Computational Linguistics and Chinese Language Processing
Vol. 6, No. 1, February 2001, pp. 1-26

Improving Translation Selection with a New Translation Model Trained by Independent Monolingual Corpora

Ming Zhou[*], Yuan Ding[+1], Changning Huang[*]

Abstract

We propose a novel statistical translation model to improve translation selection of collocation. In the statistical approach that has been popularly applied for translation selection, bilingual corpora are used to train the translation model. However, there exists a formidable bottleneck in acquiring large-scale bilingual corpora, in particular for language pairs involving Chinese. In this paper, we propose a new approach to training the translation model by using unrelated monolingual corpora. First, a Chinese corpus and an English corpus are parsed with dependency parsers, respectively, and two dependency triple databases are generated. Then, the similarity between a Chinese word and an English word can be estimated using the two monolingual dependency triple databases with the help of a simple Chinese-English dictionary. This cross-language word similarity is used to simulate the word translation probability. Finally, the generated translation model is used together with the language model trained with the English dependency database to realize translation of Chinese collocations into English. To demonstrate the effectiveness of this method, we performed various experiments with verb-object collocation translation. The experiments produced very promising results.

Keywords: Translation selection, Statistical machine translation, Chinese-English machine translation, Cross language word similarity

1. Introduction

Selecting the appropriate word translation among several options is a key technology of

---

[*] Microsoft Research, Asia.

[+] Tsinghua University.

[1] Currently is studying at University of Pennsylvania as Ph.D. student. This work was done while visiting Microsoft Research Asia as a visiting student.

machine translation. For example, the Chinese verb “订” is translated in different ways in terms of objects, as shown in the following:

订 报纸 →subscribe to a newspaper
订 计划 →make a plan
订 旅馆 →book a hotel
订 车票 →reserve a ticket
订 时间 →determine the time

In recent years, there has been increasing interest in applying statistical approaches to various machine translation tasks, from MT system mechanisms to translation knowledge acquisition. For translation selection, most researches applied statistical translation models. In such statistical translation models, to get the word translation probability as well as translation templates, bilingual corpora are needed. However, for quite a few languages, large bilingual corpora rarely exist, while large monolingual corpora are easy to acquire. It will be helpful to alleviate the burden of collecting bilingual corpus if we can use monolingual corpora to estimate the translation model and find alternative to translation selection.

We propose a novel approach to this problem in the Chinese-English machine translation module which is to be used for cross-language information retrieval. Our method is based on the intuition that although the Chinese language and the English language have different definitions of dependency relations, the main dependency relations like subject-verb, verb-object, adjective-noun and adverb-verb tend to have strongly direct correspondence. This assumption can be used to estimate the word translation probability. Our proposed method works as follows. First, a Chinese corpus and an English corpus are parsed, respectively, with a Chinese dependency parser and an English dependency parser, and two dependency triple databases are generated as the result. Second, the word similarity between a Chinese word and an English word are estimated with these two monolingual dependency triple databases with the help of a simple Chinese-English dictionary. This cross-language word similarity is used as the succedaneum of the word translation model. At the same time, the probability of a triple in English can be estimated with the English triple database. Finally, the word translation model, working together with the triple probability, can realize a new translation framework. Our experiments showed that this new translation model achieved promising results in improving translation selection. The unique characteristics of our method include: 1) use of two monolingual corpora to estimate the translation model. 2) use of dependency triples as basis for our method.

The remainder of this paper is organized as follows. In Section 2, we give a detailed description to our new translation model. In section 3, we describe the training process of our new model, focusing on the process of constructing the dependency triple database for English and Chinese. The experiments and evaluation of this new method are reported in Section 4. In Section 5, some related works are introduced. Finally in Section 6, we draw conclusions and discuss future work.

2. A New Statistical Machine Translation Model

In this section, we will describe the proposed translation model. First, we will report our observations from a sample word-aligned bilingual corpus in order to verify our assumption. After that, we will introduce the method for estimating the cross-language word similarity by means of two monolingual corpora. Finally, we will give a formal description of the new translation model.

2. 1 Dependency Correspondence between Chinese and English

A dependency triple consists of a head, a dependant, and a dependency relation between the head and the dependant. Using a dependency parser, a sentence can be analyzed to obtain a set of dependency triples in the following form:

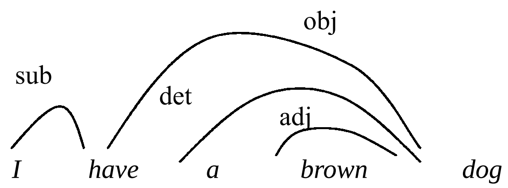
$$trp = (w_1, rel, w_2),$$

which means that word w_1 has a dependency relation of rel with word w_2.

For example, for the English sentence *"I have a brown dog"*, a dependency parser obtains a set of triples as follows:

(1)

a.



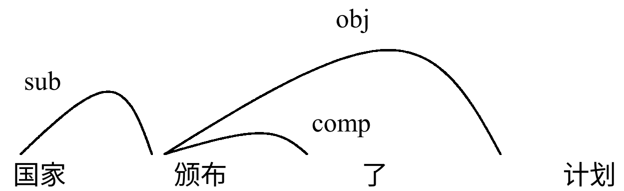
b. (*have, sub, I*), (*I, sub-of, have*), (*have, obj, dog*), (*dog, obj-of, have*), (*dog, adj, brown*), *(brown, adj-of, dog), (dog, det, a), (a, det-of, dog)*[2]

Similarly, for the Chinese sentence “国家颁布了计划”, we can get the following dependency triples with a dependency parser:

[2] The standard expression of the dependency parsing result is: (*have, sub, I*), (*have, obj, dog*), (*dog, adj, brown*), *(dog, det, a).*

(2)

a



.b. (颁布, sub, 国家), (国家, sub-of, 颁布), (颁布, obj, 计划), (计划, obj-of, 颁布), (颁布, comp, 了), (了,comp-of, 颁布)[3]

Among all the dependency relations in Chinese and in English, the key dependency relations are subject-verb (denoted as sub), verb-object (denoted as obj), adjective-noun(denoted as adj) and adverb-verb(denoted as adv). Our intuitive assumption is that although Chinese language and English language have different schemes of dependency relations, these key dependency relations tend to have strong correspondence. For instance, normally, a word pair with subject-verb relation in Chinese can be translated into a subject-verb relation pair in English. Formally speaking, for a triple *(A, D, B)* in Chinese, where *A and B* are words, and *D* is one of the key dependency relations mentioned above, the translation of the triple *(A, D, B*) in English, can be expressed as *(A', D', B')*, where *A'* and *B'* are the translations of *A* and *B,* respectively, and *D'* is the dependency relation between *A'* and *B'* in the English language[4]. Our assumption is that although D and D' may be different in denotation, they can be mapped directly in most cases.

In order to verify our assumption, we conducted an investigation with a Chinese-English bilingual corpus[5]. The bilingual corpus, consisting of 60,000 pairs of Chinese sentences and English sentences selected from newspapers, novels, general bilingual dictionaries and software product manuals, was aligned manually at the word level. An example of the word aligned corpus is given in Table 1. Each word is identified with a number in order to indicate the word alignment information.

