An Approach to Automatic Acquisition of Translation Templates Based on Phrase Structure Extraction and Alignment

Rile Hu, Chengqing Zong, *Associate Member, IEEE*, and Bo Xu

Abstract—**In this paper, we propose a new approach for automatically acquiring translation templates from unannotated bilingual spoken language corpora. Two basic algorithms are adopted: a grammar induction algorithm, and an alignment algorithm using bracketing transduction grammar. The approach is unsupervised, statistical, and data-driven, and employs no parsing procedure. The acquisition procedure consists of two steps. First, semantic groups and phrase structure groups are extracted from both the source language and the target language. Second, an alignment algorithm based on bracketing transduction grammar aligns the phrase structure groups. The aligned phrase structure groups are post-processed, yielding translation templates. Preliminary experimental results show that the algorithm is effective.**

Index Terms—**Bilingual grammar induction, machine translation, structure alignment, translation template acquisition.**

I. Introduction

WITH THE development of corpus processing technology, more and more bilingual corpora are becoming available for knowledge acquisition in machine translation and many other natural language processing tasks. Translation templates provide one especially useful kind of knowledge for machine translation systems. At the same time, phrasal translation examples are an essential resource for many MT and machine-assisted translation architectures. In this paper, we bring together this need and this resource. We present a new approach for acquiring translation templates automatically from a sentence-aligned parallel English–Chinese corpus through phrase structure extraction and alignment.

In some example-based machine translation systems, the translation templates are extracted manually from the corpus. For example, [10] manually encodes translation rules in this way. Similarly, [18] has also proposed an example-based system which employs manually-built matching expressions as the translation templates. However, as the size of corpus grows larger, the manual template extraction becomes increasingly difficult and error-prone.

Some methods for automatically acquiring translation templates have also been proposed. For instance, in [3], [16], and [1], the analogical models are adopted for learning translation templates from bilingual parallel corpus. Templates are obtained by grouping the similar translation examples and replacing their differences with variables. However, such methods need a very large bilingual parallel corpus which contains many similar sentences. By contrast, other methods for template acquisition are instead based on structure alignment, e.g. those of [5], [9], [20]. These approaches follow a procedure which may be termed "parse-parse-match" [23]. In these methods, each language of the parallel corpus is first parsed separately using monolingual grammars, and then the corresponding constituents are matched using some heuristic procedures. Such methods, however, need two high-performance parsers, one for the source language and one for the target language. In a similar vein, [8] has proposed a method based on bilingual language modeling: bilingual sentence pairs are first aligned with respect to syntactic structure by combining a parser with a statistical bilingual language model. The alignment results are then used to extract translation templates. This method, too, needs a high-performance parser. It also requires the part-of-speech tagging systems for both the source and the target language.

And some other statistical methods are also proposed to perform the task of translation template acquisition [23] introduced the bracketing transduction grammar (BTG). It uses no language specific syntactic grammar, and employs a maximum-likelihood parser to select the parse tree that best satisfies the combined lexical translation preference. This method achieves encourage results for bilingual bracketing using a word-translation lexicon alone [13], [14] proposed the alignment template translation model. It explicitly takes shallow phrase structures into account, using two different alignment levels: a phrase level alignment between phrases and a word level alignment between single words. This method can learn fully automatically by using a bilingual training corpus and are capable of achieving better translation results on a limited-domain task than other example-based or rule-based translation systems. And [24] presented an integrated phrase segmentation and alignment algorithm, which segments the sentences into phrases and finds their alignments simultaneously without building an initial word-to-word alignment.

Manuscript received June 18, 2004; revised May 16, 2005. This work was supported by the National Natural Science Foundation of China under Grants 60175012, 60272041, and 60121302, the Outstanding Overseas Chinese Scholars Fund of Chinese Academy of Sciences under Grant 2003-1-1, and by PRA project under Grant SI02-05. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yuqing (G. E.) Gao

R. Hu was with National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China. He is now with Nokia (China) Research Center, Beijing 100013, China (e-mail: ext-rile.hu@nokia.com).

C. Zong and B. Xu are with National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China (e-mail: cqzong@nlpr.ia.ac.cn; xubo@nlpr.ia.ac.cn).

Digital Object Identifier 10.1109/TSA.2005.858521

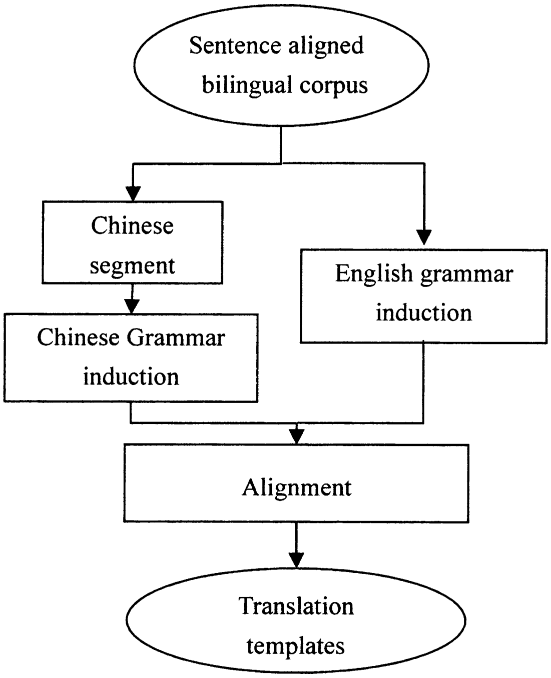

Sentence aligned bilingual corpus
Chinese segment
English grammar induction
Chinese Grammar induction
Alignment
Translation templates

Fig. 1. Architecture of the proposed translation template acquisition system.

