Chapter 3: Bilingual NE Alignment

Bilingual NE alignment plays a vital role in extracting key information from bilingual corpora, which is essential to multilingual language processing. However, traditional statistical MT approaches, such as Brown et al. (1993), mainly focus on aligning words in parallel corpora. These approaches cannot achieve the goal of NE alignment due to their inability to handle many-to-many alignment within bilingual NE phrases. In addition, the approaches based on individual word alignment usually cannot handle Chinese well due to the ambiguity in Chinese word segmentation. To achieve the goal of aligning NE pairs, we adopt a many-to-many word alignment scheme based on statistical models and additional knowledge clues. Moreover, we get rid of the error-prone process of Chinese word segmentation by using an asymmetric approach to aligning English-Chinese NE pairs.

3.1 Problem Statement

We aim at returning a list of English-Chinese NE pairs automatically extracted from parallel corpora. As mentioned previously, without relying on NE identification in the Chinese part, we focus on aligning an English NE (which has been automatically or manually labeled) with its Chinese equivalent in the aligned sentence. Formal statement of the problem is addressed next.

Problem Statement: We are given an English-Chinese parallel corpus and a set of English (i.e. source) NE phrases $\{e\}$ labeled in the corpus. Our goal is to extract the bilingual NE pairs $\{(e, f)\}$ from the corpus, where f is the translation equivalent of e. For this purpose, each e is transformed into a set of translation candidates $\{f^*\}$ via our proposed models, such that $\{f^*\}$ is likely to contain the translation equivalent f of e.

Our solution to the above problem is described in the subsequent chapters. In contrast to previous work, the proposed method integrates a statistical phrase translation model (SPTM) (Chapter 3.3), a transliteration model (TM) (Chapter4), and other language-specific modules in a unified way. The language-specific modules include abbreviation handling (AH) (Chapter 3.4), Chinese person name recognition (CPNR) (Chapter 3.5), and acronym expansion (AE) (Chapter 3.6). A framework, which incorporates SPTM, TM, AH, CPNR and AE to extract NE pairs from parallel corpora, is presented in Chapter 5. We will describe the above modules in detail in the rest of this dissertation.

3.2 Outline of the Proposed Approach

We attempt to align source NEs with their translation equivalents in parallel corpora. The outline of the NE alignment process is shown in Figure 3.1.

(1) Employ a sentence alignment procedure and a source NE identifier to align parallel texts at the sentence level and to label NEs in each source sentence, respectively.

(2) Utilize phrase translation and transliteration models to generate a set of translation candidate strings that appear in the target sentence for each source NE *e*.

(3) Sort the translation scores associated with the set of translation candidates in descending order. Choose top-1 candidate *f* with the highest score as the target NE.

(4) Denote $\{(e, f)\}$ as NE pairs.

Figure 3.1 The outline of the NE alignment process in parallel corpora.

As we mentioned previously, NE translation involves word translation and word reordering. Thus, to translate a source NE *e* into its target NE *f*, we propose using a phrase translation model to approximate the translation score function, $Score(f \mid e)$, by decomposing $Score(f \mid e)$ into a lexical translation score function, $Score_{LEX}(f \mid e)$, and a position alignment score function, $Score_{ALI}(f \mid e)$, as show in Eq. (3.1):

$$Score(f \mid e) = Score_{LEX}(f \mid e) + Score_{ALI}(f \mid e). \tag{3.1}$$

Translating NEs also involves transliteration, especially in the case of person names. To do so, we propose a transliteration model for modeling proper noun transliteration (Chapter 4). In the rest of this thesis, we will formally explain how the phrase translation and transliteration models and language-specific knowledge sources can be integrated in a unified way to align NE pairs in parallel corpora.

3.3 Statistical Phrase Translation Model (SPTM)

In this section, we describe a modified phrase translation model and a scoring formula to determine the alignment scores of NE pairs. In the noisy channel approach to machine translation proposed by Brown et al. (1993), a source sentence *e* is fed into a noisy channel and translated to a target sentence *f*. Following Brown et al. (1993), we model the probability of translating an English phrase *e* with *l* words into a Mandarin Chinese phrase *f* with *m* words by decomposing the channel function into two independent probabilistic functions: (a) a lexical translation probability (LTP) function, $P(f_{a_i} \mid e_i)$, where e_i is the *i*-th word in *e* and e_i is aligned with f_{a_i} in *f* under the alignment *a*, and (b) a position alignment probability (PAP) function, $P(a \mid l, m)$.

