Computational Linguistics and Chinese Language Processing
Vol. 6, No. 1, February 2001, pp. 83-108

Automatic Translation Template Acquisition Based on Bilingual Structure Alignment[1]

Yajuan Lü[*], Ming Zhou[+], Sheng Li[*],

Changning Huang[+], Tiejun Zhao[*]

Abstract

Knowledge acquisition is a bottleneck in machine translation and many NLP tasks. A method for automatically acquiring translation templates from bilingual corpora is proposed in this paper. Bilingual sentence pairs are first aligned in syntactic structure by combining a language parsing with a statistical bilingual language model. The alignment results are used to extract translation templates which turn out to be very useful in real machine translation.

Keywords: Bilingual corpus, Translation template acquisition, Structure alignment, Machine translation

1. Introduction

Bilingual corpora have been recognized as a valuable resource for knowledge acquisition in machine translation and many other NLP tasks. To make better use of them, bilingual corpora are often aligned first. Intensive researches have been done on sentence and word level alignment [Brown *et al.* 1991, Church 1993, Ker *et al.* 1997, Huang *et al.* 2000]. These alignments have been proven to be very useful in machine translation, word sense disambiguation, information retrieval, translation lexicon extraction, and so on. With a sentence aligned parallel English-Chinese corpus ready in hand, this paper extends word-level alignment to syntactic structure alignment with the aim of acquiring structural translation templates automatically.

---

[1] The work was partially done at Microsoft Research Asia.

[*] Dept. of Computer Science & Engineering, Harbin Institute of Technology, Harbin, 15001
E-mail: Lü Yajuan: lyj@mtlab.hit.edu.cn Zhao Tiejun: tjzhao@mtlab.hit.edu.cn

[+] Microsoft Research Asia, Beijing, 10080
E-mail: Zhou Ming: mingzhou@microsoft.com Huang Changning: cnhuang@microsoft.com

Numerous researches have been done to acquire knowledge from bilingual corpora. Many of these studies aimed to acquire word or phrase translation lexicons [Shin *et al*. 1996, Fung *et al*. 1997, Ralf 1997, Turcato 1998]. This paper focuses on the automatic learning of translation templates, which are especially useful for machine translation. In [Guvenir *et al*. 1998], [Malavazos *et al*. 2000] and [Cicekli *et al*. 2001], analogical models were proposed to learn translation templates. By grouping similar translation examples and replacing their difference with a variable, they could obtain translation templates. Structure alignment has been studied by several researchers for use in structural translation template acquisition. Most of the approaches have followed what may be called a "parse-parse-match" procedure [Wu 1997]. The main idea is that each language of the parallel corpus is first parsed individually using a monolingual grammar, and then the corresponding constituents are matched using some heuristic procedures. The works by [Kaji *et al*. 1992], [Almuallim *et al*. 1994], [Grishman *et al*. 1994], [Matsumoto *et al*. 1995], [Meyers *et al*. 1998], [Watanabe *et al*. 2000] etc. can be considered such approaches. Differences between them are in their parsing grammars and heuristic procedures. Kaji and Watanabe used phrase structure grammar, while Grishman employed a regularized syntactic structure. The dependency structure is used in most of the other systems. In [Watanabe 1993], bilingual structure matching was used to improve the existing transfer rules by comparing in incorrect translation and correct translation. Wu [Wu 1995a, Wu 1997] proposed a bilingual language model to represent a bilingual corpus and parse bilingual sentences simultaneously. Because of the lack of a suitable bilingual grammar, their system is used to acquire phrase translation examples, not templates. In all these studies, structure-aligned bilingual corpora were shown to be very useful for translation knowledge acquisition.

The method proposed in this paper differs from the previous approaches in two ways: (1) The bilingual structure alignment is based on a bilingual language model and uses only one language parsing result. Compared with the "parse-parse-match" procedure, monolingual parsing is particularly suitable when there is no robust parser for one of the languages (such as Chinese). (2) The translation templates we acquire are integrated with the processes of transfer and generation, which are the usual two phases in machine translation systems. Two types of templates are obtained: structure translation templates and word selection templates.

This paper is organized as follows: In the next section, we propose a bilingual structure alignment algorithm by combining a language parsing with a statistical bilingual language model. Then, the learning of translation templates is described in section 3. A translation experiment based on the acquired knowledge is described in section 4. We conclude our work in section 5. Although this paper is related to English-Chinese structure alignment and template acquisition, the proposed method is also applicable to other language pairs because it is language independent.

2. Bilingual structure alignment using monolingual parsing

The "parse-parse-match" procedure for bilingual structure alignment is susceptible to three weaknesses: [Wu 1995a]

- Appropriate, robust, monolingual grammars may not be available for both languages. This is the case when Chinese is one of the languages.
- The parsing grammars used in the two languages may be incompatible.
- The process of selecting between multiple possible arrangements may be arbitrary.

To overcome these weaknesses, Wu [Wu 1995d, Wu 1997] has proposed a bilingual language model called the Inversion Transduction Grammar (ITG), which can be used to parse bilingual sentence pairs simultaneously. Subsection 2-1 will give a brief description. For details please refer to [Wu 1995a, Wu 1995b, Wu 1995c, Wu 1995d, Wu 1997]. Based on this model, a bilingual structure alignment algorithm guided by one language parsing will be presented in subsection 2-2.

2.1 ITG bilingual language model

The Inversion Transduction Grammar is a bilingual context-free grammar that generates two matched output languages (referred to as L_1 and L_2). It also differs from standard context-free grammars in that the ITG allows right-hand side production in two directions: straight or inverted. The following examples are two ITG productions:

$$C \text{ -> } [A\ B],$$
$$C \text{ -> } \langle A\ B \rangle.$$

In the above productions, each nonterminal symbol stands for a pair of matched strings. For example, the nonterminal A stands for the string-pair (A_1, A_2). A_1 is a sub-string in L_1, and A_2 is A_1' s corresponding translation in L_2. Similarly, (B_1, B_2) denotes the string-pair generated by B. The operator [ ] performs the usual concatenation, so that $C \text{ -> } [A\ B]$ yields the string-pair (C_1, C_2), where $C_1=A_1B_1$ and $C_2=A_2B_2$. On the other hand, the operator <> performs the straight concatenation for language 1 but the reversing concatenation for language 2, so that $C \text{ -> } \langle A\ B \rangle$ yields $C_1=A_1B_1$, but $C_2=B_2A_2$. The inverted concatenation operator permits the extra flexibility needed to accommodate many kinds of word-order variation between source and target languages [Wu 1995b].

There are also lexical productions of the following form in ITG:

$$A \text{ -> } x/y,$$

which means that a symbol *x* in language L_1 is translated by the symbol *y* in language L_2. The *x*, *y* may be a null symbol *e*, which means there may be no counterpart string in the other language.