In this paper, we propose a statistical, data-driven approach which acquires translation templates from unannotated bilingual corpora based on the bilingual grammar induction and BTG. The remainder of the paper is organized as follows. In Section II, our motivations are introduced in detail. In Section III, we survey the basic algorithms for both grammar induction and alignment using BTG. In Section IV, the experimental results and analysis are shown. Finally, some concluding remarks are given in Section V.

II. Our Motivations

The translation template acquisition based on structural alignment is a popular method in the area of statistical machine translation. Considerable research has been carried out on this topic. We focus here on the methods based on unsupervised machine learning; propose a translation template acquisition method based on statistical phrase structure extraction and alignment.

The main ideas of our approach to translation template acquisition are shown in Fig. 1.

The input of our approach is the sentence-aligned bilingual corpus. Here, an English-Chinese bilingual corpus is used. The Chinese sentences are first segmented, then the grammar induction procedure is performed on both English and Chinese sentences. Next, the semantic groups (labeled SCi) and phrasal groups (labeled PCi) are obtained from the corpus for both English and Chinese. Finally, the phrase structures of the languages are aligned, using a modified BTG. These aligned phrase structures are post-processed to create the translation templates, which are the results of our approach.

We now give a simple example to explain how the translation templates are acquired from the unannotated corpora.

Suppose some SCi and PCi groups are obtained from the corpus, as shown in Table I.

TABLE I
Examples of the Grammars Acquired From the Experimental Corpus

Chinese part:
SCC10 → 单人间 | 双人间 | 标准间
PCC3 → 一个
PCC8 → PCC3 SCC10
PCC12 → 我 想 预订
PCC20 → PCC12 PCC8
English part:
SCE5 → single | double | standard
PCE2 → want to
PCE4 → a SCE5 room
PCE8 → I PCE2 reserve
PCE14 → PCE8 PCE4

With the grammars shown in Table I, the system aligns the phrase structures as follows:

[[I/我 [want/想 to/ε] reserve/预订] [a/一 ε/个 N* room/N]].

where to/ε means the Chinese word aligned with the word "to" is null, and

N=单人间 ⇔ N^*=single;
N=双人间 ⇔ N^*=double;
N=标准间 ⇔ N^*=standard.

Thus, we can obtain the following translation templates:

我 想 预订 ⇔ I want to reserve ①
一个N ⇔ a N^* room ②

in ②

N=单人间 ⇔ N^*=single;
N=双人间 ⇔ N^*=double;
N=标准间 ⇔ N^*=standard

Here, ① and ② exemplify two kinds of translation templates in our approach. ① is a *constant template*, since all of its elements are constants. By contrast, ② contains at least one variable element, so we call such kind of temples *the variable template*.

III. Basic Algorithms

In this section, we provide a brief overview of our basic algorithms for grammar induction and alignment using BTG.

A. Grammar Induction Algorithm

This clustering method consists of two steps, spatial clustering and temporal clustering. In the clustering procedure, we consider entities as processing unit. The entities include single words and the semantic group labels, the phrasal structure group labels got from the procedure of clustering. For example, in Table I, the single word "single," "want," and etc.

are single-word-level entities; the semantic group labels SCE5 and SCC10 are the semantic-group-level entities, each entity contains a group of single words; and the phrasal structure group labels PCC8, PCE14 and etc. are the phrasal- structure-group-level entities, each entity contains a sequence of words or a sequence of entities. In the spatial clustering step, the entities which have similar left and right contexts are grouped together. These entities generally have similar semantics. In the temporal clustering step, the entities which frequently co-occur are clustered into groups. These entity groups tend to be commonly-used phrases.

In spatial clustering, the Kullback-Leibler distance is used to describe the similarity of the distributions of the local contexts of entities, where an entity's local context consists of the entity immediately before it and the entity immediately after it (1)

$$D(p_1 \| p_2) = \sum_{w_i \in V} p_1(w_i) \log \frac{p_1(w_i)}{p_2(w_i)}. \tag{1}$$

Here, p_1 denotes the unigram distribution of the words which appear in the local context of the entity e_1, p_2 denotes the same distribution for entity e_2, and w_i denotes the word which appears in the local contexts of the entities e_1 and that of e_2, and V denotes the union of w_i.