Based on the above model, finding the best translation f^* for a given *e* is expressed as follows:

$$f^* = \arg\max_f P(f \mid e) = \arg\max_f \sum_a P(f, a \mid e). \qquad (3.2)$$

For simplicity, the best alignment with the highest probability is chosen to decide the most probable translation f^*, instead of summing all possible alignments a. Eq. (3.2) can, thus, be expressed as

$$f^* = \arg\max_f \max_a P(f, a \mid e) = \arg\max_f \max_a P(a \mid l, m) \times \prod_{i=1,l} P(f_{a_i} \mid e_i). \qquad (3.3)$$

In the original formulation, Brown et al. decomposed the probability of alignment for a sentence as the product of the alignment of the i-th word for i = 1 to l. Since the number l is usually quite small for phrase to phrase translation, it might be better to compute the phrase alignment probability as a whole instead of as the product of individual word alignment, $P(a_i \mid i, l, m)$. Therefore, we have

$$P(a \mid l, m) \equiv P(a_1, a_2, ..., a_l \mid l, m). \qquad (3.4)$$

For example, consider the case where the source phrase e = "Ichthyosis Concern Association" and its translation equivalent f = "關懷魚鱗癬協會." Reasonable word segmentations for f, in this case, are "關懷," "魚鱗癬," and "協會." The correct alignment is (a_1 = 2, a_2 = 1, a_3 = 3). Thus, the phrase translation probability is represented as

$$\begin{aligned} &P(\text{關懷 魚鱗癬 協會} \mid Ichthyosis\ \ Concern\ \ Association) \approx \\ &\quad P(\text{魚鱗癬} \mid Ichthyosis) \times P(\text{關懷} \mid Concern) \times \\ &\quad P(\text{協會} \mid Association) \times P(2,1,3 \mid 3,3). \end{aligned} \qquad (3.5)$$

Based on the modified formulation for alignment probability, Eq. (3.3) can be written as

$$\begin{aligned} f^* &= \arg\max_f \max_a P(f, a \mid e) \\ &= \arg\max_f \max_a P(a_1 a_2 ... a_n \mid l, m) \times \prod_{i=1,l} P(f_{a_i} \mid e_i). \end{aligned} \tag{3.6}$$

The integrated score function for the target phrase *f*, given *e*, is defined as follows by regarding the score function as a log probability function:

$$\begin{aligned} Score(f \mid e) &\equiv \max_a \log(P(a \mid l, m) \prod_{i=1,l} P(f_{a_i} \mid e_i)) \\ &= \max_a (Score_{ali}(a \mid l, m) + \sum_{i=1,l} Score_{lex}(f_{a_i} \mid e_i)). \end{aligned} \tag{3.7}$$

Accordingly, $Score_{LEX}(f \mid e)$ and $Score_{ALI}(f \mid e)$ in Eq. (3.1) can be defined as follows:

$$Score_{LEX}(f \mid e) \equiv \max_a \sum_{i=1,l} Score_{lex}(f_{a_i} \mid e_i), \tag{3.8}$$

$$Score_{ALI}(f \mid e) \equiv \max_a Score_{ali}(a \mid l, m). \tag{3.9}$$

3.3.1 An Illustrative Example

In the following, we demonstrate how we can use SPTM to generate a set of potential translation candidates from a source NE. For instance, given the source NE *e* = “Ichthyosis Concern Association,” the proposed SPTM can generate the top-10 candidates with associated scores, as shown in Figure 3.2. Some relevant probabilities for this example are shown in Tables 3.1 and 3.2. As will be describe later, the LTP and PAP probabilities in SPTM can be estimated from training corpora. In this example, the proposed SPTM can effectively help to limit the number of generated candidates.

