%)     | F1(%)    |
|-----------|----------|----------|----------|
| StanfordNER | 82.75 / 75.90 | 69.90 / 30.13 | 75.75 / 42.83 |
| Che2013   | **83.44 / 79.99** | **71.66 / 45.91** | **77.08 / 57.65** |
| TopK      | 83.14 / 72.32 | 69.71 / **51.62** | 75.79 / **59.84** |
| Threshold | 83.16 / 72.44 | 69.71 / 51.20 | 75.80 / 59.57 |

% (Che2013) to 60.22%. As for the English side, the NER performance is almost the same as Bilingual NER without Distortion feature (Table 2).

Table 3. The Bilingual NER Results with Distortion of English / Chinese

| Method    | P(%)     | R(%)     | F1(%)    |
|-----------|----------|----------|----------|
| StanfordNER | 82.75 / 75.90 | 69.90 / 30.13 | 75.75 / 42.83 |
| Che2013   | **83.44 / 79.99** | **71.66 / 45.91** | **77.08 / 57.65** |
| TopK      | 83.14 / 70.95 | 69.72 / 52.73 | 75.80 / 60.12 |
| Threshold | 83.14 / 70.22 | 69.71 / **53.28** | 75.80 / **60.22** |

3.2.4. Bilingual NER with O-type Penalty In this section, we demonstrate the effectiveness of using O-Type Penalty to further improve our proposed method as in Table 4. Similar to Bilingual NER with Distortion, the mono weight $\lambda$ of Chinese is small (0.1625) which showed the importance of using information from English side. Chinese NER improve significantly by 20.17% compared to StanfordNER and 5.35 % compared to Che2013 in term of F1-score. On the other hand, English NER improved slightly from 75.75% (StanfordNER) to 76.03% but this result is still lower than Che2013 (77.08%).

Table 4. The Bilingual NER Results with O-type Penalty of English / Chinese

| Method    | P(%)     | R(%)     | F1(%)    |
|-----------|----------|----------|----------|
| StanfordNER | 82.75 / 75.90 | 69.90 / 30.13 | 75.75 / 42.83 |
| Che2013   | **83.44 / 79.99** | **71.66 / 45.91** | **77.08 / 57.65** |
| TopK      | 82.83 / 71.14 | 70.32 / **57.00** | 76.03 / 63.00 |
| Threshold | 82.84 / 71.35 | 70.40 / 56.95 | 76.08 / **63.04** |

3.2.5. Bilingual NER with Distortion and O-type Penalty We illustrate the effectiveness of using both Distortion feature and O-type Penalty to further improve our proposed method as in Table 5. As shown in the table, using Distortion feature and O-type Penalty with Threshold strategy yields the best results, achieving $F_1$-score of 63.64% for Chinese NER. These results agree with our intuition that combining these two additional bilingual features could refine the performance of our initial Bilingual NER system. As for English NER, the $F_1$-score is 76.08%, which shows no further improvement when compared with Bilingual NER with O-type Penalty (Table 4). In general, our system achieved the best performance with Distortion and O-Penalty features, which is 76.08% for English NER and 63.64% for Chinese NER in terms of $F_1$-score.

4. Related Work

Bilingual corpora have long been exploited in improving different monolingual as well as cross-lingual Natural Language Processing tasks. In 2001, Yarowsky et al. [21] presented a detailed
Table 5. The Bilingual NER Results with O-type Penalty and Distortion of English / Chinese

| Method     | P(%)     | R(%)     | F1(%)    |
|------------|----------|----------|----------|
| StanfordNER| 82.75 / 75.90 | 69.90 / 30.13 | 75.75 / 42.83 |
| Che2013    | 83.44 / 79.99 | 71.66 / 45.91 | 77.08 / 57.65 |
| TopK       | 82.90 / 70.41 | 70.28 / 58.24 | 76.03 / 63.47 |
| Threshold  | 82.80 / 69.69 | 70.42 / 58.99 | 76.08 / 63.64 |

survey for cross-language annotation projection for four different NLP tasks: POS-tagging, noun phrase parsing, named entity recognition, and morphological analysis. The work showed that high-quality results of English could be transferred across languages through word alignment. In 2009, Huang et al. [22] incorporated bilingual constraints into a shift-reduce parser to improve parsing performance in dependency parsing task. Using only three simple alignment features, the authors managed to resolve shift-reduce conflicts with negligible overhead.

Specifically for Named Entity Recognition task, there were various works which made use of existing named entity annotated bilingual resources. Li et al. (2012) [11] proposed a bilingual CRFs framework which incorporates bilingual constraints to jointly tag named entities in a parallel English-Chinese corpus. Besides the use of bilingual corpora, other bilingual resources have proven to be useful in bilingual NER task. The work of Huang and Vogel (2002) [23] used bilingual named entity dictionaries to support the process of extracting named entity pairs from an unannotated corpus. Kim et al. (2012) [24] utilized Wikipedia metadata in combined with bilingual corpora through CRF model to label named entities in sentences of different foreign languages. Despite high performance, these methods required annotated bilingual corpora which are costly to obtain.