In order to acquire a symmetric measure of the distance, or degree of difference, between two local context distributions, we use the *divergence* of the distributions, as shown in (2)

$$Div(p_1, p_2) = D(p_1 \| p_2) + D(p_2 \| p_1). \tag{2}$$

Then, the distance between two entities e_1 and e_2 is defined as (3):

$$Dist(e_1, e_2) = Div\left(p_1\text{left}, p_2\text{left}\right) + Div\left(p_1^{\text{right}}, p_2^{\text{right}}\right). \tag{3}$$

Distance between entities is thus the sum of the divergences of the distributions of the entities' left and right contexts.

In order to increase the clustering accuracy, we introduce the *extended distance contexts* into the measurement of distance between entities: we consider the words next to the entities' contexts, called *extended contexts*. The Kullback-Leibler distance of extended contexts is calculated as the similarity of their distributions, using (1). The symmetric distance between words can also be calculated using (2).

Finally, the distance between two entities is computed as the sum of the distance of the contexts and that of the extended contexts. Thus the distance between entities e_1 and e_2 can be described using (4)

$$\begin{aligned} Dist^*(e_1, e_2) = {} & Div\left(p_1^{\text{left}}, p_2^{\text{left}}\right) + \frac{1}{2} Div_2\left(p_1^{\text{left}}, p_2^{\text{left}}\right) \\ & + Div\left(p_1^{\text{right}}, p_2^{\text{right}}\right) + \frac{1}{2} Div_2\left(p_1^{\text{right}}, p_2^{\text{right}}\right). \end{aligned} \tag{4}$$

Here, the expression $Div_2(p_1, p_2)$ denotes the symmetric distance of the extended contexts of the two entities e_1 and e_2.

The maximally similar entities are gathered into a semantic group, labeled SCi. That is, we cluster the pairs of entities which have the minimal distance between them (as calculated by (4)).

Other measures which can be used to calculate the similarity between two entities have also been considered. We use feature vectors to describe the contexts of an entity, and these can be used to calculate the similarity between two entities. If an entity e appears in the context of another given entity, this relationship can be described using the expression (*posi*, e), where *posi* has the value *left* if e appears to the left side of the entity, or *right* if e appears to the entity's right. The value of each feature is the frequency count of the feature in the corpus.

$(u_1, u_2, \ldots, u_n)$ and $(v_1, v_2, \ldots, v_n)$ denote the feature vectors for the entity u and v, n is the number of feature types extracted from the corpus, and $f(i)$ is the ith feature.

Three other similarity measures are also used in the spatial clustering step, the Cosine Measure, Cosine of Pointwise Mutual Information, and Dice Co-efficient.

The Cosine Measure computes the cosine of two entities' feature vectors (5)

$$Cos(u, v) = \frac{\sum_{i=1}^{n} u_i \times v_i}{\sqrt{\sum_{i=1}^{n} u_i^2} \times \sqrt{\sum_{i=1}^{n} v_i^2}}. \tag{5}$$

The pointwise mutual information (PMI) between a feature $f(i)$ and an entity u measures the strength of the association between them, as defined in (6)

$$PMI\left(f(i), u\right) = \log\left(\frac{P\left(f(i), u\right)}{P\left(f(i)\right) \times P(u)}\right). \tag{6}$$

Here, $P(f(i), u)$ is the probability of $f(i)$ co-occurring with u; $P(f(i))$ is the probability of $f(i)$ co-occurring with any entity; and $P(u)$ is the probability of any feature co-occurring with u. For example, if all the features occur 1,000 times in the corpus, $f(i)$ occurs 50 times, and $f(i)$ co-occurs with u for 10 times, any feature co-occurs with u for 100 times, then $P(f(i), u) = 10/1000 = 0.01$, $P(f(i)) = 50/1000 = 0.05$, $P(u) = 100/1000 = 0.1$.

The *Cosine of Pointwise Mutual Information* (CosPMI) is defined in (7)

$$\begin{aligned} & CosPMI(u, v) \\ & = \frac{\sum_{i=1}^{n} PMI\left(f(i), u\right) \times PMI\left(f(i), v\right)}{\sqrt{\sum_{i=1}^{n} PMI\left(f(i), u\right)^2} \times \sqrt{\sum_{i=1}^{n} PMI\left(f(i), v\right)^2}}. \end{aligned} \tag{7}$$

This formula computes the cosine between two entities' pointwise mutual information.

The Dice Co-efficient is defined in (8). It is a simple measure of the difference between zero and nonzero frequency counts

$$Dice(u, v) = \frac{2 \times \sum_{i=1}^{n} s(u_i) \times s(v_i)}{\sum_{i=1}^{n} s(u_i) + \sum_{i=1}^{n} s(v_i)}. \tag{8}$$

Here, $s(x) = 1$ if $x > 0$ and $s(x) = 0$ otherwise.

After the spatial clustering, we substitute a category label throughout the corpus for the words that have been grouped. Then the temporal clustering is computed.