---

[3] The standard expression of the dependency parsing result is: (颁布, sub, 国家), (颁布, obj, 计划), (颁布, comp, 了).

[4] Sometimes to get a better translation, a triple in one language is not translated into a triple in other language, but except in very extreme cases, it will still be acceptable if it is translated into a triple.

[5] This corpus, produced by Microsoft Research Asia, is currently reserved for Microsoft internal use only.

Table 1. *The word aligned bilingual corpus*

Chinese sentence	当/1 斯科特/2 抵达/3 南极/4 的/5 时候/6 ，/7 他/8 发现/9 阿蒙森/10 比/11 他/12 领先/13 。/14
English sentence	When/1 Scott/2 reached/3 the/4 South/5 Pole/6 , /7 he/8 found/9 Amundsen/10 had/11 anticipated/12 him/13 ./14
Aligned word pair	(1,5,6:1); (2:2); (3:3); (4:4,5,6); (7:7); (8:8); (9:9); (10:10); (11:nil); (12:13); (13:12); (14:14);

To obtain statistics of the dependency relation correspondence, we parsed 10,000 sentence pairs with the English parser Minipar [Lin 1993, Lin 1994] and the Chinese parser BlockParser [Zhou 2000]. The parsing results were expressed in dependency triples. We then mapped the dependency relations so that we could count the correspondences between an English dependency relation and a Chinese dependency relation. More than 80% of subject-verb, adjective-noun and adv-verb dependency relations could be mapped, while verb-object correspondence was not so high. We show the verb-object correspondence results in Table 2.

Table 2. *Triple correspondence between Chinese and English.*

Dependency Type	E-C Positive	E-C Negative	Mapping Rate	C-E Positive	C-E Negative	Mapping Rate
Verb-Object	7,832	4,247	64.8%	6,769	3,751	64.3%

"*E-C Positive*" means an English verb-object was translated into a Chinese verb-object. "*E-C Negative*" means an English verb-object was not translated into a Chinese verb-object. The *E-C Positive Rate* reached 64.8% and the *C-E Positive Rate* reached 64.3%. These statistics show that our correspondence assumption is reasonable but not strong. Now we will examine the reasons why some of the dependency relations cannot be mapped directly.

Table 3. *Negative examples of triple mapping.*

Chinese verb-object triple	English translation
够 开销	be enough for
用 数字	in numeral characters
用 货币	Change to currency
名叫 威廉_・_罗	an Englishman, Willian Low
…觉得逃避到生活虽艰苦但比较简朴的年代里是件愉快的事。	…found it pleasant to escape to a time when life, though hard, was relatively simple.

From Table 3, we can see that "negative" mapping has several causes. The most important reasons are: a Chinese verb-object can be translated into a single English verb (e.g., an intransitive verb) or can be translated into verb+prep+obj. If these two mappings (as shown in Table 4) are also considered reasonable correspondences, then the mapping rate will increase significantly. As seen in Table 5, the *E-C Positive rate* and the *C-E Positive rate* reached 82.71% and 83.87% respectively.

Table 4. *Extended mapping.*

Chinese triple	English triple	Examples
Verb-Object	Verb(usually intransitive verb)	读-书 →read
Verb-Object	Verb+Prep-Object	用-货币→change to – currency

Table 5. *Triple correspondence between Chinese and English.*

Type	E-C Positive	E-C Negative	Mapping rat	C-E Positive	C-E Negative	Mapping Rate
Verb-Object	9991	2088	82．71%	8823	1697	83．87%

This implies that all four key dependency relations can be mapped very well, showing that our assumption is correct. This fact will be used to estimate the word translation model using two monolingual corpora. The method will be given in the following subsections.

2.2 Cross-Language Word Similarity

We will next describe our approach to estimating the word translation likelihood based on the triple correspondence assumption with the help of a simple Chinese-English dictionary. The key idea is to calculate “cross-language similarity”, which is an extension of word similarity within one language.

Several statistical approaches to computing word similarity have been proposed. In these approaches, a word is represented by a word co-occurrence vector in which each feature corresponds to one word in the lexicon. The value of a feature specifies the frequency of joint occurrence of the two words in some particular relations and/or in a certain window size in the text. The degree of similarity between a pair of words is computed using a certain similarity (or distance) measure that is applied to the corresponding pairs of vectors. This similarity computation method relies on the assumption that the meanings of the words are related to their co-occurrence patterns with other words in the text. Given this assumption, we can expect that words which have similar co-occurrence patterns will resemble each other in meaning.

Different types of word co-occurrences have been examined with respect to computing word similarity. They can in general be classified into two types, which refer to the co-occurrence of words within the specified syntactic relations, and the co-occurrence of words that have non-grammatical relations in a certain window in the text. The set of co-occurrences of a word within syntactic relations strongly reflects its semantic properties. Lin [1998b] defined lexical co-occurrences within syntactic relations, such as subject-verb, verb-object, adj-noun, etc. These types of co-occurrences can be used to compute the similarity of two words.

While most methods proposed up to now are for computing the word similarity within one language, we believe that some of these ideas can be extended to computation of "cross-language word similarity". Cross-language word similarity denotes the commonality between one word in a language and one word in another language. In each language, a word is represented by a vector of features in which each feature corresponds to one word in the lexicon. The key to computing cross-language similarity is to determine how to calculate the similarity of two vectors which are represented by words in different languages.

Based on the triple correspondence assumption which we have made in 2.1, dependency triples can be used to compute the cross language similarity. In each language, a word is represented by a vector of dependency triples which co-occur with the word in the sentence. Our approach assumes that a word in one language is similar to a word in another language if their vectors are similar in some sense. In addition, we can use a bilingual lexicon to bridge the words in the two vectors to compute cross-language similarity.

Our similarity measure is an extension of the measure proposed in [Lin, 1998b], where the similarity between two words is defined as the amount of information contained in the commonality between the words and is divided by the sum of information in the descriptions of the two words in each language respectively.

In Lin [1998b]'s work, a dependency parser was used to extract dependency triples. For a word w_1, a triple (w_1, rel, w_2) represents a feature of w_1, which means w_1 can be used in relation of rel with word w_2. The description of a word *w* consists of the frequency counts of all the dependency triples that match the pattern *(w,* , *).*

An occurrence of a dependency triple (w_1, rel, w_2) can be regarded as the co-occurrence of three events [Lin, 1998b]:

A: a randomly selected word is w_1;

B: a randomly selected dependency type is rel;

C: a randomly selected word is w_2.