Parsing, in the case of an ITG, means building matched constituents for an input sentence-pair. For example, Figure 1 shows an ITG parsing tree for an English-Chinese sentence-pair. The inverted production is indicated by a horizontal line in the parsing tree. The English text is read in the usual depth-first left to right order, but for the Chinese text, a horizontal line means the right sub-tree is traversed before the left. The generated parsing results are:

(1) a. [[[The game]$_{BNP}$ [[will start]$_{VBP}$ [on Wednesday]$_{PP}$]$_{VP}$]$_{S}$.]$_{S}$

b. [[**比赛 [星期三 开始**]$_{VP}$]$_{S}$ 。]$_{S}$

We can also represent the common structure of the two sentences more clearly and compactly with the aid of <> notation:

(2) [[[The/e game/**比赛**]$_{BNP}$ < [will/e start/**开始**]$_{VBP}$ [on/e Wednesday/**星期三**]$_{PP}$ >$_{VP}$]$_{S}$ ./。]$_{S}$

where the horizontal line from Figure 1 corresponds to the <> level of bracketing.

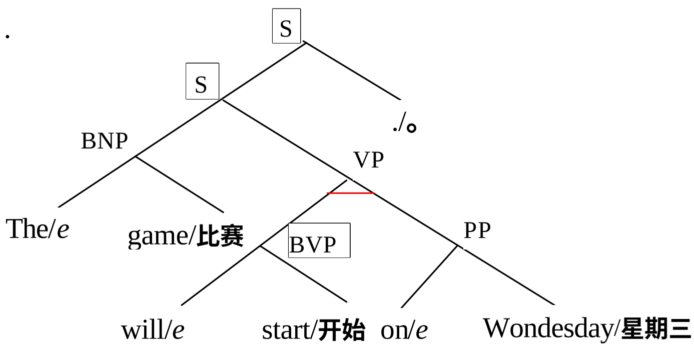


Figure 1 *Inversion transduction grammar parsing tree.*

Any ITG can be converted to a normal form, where all productions are either lexical productions or binary-fanout nonterminal productions [Wu 1995b, Wu 1995c, Wu 1997]. If probability is associated with each production, the ITG is called the Stochastic Inversion Transduction Grammar (SITG).

Because of the difficulty of finding a suitable bilingual syntactic grammar, a practical ITG is a generic Bracketing Inversion Transduction Grammar (BTG), which has been used by Wu in several experiments on bilingual bracketing and to extract phrasal translation examples

[Wu 1995a, Wu 1995b, Wu 1995c]. BTG is a simplified ITG that has only one nonterminal and does not use any syntactic grammar. A Statistical BTG (SBTG) grammar is as follows:

$$A \xrightarrow{a} [AA]; \quad A \xrightarrow{a} < AA >; \quad A \xrightarrow{b_{ij}} u_i / v_j\,; \quad A \xrightarrow{b_{ie}} u_i / e; \quad A \xrightarrow{b_{ej}} e / v_j\,.$$

SBTG employs only one nonterminal symbol *A* that can be used recursively. Here, "*a*" denotes the probability of syntactic rules. However, since those constituent categories are not differentiated in BTG, it has no practical effect here and can be set to an arbitrary constant. The remaining productions are all lexical. b_{ij} is the translation probability that source word u_i translates into target word v_j. b_{ij} can be obtained using a statistical word-translation lexicon[Wu 1997] or statistical word alignment[Lü *et al*. 2001]. The last two productions denote that the word in one language has no counterpart in another language. A small constant can be chosen for the probabilities b_{ie} and b_{ej}.

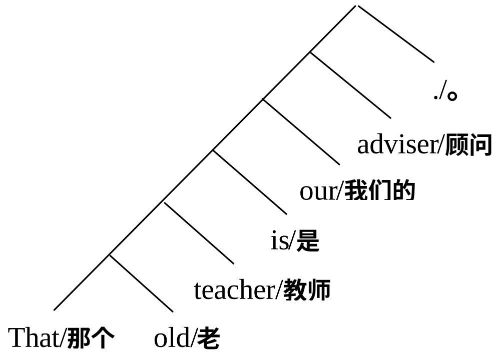
In BTG, no language specific syntactic grammar is used. The maximum-likelihood parser selects the parse tree that best satisfies the combined lexical translation preferences, as expressed by the b_{ij} probabilities. Because the expressiveness characteristics of ITG naturally constrain the space of possible matching in a highly appropriate fashion, BTG achieves encouraging results for bilingual bracketing using a word-translation lexicon alone [Wu 1995a].

Since no syntactic knowledge is used in SBTG, output grammaticality can not be well guaranteed. In particular, if the corresponding constituents appear in the same order in both languages, both straight and inverted, then lexical matching does not provide the discriminative leverage needed to identify the sub-constituent boundaries. For example, consider an English-Chinese sentence pair:

(3) English: That old teacher is our adviser.
Chinese: **那个老教师是我们的顾问**。

The SBTG parsing tree is shown in Figure 2(a), and the corresponding bracketing result is shown in Figure 2(b). The result does not accord with the syntactic structure as we expected. In this case, grammatical information about one or both of the languages can be very helpful. For example, if we know the English parsing result shown in (a) in Figure 3, then the bilingual parsing can be determined easily; the result should be that shown in (b), and the corresponding bracketing result is that shown in (c).

(a)

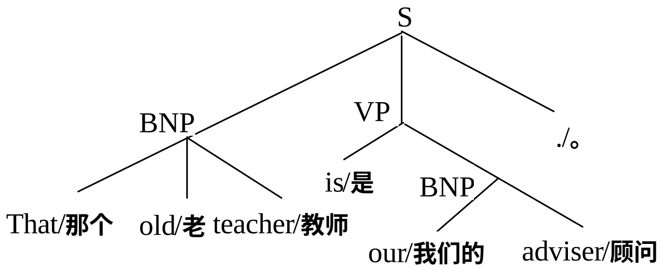


(b) [[[[[[The/那个 old/老] teacher/教师] is/是] our/我们的] adviser/顾问] ./。]

Figure 2 *Bilingual parsing with SBTG.*

(a) English parsing: [[That old teacher]$_{BNP}$ [is [our adviser]$_{BNP}$]$_{VP}$.]$_{S}$

(b)



(c) [[That/那个 old/老 teacher/教师]$_{BNP}$ [is/是 [our/我们的 adviser/顾问]$_{BNP}$]$_{VP}$ ./。]$_{S}$

Figure 3 *Bilingual parsing guided by English parsing.*

Statistics in a corpus of 20,000 word-aligned sentence-pairs indicates that nearly 72% of the sentence-pairs contain the corresponding constituents, which include more than three continuous sub-constituents in identical order. These constituents often lead to ungrammatical parsing with SBTG. Therefore, it is necessary to introduce a language grammar in ITG instead of not using any grammar as in BTG.

2.2 Integrating monolingual parsing with a bilingual language model

From the above discussion, we can see that if one language parser is available, then the bilingual bracketing result can be more grammatical. This is important for syntactic translation template acquisition.

English parsing methods have been well studied. We have also developed an incremental English parser using statistic and learning methods [Meng *et al*. 2001]. A structure alignment algorithm guided by English parsing will be described in this section.

Here, structure alignment guided by English parsing means using an English parser's bracketing information as a boundary restriction in the ITG language model. But this does not necessarily mean parsing the other language completely according to the same parsing boundary. If a parsing structure is fixed according to one language, it is possible that the structure is not linguistically valid for the other language under the formalism of Inversion Transduction Grammar. To illustrate this, see the example shown in Figure 4.