In the temporal clustering step, the Mutual Information (MI) is used to describe the degree of co-occurrence of two entities e_1 and e_2 in the same sentence of the corpus, and it becomes the metric used for clustering. MI is defined in (9)

$$MI(e_1, e_2) = P(e_1, e_2) \log \frac{P(e_2|e_1)}{P(e_2)}. \tag{9}$$

The entities which have the highest MI are clustered into phrasal groups labeled PCi. Next, PC labels are substituted for these entity pairs. Then another iteration of spatial clustering can be started. After application of the clustering algorithm, the semantic groups and phrase structure groups will be extracted from the corpus.

After each iteration of the clustering algorithm, more words are clustered into semantic groups and phrasal structure groups. The coverage of the clustering algorithm can be measured in terms of the percentage of words in the input corpus that are captured in the clustering groups. A stopping criterion (STC) is defined as the relative increment of the clustering coverage. For example, if the coverage after iteration is 80% and that of next iteration is 82%, then the STC between these two iterations is $(82 - 80)/80 = 3.75\%$. When the STC is below 1%, the clustering algorithm will be stopped.

We now describe our grammar induction approach. Importantly, it can capture semantic and phrase structures from unannotated corpora.

The grammar induction algorithm is described in Fig. 2.

The input of the algorithm is the English part or the Chinese part of the bilingual corpus.

Step 1: If the distance measure is used, for each two entities e_1 and e_2 in the corpus, calculate the distance between them using (4). If other similarity measures are used, for each two entities e_1 and e_2 in the corpus, calculate the similarity between them using (5) or (7) or (8).

Step 2: Group the N pairs which have the minimum distance or the maximum similarity into a semantic class.

Step 3: Replace the entities in Step2 with their semantic class label SCi.

Step 4: For each two entities e_1 and e_2 in the corpus, calculate the MI between them using (9).

Step 5: Select the N pairs of entities with the highest MI to form the phrasal structure groups.

Step 6: Replace the entities in Step5 with their phrasal structure class label PCi.

Step 7: Calculate STC. If STC is lower than 1%, stop the procedure of the clustering algorithm, else go to Step 1.

The Output of the algorithm is a list of semantic groups and phrasal structure groups.

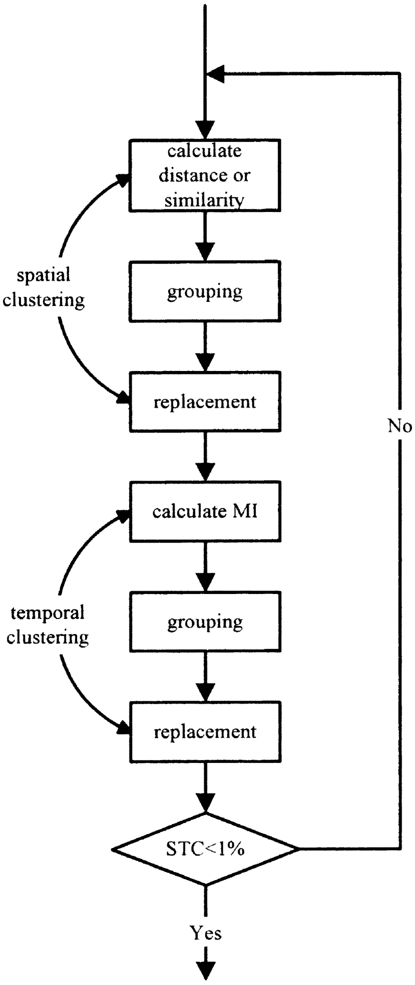


Fig. 2. Flow chart of the grammar induction algorithm.

B. Alignment Using Bracketing Transduction Grammar

A bilingual model called inversion transduction grammar (ITG), proposed by Wu [23], parses bilingual sentence pairs simultaneously. As it is difficult to find suitable bilingual syntactic grammars for English and Chinese, we employ a simplified ITG called BTG [22]. A BTG contains only one nonterminal symbol A. This symbol can be rewritten either recursively as a pair of A's or as a single terminal pair

$$A \xrightarrow{a} [A\,A]$$
$$A \xrightarrow{a} \langle A\,A \rangle$$
$$A \xrightarrow{b_{ij}} u_i/v_j \quad \text{for all English} - \text{Chinese lexical translations } i, j.$$
$$A \xrightarrow{b_{i\varepsilon}} u_i/\varepsilon \quad \text{for all English vocabulary items } i.$$
$$A \xrightarrow{b_{\varepsilon j}} \varepsilon/v_j \quad \text{for all Chinese vocabulary items } j.$$

Here, the lower-case a denotes the probability of the syntactic rules. It has no practical effect $b_{i\varepsilon}$, and $b_{\varepsilon j}$ can be chosen to be unrelated to the probabilities b_{ij} of lexical translations pairs. At present, it is set to an arbitrary constant. The operator "[ ]" is used to represent the concatenation when both languages use the same ordering. The operator "$\langle\,\rangle$" is used when language one concatenates in the order shown, but language two concatenates in the reverse order. b_{ij} represents the probability that the source language word u_i is translated by the target language word v_j, as trained by the Expectation–Maximization word-translation algorithm. The last two singleton productions denote that, in the parallel sentences, the word in one language has no counterpart in the other language. A small constant can be chosen for the probabilities $b_{i\varepsilon}$ and $b_{\varepsilon j}$ in these productions, so that the optimal bracketing falls back on these productions only when it is otherwise impossible to match the relevant singletons.