According to Lin [1998b], if we assume that A and C are conditionally independent given B, then the information contained in $\| w_1, rel, w_2 \| = f(w_1, rel, w_2) = c$ can be computed as follows[6]:

$$I(w_1, rel, w_2) = -\log(P_{MLE}(B)P_{MLE}(A \mid B)P_{MLE}(C \mid B)) - (-\log P_{MLE}(A,B,C)); \qquad (1)$$

where:

$$P_{MLE}(A \mid B) = \frac{f(w_1, rel, *)}{f(*, rel, *)}; \qquad (2)$$

$$P_{MLE}(C \mid B) = \frac{f(*, rel, w_2)}{f(*, rel, *)}; \qquad (3)$$

$$P_{MLE}(B) = \frac{f(*, rel, *)}{f(*, *, *)}; \qquad (4)$$

$$P_{MLE}(A,B,C) = \frac{f(w_1, rel, w_2)}{f(*, *, *)}; \qquad (5)$$

where $f(x)$ denotes the frequency of x; * is a wildcard for all possible combinations.

Finally, we have [Lin, 1998b]

$$I(w_1, rel, w_2) = \log_2 \frac{f(w_1, rel, w_2) f(*, rel, *)}{f(w_1, rel, *) f(*, rel, w_2)} \qquad (6)$$

Let $T(w)$ be the set of (rel, w') such that $\log_2 \frac{f(w, rel, w') f(*, rel, *)}{f(w, rel, *) f(*, rel, w')}$ is positive.

Then the similarity between two words, w_1 and w_2, within one language is defined as follows [Lin, 1998b]:

$$Sim(w_1, w_2) = \frac{\sum_{(rel, w) \in T(w_1) \cap T(w_2)} (I(w_1, rel, w) + I(w_2, rel, w))}{\sum_{(rel, w) \in T(w_1)} I(w_1, rel, w) + \sum_{(rel, w) \in T(w_2)} I(w_2, rel, w)} \qquad (7)$$

Now, let us see how we can extend to cross language. Similarly, for a Chinese word w_C and an English word w_E, let $T(w_C)$ be the set of pairs (rel_C, w_C') such that $\log_2 \frac{f(w_c, rel_c, w_c') f(*, rel_c, *)}{f(w_c, rel_c, *) f(*, rel_c, w_c')}$ is positive, and let $T(w_E)$ be the set of pairs (rel_E, w_E') such that $\log_2 \frac{f(w_E, rel_E, w_E') f(*, rel_E, *)}{f(w_E, rel_E, *) f(*, rel_E, w_E')}$ is positive. Then we can similarly define cross-language word similarity as follows:

[6] Please see [Lin, 1998b] for the detailed derivation process of this formula.

$$Sim(w_C, w_E) = \frac{I_{common}(w_C, w_E)}{\sum_{(rel_C, w'_C) \in T(w_C)} I(w_C, rel_C, w'_C) + \sum_{(rel_E, w'_E) \in T(w_E)} I(w_E, rel_E, w'_E)} \quad (8)$$

where $I_{common}(w_C, w_E)$ denotes the total information contained in the commonality of the features of w_C and w_E. Actually, we have three different methods for calculating $I_{common}(w_C, w_E)$.

1) Map Chinese into English

We define

$$T_{C \to E}(w_E) = \{(rel_E, w'_E) \mid rel_E = correspondence(rel_C), w'_E \in Tran(w'_C)\} \cap T(w_E), where\ (rel_C, w'_C) \in T(w_C)$$

$$T_{C \to E}(w_C) = \{(rel_C, w'_C) \mid (rel_E, w'_E) \in T(w_E), where\ rel_E = correspondence(rel_C), w'_E \in Tran(w'_C)\}$$

Here,

$Tran(x)$ denotes the set of possible translations of word x which are defined in the bilingual lexicon and $rel_E = correpondence(rel_C)$ is the English dependency type corresponding to a Chinese dependency type rel_C.

2) Map English into Chinese

Similarly, we define

$$T_{E \to C}(w_C) = \{(rel_C, w'_C) \mid rel_C = correspondence(rel_E), w'_C \in Tran(w'_E)\} \cap T(w_C), where\ (rel_E, w'_E) \in T(w_E)$$

$$T_{E \to C}(w_E) = \{(rel_E, w'_E) \mid (rel_C, w'_C) \in T(w_C), where\ rel_C = correspondence(rel_E), w'_C \in Tran(w'_E)\}$$

Here,

$rel_C = correpondence(rel_E)$ is the Chinese triple type with rel_C corresponding to an English triple type rel_E.

3) Map both English into Chinese and Chinese into English

Similarly, we define

$$T_{C \leftrightarrow E}(w_C) = T_{E \to C}(w_C) \cup T_{C \to E}(w_C)$$
$$T_{C \leftrightarrow E}(w_E) = T_{E \to C}(w_E) \cup T_{C \to E}(w_E)$$

Then, we can define the cross-language word similarity of w_C and w_E in the following three ways:

$$Sim_{C \to E}(w_C, w_E) = \frac{\sum\limits_{(rel_C, w'_C) \in T_{C \to E}(w_C)} I(w_C, rel_C, w'_C) + \sum\limits_{(rel_E, w'_E) \in T_{C \to E}(w_E)} I(w_E, rel_E, w'_E)}{\sum\limits_{(rel_C, w'_C) \in T(w_C)} I(w_C, rel_C, w'_C) + \sum\limits_{(rel_E, w'_E) \in T(w_E)} I(w_E, rel_E, w'_E)} \tag{9}$$

$$Sim_{E \to C}(w_C, w_E) = \frac{\sum\limits_{(rel_C, w'_C) \in T_{E \to C}(w_C)} I(w_C, rel_C, w'_C) + \sum\limits_{(rel_E, w'_E) \in T_{E \to C}(w_E)} I(w_E, rel_E, w'_E)}{\sum\limits_{(rel_C, w'_C) \in T(w_C)} I(w_C, rel_C, w'_C) + \sum\limits_{(rel_E, w'_E) \in T(w_E)} I(w_E, rel_E, w'_E)} \tag{10}$$

$$Sim_{E \leftrightarrow C}(w_C, w_E) = \frac{\sum\limits_{(rel_C, w'_C) \in T_{E \leftrightarrow C}(w_C)} I(w_C, rel_C, w'_C) + \sum\limits_{(rel_E, w'_E) \in T_{E \leftrightarrow C}(w_E)} I(w_E, rel_E, w'_E)}{\sum\limits_{(rel_C, w'_C) \in T(w_C)} I(w_C, rel_C, w'_C) + \sum\limits_{(rel_E, w'_E) \in T(w_E)} I(w_E, rel_E, w'_E)} \tag{11}$$

Similarity (9) can be seen as the likelihood of translating a Chinese word into an English word, similarity (10) can be seen as the likelihood of translating an English word into a Chinese word, and similarity (11), a balanced and asymmetry formula, can be seen the "neural" similarity of a Chinese word and an English word.

2.3 Translation Selection Model Based on Cross-Language Similarity

We will next discuss how we can build a translation model in order to solve the translation selection problem in dependency triple translation. Suppose we want to translate a Chinese dependency triple $c = (w_{C1}, rel_C, w_{C2})$ into an English dependency triple $e = (w_{E1}, rel_E, w_{E2})$; this is equivalent to finding $e_{\max}$ that will maximize the value $P(e \mid c)$ according to the statistical translation model [Brown, 1993].