The sub-trees for each blacked underlined part are shown in Figure 4(a) and (b). We can see that the Chinese constituents do not match the English counterparts in the English structure. In this case, our solution is that shown in Figure 4 (c): the whole English constituent of "VP" is aligned with the whole Chinese correspondence; i.e., "eat less bread" is matched with "**少吃面包**." At the same time, we give the inner structure matching according to SITG regardless of the English parsing constraint. An "X" tag is used to indicate that the sub-bilingual-parsing-tree is not consistent with the given English sub-tree.

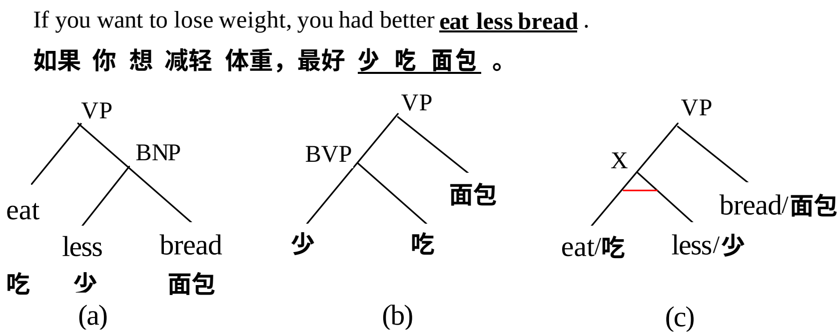


Figure 4 *An example of mismatched sub-trees.*

The main idea is that the given parser is only used as an boundary constraint for bilingual parsing. When the constraint is incompatible with the bilingual model ITG, we use ITG as the default result. This process enables parsing go on regardless of some failures in matching.

We heuristically define a constraint function $F_e(s, t)$ to denote the boundary constraint, where s is the beginning position and t is the end. There are three cases of structure matching: *violate match, exact match* and *inside match. Violate match* means the bilingual parsing conflicts with the given English bracketing boundary. (1,2), (1,3), (2,3), (2,5) etc. in the following English sentence (4) are examples. We assign a minimum $F_e(s, t)$ (0.0001 at present) to prevent the structure match from being chosen when an alternative match is available. *Exact match* means the match falls exactly on the English parsing boundary, and we assign a high $F_e(s, t)$ value (10 at present) to emphasize it. (1,6), (2,5), (3,5) are examples. (3,4), (4,5) are examples of *inside match*, and the value 1 is assigned to these $F_e(s, t)$ functions.

(4) [She/1 [is/2 [a/3 lovely/4 girl/5]] ./6]

[Wu 1997] introduced an algorithm to compute an optimal parsing tree for a given sentence-pair using dynamic programming (DP). This algorithm is similar in spirit to the recognition algorithm of HMM [Rabiner 1989] and to the parsing algorithm of PCFG [Lari *et al*. 1990]. The difference from the usual PCFG parsing is that the DP in SITG parses a sentence-pair simultaneously rather than a sentence only. The basic idea of DP is to divide a problem into several sub-problems, and to calculate the final solution according to the solutions of the sub-problems. In bilingual parsing, dynamic programming is used to calculate the bilingual parsing tree of a sentence-pair by decomposing it into several sub-bilingual-parsing-trees of sub-string-pairs. The whole process is that of calculating the local optimization function from the sub-parsing-tree to the whole parsing tree, recording the preceding path and back tracking along the best path in the end.

Let the input English and Chinese sentences be $e_1,...e_T$ and $c_1,...c_V$. As an abbreviation we write $e_{s...t}$ for the sequence of words $e_{s+1}, e_{s+2,} ...e_t$, and similarly write $c_{u...v}$. The local optimization function $d(s,t,u,v) = \max P[e_{s..t} / c_{u..v}]$ denotes the maximum probability of sub-parsing-tree of node q and that both the sub-string $e_{s...t}$ and $c_{u...v}$ derive from node q. Thus, the best parser has the probability $d(0,T,0,V)$. In [Wu 1995b], $d(s,t,u,v)$ was calculated as the maximum probability combination of all possible sub-tree combinations as given below:

$$
\begin{aligned}
&d(s,t,u,v) = \max[d^{[]}(s,t,u,v), d^{<>}(s,t,u,v)],\\
&d^{[]}(s,t,u,v) = \max_{\substack{s \le S \le t \\ u \le U \le v \\ (S-s)(t-S)+(U-u)(v-U) \ne 0}} d(s,S,u,U)\,d(S,t,U,v),\\
&d^{<>}(s,t,u,v) = \max_{\substack{s \le S \le t \\ u \le U \le v \\ (S-s)(t-S)+(U-u)(v-U) \ne 0}} d(s,S,U,v)\,d(S,t,u,U),
\end{aligned}
\tag{1}
$$

where S is the split point used to break $e_{s \ldots t}$ into two constituent sub-trees, $e_{s \ldots S}$ and $e_{S \ldots t}$. U is the split point used to break $c_{u \ldots v}$ into two constituent sub-trees, $c_{u \ldots U}$ and $e_{U \ldots v}$.The condition $(S-s)(t-S)+(U-u)(v-U) \ne 0$ serves to specify that the sub-string in one, but not both languages may be split into an empty string. Because ITG permits production in two directions, the combination of sub-trees has two corresponding directions. We use [] and <> to denote the straight and reverted production, respectively.

We integrate the constraint function $F_e(s, t)$ into the local optimization function to insert English parsing constraints in bilingual parsing. The computation of the local optimization function is modified as follows:

$$
\begin{aligned}
&d(s,t,u,v) = \max[d^{[]}(s,t,u,v), d^{<>}(s,t,u,v)],\\
&d^{[]}(s,t,u,v) = \max_{\substack{s \le S \le t \\ u \le U \le v \\ (S-s)(t-S)+(U-u)(v-U) \ne 0}} F_e(s,t)\,d(s,S,u,U)\,d(S,t,U,v),\\
&d^{<>}(s,t,u,v) = \max_{\substack{s \le S \le t \\ u \le U \le v \\ (S-s)(t-S)+(U-u)(v-U) \ne 0}} F_e(s,t)\,d(s,S,U,v)\,d(S,t,u,U).
\end{aligned}
\tag{2}
$$

The other symbols in the algorithm are defined as follows: $b(e_t / c_v)$ is the probability of translating English word e_t into Chinese word c_v obtained from word alignment [Lü *et al.* 2001]. We assign a minimal probability (0.0001 at present) to empty word alignment $b(e_t / e)$ and $b(e / c_v)$. $q(s,t,u,v)$, $s(s,t,u,v)$ and $g(s,t,u,v)$ are variables used to record the production direction, the split point in English and the split point in Chinese, respectively, when $d(s,t,u,v)$ is achieved. These variables are used to reconstruct the bilingual parsing tree in the final step. Suppose node $q = (s,t,u,v)$; then, $\lambda(s,t,u,v) = \lambda(q)$ is the nonterminal label of q. LEFT(q) is the left sub-tree of q, and RIGHT(q) is the right sub-tree of q.