The expressive characteristics of ITG grammars naturally constrain the space of possible matching in a highly appropriate fashion. As a result, BTG grammars achieve encouraging results for bilingual bracketing using a word-translation lexicon alone [23]. However, since no language specific syntactic knowledge is used in BTG's, the grammaticality of the output can not be guaranteed [8].

Our main idea in the present work is to use phrase structure information acquired by the grammar induction algorithm as a boundary restriction in the BTG language model. When the constraint is in compatible with BTG, BTG is used as the default result. This procedure makes the alignment going on regardless of some failure in the matching process. Then a dynamic programming algorithm is used to compute the maximally probable alignment of all possible phrase structures.

A constraint heuristic function $F_e(s,t)$ is defined to denote the English boundary constraint. Here, s denotes the beginning position of the phrase structure and t denotes its end position. Phrase structure matching can yield three cases: invalid match, exact match, and inside match. An invalid match occurs when the alignment conflicts with phrasal boundaries. Examples appear in (1,2), (3,4) and (4,5) etc. in the sample sentence below. (The constraint function is set at a minimum value 0.0001 to prevent selection of such matches when an alternate match is available.) An exact match means that the match falls exactly on the phrase boundaries, as in (2,3), (1,4) and (5,7) below. (When this condition is met, the function is set at a high value 10 for weighting.) Examples of inside matches are seen in (5,6) and (6,7) below. (The value of these functions is set to 1.)

Example:
[[I/1 [want/2 to/3] reserve/4][a/5 single/6 room/7]].

The Chinese constraint function $F_c(u,v)$ is defined similarly.

Now let the input English sentence be $e_1,\ldots,e_T$ and let the corresponding Chinese sentence be $c_1,\ldots,c_V$. As an abbreviation, we write $e_{s\ldots t}$ for the sequence of English words e_{s+1}, $e_{s+2},\ldots,e_t$; Similarly, we write $c_{u\ldots v}$ for the Chinese word sequence. Further, the expression $q=(s,t,u,v)$ identifies all possible matched structures, where the substrings $e_{s\ldots t}$ and $c_{u\ldots v}$ both derive from the node q. The local optimization function is shown in (10)

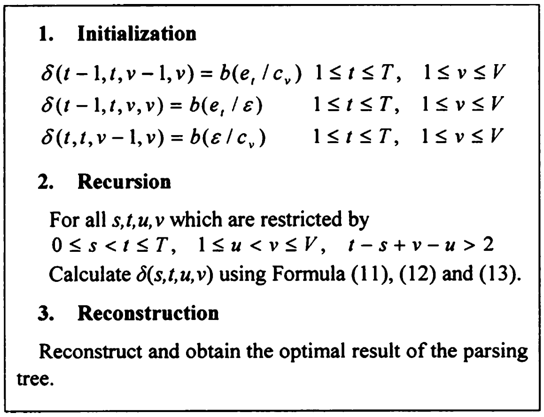
$$\delta(s,t,u,v) = maxP[q]. \tag{10}$$

Equation (10) denotes the maximally probable alignment of the phrase structures. Then the best combination of the phrase structures has the probability $\delta(0,T,0,V)$.

To insert the English and Chinese constraints into the alignment procedure, we integrate the constraint functions $F_e(s,t)$ and $F_c(u,v)$ into the local optimization function. For this purpose, the function is split into three functions, as in (11)–(13)

$$\delta(s,t,u,v) = max\left[\delta^{[]}(s,t,u,v), \delta^{\langle\rangle}(s,t,u,v)\right] \tag{11}$$

TABLE II
ALIGNMENT ALGORITHM



1. Initialization

$\delta(t-1,t,v-1,v) = b(e_t/c_v)$ $1\le t\le T,$ $1\le v\le V$
$\delta(t-1,t,v,v) = b(e_t/\varepsilon)$ $1\le t\le T,$ $1\le v\le V$
$\delta(t,t,v-1,v) = b(\varepsilon/c_v)$ $1\le t\le T,$ $1\le v\le V$

2. Recursion

For all s,t,u,v which are restricted by
$0\le s<t\le T,$ $1\le u<v\le V,$ $t-s+v-u>2$
Calculate $\delta(s,t,u,v)$ using Formula (11), (12) and (13).

3. Reconstruction

Reconstruct and obtain the optimal result of the parsing tree.

$$\delta^{[]}(s,t,u,v) = \max_{\substack{s\le S\le t\\ u\le U\le v\\ (S-s)(t-S)+(U-u)(v-U)\neq 0}} F_e(s,t)F_c(u,v)\delta_1\delta_2 \tag{12}$$

$$\delta^{\langle\rangle}(s,t,u,v) = \max_{\substack{s\le S<t\\ u\le U\le v\\ (S-s)(t-S)+(U-u)(v-U)\neq 0}} F_e(s,t)F_c(u,v)\delta_3\delta_4. \tag{13}$$

Here,

$$\begin{aligned}\delta_1 &= \delta(s,S,u,U)\\ \delta_2 &= \delta(S,t,U,v)\\ \delta_3 &= \delta(s,S,U,v)\\ \delta_4 &= \delta(S,t,u,U).\end{aligned}$$

In (12) and (13), the condition $(S-s)(t-S)+(U-u)(v-U)\neq 0$ specifies that only one of the language strings, not both, may be split into an empty string.