Using *Bayes' theorem,* we can write

$$P(e \mid c) = \frac{P(e)P(c \mid e)}{P(c)} \tag{12}$$

Since the denominator $P(c)$ is independent of *e* and is a constant for a given Chinese triple, we have

$$e_{\max} = \arg\max_{e}(P(e)P(c \mid e)) \tag{13}$$

Here, the $P(e)$ factor is a measure of the likelihood of the occurrence of a dependency triple e in the English language. It makes the output of e natural and grammatical. $P(e)$ is usually called the language model, which depends only on the target language. $P(c \mid e)$ is usually called the translation model.

In single triple translation, $P(e)$ can be estimated using formula (5), which can be rewritten as

$$P_{MLE}(w_{E1}, rel_E, w_{E2}) = \frac{f(w_{E1}, rel_E, w_{E2})}{f(*,*,*)}$$

In addition, we have

$$P(c \mid e) = P(w_{C1} \mid rel_C, e) \times P(w_{C2} \mid rel_C, e) \times P(rel_C \mid e)$$

We suppose that the selection of a word in translation is independent of the type of dependency relation, therefore we can assume that w_{C1} is only related to w_{E1}, and that w_{C2} is only related to w_{E2}. Here, we use cross-language word similarity $Sim_{E \to C}$ (see formula 10) to simulate the translation probability from an English word into a Chinese word. Using $Likelihood(c \mid e)$[7] to replace $P(c \mid e)$, we define

$$Likelihood(c \mid e) = Sim_{E \to C}(w_{C1}, w_{E1}) \times Sim_{E \to C}(w_{C2}, w_{E2}) \times P(rel_C \mid e) \tag{14}$$

$P(rel_C \mid e)$ is a parameter which mostly depends on specific word. But this can be simplified as

$$P(rel_C \mid e) = P(rel_C \mid rel_E)$$

Then we have

$$Likelihood(c \mid e) = Sim_{E \to C}(w_{C1}, w_{E1}) \times Sim_{E \to C}(w_{C2}, w_{E2}) \times P(rel_C \mid rel_E)$$

According to our assumption of correspondence between Chinese dependency relations and English dependency relations, we have $P(rel_C \mid rel_E) \approx 1$. Then we have

---

[7] Since $Likelihood$ is not normalized in [0,1], we do not call it probability to avoid confusion.

$$Likelihood(c \mid e) = Sim_{E \to C}(w_{C1}, w_{E1}) \times Sim_{E \to C}(w_{C2}, w_{E2})$$

Therefore, we have

$$\begin{aligned} e_{\max} &= \arg\max_{e}(P(e) \times P(c \mid e)) \\ &= \arg\max_{e}(P(e) \times Likelyhood(c \mid e)) \\ &= \arg\max_{w_{E1}, w_{E2}}(P(e) \times Sim_{E \to C}(w_{C1}, w_{E1}) \times Sim_{E \to C}(w_{C2}, w_{E2})) \end{aligned} \quad (15)$$

In this formula, we use the English dependency triple sets to estimate $P(e)$, and use the English dependency sets and Chinese dependency sets which are independent of each other, to estimate the translation model based on our dependency correspondence assumption. In the whole process, no manually aligned or tagged corpus is needed.

3. Model Training

To estimate the cross-language similarity and the target language triple probability, both Chinese and English dependency triple sets are required to build. Similar to [Lin 1998b], we also use parsers to extract dependency triples from the text corpus. The workflow of constructing the dependency triple databases is depicted in Fig 1.

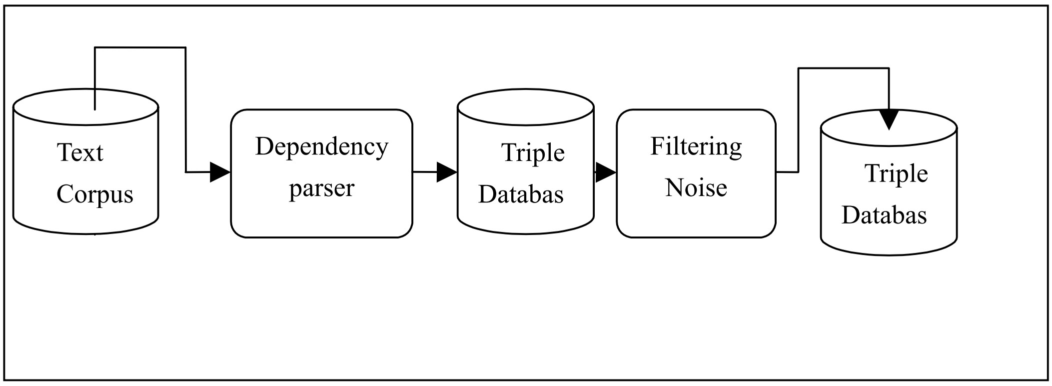


Figure 1 *The flowchart of constructing the dependency triple database.*

As shown in Fig. 1, each sentence from the text corpus is parsed by a dependency parser, and a set of dependency triples is generated. Each triple is put into the triple database. If an instantiation of a type of triple already exists in the triple database, then the frequency of this triple will increase one time. After all the sentences are parsed, we can get a triple database with a large number of triples. Since the parser can not be expected to be 100% correct, some parsing mistakes will inevitably be introduced into the triple database. It is necessary to remove the noisy triples as Lin did [1998a], but in our experiment, we did not apply any noise filtering technique.

Our English text corpus consists of 750 M (byte) of text from the Wall' Street Journal(1980-1990), and our Chinese text corpus contains 1,200 M(byte) of text from People's Daily (1980-1998). The English parser we used was Minipar [Lin 1993, Lin 1994]. Minipar is a broad-coverage, principle-based parser with a lexicon of more than 90,000 words. The Chinese parser we used here was BlockParser [Zhou 2000]. This is a robust rule parser that breaks up Chinese sentences into *"blocks"*, which are represented by *headwords.* Then syntactical dependency analysis was applied to the *"blocks"*. 17 POS tags and 19 grammatical relations were recognized by this parser, and 220,000 entries were registered in the parsing lexicon.

The 750M (byte) English newspaper corpus was parsed within 50 hours on a machine with 4 Pentium™ III 800 CPU, and the 1200 M (byte) Chinese newspaper corpus was parsed in 110 hours on the same machine. We extracted the dependency triples from the parsed corpus. There were 19 million occurrences of dependency triple in the English parsed corpus, and 33 million occurrences of dependency triples in the Chinese parsed corpus. As a result, we acquired two databases of dependency triples of the two languages. These two databases served as the information source for the translation model training and triple probability, which we have described in the above sections.

Table 6. *shows a summary of the corpora and parsers in Chinese and English.*

Language	**Description**	**Size(bytes)**	**#Triple**	**Parser**
Chinese	People's Daily 1980~1998	1,200M	33,000,000	Block Parser
English	Wall's Street Journal 1980-1990	750M	19,000,000	Minipar

The E-C and C-E dictionaries used here are the bilingual lexicon used in machine translation systems developed by Harbin Institute of Technology[8]. The E-C lexicon contains 78,197 entries, and C-E dictionary contains 74,299 entries.