$$
\begin{aligned}
Score_{SPTM}(\text{魚鱗癬關心協會} \mid e) &= Score_{lex}(\text{魚鱗癬} \mid Ichthyosis) + Score_{lex}(\text{關心} \mid Concern) \\
&\quad + Score_{lex}(\text{協會} \mid Association) + Score_{ali}(1,2,3 \mid 3,3) \\
&= -2.36932 \\
Score_{SPTM}(\text{魚鱗癬關懷協會} \mid e) &= Score_{lex}(\text{魚鱗癬} \mid Ichthyosis) + Score_{lex}(\text{關懷} \mid Concern) \\
&\quad + Score_{lex}(\text{協會} \mid Association) + Score_{ali}(1,2,3 \mid 3,3) \\
&= -2.49761 \\
Score_{SPTM}(\text{魚鱗癬關心會} \mid e) &= Score_{lex}(\text{魚鱗癬} \mid Ichthyosis) + Score_{lex}(\text{關心} \mid Concern) \\
&\quad + Score_{lex}(\text{會} \mid Association) + Score_{ali}(1,2,3 \mid 3,3) \\
&= -2.91506 \\
Score_{SPTM}(\text{關心魚鱗癬協會} \mid e) &= Score_{lex}(\text{魚鱗癬} \mid Ichthyosis) + Score_{lex}(\text{關心} \mid Concern) \\
&\quad + Score_{lex}(\text{協會} \mid Association) + Score_{ali}(2,1,3 \mid 3,3) \\
&= -2.94252 \\
Score_{SPTM}(\text{魚鱗癬關懷會} \mid e) &= Score_{lex}(\text{魚鱗癬} \mid Ichthyosis) + Score_{lex}(\text{關懷} \mid Concern) \\
&\quad + Score_{lex}(\text{會} \mid Association) + Score_{ali}(1,2,3 \mid 3,3) \\
&= -3.04335 \\
Score_{SPTM}(\text{關懷魚鱗癬協會} \mid e) &= Score_{lex}(\text{魚鱗癬} \mid Ichthyosis) + Score_{lex}(\text{關懷} \mid Concern) \\
&\quad + P(\text{協會} \mid Association) + Score_{ali}(2,1,3 \mid 3,3) \\
&= -3.07081 \\
Score_{SPTM}(\text{魚鱗癬協會關心} \mid e) &= Score_{lex}(\text{魚鱗癬} \mid Ichthyosis) + Score_{lex}(\text{關心} \mid Concern) \\
&\quad + Score_{lex}(\text{協會} \mid Association) + Score_{ali}(1,3,2 \mid 3,3) \\
&= -3.07162 \\
Score_{SPTM}(\text{魚鱗癬協會關懷} \mid e) &= Score_{lex}(\text{魚鱗癬} \mid Ichthyosis) + Score_{lex}(\text{關懷} \mid Concern) \\
&\quad + Score_{lex}(\text{協會} \mid Association) + Score_{ali}(1,3,2 \mid 3,3) \\
&= 3.19991 \\
Score_{SPTM}(\text{魚鱗癬關心公會} \mid e) &= Score_{lex}(\text{魚鱗癬} \mid Ichthyosis) + Score_{lex}(\text{關心} \mid Concern) \\
&\quad + Score_{lex}(\text{公會} \mid Association) + Score_{ali}(1,2,3 \mid 3,3) \\
&= -3.46717
\end{aligned}
$$

Figure 3.2 Translation candidates with associated scores.

Table 3.1 A LTP table for “Ichthyosis Concern Association.”

English word (e)	Chinese word (f)	$Score_{LEX}(f \mid e)$
Ichthyosis	魚鱗癬	-0.39793
Concern	關心	-0.84185
Concern	關懷	-0.97014
Association	協會	-0.32931
Association	會	-0.87505
Association	公會	-1.42716

Table 3.2 A PAP table for “Ichthyosis Concern Association.”

a_1	a_2	a_3	$Score_{ALI}(f \mid e)$
1	0	2	-0.57147
1	2	3	-0.80023
2	0	1	-0.82583
2	1	3	-1.37343
1	3	2	-1.50253

3.3.2 Estimation of LTP and PAP

To estimate parameters LTP and PAP of the SPTM model, an EM algorithm for maximizing the likelihood of generating the target f given e is adopted. We use a bilingual dictionary and parallel corpora as the training data for learning SPTM. First, we make initial estimates, such as by using uniform segmentation scheme. Then, we perform an iterative learning process to find the optimal phrase alignment under the current model and re-estimate parameters according to the optimal alignment just found. Further details about the process of estimating parameters can be found in Chang et al. (2001).