Other symbols in the algorithm are defined as follows: $\theta(s,t,u,v)$, $\sigma(s,t,u,v)$ and $\gamma(s,t,u,v)$ are the variables used to record the production direction, the spilt points in English, and the split points in Chinese, when $\delta(s,t,u,v)$ is achieved. These variables are used to reconstruct the bilingual alignment tree in the final step. $\lambda(s,t,u,v) = \lambda(q)$ is the nonterminal label of the node q. $LEFT(q)$ is the left side of q, and $RIGHT(q)$ is its right side.

The optimal bilingual parsing tree for a given sentence-pair is then computed using the dynamic programming (DP) algorithm [23] shown in Table II.

IV. EXPERIMENT AND DISCUSSION

A. Experimental Corpus

Our experiments employed an English-Chinese parallel spoken language corpus, collected in the travel information

TABLE III
EXPERIMENTAL RESULT COMPARE WITH BTG

Experiment	*P*(%)	*R*(%)	*F*(%)
Only BTG	63.58	69.70	66.50
Parse-parse-match	76.58	79.09	77.81
Our framework	76.77	80.83	78.75

domain, and consisting of 2,950 bilingual parallel utterances. The Chinese vocabulary size of the corpus is 989, and the English size is 1074. The average length of sentences is 7.8 words for Chinese, and 6.5 words for English.

B. Experimental Results

We define the $\mathrm{precision}(P)$, $\mathrm{recall}(R)$ and $\mathrm{F-measure}(F)$ of the results as follows:

$$P = \frac{Nr}{N} \times 100\% \tag{14}$$

$$R = \frac{N_r}{N_a} \times 100\% \tag{15}$$

$$F = \frac{2 \times P \times R}{P + R} \times 100\%. \tag{16}$$

Here, Nr denotes the number of correct translation templates acquired by our algorithm (judged manually); N denotes the total number of the translation templates acquired by the algorithm, and N_a denotes the number of the translation templates which we manually extracted from the whole corpus.

The manual work done by us follow the rules as follows.

1) The templates extracted must be grammatical on both side of Chinese and English.
2) Both sides of the templates must be integral. The commonly used phrase structures can not be split into several parts. These commonly used phrases are included in the phrase dictionary we collected and compiled, which contains 340 000 phrases in it.
3) The alignment of the phrase structures must be correct.

BTG grammars can be used without our framework to segment and align bilingual corpora at the phrasal level. Accordingly, as a control, we performed the experiment using BTG only for translation template acquisition. Then, for comparison, we performed it with our framework on the same task.

We also performed the same task followed the procedure which is called "parse-parse-match." The English sentences were parsed by the Stanford parser [6]. The Chinese parser we used here is developed by our research group [7]. The parsing results have not been manually revised.

The experimental results are shown in Table III.

To determine the effects of different similarity measures, we also carried out the experiments using: 1) the Distance measure in (4); 2) the Cosine Measure; 3) the Cosine of Pointwise Mutual Information; and 4) the Dice Co-efficient. The numbers of clustering semantic groups and phrasal structure groups are shown in Table IV.

The results of precision of the clustering got by different similarity measurement are shown in Fig. 3.

TABLE IV
NUMBERS OF CLUSTERING GROUPS FOR DIFFERENT SIMILARITY MEASURES

Similarity measure	Number of the groups (E)	Number of the groups (C)
*Dist**	5,211	5,098
Cosine Measure	5,138	5,002
Cosine of Pointwise Mutual Information	5,204	5,074
Dice Co-efficient	5,176	5,022

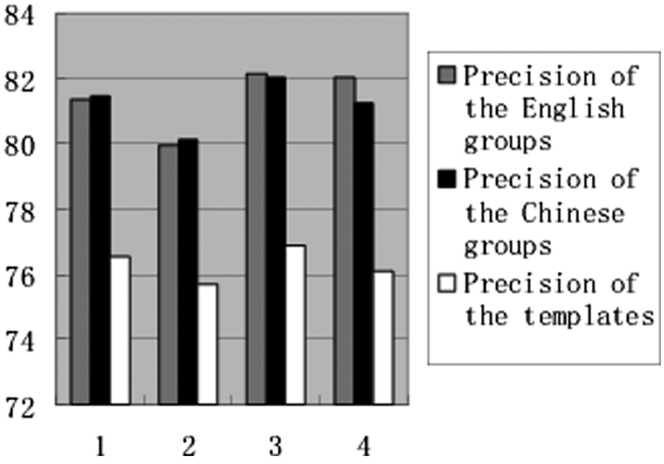


Fig. 3. Experimental results of the precision for different similarity measures.