Since in this paper, we are primarily interested in the selection of translations of verbs, we utilized only three types of dependency relations for similarity estimation, i.e., verb-object, verb-adverb and subject-verb. The symmetric triples "object-of", "adverb-of" and "subject-of" were also used in calculating the translation model and the triple probability. Table 7 shows the statistics of occurrences of the three kinds of dependency relations.

[8] These two lexicons are not publicly available.

Table 7. *Statistics of the three main triples*

Language	Verb-Object	Verb-Adverb	Subject-Verb
Chinese	14,327,358	10,783,139	8,729,639
English	6,438,398	3,011,767	5,282,866

Therefore, a word w is represented by a co-occurrence vector $\{(rel, w_1^{'}, \#), (rel, w_2^{'}, \#)...\}$, where $rel \in \{verb-object, verb-adverb, subj-verb\}$ [9] in which each feature $(rel, w_1^{'}, \#)$ consists of the dependency relation rel, another word $w_1^{'}$ that constructs the dependency relation, and the frequency count #. Then we extracted the word lists from the Chinese triple sets and the English triple sets, and calculated the similarity of each Chinese word and each English word. For similarity, we only calculated the similarity between verbs and between nouns of the two languages. As a result, a large table was constructed recording the cross-language similarity as shown in table 8. S (i,j) is the similarity between a Chinese word C_i and an English word E_j. Please note that we only apply similarity formula (10) since we were interested in the translation likelihood from an English word to a Chinese word, as explained in the previous section.

Table 8. *Cross-language word similarity matrix*

	E_1	E_2	$\cdots$	E_m
C_1	S_{11}	S_{12}	$\cdots$	S_{1m}
C_2	S_{21}	S_{22}	$\cdots$	S_{2m}
…	…	…	…	…
C_n	S_{n1}	S_{n2}	$\cdots$	S_{nm}

4. Translation Experiments

Please note that in this paper, we only focus on the verb-object triple translation experiments to demonstrate how to improve translation selection. We conducted a set of experiments with several translation models on the verb-object translation. As the baseline experiment, Model A selected the translation of a verb and its object with the highest frequency as the translation output. Model B utilized the target language triple probability but did not apply the translation model. Model C utilized both the target language triple probability and the translation model.

The verb-object translation answer sets were built manually by English experts from the Department of Foreign Languages of Beijing University. For a certain triple, all the plausible translations are given in building the translation evaluation set. Samples of the evaluation sets are shown in Table 9.

[9] We didn't use the dependency relation of adj-noun.

Table 9. *Evaluation sets prepared by human translators*

Verb	Noun	Translation
说	事	talk business
用	手	use hand
看	电影	see film, see movie
看	电视	watch TV
作	贡献	make contribution

The performance was evaluated based on precision, which is defined as

$$precision = \frac{\#correct\ translaion}{\#total\ verb-obj\ triples} \times 100\%$$

4.1 Various Translation Models

Suppose we want to translate the Chinese dependency triple $c = (w_{C1}, rel_C, w_{C2})$ into the English dependency triple $e = (w_{E1}, rel_E, w_{E2})$; this is equivalent to finding $e_{\max}$ that would maximize translation model we have proposed. To test our method, we conducted a series of translation experiments with incrementally enhanced resources. All the translation experiments reported in this paper were conducted with Chinese-English verb-object triple translation.

Model A (selecting the highest-frequency translation)

As the baseline for our experiment, *Model A* simply selected the translation word in the bilingual lexicon which had the highest frequency in the English corpus. It translated *verb* and *object* separately. *Model A* did not utilize the triple probability or the translation model. Formally, Model A can be expressed as

$$e_{\max} = (\underset{w_{e1}e \in Trans(w_{c1})}{\arg\max}\ (freq(w_{E1})), verb-object, \underset{W_{e2} \in Trans(w_{C2})}{\arg\max}\ (freq(w_{E2}))$$

Model B (selecting the translation with the maximal triple probability)

Model B only used the triple probability in target language, neglecting the translation model. It selected the translation of the triple which was most likely to occur in the target language. We have

$$e_{\max} = \underset{e}{\arg\max}\, P(e) = \underset{\substack{w_{E1} \in Trans(w_{Cq}), \\ w_{E2} \in Trans(w_{C2})}}{\arg\max}\ P(w_{E1}, verb-obj, w_{E2})$$

Model C (selecting the translation which fits both the triple probability and the translation model best)

In *Model C*, both the translation model and triple probability were considered. We have

$$e_{\max} = \arg\max_{e} P(e) \times Likelyhood(c \mid e)$$

$$= \arg\max_{\substack{w_{E1} \in Tran(w_{C1}) \\ w_{E2} \in Tran(w_{C2})}} P(w_{E1}, verb-obj, w_{E2}) \times sim_{E \to C}(w_{C1}, w_{E1}) \times sim_{E \to C}(w_{C2}, w_{E2})$$

4.2 Evaluation

We designed a series of evaluations to test the above models. In this subsection, the evaluation results will be reported. To achieve an objective evaluation, we designed three kinds of testing set, 1) high frequency verb and its object, 2) a low frequency verb and its object, and 3) a low frequency verb-object triple. Please note that each selected verb should take a simple noun as its object, the verbs like “是”(be)，”使”(make), “请”(invite), “认为” were not used since their translations were not directly relied on their objects.

Case-I: High-frequency verbs with their objects

We wanted to observe the performance of these models in the translation of verb-objects in which the verbs were high frequency ones. We randomly selected 53 high-frequency verbs (see Appendix I), and randomly extracted certain number of triples of verb-object relation from the Chinese triple database. Totally 730 triples are extracted. The translation results obtained using the various models are shown in Table 10.

Table 10. *Evaluation on verbs of high frequency*

Model	**#Correct**	**Percentage**
Model A	393	53.8%
Model B	512	70.1%
Model C	519	71.1%

From these results, we can see that Model B and Model C achieved considerably better translation precision than did Model A. Model C worked a little better than Model B.

Case-II: Translation of low-frequency verbs with their objects

We tested the translation of the triples composed of low-frequency verbs and a noun. We randomly selected 23 low frequency verbs (see Appendix II) and randomly extracted 108 verb-object triples containing these words from the Chinese triple database. The translation results obtained using the various models are shown in Table 11.

Table 11. *Evaluation of verbs of low frequency*

Model	**#Correct**	**Percentage**
Model A	61	56.5%
Model B	85	78.7%
Model C	88	81.5%

Case III: Translation of low-frequency triples

We also tested the translation of low-frequency triples. First we selected the following objects: "国家, 同志, 企业, 政府, 记者, 会议, 经济, 群众, 农民, 市场, 政策, 公司, 家, 条件, 地区, 基础, 书, 时间, 项目, 人员, 利益". Then we selected triples which contained the above words and occurred less than 5 times. Since the set of such low-frequency triples was very large, we randomly selected 340 triples as the evaluation sets. The results are shown in Table 12.