Other approaches explored the use of unannotated bilingual corpora which are aligned at sentence level. In 2014, Fu et al. [25] presented intuitive and effective heuristics to project English named entities into Chinese ones. Results showed that the generated corpus achieved comparable results to a manually annotated corpus in Named Entity Recognition task. This method could be expanded to different domains to solve the common domain over-fitting problem. Che et al. (2013) [12] used integer linear programming to enforce entities to agree through bilingual constraints. This method could jointly tag named entities in both languages without any annotated data.

5. Conclusion
In this paper, we explore the use of bilingual resources to improve monolingual Named Entity Recognition systems of English and Chinese. In general, the main idea of our system is incorporating projecting methods with additional bilingual features to fix existing NER errors. Our method is performed on an existing unannotated bilingual corpus without any further annotations like named entity annotations or gazetteers. Experimental results show that the proposed system managed to make an improvement in Chinese NER performance. In particular, F1-score of Chinese NER increase significantly from 42.83% (StanfordNER) and 57.65% (Che2013) to 63.64%. Regarding the English side, we managed to outperform StanfordNER, in which F1-score increase from 75.75% to 76.08%. The proposed system could be applied to broader domains to recognize more entity types such as MONEY, TIME, PRODUCT or PROTEIN and GENE in Bio-informatics. Moreover, a modified version of our method can be applied to other tasks like Part-of-speech Tagging, Co-reference Resolution, Semantic Role Labeling, as well as other resource-limited languages.
Acknowledgements
This research is supported by research funding from Advanced Program in Computer Science, University of Science, Vietnam National University - Ho Chi Minh City.

6. References
[1] Chen M, Mao S and Liu Y 2014 Big data: A survey Mobile Networks and Applications ISBN 1383-469X
[2] Labrinidis A and Jagadish H V 2012 Proceedings of the VLDB Endowment 5 2032–2033
[3] Grishman R and Sundheim B 1995 Design of the nuc-6 evaluation Proceedings of the 6th conference on Message understanding (Association for Computational Linguistics) pp 1–11
[4] Babych B and Hartley A 2003 Improving machine translation quality with automatic named entity recognition Proceedings of the 7th International EAMT workshop on MT and other Language Technology Tools, Improving MT through other Language Technology Tools: Resources and Tools for Building MT (Association for Computational Linguistics) pp 1–8
[5] Molla D, Van Zaamen M, Smith D et al. 2006
[6] Chen H H, Ding Y W and Tsai S C 1998 Computer Processing of Oriental Languages 12 75–85
[7] Bagga A and Baldwin B 1998 Entity-based cross-document coreferenceing using the vector space model Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1 (Association for Computational Linguistics) pp 79–85
[8] Santos C N d and Guimaraes V 2015 arXiv preprint arXiv:1505.05008
[9] Chiu J P and Nichols E 2015 arXiv preprint arXiv:1511.08308
[10] Nguyen L H, Dinh D and Tran P 2016 ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 16 9
[11] Li Q, Li H, Ji H, Wang W, Zheng J and Huang F 2012 Joint bilingual name tagging for parallel corpora Proceedings of the 21st ACM international conference on Information and knowledge management (ACM) pp 1727–1731
[12] Che W, Wang M, Manning C D and Liu T 2013 Named entity recognition with bilingual constraints Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies pp 52–62
[13] Nemhauser G L and Wolsey L A 1988 Wiley, Chichester. GL Nemhauser, MWP Savelsbergh, GS Sigismondi (1992). Constraint Classification for Mixed Integer Programming Formulations. COAL Bulletin 20 8–12
[14] Tseng H, Chang P, Andrew G, Jurafsky D and Manning C 2005 A conditional random field word segmenter for sighan bakeoff 2005 Proceedings of the fourth SIGHAN workshop on Chinese language Processing
[15] Manning C D, Surdeanu M, Bauer J, Finkel J, Bethard S J and McClosky D 2014 The Stanford CoreNLP natural language processing toolkit Association for Computational Linguistics (ACL) System Demonstrations pp 55–60 URL http://www.aclweb.org/anthology/P/P14/P14-5010
[16] Lafferty J, McCallum A and Pereira F C 2001
[17] Brown P F, Pietra V J D, Pietra S A D and Mercer R L 1993 Computational linguistics 19 263–311
[18] Och F J 2003 Minimum error rate training in statistical machine translation Proceedings of the 41st Annual Meeting on Computational Linguistics-Volume 1 (Association for Computational Linguistics) pp 160–167
[19] Zhang Y and Clark S 2011 Computational linguistics 37 105–151
[20] BerkeleyAligner Google Code Archive URL https://code.google.com/archive/p/berkeleyaligner/
[21] Yarowsky D, Ngai G and Wicentowski R 2001 Inducing multilingual text analysis tools via robust projection across aligned corpora Proceedings of the first international conference on Human language technology research (Association for Computational Linguistics) pp 1–8
[22] Huang L, Jiang W and Liu Q 2009 Bilingually-constrained (monolingual) shift-reduce parsing Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3 (Association for Computational Linguistics) pp 1222–1231
[23] Huang F and Vogel S 2002 Improved named entity translation and bilingual named entity extraction Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference on (IEEE) pp 253–258
[24] Kim S, Toutanova K and Yu H 2012 Multilingual named entity recognition using parallel data and metadata from wikipedia Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1 (Association for Computational Linguistics) pp 694–702
[25] Fu R, Qin B and Liu T 2014 Frontiers of Computer Science 8 629–641