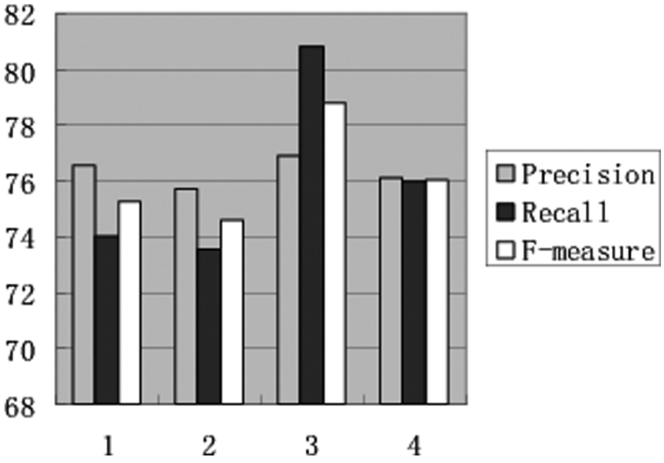


Fig. 4. Experimental results of the precision, recall, and F-measure for different similarity measures.

And the precision, recall and F-measure of the output templates got by different similarity measurement are shown in Fig. 4.

C. Analysis of the Experimental Results

Table III shows that the results when using our framework for translation template acquisition are much better than those when using only BTG to segment and align phrase structures. And the results using our framework are also comparable with that of the procedure which is called "parse-parse-match". Since our framework need fewer resources, it is clearly helpful to use our framework following bilingual grammar induction and phrasal alignment.

Table IV shows the numbers of the clustering groups got by the different similarity measurement.

Fig. 3 shows that the different similarity measures yield different performance results for translation template acquisition. And it also shows that the higher the performance of the clustering is, the higher that of the templates acquisition will be.

Fig. 4 shows that using the Cosine of Pointwise Mutual Information as a similarity measure gives the highest performance.

Two kinds of errors appear in the experimental results. First, some errors occur in the grammar induction step: because the induction algorithm does not adequately use the information contained in the corpus, unrelated entities are sometimes clustered into a single group. The second sort of errors occurs in the alignment step when idiomatic translations are compared.

According to the errors in the experiments, in the future research, we plan to use a synonym dictionary to reduce the number of errors during grammar induction. To reduce the number of idiom-related errors, we will introduce some dictionary information and additional pre-processing.

V. Conclusion

In this paper, we present an approach to automatic acquisition of translation templates from unannotated bilingual parallel corpora. The method is statistical and data-driven, and requires no parser. A grammar induction algorithm extracts from the corpus semantic and phrase structure grammars for both source and target languages. Based on these grammars, the phrase structures are aligned using BTG. Finally, the aligned structures are treated as translation templates. This method needs fewer resources than the method which is called as "parse-parse-match." And it can get comparable results as those of the "parse-parse-match" method. These results of the preliminary experiments show that our approach is viable.

However, we still face many difficult tasks, including the improvement of grammar induction and alignment. In the future, we will introduce more information such as some dictionary information (including a synonym dictionary) and some additional preprocessing.

Acknowledgment

The authors would like to thank Dr. M. Seligman and the anonymous reviewers for their very useful suggestions and very careful proofreading.

References

[1] I. Cicekli and H. Altay Guvenir, "Learning translation templates from bilingual translation exmples," *Appl. Intell.*, vol. 15, no. 1, pp. 57–76, 2001.

[2] K. Goodman and H. Nirenburg, *KBMT-89: A Case Study in Knowledge Based Machine Translation*. San Mateo, CA: Morgan Kaufmann, 1992.

[3] H. Altay Güvenir and I. Cicekli, "Learning translation templates from examples," *Inform. Syst.*, vol. 23, no. 6, pp. 353–363, 1998.

[4] R. Hu, C. Zong, and B. Xu, "Semiautomatic acquisition of translation templates from monolingual unannotated corpora," in *Proc. Int. Conf. Natural Language Processing and Knowledge Engineering*, Beijng, China, Oct. 2003, pp. 163–167.

[5] K. Imamura, "Hierarchical phrase alignment harmonized with parsing," in *Proc. 6th Natural Language Processing Pacific Rim Symp.*, 2001, pp. 377–384.

[6] D. Klein and C. D. Manning, "Accurate unlexicalized parsing," in *Proc. 41st Meeting of the Association for Computational Linguistics*, Sapporo, Japan, 2003, pp. 423–430.

[7] X. Li and C. Zong, "An effective framework for chinese syntactic parsing," in *Proc. Int. Conf. Signal Processing*, Istanbul, Turkey, Dec. 2004.

[8] Y. Lü, M. Zhou, S. Li, C. Huang, and T. Zhao, "Automatic translation template acquisition based on bilingual structure alignment," *Comput. Linguist. Chinese Language Processing*, vol. 6, no. 1, pp. 83–108, Feb. 2001.