Table 12. *Evaluation of triples of low frequency*

Model	**#Correct**	**Percentage**
Model A	182	53.5%
Model B	283	83.2%
Model C	289	85.0%

We can see that our methods obtained very promising results in all the cases.

4.3 Accommodating Lexical Gaps (OOV)

One of the reasons for translation mistakes is the OOV problem, *i.e.*, the best translation is out of vocabulary. Therefore, the translation quality is seriously affected. For example, "展开" has two translations in the translation lexicon: "unfold" and "develop". However, the triple "展开, verb-object, 进攻", which should be translated as "launch, verb-object, attack", cannot be properly produced with the translations given by the dictionary. To solve this problem, we used new methods to get a number of possible translations based on the translations defined in the dictionary and obtained very interesting results.

Model D (Translation expansion using a bilingual lexicon)

For the Chinese verb-object triple $c = (w_{C1}, verb-object, w_{C2})$, we can expand new translations by employing an E-C lexicon and the C-E lexicon circles:

$$Tran1(x) = \{x''' \mid x''' \in Tran(x''), x'' \in Tran(x'), x' \in Tran(x)\} \cup Tran(x)$$

Let x be a Chinese words, let x' be the English translation of x defined in the C-E lexicon, let x'' be the Chinese translation of x' defined in E-C lexicon, and let x''' be the English translation of x'' defined in C-E lexicon. Taking "说" as an example, "talk" is one translation based on the C-E lexicon. Then looking up in the E-C lexicon, "说话" is one translation of “talk”. Looking up in the C-E dictionary again, “speak” is one translation of “说话”. In this way, “说” is translated as “speak” in addition to the original translation “talk”. Model D can be described formally as follows:

$$\begin{aligned} e_{\max} &= \arg\max_{e} P(e) \times Likelyhood(c \mid e) \\ &= \arg\max_{\substack{w_{E1} \in Tran1(w_{C1}) \\ w_{E2} \in Tran1(w_{C2})}} P(w_{E1}, verb-obj, w_{E2}) \times sim_{E \to C}(w_{C1}, w_{E1}) \times sim_{E \to C}(w_{C2}, w_{E2}) \end{aligned}$$

Model E (Translation expansion using dependency triple database)

For a Chinese verb-object triple $c = (w_{C1}, verb-object, w_{C2})$, we assume that the translation of object w_{C2} is expanded by *Model D*, *i.e.*,

$$Tran1(w_{C2}) = \{x''' \mid x''' \in Tran(x''), x'' \in Tran(x'), x' \in Tran(w_{C2})\} \cup Tran(w_{C2})$$

However, we expand the verb w_{C1} translation in a new way as shown below:

$$Tran2(w_{C1}) = \{w_{E1} \mid I(w_{E1}, verb-object, w_{E2}) \succ 0, where\ w_{E2} = Tran1(w_{C2})\} \cup Tran(w_{C1})$$

To reduce the bad impact of the blind translation expansion of Model E, we try to assign lower probability to the verbs that are expanded out of the bilingual lexicon. We use the following method: the translations given by the bilingual lexicon share a probability of 0.6 and the other possible translations that are expanded using Model E share a probability of 0.4. Suppose P^* is the additionally assigned probability, and suppose there are m translations given by the bilingual lexicon and n translations expanded by model E. We have the following:

$P^* = \frac{0.6}{m}$	If the translation is obtained from the C-E lexicon
$P^* = \frac{0.4}{n}$	If the translation is obtained through expansion of Model E

Then *Model E* can be described as:

$$\begin{aligned} e_{\max} &= \arg\max_{e} P(e) \times Likelyhood(c \mid e) \\ &= \arg\max_{\substack{w_{E1} \in Tran2(w_{C1}) \\ w_{E2} \in Tran1(w_{C2})}} P(w_{E1}, verb-obj, w_{E2}) \times sim_{E \to C}(w_{C1}, w_{E1}) \times P^* \times sim_{E \to C}(w_{C2}, w_{E2}) \times P^* \end{aligned}$$

The evaluation results obtained using Case-I testing set are shown in Table 13. We can find that both Model D and Model E improved the translation precision. Model E is more powerful than Model D.

Table 13. *Evaluation on verbs of high frequency*

Model	**#Correct**	**Percentage**
Model D	526	71.8%
Model E	587	80.1%

Using Model C, "展开进攻" could not be translated correctly, while Model E correctly gave the answer "launch attack". In table 14 and Appendix III, there are more examples showing the cases in which Model E correctly selected translations. (The English translations marked with * are cases where the translations could not be found in the translation lexicon but were generated with Model E only.)

Table 14. *The translation result overcoming OOV*

展开进攻	launch* attack	打主意	make plan
采取行动	Take action	打基础	make foundation
采取办法	adopt* method	打球	play ball
看电视	watch television	打洞	make hole
看书	Read book	打折扣	offer* discount
看节目	See program	打锣	strike gong
打电报	send telegram	博取同情	evoke* sympathy

We also found that the translation performance was influenced by data sparseness of the triple database. Typically, when an English counterpart for a verb-object triple in Chinese could not be found, Model E will yielded 0 for $P(w_{E1}, verb-object, w_{E2})$. For example, "eat twisted crullers", which corresponds to "吃油条" did not appeared anywhere in the English triple set. This will generate very big influence. We shall tackle this problem in the future.

5. Related Works

The key to improving translation selection is to incorporate human translation knowledge into a computer system. One way is for translation experts to handcraft the translation selection knowledge in the form of selection rules and lexicon features. However, this method is time-consuming and cannot ensure high quality in a consistent way. Current commercial MT systems mainly rely on this method. Another way is to let the computer learn the translation selection knowledge automatically by using a large parallel text. A good survey on this research is that of McKeown & Radev [2000]. Some of the contents are quoted here in a condensed way. Smadja *et al.* [1996] created a system called Champolion, which is based on Smadja's collocation extractor, Xtract. Champollion uses a statistical method to translate both flexible and rigid collocations between English and French using the Canadian Hansard corpus. Champollion's output is a bilingual list of collocations ready for use in a machine translation system. Smadja et al. indicated that 78% of the French translations of valid English collocations were judged to be correct based on three evaluations by human experts. Kupiec [1993] described an algorithm for the translation of a specific kind of collocations, namely, noun phrases. An evaluation of his algorithm has shown that 90% of the 100 highest ranking correspondences are correct.

Selecting the right word translation is related to word sense disambiguation. Most of the research has reported on using supervised methods, which use sense-tagged corpora. Mooney [1996] gave a good quantitative comparison of various methods. Yarowsky [1995] reported an impressive unsupervised-learning result that trains decision lists for binary sense disambiguation. Schutze [1998] also proposed an unsupervised method, which in essence clusters usages of a word. However, although both Yarowsky and Schutze minimized the amount of supervision, their reported results only for very few examples.