The algorithm is as follows:

1. Initialization

$$
\begin{aligned}
&d(t-1,t,v-1,v) = b(e_t / c_v), && 1 \le t \le T,\ 1 \le v \le \mathrm{V}\\
&d(t-1,t,v,v) = b(e_t / e), && 1 \le t \le T,\ 1 \le v \le \mathrm{V}\\
&d(t,t,v-1,v) = b(e / c_v), && 1 \le t \le T,\ 1 \le v \le \mathrm{V}
\end{aligned}
$$

2. Recursion

For all *s,t,u,v* ($0 \le s < t \le T, \quad 1 \le u < v \le V, \quad t - s + v - u > 2$),

$$d(s,t,u,v) = \max[d^{[]}(s,t,u,v), d^{<>}(s,t,u,v)],$$
$$q(s,t,u,v) = \begin{cases} [\,] & if \quad d^{[]}(s,t,u,v) \ge d^{<>}(s,t,u,v), \\ <> & otherwise \end{cases}$$

where

$$d^{[]}(s,t,u,v) = \max_{\substack{s \le S \le t \\ u \le U \le v \\ (S-s)(t-S)+(U-u)(v-U) \ne 0}} F_e(s,t) d(s,S,u,U) d(S,t,U,v),$$
$$d^{<>}(s,t,u,v) = \max_{\substack{s \le S \le t \\ u \le U \le v \\ (S-s)(t-S)+(U-u)(v-U) \ne 0}} F_e(s,t) d(s,S,U,v) d(S,t,u,U),$$

$$s^{[]}(s,t,u,v) = \arg_S \max_{\substack{s \le S \le t \\ u \le U \le v}} d(s,S,u,U) d(S,t,U,v),$$
$$s^{<>}(s,t,u,v) = \arg_S \max_{\substack{s \le S \le t \\ u \le U \le v}} d(s,S,U,v) d(S,t,u,U),$$

$$g^{[]}(s,t,u,v) = \arg_U \max_{\substack{s \le S \le t \\ u \le U \le v}} d(s,S,u,U) d(S,t,U,v),$$
$$g^{<>}(s,t,u,v) = \arg_U \max_{\substack{s \le S \le t \\ u \le U \le v}} d(s,S,U,v) d(S,t,v,U).$$

3. Reconstruction

The root of the parsing tree is (0,T,0,V), and its nonterminal label is set to $\lambda(0,T,0,V) = \lambda_e(0,T)$, where $\lambda_e(s,t)$ is the English sub-tree tag that sub-string $e_{s...t}$ are derived from this sub-tree. If $e_{s...t}$ is not a sub-tree in the English parsing tree, then $\lambda_e(s,t)$ is given a tag "X". The remaining node q = *(s,t,u,v)* in the optimal parsing tree is calculated recursively as follows:

$$if \quad q(s,t,u,v) = [\,], \begin{cases} LEFT(q) = (s, s^{[]}(s,t,u,v), u, g^{[]}(s,t,u,v)) \\ RIGHT(q) = (s^{[]}(s,t,u,v), t, g^{[]}(s,t,u,v), v) \\ \lambda(\text{LEFT(q)}) = \lambda(s, s^{[]}(s,t,u,v), u, g^{[]}(s,t,u,v)) = \lambda_e(s, s^{[]}(s,t,u,v)) \\ \lambda(\text{RIGHT(q)}) = \lambda(s^{[]}(s,t,u,v), t, g^{[]}(s,t,u,v), v) = \lambda_e(s^{[]}(s,t,u,v), t) \end{cases},$$

$$if \quad q(s,t,u,v) = <>, \begin{cases} LEFT(q) = (s, s^{<>}(s,t,u,v), g^{<>}(s,t,u,v), u) \\ RIGHT(q) = (s^{<>}(s,t,u,v), t, u, g^{<>}(s,t,u,v)) \\ \lambda(\text{LEFT(q)}) = \lambda(s, s^{<>}(s,t,u,v), g^{<>}(s,t,u,v), u) = \lambda_e(s, s^{<>}(s,t,u,v)) \\ \lambda(\text{RIGHT(q)}) = \lambda(s^{<>}(s,t,u,v), t, u, g^{<>}(s,t,u,v)) = \lambda_e(s^{<>}(s,t,u,v), t) \end{cases}.$$

After the bilingual parsing tree is created, the post-process consisting of rotation and flattening operations is used to restore the fanout flexibility [Wu 1997].

Using this improved SITG (ISITG), we can obtain the bilingual parsing result shown in Figure 3(b) for the given sentence-pair (3); when SBTG is used, the parsing result is that shown in Figure 2. Comparing the two results, we can see that by integrating English parsing constraints into ITG, the bilingual parsing becomes more grammatical. In the next section, we will give a quantitative experimental comparison of SBTG with ISITG.

It should be pointed out that the proposed algorithm can also be used with one-language-partial parsing, as well as with both-language parsing.

2.3 Experiments on bilingual structure alignment

To find out how important it is to include at least one language parsing, four experiments were carried out using (1) no parser (E+C); (2) only an English parser (E-parsing+C); (3) an English parser and a Chinese base phrase parser (E-parsing+C-base); (4) an English parser and a Chinese parser (E-parsing+ C-parsing). Experiment (1) followed the model of SBTG, and the other three experiments used ISITG.

The test set consisted of 2,000 English-Chinese bilingual sentence-pairs. 1,000 of the sentence pairs were collected from English textbooks for junior and senior middle school or college. The others came from the machine translation evaluation corpus of the Institute of Computational Linguistics at Peking University [Duan *et al.* 1996]. The lengths of the English sentences varied from 4 to 25 words. The test sentence pairs were first aligned at the word level based on statistics and a lexicon [Lü *et al.* 2001]. The English sentences were parsed using an incremental parser [Meng *et al.* 2001]. Both the word alignment and the English parsing were post revised manually. The Chinese parser used here is being developed by our research group. The whole parsing results are not yet robust with a precision of less than 80%. But its first stage—base phrase parsing— is quite good with a precision rate of 91.1%[Zhao *et al.* 2000]. The Chinese parsing results were not manually revised.

We evaluated the structure alignment results using a syntactic criterion. This means the matching must be grammatical. For example, for the sentence pair shown below:

(5) English: The student will get a pen .
Chinese: **这学生将得到一支钢笔**。

the matchings "The student <--> **这学生**", "will get<-->**将得到**", and "a pen <-->**一支钢笔**" are grammatical, while "student will<-->**学生将**" and "get a<-->**得到一支**" are ungrammatical.

All the phrases in the test set with grammatical structure matching were manually edited. These phrases were regarded as the standard structure correspondences in the evaluation. We obtained 7,812 standard structure pairs in total. The accuracy rate is defined as

$$Accuracy\ rate = \frac{standard\ structure\ numbers\ obtained\ in\ test}{total\ numbers\ of\ standard\ structures}. \tag{3}$$

Table 1. *Comparison of accuracy in bilingual structure alignment.*

Experiment type	E+C	E-parsing +C	E-parsing+C-base	E-parsing+C-pasing
Accuracy rate(%)	64.62	85.05	90.55	88.25

Table 1 shows the results of the four experiments. From the comparison of accuracy, we can see that when no parsing was conducted, the quality of alignment could not be guaranteed. The result is hardly usable for syntactic translation template acquisition. An English parsing could improve the result greatly. When a Chinese base parsing was also used, the result was even better. However, if both English and Chinese parsing were used, the result worsened slightly. This is not surprising. One reason is that Chinese parsing is still not robust. Another reason is that the two languages are parsed separately in different grammars, which may be incompatible in some respects. In the general "parse-parse-match" approach, this problem cannot be avoided.