Although, in bilingual NE translation, a general bilingual dictionary contains translations for many common words, many domain-specific words are usually not covered, such as terminology and proper nouns. However, manually constructing bilingual lexicons is a labor-intensive and time-consuming process. Therefore, a method for automatically exploiting domain-specific bilingual lexicons from relevant bilingual corpora has also been developed (Wu and Chang, 2004).

For a source word e_i in a given source NE e, the probability $P(f_{a_i} | e_i)$ for a translation candidate f_{a_i} is estimated using a weighted average strategy as follows:

$$P(f_{a_i} | e_i) = \lambda_1 P_{gen}(f_{a_i} | e_i) + \lambda_2 P_{ne}(f_{a_i} | e_i) + \lambda_3 P_{cor}(f_{a_i} | e_i), \tag{3.10}$$

$$\lambda_1 + \lambda_2 + \lambda_3 = 1, \tag{3.11}$$

where $P_{gen}(f_{a_i} | e_i)$, $P_{ne}(f_{a_i} | e_i)$, and $P_{cor}(f_{a_i} | e_i)$ are estimated from a general bilingual dictionary, an NE-pair list, and a domain-relevant corpus, respectively, and λ_1, λ_2, and λ_3 are weighting factors to be determined empirically.

Similarly, the probability $P(a | l, m)$ is also estimated using a weighted average strategy:

$$P(a \mid l,m) = \lambda_4 P_{gen}(a \mid l,m) + \lambda_5 P_{ne}(a \mid l,m) + \lambda_6 P_{cor}(a \mid l,m), \tag{3.12}$$

$$\lambda_4 + \lambda_5 + \lambda_6 = 1, \tag{3.13}$$

where $P_{gen}(a \mid l,m)$, $P_{ne}(a \mid l,m)$, and $P_{cor}(a \mid l,m)$ are estimated from a general bilingual dictionary, an NE-pair list, and a domain-relevant corpus, respectively, and λ_4, λ_5, and λ_6 are weighting factors to be determined empirically.

3.4 Abbreviation Handling (AH)

In practice, the transformation of NEs from English to Chinese is more complicated than the transformation process mentioned above. Usually, an English NE phrase may have several equally acceptable Chinese NE candidates. For example, the NE "International Commercial Bank of China" can be translated as "中國國際商業銀行," "中國商業銀行," or "中國商銀." We can simply measure the similarity between two Chinese NE candidates when estimating the phrase translation probability. For example, a high probabilistic value for *P*("中國商業銀行"| "International Commercial Bank of China") and a high similarity measure between "中國商業銀行" and "中國商銀" imply that we should also give a high probabilistic value for *P*("中國商銀" | "International Commercial Bank of China"). Therefore, we can enhance the lexical score function by using approximate string matching (Damerau, 1964). Suppose that there is an entry (e_i, w_f) with probability p (or score c_p = log p) in the derived bilingual lexicon based on Eq. (3.10). The score function $Score_{lex}(f_{a_i} \mid e_i)$ in Eq. (3.8) can be modified as follows:

$$Score_{lex}(f_{a_i} \mid e_i) = \begin{cases} c_p, & \text{if } f_{a_i} = w_f \\ c_p \times (\frac{I}{J}) - c_{\gamma 1}, & \text{if } 0 < I < J \\ Score_{tm}(R(f_{a_i}) \mid e_i), & \text{if } I = 0 \text{ and } \\ & Score_{tm}(R(f_{a_i}) \mid e_i) \geq Thr_1 \\ c_{\gamma 2}, & \text{otherwise,} \end{cases} \tag{3.14}$$

where J is the number of Chinese characters in w_f, I is the number of matched Chinese characters between f_{a_i} and w_f, $c_{\gamma 1}$ and $c_{\gamma 2}$ are floor score values, $R(f_{a_i})$ is the romanization of f_{a_i}, $Score_{tm}(R(f_{a_i}) \mid e_i)$ is the transliteration score function for f_{a_i}, given e_i (which will be described in more detail in Chapter 4), and Thr_1 is a threshold value.