Another related field is computer assisted bilingual lexicon (term) construction. A tool for semi-automatic translation of collocations, Termight, wa described by Dagan and Church [1994]. It can be used to aid translators in finding technical term correspondences in bilingual corpora. The method proposed by Dagan and Church uses extraction of noun phrases in English and word alignment to align the head and tail words of noun phrases with words in the other language. A word sequence of words corresponding to the head and tail is produced as the translation. Because it does not rely on statistical correlation metrics to identify the words of the translation, this method allows the identification of infrequent terms that would otherwise be missed owing to their low statistical significance. Fung [1995] used a pattern-matching algorithm to compile a lexicon of nouns and noun phrases between English and Chinese. Wu and Xia [1994] computed a bilingual Chinese-English lexicon. They used the EM algorithm to produce word alignment across parallel corpora and then applied various linguistic filtering techniques to improve the results.

Since large aligned bilingual corpora are hard to acquire due to copyright restrictions and construction expenses, some researchers have proposed methods which do not rely on parallel corpora. Tanaka and Iwasaki [1996] demonstrated how to use nonparallel corpora to choose the best translations among a small set of candidates. Fung [1997] used similarities in the collocates of a given word to find its translation in the other language. Fung [1998] also explored using an IR approach to get translations of new words using non-parallel but comparable corpora. Dagan and Itai [1994] use a second language monolingual corpus for word sense disambiguation. They used a target language model to find the correct word translations.

Most of the methods for statistical machine translation obtain word translation probability by learning from large parallel corpora [Brown *et al.*, 1993]. Very few researchers have tried to use monolingual corpora to train word translation probability. The most similar work to our approach is that of [Koehn and Knight. 2000]. Using two completely unrelated monolingual corpora and a bilingual lexicon, they constructed a word translation model for 3830 German and 6147 English noun tokens by estimating word translation probabilities using the EM algorithm. In their experiment, they assumed that the word sequence of English and German was the same, so that in the EM iteration step, the language model of the target language could be used. However, their model was only used to test the translation of nouns; they did not conduct experiments on verb translation. They also did not consider syntactic relations. In addition, it is hard to extend their model to other language pair like Chinese and English.

6. Conclusion

We have proposed a new statistical translation model. The unique characteristics of our model are:

1) The translation model is trained using two unrelated monolingual corpora. We have defined the cross- lingual word similarity, which enable us to compute the similarity between a source language word and a target language word with a simple bilingual lexicon, without using bilingual corpora.

2) The translation model is based on dependency triples, not on word level, which is typically used. It can overcome the long distance dependence problem to some extent. Since the translation of a word is often decided based on a syntactic member that may not be adjacent to the word, this method can hopefully improve translation precision compared with the existing word-based model.

3) Based on the new translation model, we have further proposed new models for tackling OOV issue. The experiments showed that Model E, which expands translations using an English triple database, is a promising model for solving the OOV issue. This is very promising too for the application of cross language information retrieval.

Our approach is completely unsupervised, so it is not necessary for the two corpora to be aligned in any way or to be tagged manually with any information. Such monolingual corpora are readily available for most languages, while parallel corpora rarely exist even for common language pairs. So our method can help overcome the bottleneck of acquiring large-scale parallel corpora. Since this method does not rely on specific dependency triples, it can be used to translate other types of triples such as adjective-noun, adverb-verb and verb-complement in the same way. In addition, our method can be used to build a collocation translation lexicon for an automatic translation system.

This triple based translation approach can be further extended to sentence level translation. Given a sentence, the main dependency triple can be extracted with a parser, and then each triple can be translated using our method. Then, for dependency triples which are specific to the source language, we can apply a rule-based approach. After all the main triples are correctly translated, a target language grammar can be introduced to realize target language generation. This hopefully will enable us to realize sentence skeleton translation system.

There are some interesting topics for future research. First, since we use parsers which inevitably introduce some parsing mistakes into the generated dependency triple databases, we need to find an effective way to filter out mistakes and perform necessary automatic correction. Second, we need to find a more precise translation expansion method to overcome the OOV issue which is caused by the limited coverage of the lexicon. For instance, we can try using translation expansion by employing a thesaurus that is trained automatically with a large corpus or employ a pre-defined thesaurus like WORDNET. Third, triple data sparseness is a big problem; to solve it, we need to apply some approaches used in statistical language models, such as smoothing methods and the class based models.

References

Brown P.F., Stephen A. Della Pietra, Vincent J. Della Pietra, Robert L. Mercer, "The mathematics of machine translation: parameter Estimation". *Computational Linguistics*, 19(2), 1993, pp. 263-311.

Tanaka K, H Iwasaki, "Extraction of lexical translation from nonaligned corpora." *COLING-96*: *The 16th International Conference on Computational Linguistics*, Copenhagen, Denmark, 1996, pp. 580-585.

Dagan I, K Church, "TERMIGHT: identifying and translating technical terminology". 4th *Conference on Applied Natural Language Processing*, Stuttgart, Germany, 1994, pp. 34-40.

Dagan I, Itai, A, "Word sense disambiguation using a second language monolingual corpus", *Computational Linguistics*, 20(4), 1994, pp. 563-596.

Fung P, " A pattern matching method for finding noun and proper noun translations from noisy parallel corpora", *33rd Annual Conference of the Association for Computational Linguistics*, Cambridge, MA, 1995, pp. 236-233.

Fung P, "Using word signature features for terminology translation from large corpora". Ph.D dissertation, Columbia University, 1997, New York.

Fung P and LO Yuen Yee, "An IR approach for translating new words from nonparallel, comparable Texts". *The 36th Annual Conference of the Association for Computational Linguistics*, Montreal, Canada, August 1998, pp. 414—420.

Koehn. P and K. Knight, "Estimating word Translation probabilities from unrelated monolingual corpora using the EM Algorithm", *National Conference on Artificial Intelligence (AAAI)*, 2000, Austin, Texas.

Lin D., "Principle-based parsing without over-generation", *Proceedings of ACL-93*, 1993, pp 112-120, Columbus, Ohio.

Lin D., "Principar-an efficient, broad-coverage, principle-based parser". *Proceedings of COLING-94*, pp. 482-488, Kyoto, Japan, 1994.

Lin D..1998a, "Extracting collocations from test corpora", *First Workshop on Computational Terminology*, Montreal, Canada, 1998.

Lin D., 1998b, "Automatic retrieval and clustering of similar words", *COLING-ACL98,* Montreal, Canada, 1998.

Mckeown, K R, D R Radev, "Collocations", *Handbook of Natural Language Processing*, pp507-523, Edited by Robert Dale, Hermann Moisl, Harold Somers, 2000.

Mooney, R., "Comparative experiments on disambiguation word senses: An illustration of bias in machine learning", *In Proceedings of the Conference on Empirical Methods in Natural Language Processing*, *EMNLP, 1996*.

Smadja F., K. R. Mckeown, V Hatzivassiloglou, "Translation collocations for bilingual lexicons: a statistical approach". *Computational Linguistics*, 22:1-38, 1996.