Following is an example to illustrate the changes of the bilingual structure alignments obtained from the four experiments (Here we use the bracketing format and do not show the parsing tree in figures to save space. Readers can draw bilingual parsing trees easily according to the bracketing results.)

(6) English: This new method was brought into existence in the fifties.

Chinese: **这一新方法出现于五十年代。**

English parsing: [[This new method]$_{BNP}$ [[was brought into existence]$_{VBD}$ [in [the fifties]$_{BNP}$]$_{PP}$]$_{VP}$.]$_{S}$

Chinese base phrase parsing : [**这 一** [**新 方法**]$_{BNP}$]$_{BNP}$ **出现 于** [**五十 年代**]$_{BNP}$ **。**

Chinese pasrsing: [[[**这 一** [**新 方法**]$_{BNP}$]$_{BNP}$ **出现**]$_{SS}$ [**于** [**五十 年代**]$_{BNP}$]$_{PP}$ **。**]$_{S}$

Result 1 (E+C): [[[[[[[[[[This/**这** *e*/**一**] new/**新**] method/**方法**] was/*e*] brought into existence/**出现**] in/**于**] the/*e*] fifties/**五十**] *e*/**年代**] ./**。**]

Result 2 (E-pasing+C): [[This/**这** *e*/**一** [new/**新** method/**方法**]$_{BNP}$]$_{BNP}$ [[was/*e* brought

into existence/**出现**]$_{VBD}$ [in/**于** [the/e fifties/**五十**]$_{BNP}$]$_{PP}$ *e*/**年代**]$_{VP}$ ./。]$_S$
Result 3 (E-parsing+C-base): [[This/**这** *e*/**一** [new/**新** method/**方法**]$_{BNP}$]$_{BNP}$ [[was/e brought into existence/**出现**]$_{VBD}$ [in/**于** [the/*e* fifties/**五十** *e*/**年代**]$_{BNP}$]$_{PP}$]$_{VP}$ ./。]$_S$
Result 4 (E-parsing+C-parsing): [[[This/**这** *e*/**一** [new/**新** method/**方法**]$_{BNP}$]$_{BNP}$ [was/e brought into existence/**出现**]$_{VBD}$]$_{SS}$ [in/**于** [the/*e* fifties/**五十** *e*/**年代**]$_{BNP}$]$_{PP}$ ./。]$_S$

In experiment 1, since no grammar was used, result 1 is ungrammatical. English parsing was a big help in determining the syntactic boundary of structure alignments in experiment 2. Result 2 is much better than result 1. When the Chinese base phrase parsing was also added, it helped eliminate some Chinese boundary errors(such as "[**五十 年代**]$_{BNP}$" in result 3). But for experiment 4, the result contradicts the English parsing result because the given Chinese parsing result is incompatible with the English parsing result.

The errors in structure alignment were mainly due to empty word alignment, where a word in one language has no counterpart string in another language. Idiomatic expressions and paraphrases usually introduce many empty word alignment errors. For example, the following two sentence-pairs, (7) and (8), can not be parsed correctly because no word is aligned in the paraphrases "has an eye ←→**有鉴赏力**" and "in hunger and cold ←→ **在饥寒交迫中**". We can not recover these structure alignments using our algorithm for the time being.

(7) English: She has an eye for color.
Chinese: **她对颜色很有鉴赏力**。

(8) English: Before liberation, peasants were struggling in hunger and cold.
Chinese: **解放前，农民在饥寒交迫中挣扎着**。

Another limitation of the formalism is that it can not deal with separate two-part matches, such as the "when" match with "**当**.....**时**" in the follow example:

(9) English: Water freezes when the temperature falls below 0℃.
Chinese: **当温度下降至摄氏零度以下时，水会结冰**。

It is necessary to build special productions to handle these match patterns.

Table 2. *Some examples of bilingual structure alignment.*

[<Mr./**先生** Wu/**吴**>$_{BNP}$ <[play/**拉** accordion/**手风琴**]$_{VP}$ [very/**很** well/**会**]$_{ADVP}$ >$_{VP}$ ./**。**]$_{S}$
S[He/**他** [will/**将** <come/**来** [in/**在** [the/*e* afternoon/**下午**]$_{BNP}$]$_{PP}$ >$_{VP}$]./**。**]
[<Will/**愿意** you/**你**>$_{X}$ [tell/**告诉** me/**我** [your/**你的** age/**年龄**]$_{BNP}$ *e*/**吗**]$_{VP}$?/**？**]$_{SQ}$
[[His/**他的** punishment/**判刑**]$_{BNP}$ <[[was/*e* commuted/**减轻**]$_{VBD}$ [to/**为** life imprisonment/**无期徒刑**]$_{PP}$]$_{X}$ [by/**由** [the/*e* judge/**法官**]$_{BNP}$]$_{PP}$ >$_{VP}$ ./**。**]$_{S}$
[<[We/**我们** *e*/**还是** had/*e* *e*/**度过** *e*/**了** <quite/**相当** an/**一个**>$_{X}$ enjoyable/**愉快的** holiday/**假日**]$_{S}$,/**，** [in spite of/**尽管** <the/**如此** weather/**气候**>$_{BNP}$]$_{PP}$ >$_{S}$ ./**。**]$_{S}$

Some bilingual alignment results based on E-parsing+C-base are given in table 2. The syntactic structure alignments obtained with this method were later used to extract translation templates as described in the next section.

3. Translation template acquisition

When a sentence-pair is aligned using the proposed bilingual structure alignment method, the corresponding words and syntactic structures are determined. These correspondences can be used directly in translation template acquisition.

A translation template is a bilingual translation pair in which the corresponding units (words or phrases) may be replaced by variables. Two types of templates are extracted: structure translation templates and word selection templates. We take phrase or POS tag categories of noun(NN, NNS in our POS tag), verb(VB,VBP,VBZ,VBD,VBN), pronoun(PRP, PRP$)， adjective(JJ) and adverb(RB) as variables. (Our phrase symbols and POS tags are the same as those of the Penn Treebank [Marcus *et al.* 1993].)

Structure translation templates are created from phrase nodes. Each phrase node corresponds to a template. A structure translation template consists of two parts: the left side contains the component conditions of the phrase in the source language, and the right side contains the structure transfer and the translation pattern in the target language. The phrase itself is used as an index.

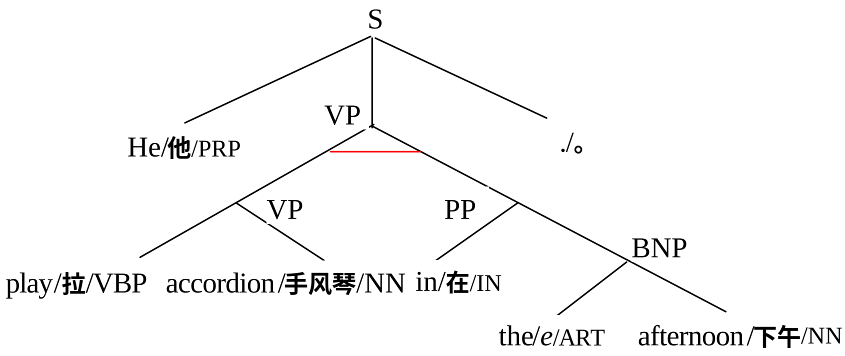


Figure 5 *A bilingual parsing tree used in translation template acquisition.*

For a bilingual structure alignment like that shown in Figure 5, five translation templates can be extracted corresponding to the five phrase nodes:

#S： 1:C=PRP+2:VP+3:W=. -> T(1)+T(2)+。;
#VP: 1:VP+2:PP ->T(2)+T(1);
#VP: 1:C=VBP+2:C=NN ->T(1)+T(2);
#PP: 1:W=in+2:BNP->在+T(2);
#BNP: 1:W=the+2:C=NN ->T(2).