To reduce the risk of oversimplification in abbreviation handling, a bilingual gazetteer which consists of well known place names and country names in English and corresponding full and abbreviated names in Chinese is constructed manually. For example, the abbreviated forms of the NE pairs (Taiwan, 台灣) and (England, 英格蘭) are (Taiwan, 台) and (England, 英), respectively. As a common practice, “台灣” cannot be abbreviated into “灣” and “英格蘭” cannot be abbreviated into “格,” “蘭,” or “格蘭.”

Figure 3.3 shows the transformation of the NE pair (International Commercial Bank of China, 中國商銀) with approximate matching operations.

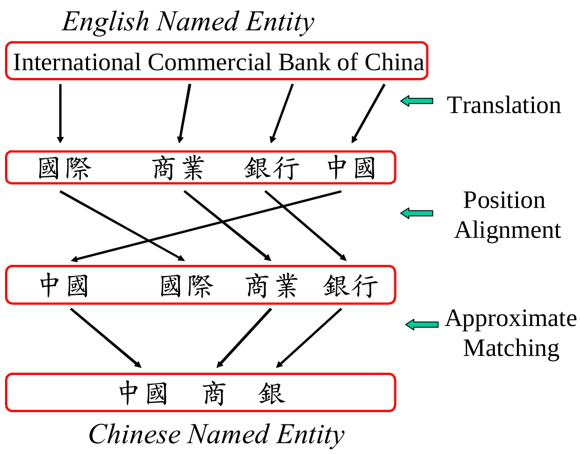

Figure 3.3 Transformation of an NE pair with approximate matching operations.

3.5 Chinese Person Name Recognition (CPNR)

In some cases, the association between the members of a bilingual NE pair is hard to obtain through only phrase translation, transliteration, and abbreviation handling. More language-dependent features, such as CPNR, can be introduced to improve the performance. The following example is taken from the magazine *Sinorama*:

English sentence:

"…But Sinorama boldly broke through, under the direction of then publisher King-yuh Chang, and under the planning of editors Wang Chi and Gypsy Chang, produced the "Greater China" special report. This expressed our concern about <u>mainland China</u> by reporting on the evaporation of <u>Tungting Lake</u> and the desertification of the <u>Huangtu Plateau</u>."

Chinese sentence:

"...，但「<u>光華</u>」勇於突破，在當時發行人<u>張京育</u>指示、總編輯<u>汪琪</u>與編輯<u>張靜茹</u>策劃下，推出「大地中國」專題，以報導<u>洞庭湖</u>日益淤淺、<u>黃土高原</u>沙漠化的現況，來表達我們對<u>大陸</u>的關懷。"

In the above example, there are seven labeled bilingual pairs of entity names. Among them, (King-yuh Chang, 張京育), (Wang Chi, 汪琪), and (Gypsy Chang, 張靜茹) are person names. In this example, (King-yuh Chang, 張京育) and (Wang Chi, 汪琪) are well aligned via the proposed phrase translation and transliteration methods. However, in (Gypsy Chang, 張靜茹), "Gypsy" is an English name which does not have any direct relationship with "靜茹," a traditional Chinese name. To deal with the mapping between a foreign name and a Chinese name in parallel corpora, we apply a Chinese person name recognizer to extract the Chinese part of the PER-typed NE.

Chinese person names consist of surnames and given names. In most cases, surnames and given names are composed of one or two characters. Our CPNR model is automatically trained from a large person name corpus consisting of one million entries. We use Chinese surnames as anchor points and then determine if the following one or two characters is a Chinese given name or not. Suppose that c_1c_2 are two subsequent Chinese characters. The decision function $d(c_1c_2)$ for the two-character given name is defined as follows:

$$d(c_1c_2) = \begin{cases} true, & if \quad P(c_1c_2 \mid GN_{12}^2) > Thr_2 \quad or \\ & \quad P(c_1 \mid GN_1^2) \times P(c_2 \mid GN_2^2) > Thr_3, \\ false, & otherwise, \end{cases} \tag{3.15}$$

where GN_{12}^2, GN_1^2, and GN_2^2 stand for the two-character given name, the first character of the two-character given name, and the second character of the two-character given name, respectively, and Thr_2 and Thr_3 are constants.