Schutze, H. "Automatic word sense disambiguation rivaling supervised methods", *Computational Linguistics, 24(1):97-123, 1998.*

Wu D., X. Xia, "Learning an English-Chinese lexicon from a parallel corpus", Technology partnerships for Crossing the Language Barrier: *Proceedings of the First Conference of the Association for Machine Translation in the Americas*, Columbia, MD, pp206-213, 1994.

Yarowsky, D. "Unsupervised word sense disambiguation rivaling supervised methods". *In Proceedings of ACL- 33*, pp. 189-196, 1995.

Zhou M., "A Block-Based Robust Dependency Parser for Unrestricted Chinese Text", *2nd workshop on Chinese language processing*, Hong Kong, 2000.

Appendix I High frequency verb list

Frequency	Word	Frequency	Word	Frequency	Word	Frequency	Word
899835	说	380677	来	322078	用	283612	去
211744	看	199602	作	181205	做	175761	想
175658	出	173802	要	129595	占	124164	上
112368	走	111260	问	92357	打	91020	叫
89115	开	84744	吃	83394	下	81221	搞
75946	讲	75753	办	73911	送	68651	找
68639	发	67103	抓	65796	听	64017	买
63468	住	62936	入	61695	拉	61695	订
384590	进行	362678	发展	228207	举行	223702	参加
214557	通过	204081	加强	195157	提出	172647	解决
151354	组织	133191	采取	126557	开展	110076	发挥
103009	达到	99867	完成	91401	介绍	68801	扩大
68588	计划	67446	引起	60426	恢复	60237	减少
60087	制定						

Appendix II Low frequency verb list

Frequency	Word	Frequency	Word	Frequency	Word	Frequency	Word
2108	践踏	2087	施加	2056	逼近	1555	调配
1549	共享	1498	扣押	1420	反驳	1402	高唱
1389	迷惑	1368	窃	460	遨游	458	规劝
457	胁迫	439	修剪	438	抄袭	304	驯服
294	调遣	278	描摹	270	剽窃	262	吸吮
158	赎回	156	暗藏	153	博取		

Appendix III Some translation results obtained with model E

√	打|锣→strike|gong	√	订|约会→order|appointment	√	做|翻译→make|translation
×	打|鼓→have|drummer	√	订|条约→sign|pact	×	做|演员→do|actor
√	打|钟→play|bell	√	订|计划→make|plan	×	做|保姆→get|housekeeper
√	打|铃→play|bell	√	订|措施→order|measure	×	做|教师→give|teacher
√	打|铁→produce|iron	×	订|日期→order|date	×	做|厨房→do|kitchen
√	打|人→beat|person	√	订|指标→order|target	√	做|纸→make|paper
√	打|仗→do|fight	×	订|制度→order|system	√	看|电影→see|film
×	打|架→buy|shelf	√	订|合同→sign|contract	√	看|电视→watch|television
√	打|脸→beat|face	√	订|契约→sign|charter	√	看|京剧→watch|Bejing opera
×	打|手→play|hand	√	订|公约→sign|pact	√	看|展览→see|exhibition
√	打|头→strike|head	√	订|条件→order|condition	√	看|人→see|person

√	打|枪→fire|gun	√	订|同盟→form|alliance	√	看|书→read|book
√	打|炮→use|cannon	×	订|婚→attend|wedding	√	看|报→read|newspaper
√	打|雷→bring|thunder	√	订|书→order|book	√	看|小说→read|novel
√	打|信号→send|signal	√	订|报→order|newspaper	√	看|文件→see|document
√	打|电话→make|telephone	√	订|杂志→order|magazine	√	看|朋友→see|friend
×	打|靶→hit|target	√	订|票→order|ticket	√	看|学生→see|student
×	打|气→strike|air	√	订|机器→order|machine	×	看|眼睛→see|eye
×	打|针→share|needle	√	订|货→order|goods	√	看|问题→see|problem
√	打|鸟→catch|bird	×	订|本子→carry|notebook	√	看|现象→see|phenomenon
√	打|鱼→catch|fish	×	订|报纸→publish|newspaper	√	看|脸色→see|expression
×	打|老虎→buy|tiger	√	作|打算→make|plan	√	看|本质→see|nature
√	打|蜡→strip|wax	√	作|结论→make|conclusion	×	出|大门→put forth|front door
√	打|草稿→make|draft	√	作|报告→write|report	×	出|国→produce|country
√	打|基础→make|foundation	×	作|斗争→have|struggle	√	出|院→leave|yard
√	打|主意→catch|decision	√	作|曲→write|melody	×	出|城→issue|city
×	打|算盘→work out|abacus	√	作|诗→write|poem	√	出|海→go|sea
×	打|伞→buy|umbrella	√	作|文章→write|article	√	出|境→leave|state
×	打|旗子→play|banner	√	做|鞋→make|shoes	×	出|洞→fill|cavity
×	打|灯笼→sell|lantern	√	做|衣服→make|clothes	×	出|厂→include|works
×	打|饭→ work out|cooked rice	√	做|裤子→make|trousers	×	出|站→make|stop
√	打|酒→buy|wine	√	做|活→do|work	×	出|场→issue|place
√	打|酱油→buy|soy	√	做|菜→make|food	×	出|血→produce|blood
√	打|票→buy|ticke	×	做|饭→make|cooked-rice	×	出|轨→build|rail
×	打|醋→prefer|vinegar	√	做|面包→make|bread	√	出|界→exceed|limit
√	打|柴→collect|firewood	√	做|点心→ make|refreshments	√	出|格→exceed|standard
√	打|草→pack|straw	√	做|工→do|work	√	出|范围→exceed|scope
×	打|麦子→buy|wheat	×	做|沙发→sit|sofa	√	出|主意→produce|idea
√	打|粮食→collect|grain	√	做|生意→make|trade	√	出|题目→issue|subject
√	打|牌→play|cards	√	做|买卖→do|business	√	出|证明→produce|proof
×	打|拳→make|fist	√	做|工作→do|work	×	出|力→produce|power
√	打|哈欠→draw|yawn	√	做|试验→do|test	×	出|钱→issue|money
√	打|盹→have|doze	√	做|事情→do|business	√	出|广告→ produce|advertisement
×	打|冷战→work out|cold war	√	做|功课→do|homework	√	出|劳动力→put forth|labour

√	打|官司→fight|lawsuit	√	做|作业→do|homework	√	出|通知→issue|notice
√	打|井→dig|well	√	做|练习→do|exercise	√	出|节目→produce|program
√	打|洞→make|hole	√	做|学生→become|student	√	出|榜→issue|announcement
√	打|包裹→work out|parcle	×	做|老师→give|teacher	√	出|煤→produce|coal
√	打|行李→pack|luggage	×	做|父亲→do|father	√	出|棉花→produce|cotton
×	打|毛衣→ work out|woolen clothes	√	做|主席→become|chairman	√	出|花生→produce|peanut
√	打|比方→use|analogy	×	做|官→ make|government offcial	√	出|英雄→become|hero

*The √ means correct translation or sometimes acceptable translation, while × means wrong translation.