The left side of the template (before ->) contains component conditions of the phrase in the source language connected with "+". "+" denotes the relation of "and", which means that the left side of the template is satisfied only when all the sub-conditions are satisfied. The numbers before ":" represent the order of the node. "W="means the word itself; "C=" means the POS category; otherwise, it is a phrase tag. The right side of the template contains the corresponding translation pattern in the target language. The function *T*(*order*) means the translation of the node "*order*". If the node is a phrase, the function returns the phrase translation by calling a structure translation template. If the node is a word, the function returns the word translation by calling a word selection template. Thus, a template "*#S: 1:C=PRP+2:VP+3:W=. -> T(1)+T (2) + 。*" means that if the phrase tag is "S" and its components satisfy the conditions that 1) the first node's category is "PRP", 2) the second node is a phrase with tag "VP" and 3) the third word is ".", then the translation should be the first node's translation plus the second node's translation, plus the punctuation mark "。". If the bilingual structure is inversely matched (with a horizontal line or "<>" notation), we write the right hand side of the template in inverse order, too. As in template "*#VP: 1:VP+2:PP->T(2)+T(1)*", the translation should be the second node's translation, followed by the first node's translation.

It can be seen that the translation templates transfer a source structure to a target structure by changing the order of nodes on the right side. At the same time, by connecting node translation on the right side, the target translation can also be generated. Therefore, the template is a union of transfer and generalization.

The word selection template is created from the leaf node. We first get the default translation—statistically the most frequent translation in a bilingual corpus. If the current leaf node translation is not the same as the default one, we create a word selection template. For example, the word "play" has the default translation "玩" when it is a verb, while in the given example, the translation is "拉", so we get a new word selection template as follows:

#play: -1:C=PRP+0:C=VBP+1:W= accordion ->拉 .

The format of a word select template is similar to that of a structure translation template except that 1) the index entry is a word; 2) the left side of the template contains the context conditions of the word. A negative number indicates that the node is to the left of the word; 3) the right side of the template contains the translation of the word. We resolve ambiguities by adding more context words as constraints on the left side. This strategy is also used in the structure translation template.

Using the previous structure alignment corpus for the test set, we obtained a total of 7,266 templates, including 4,805 structure translation templates and 2,461 word selection templates. At present, we assume that specific templates (having the "W=" condition on the left side) have higher priority than the common templates. The frequency information of templates is also used to solve ambiguities. These acquired templates are stored in a template base. Structure translation templates and word selection templates are indexed individually by means of phrases and words. The system deals with structure translation templates and word selection templates in the same way during translation.

Translation is a recursive template matching procedure as shown in Figure 6. The input is an English parsing tree. The translation starts from the root node and works recursively top-down and from left to right. The output in the target language is generated bottom-up. It is a post-order-traverse process. When the current node is processing, all its child nodes have been processed and their translations have been determined. If no translation templates can be matched, the system uses the bilingual dictionary as the default word translation, and the structure is translated from left to right. The translation result is generated in the root node's translation field after the recursive procedure is performed.

Because the transfer and generation are combined in structure of a translation template, the translation architecture is simpler than those of most existing translation systems, which include two separate processes for transfer and generation. The obtained translation templates are similar in format with manually edited rules, and the templates are easy to understand, so they can be modified easily and integrated into an existing machine translation system.

```
procedure Translation(ParsingTree * pnode)    // pnode is the current translation node
{
     if( IsLeafNode(pnode) )                  // decide if pnode is a leaf node
    {                                          // process for leaf node
        if ( MatchWordSelRule(pnode, rule)) //find word selection template, success return true
            pnode->translation=GetTrans(pnode, rule); //get translation according to the rule
        else
            pnode->translation=GetDefaultTrans(pnode); //get default translation
        return;
    }
    for(all pcnode, pcnode is pnode's child node ) // translate all child node
        Translation(pcnode);
    If(MatchStructureTransRule(pnode,rule))        //Find structure translation template
        pnode->translation=GetTrans(pnode, rule); //Get translation according to the rule
    else
        pnode->translation=GetDefaultTrans(pnode); //Get default translation
}
```

Figure 6 *Translation procedure.*

4. Experiments on translation using the acquired templates

In this section, we will describe translation experiments conducted based on the acquired templates to evaluate the quality of these templates.

4.1 System architecture

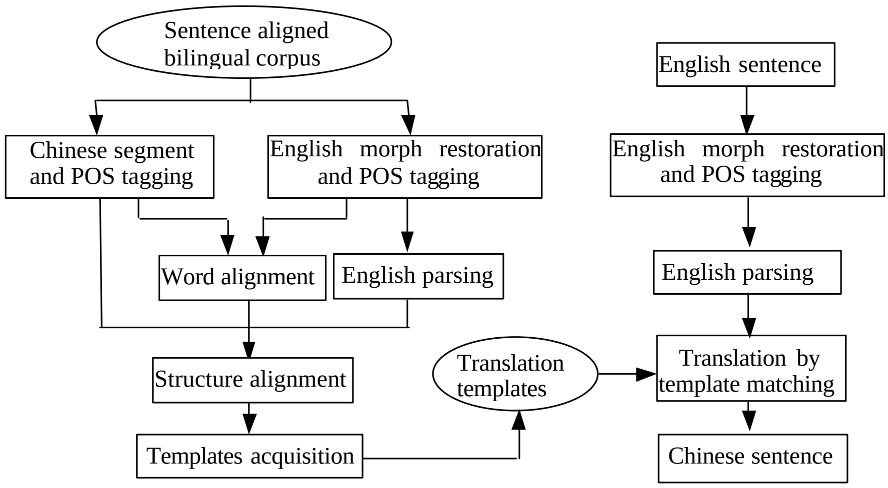


Figure 7 *System Architecture.*

An overview of the machine translation system with auto acquired translation templates is shown in Figure 7. The left part contains the learning process for translation template acquisition. The right part contains a machine translation process that uses the acquired templates. In the learning process, the bilingual sentence pairs are first aligned at the word level based on statistics and a lexicon [Lü *et al*. 2001]. Then, word alignment is extended to structure alignment as mentioned in section 2. Based on the structure alignment, translation templates are acquired and stored in a template base. In the translation process, an English sentence is parsed first; then, the template matching procedure as shown in Figure 6 is used to translate the English sentence into Chinese.

4.2 Translation experiments and evaluation

Translation experiments were conducted on the 2,000 English sentences in the test set. Some translation results and the templates used are presented in the following examples. The first line in each example is the original English sentence to be translated. The second line is the sentence's parsing result. The last line is the Chinese translation result, and the other lines are the templates used in the translation procedure.

1) He abandoned the plan of going abroad.