[9] H. Kaji, Y. Kida, and Y. Morimoto, "Learning translation templates from bilingual texts," in *Proc. 14th Int. Conf. Computational Linguistics*, 1992, pp. 672–678.

[10] H. Kitano, "A comprehensive and practical model of memory-based machine translation," in *Proc. IJCAI*, Chambery, France, 1993.

[11] S. Kumar and W. Byrne, "A weighted finite state transducer implementation of the alignment template model for statistical machine translation," in *Proc. Conf. Human Language Technology*, Edmonton, AB, Canada, 2003, pp. 142–149.

[12] H. M. Meng and K.-C. Siu, "Semi-automatic acquisition of domain-specific semantic structures," *IEEE Trans. Knowledge Data Eng.*, vol. 14, no. 1, pp. 172–180, Jan./Feb. 2002.

[13] F. J. Och, C. Tillmann, and H. Ney, "Improved alignment models for statistical machine translation," in *Proc. Joint Conf. Empirical Methods in Natural Language Processing and Very Large Corpora*. College Park, MD, Jun. 1999, pp. 20–28.

[14] F. J. Och, "Statistical machine translation: From single word models to alignment templates," Ph.D. dissertation, RWTH Aachen, Aachen, Germany, 2002.

[15] Z. Orhan, "Confidence factor assignment to translation templates," M.S. Thesis, Bilkent Univ., Ankara, Turkey, Sep. 1998.

[16] Z. Oz and I. Cicekli, "Ordering translation templates by assigning confidence factors," in *Proc. AMTA'98—Conf. Assoc. Machine Translation in the Americas*, Langhorne, PA, Oct. 1998, pp. 51–61.

[17] S. Sato, "Example-based translation approach," in *Proc. ATR Workshop*, Jul. 1991.

[18] ——, "MBT2: a method for combining fragments of examples in example-based translation," *Artif. Intell.*, vol. 75, pp. 31–50, 1995.

[19] Y. Wang, "Grammar inference and statistical machine translation," Ph.D. dissertation, Carnegie Mellon Univ., Pittsburgh, PA, 1998.

[20] H. Watanabe, S. Kurohashi, and E. Aramaki, "Finding structural correspondences from bilingual parsed corpus for corpus-based translation," in *Proc. 18th Int. Conf. Computational Linguistics*, 2000, pp. 906–912.

[21] C.-C. Wong, H. Meng, and K.-C. Siu, "Learning strategies in a grammar induction framework," in *Proc. 6th Natural Language Processing Pacific Rim Symp.*, 2001, pp. 153–157.

[22] D. Wu, "An algorithm for simultaneously bracketing parallel texts by aligning words," in *Proc. 33th Annu. Meeting Association for Computational Linguistics*, 1995, pp. 244–251.

[23] ——, "Stochastic inversion transduction grammars and bilingual parsing of parallel corpora," *Comput. Linguist.*, vol. 23, no. 3, pp. 377–403, 1997.

[24] Y. Zhang, S. Vogel, and A. Waibel, "Integrated phrase segmentation and alignment model for statistical machine translation," in *Proc. Conf. Natural Language Processing and Knowledge Engineering*, Beijing, China, 2003, pp. 567–573.

[25] S. Zhao and D. Lin, "A nearest-neighbor method for resolving PP-attachment ambiguity," in *Proc. 1st Int. Joint Conf. Natural Language Processing (IJCNLP2004)*, Mar. 2004, pp. 428–434.

Rile Hu received the Bachelor and Master degrees from the Department of Dynamic Engineering, North China Electric Power University, in 1998 and 2001, respectively. He received the Ph.D. degree from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, in 2005.

He is now a Postdoctoral Researcher at the Nokia (China) Research Center, Beijing. His research interests are in machine translation and machine learning.

Chengqing Zong (A'04) received the Bachelor and Master degrees from Shandong University of Technology (SUT), China, in 1987 and 1990, respectively. He received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences (CAS), in March 1998.

From 1990 to 1995, he was with SUT as a Teacher. From May 1998 to April 2000, he was with the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, CAS, as a Postdoctoral Research Fellow. After his postdoctoral program, he joined NLPR. In 1999 and 2001, he visited the Laboratory of Spoken Language Translation, ATR of Japan, as a Guest Researcher. From November 2004 to January 2005, he visited GETA, CLIPS-IMAG, France, as a Visiting Scholar. His research interests include approaches to spoken language translation, spoken language parsing, and document classification. Please see his Web homepage at http://www.nlpr.ia.ac.cn/English/cip/cqzong.htm.

Bo Xu received the Bachelor degree from Zhejiang University, China, in 1988. In 1988, he joined the Speech Recognition Research Center, National Lab of Pattern Recognition (NLPR), and received the Master and Ph.D. degrees in 1992 and 1997, respectively.

His current interests mainly focus on multimedia content management and statistical speech translation.

Dr. Xu is the Associate President of CASIA and Deputy Director of the NLPR. He was the Coordinator of International Consortium for Speech Translation Advanced Research (C-Star). He is on the steering committee of National High-tech Program in fields of Chinese information processing and intelligent interfaces.