The decision function $d(c_1)$ for a single-character given name is defined as follows:

$$d(c_1) = \begin{cases} true, & if \quad P(c_1 \mid GN^1) > Thr_4, \\ false, & otherwise, \end{cases} \tag{3.16}$$

where GN^1 is the one-character given name and Thr_4 is a constant.

The threshold values Thr_2, Thr_3, and Thr_4 were empirically determined so as to let 95% of the training set pass the verification test. Since the bilingual sentences are well aligned and surnames are used as the anchor points, this approach works quite well for aligning foreign names with their corresponding Chinese names.

CPNR is applied only when the given NE is a named person and

$Score_{tm}(R(f_{a_i}) \mid e_i)$ in Eq. (3.14) is less than Thr_1. Then, given a named person, the transliteration score function is reformulated as

$$Score_{tm}(R(f_{a_i}) \mid e_i) = \begin{cases} \max\{\log(P(c_1c_2 \mid GN_{12}^2)), \log(P(c_1 \mid GN_1^2) \times P(c_2 \mid GN_2^2))\}, \\ \qquad\qquad\qquad\qquad\qquad\qquad \textit{if } d(c_1c_2) \textit{ is true}, \\ \log(P(c_1 \mid GN^1)), \qquad\qquad \textit{otherwise} \quad \textit{if } d(c_1) \textit{ is true}. \end{cases} \tag{3.17}$$

For instance, in the above example, the surname pair (Chang, 張) is detected and aligned as an anchor point. Then the potential Chinese given name "靜茹," which is the two subsequent Chinese characters following "張," is put into the proposed CPNR model to verify if it is a Chinese given name. The verification process for the Chinese given name "靜茹" based on Eq. (3.17) is shown in Figure 3.4.

(1)*Given thresholds:*
$Thr_2 = 0.0000028$
$Thr_3 = 0.000000015$
(2)*Verification:*
$P(\text{靜茹} \mid GN_{12}^2) = 0.0000038 > Thr_2$
$P(\text{靜} \mid GN_1^2) \times P(\text{茹} \mid GN_2^2) = 0.0077648 \times 0.0011218 > Thr_3$
(3)*Decision:*
$d(\text{靜茹}) = true$

Figure 3.4 Verification process for the Chinese given name "靜茹."

Thus, the NE pair (Gypsy Chang, 張靜茹) can be correctly aligned through combining CPNR and SPTM, as shown in Figure 3.5.

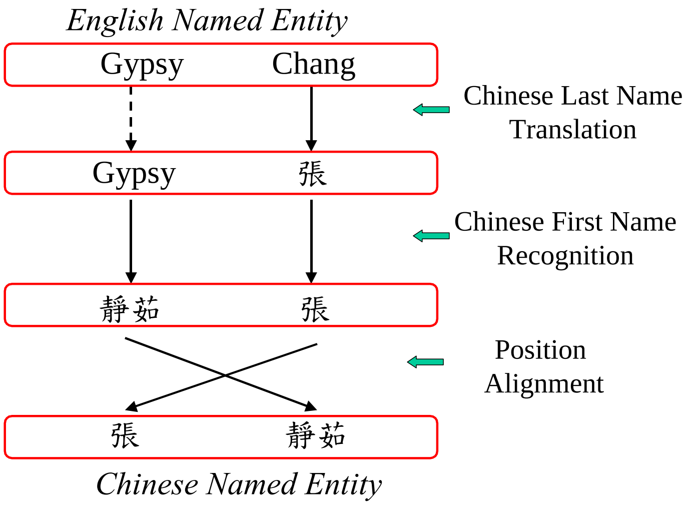

Figure 3.5 Alignment of the NE pair (Gypsy Chang, 張靜茹).