[He\PRP [abandoned\VBD [the\ART plan\NN]$_{BNP}$ [of\IN [going abroad\VBG]$_{BNP}$]$_{PP}$]$_{VP}$.\FSP]$_{S}$

#S: 1:C=PRP+2:VP+3:W=. ->T(1)+T(2)+。;

#VP: 1:C=VBD+2:BNP+3:PP ->T(1)+**了**+T(3)+T(2);

#BNP: 1:W=the+2:C=NN ->T(2);

#PP: 1: W=of+2:BNP->T(2)+**的**;

#BNP: 1:W=going abroad ->**出国**;

#abandon: -1:C=PRP+1:W=the+2:W=plan ->**放弃**;

他放弃了出国的计划。

2) We passed our time pleasantly.

[We\PRP [passed\VBD [our\PRP$ time\NN]$_{BNP}$ pleasantly\RB]$_{VP}$.\FSP]$_{S}$

#S: 1:C=PRP+2:VP+3:W=. ->T(1)+T(2)+。;

#VP: 1:C=VBD+2:BNP+3:C=RB ->T(3)+T(1)+**了**+T(2);

#BNP: 1:C=PRP$+2:C=NN ->T(1)+T(2);

#pass: -1:C=PRP+0:C=VBD+1:W=our+2:W=time ->**度过**;

我们愉快地度过了我们的时间。

3) The policeman demanded his name and address .

[[The\ART policeman\NN]$_{BNP}$ [demanded\VBD [his\PRP$ name\NN and\CC address\NN]$_{BNP}$]$_{VP}$.]$_{S}$

#S: 1:BNP+2:VP+3:W=.->T(1)+T(2)+。;

#BNP: 1:W=the+2:C=NN->T(2);

#VP: 1:C=VBD+2:BNP->T(1)+T(2)

#BNP: 1:C=PRP$+2:C=NN+3:W=and+4:C=NN->T(1)+T(2)+**和**+T(3)

#demand: -1:W=**警察**+0:C=VBD+1:W=**他的**->**询问**

警察询问他的名字和地址。

To evaluate the quality of the acquired templates, we compared the translation results based on these acquired templates with those based on our existing manually edited translation knowledge base. This translation knowledge based system has the same parsing input as the learned template based system. The difference is that the system's translation process is directed by knowledge base that is totally edited by linguistic engineers. There are more than 35,000 knowledge rules in the system's knowledge base at present. The previous test set was also used as reference translation examples when the translation knowledge base was manually defined in this knowledge-based machine translation system. The evaluation followed the standards of The National High Technology Research and Development Program (the 863 Program) machine translation evaluation project conducted in 1997 [Duan *et al*. 1996]. In the standards, translations are ranked in 6 grades, named A, B, C, D, E and F. They are defined as follows: A denotes an accurate and fluent translation; B denotes a translation that is approximately correct except for a few unimportant problems; C is a translation that can express the meaning of the source text, but some segments are ill-formed; D is a translation that is only partially correct, and separate word translations are given; E is a bad translation except that some word translations are correct; F denotes that no translation is obtained. In our evaluation, no F type translation appeared. We converted A, B, C, D and E into 100, 80, 60, 40 and 20 when calculating the average scores. 200 English sentences were random selected from test set for the manual test. These sentences were translated using the learned template-based system (LTBS) and the manually edited knowledge-based system (MEKBS), respectively. The same evaluator gave evaluations for both translations. Table 3 shows a comparison of the results. Table 4 gives some translation examples and the corresponding evaluation grades based on the acquired translation templates.

Table 3. *Translation test results.*

Type / System	A	B	C	D	E	Average score
LTBS	60%	21%	12%	4%	3%	86.2
MEKBS	48%	41.5%	8.5%	1%	1%	86.9

The results show that without any manual encoding of translation knowledge, we were able to achieve performance nearly equal to that of traditional knowledge based machine translation. The system generated more perfect translations (A) than manually constructed translation rules did. This is because the templates were all learned automatically from real translation texts, so it could produce correct translations exactly when no ambiguities occurred. Although it also produced some bad translations (D, E), the translation results seem quite promising.

Table 4. *Some translation and evaluation grades.*

English	Translation	Grade
I will not be able to go to the movies tomorrow.	我明天不能去看电影。	A
The singer was accompanied at the piano by her pupil.	演唱者由她的学生用钢琴伴奏。	A
Which of them arrived first?	他们中哪个人第一个到达的?	A
He is having his breakfast.	他正在吃他的早饭。	B
The air here is very good.	这里空气是很好。	B
They started at night.	在晚上他们开始。	C
Will you tell me your age?	你愿意告诉我你的那个时代吗?	C
The student has a pen.	这学生长一支钢笔。	D
Some fish jump out of the water to catch insects.	一些鱼跳来自水抓住昆虫。	D
You don't like him, and I don't either.	你做也喜欢它，我做不也不喜欢。	E

Bad translations were produced because there were conflicts between templates. This disambiguation between templates is a difficult problem for any knowledge-based or example-based machine translation system. In our learning process, we solve this problem in two steps: firstly, we use the template with the highest frequency as the default template; then, when a candidate template conflicts with the default template, we add context words or categories as restrictions for this template. In the translation process, specific templates that contain a word restriction are given higher priority; otherwise the templates with highest frequency are chosen. This simple strategy works well when the training corpora are small. But when the training corpora are large, conflicts will occur more frequently. Finding a more robust method for disambiguation will be a goal of future research.

4.3 Discussion

We have developed a method for learning translation templates from bilingual corpora. These learned translation templates lead to good performance in real machine translation. Our study has shown that it is possible to reduce the need for manually encoding of translation templates, which is a difficult task in traditional knowledge-based machine translation. In addition, our method also has the following advantages:

- Compared with statistic-based machine translation(SBMT)， the translation templates obtained using our method are easier to understand than the abstract probability used by Brown [Brown *et al*. 1993].

- Unlike pure example-based machine translation (EBMT), our translation templates replace the same categories of parts-of-speech and phrases with variables, making it more general than the sentence or phrase translation examples given in [Nagao 1984].
- Unlike the traditional knowledge based (KBMT) systems, our translation templates are acquired from translation examples automatically. This can reduce the effort required for manual compilation of translation rules to a minimum.
- The learning method can easily be adapted to a new domain if only domain specific bilingual corpora are provided.

5. Conclusion and future work

Translation knowledge acquisition has been a bottleneck in machine translation. This paper has presented a method for automatic acquisition of translation templates from a bilingual corpus. The bilingual corpus is first aligned in syntactic structures using an alignment algorithm that is based on a bilingual language model and only one language parsing. The algorithm is particularly useful when a full bilingual grammar is not available. It also can be used to acquire a parsing grammar for a language lacking a well-studied grammar from a second language with a well-studied grammar. Based on the alignment result, both structure translation templates and word selection templates are extracted. Application of such templates in machine translation has demonstrated their superior performance in describing translation knowledge.

Although the results we have obtained are quite promising, there is still much to do in the near future. The corpus we used in our experiments is relatively small, and its contents are normative. We will increase the scale and extend the domain of the corpus to improve the quality and quantity of acquired translation templates. In addition, disambiguation of conflicting templates is a key problem. When the training corpus becomes large, this problem becomes serious. To solve it, we will try to introduce semantic restrictions and statistical information into templates in our future work.

Acknowledgement

This research was funded by the Microsoft Machine Translation Laboratory of the Harbin Institute of Technology. We would like to thank Microsoft Research Asia and the Institute of Computational Linguistics at Peking University for providing bilingual corpora for data training. We also would like to thank Dr. Gao Jianfeng of Microsoft Research Asia for his help in developing our ideas.

References

Hussein Almuallim, Yasuhito Akiba and Takefumi Yamazaki, "A Tool for the Acquisition of Japanese-English Machine Translation Rules Using Inductive Learning Techniques," *Proceedings of the Conference on Artificial Intelligence for Applications*, 1994, pp. 194-201.

P. F. Brown, J. C. Lai and R. L. Mercer, "Aligning Sentences in Parallel Corpora," *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics, (ACL-1991)*, pp. 169-176.

P. F. Brown, S. A. Della Pietra, V. J. Della Pietra and R. L. Mercer, "The Mathematics of Statistical Machine Translation: Parameter Estimation," *Computational Linguistics*, Vol. 19, No. 2, 1993, pp. 263-311.

K. W. Church, "Char-align: a Program for Aligning Parallel Texts at the Character Level," *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics (ACL-1993)*, pp. 1-8.

Ilyas Cicekli and Halil Altay Guvenir, "Learning Translation Templates from Bilingual Translation Examples," *Applied Intelligence*, Vol. 15, No. 1, 2001, pp. 57-76.

Huiming Duan and Shiwen Yu, "Report for machine translation evaluation," *Computer World*, 1996.3:183 (in Chinese)

Pascale Fung and Kathleen McKeown, "Finding Terminology Translations from Non-parallel Corpora," *The 5th Annual Workshop on Very Large Corpora*, Hong Kong: August 1997, pp. 192-202.

Ralph Grishman, and John Sterling, "Generalizing Automatically Generated Selectional Patterns," *Proceedings of 15th International Conference on Computational Linguistic (COLING-1994)*, pp. 742-747.

Halil Altay Guvenir and Ilyas Cilekli, "Learning Translation Templates from Examples," *Information Systems*, Vol.23, No. 6, 1998, pp. 353-363.

Jin-Xia Huang and Key-Sun Choi "Chinese-Korean Word Alignment Based on Linguistic Comparison," *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL-2000)*, pp. 392-399.

Hiroyuli Kaji, Yuuko Kida, and Yasutsugu Morimoto, "Learning Translation Templates from Bilingual Texts," *Proceedings of the 14th International Conference on Computational Linguistics (COLING-1992)*, pp. 672-678.

Sue J. Ker and Jason S. Chang, "A Class-based Approach to Word Alignment," *Computational Linguistics* 23(2), 1997, pp. 313-343.

K. Lari and S. J. Young, "The estimation of stochastic context-free grammars using the Inside-Outside algorithm," *Computer Speech and Language*, 4:35-56, 1990.

Yajuan Lü, Tiejun Zhao, Sheng Li and Muyun Yang, "English-Chinese Word Alignment Based on Statistic and Lexicon," *Proceedings of 6th Joint Symposium of Computational Linguistics*, TaiYuan, China, 2001, pp. 108-115. (in Chinese)

Christos Malavazos and Stelios Piperidis, "Application of analogical Modeling to Example Based Machine Translation," *Proceedings of the 18th International Conference on Computational Linguistics (COLING-2000)*, pp. 516-522.

M.P. Marcus, B. Santorini, and M. A. Marcinkiewicz, "Building a Large Annotated Corpus of English: the Penn Treebank," *Computational Linguistics* 19(2), 1993, pp. 313-330.

Yuji Matsumoto and Mihoko Kitamura, "A Machine Translation System Based on Translation Rules Acquired from Parallel Corpora," *Recent Advances in NLP,* Bulgnira 1995, pp. 27-44.

Meng Yao, Zhao Tiejun, Li Sheng and Fang Gaolin, "Incremental English parser using Combination of Statistic and Learning," *Proceedings of 863 Conference on Intelligent Computer*, Beijing, 2001, pp. 343-348. (in Chinese)

Adam Meyers, Roman Yangarber, Ralph Grishman, Catherine Macleod and Antonio Moreno. Sandoval, "Deriving Transfer Rules from Dominance-Preserving Alignments," *Proceedings of the Conference 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL-1998)*, pp. 843-847.

Nagao, M., "A Framework of a Mechanical Tranlation between Japanese and English by Analogy Principle," *Artificial and Human Intelligence*, edited by Elithorn, A. and Banerji, R., North-Holland, 1984, pp. 173-180.

Lawrence R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, Vol. 77, No.2, February 1989. pp. 257-285.

Ralf D. Brown, "Automated Dictionary Extraction for "Knowledge-Free" Example-Based Translation," *Proceedings of the Seventh International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-1997)*, pp. 111-118.

Jung H. Shin, Young S. Han and Key-Sun Choi, "Bilingual Knowledge Acquisition from Korean-English Parallel Corpus Using Alignment Method, " *Proceeding of the 16th International Conference on Computational Linguistics (COLING-1996)*, pp. 230-235.

Davide Turcato, "Automatically Creating Bilingual Lexicons for Machine Translation from Bilingual Text," *Proceedings of the conference 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL-1998)*, pp. 1299-1305.

Hideo Watanabe, "A Method for Extracting Translation Patterns from Translation Examples," *Proceedings of the 5th International Conference on Theoretical and Methodological Issues in Machine Translation*, 1993, pp. 292-302.

Hideo Watanabe, Sadao Kurohashi, and Eiji Aramaki, "Finding Structural Correspondences from Bilingual Parsed Corpus for Corpus-based Translation," *Proceedings of the 18th International Conference on Computational Linguistics (COLING-2000)*, pp. 906-912.

Dekai Wu, "Grammarless Extraction of Phrasal Translation Examples from Parallel Texts," *Proceedings of 6th International Conference On Theoretical and Methodological Issues in Machine Translation*, Leuven, Belgium, 1995, pp. 354-372.

Dekai Wu, "An Algorithm for Simultaneously Bracketing Parallel Texts by Aligning Words," *Proceedings of the 33th Annual Meeting of the Association for Computational Linguistics(ACL-1995)*, pp. 244-251.

Dekai Wu, "Stochastic Inversion Transduction Grammars, with Application to Segmentation, Bracketing, and Alignment of Parallel Corpora," *Proceedings of 14th International Joint Conference On Artificial Intelligence*, Montreal, 1995(IJCAI-1995), pp. 1328-1335.

Dekai Wu, "Trainable Coarse Bilingual Grammars for Parallel Text Bracketing," *Proceedings of 3rd Annual Workshop on Very Large Corpora*, Cambridge, 1994, pp. 69-82.

Dekai Wu, "Stochastic Inversion Transduction Grammars and Bilingual Parsing of Parallel Corpora," *Computational Linguistics*, vol.23, No.3, 1997, pp. 377-403.

Tiejun Zhao, Muyun Yang, Fang Liu, Jianmin Yao and Hao Yu, "Statistics Based Hybrid Approach to Chinese Base Phrase Identification," *Proceedings of the 2nd Chinese Language Processing Workshop*, 2000, pp. 73-77.