3.6 Acronym Expansion (AE)

An acronym is an "abbreviation consisting of letters that form a word"; an abbreviation is "a short form of a word or phrase" (Macmillan, 2002). Thus, an acronym is a particular form of abbreviation, which is read as a single word. For example, "*WHO*" is an acronym for "*W*orld *H*ealth *O*rganization." Acronyms carry significant information in corpora. However, they are frequently created in a domain specific manner and cannot be completely covered by any existing dictionaries. If acronym-expansion pairs are not mined, it is highly difficult to align NE pairs in corpora. As far as we know, lots of researchers focus on acquiring acronym-expansion pairs (Taghva and Gilbreth, 1999; Yeates, 1999; Larkey et al., 2000; Yeates et al., 2000; Schwartz and Hearst, 2003). None of the previous methods concerns the work of acquiring the acronym-translation pairs from parallel texts simultaneously. In this dissertation, we propose an innovative approach to acquire acronym-translation pairs. A two-stage scheme is adopted in this approach. In the first stage, in which an acronym-expansion list is compiled, a simple algorithm is applied to extract a possible expansion candidate for each acronym in the source sentence. Thus, an acronym can be mapped to its expansion subsequently. In the second stage, we apply the proposed bilingual NE alignment algorithm to extract the translation of the expansion from the aligned target sentence. The strategy of acronym expansion is based on the observation that an acronym and its expansion typically appear within a sentence in a specific pattern. This usually occurs in one of the following two canonical forms:

- a pair of parentheses around an acronym, such as “World Health Organization (WHO)”;
- a pair of parentheses around an expansion, such as “WHO (World Health Organization).”

In this study, we use the algorithm proposed by Schwartz and Hearst (2003) to extract acronym-expansion pairs from the corpora we employed. For those acronyms cannot be found in the canonical forms from the corpora, we use a web search engine to extract the texts containing the canonical forms. For example, we can easily find the following instances for the acronym “GATT” in the canonical forms via Google search engine:

- “The General Agreement on Tariffs and Trade (GATT) was first signed in 1947.”
- “The General Agreement on Tariffs and Trade (GATT) covers internal trade in goods.”
- “The Uruguay Round Agreements Act (URAA) of 1994 implements the Uruguay Round General Agreement on Tariffs and Trade (GATT), which includes an agreement…”

Then, the pair (GATT, General Agreement on Tariffs and Trade) can be identified as an acronym-expansion pair. After the acronym-expansion process, a list of acronym-expansion pairs is automatically constructed. Table 3.3 shows a partial list of acronym-expansion pairs extracted from *Sinorama* Magazine. This list can be applied in subsequent bilingual NE extraction. For example, in Table 3.3, the expansion of “CLA” is “Council of Labor Affairs.” A step-by-step diagram for aligning the pair (CLA, 勞委會) is shown in Figure 3.6.

Table 3.3 A partial list of acronym-expansion pairs automatically extracted from the corpora.

Acronym	Text	Expansion
EPA	Last year the ROC Environmental Protection Administration (EPA)…	Environmental Protection Administration
CCPD	The Council for Cultural Planning and Development (CCPD) mediated negotiations…	Council for Cultural Planning and Development
CSC	Three years ago when Yeh Man-sheng, president of China Shipbuilding Corp. (CSC)…	China Shipbuilding Corp.
AOS	…a 32-year-old researcher at the Institute of Far Eastern Studies of the Academy of Sciences (AOS)…	Academy of Sciences
NTNU	…states Professor Wang Ying of the National Taiwan Normal University (NTNU)…	National Taiwan Normal University
NTU	…says Li Ling-ling, associate professor of zoology at National Taiwan University (NTU) Department of Biology.	National Taiwan University
CLA	Under a chorus of pleas, the Council of Labor Affairs (CLA)…	Council of Labor Affairs
SEM	…some doctors of emergency medicine went further, establishing the Society of Emergency Medicine (SEM)…	Society of Emergency Medicine
WTO	Hong Kong is already a member of the World Trade Organization (WTO)…	World Trade Organization
WHO	We could link up with the World Health Organization (WHO)…	World Health Organization
MOE	Therefore, as Ministry of Education (MOE) officials put it, students will be focusing on Taiwan for one third of their six years in primary school.	Ministry of Education
NHIB	On the other hand, medical institutions protest that the National Health Insurance Bureau (NHIB)…	National Health Insurance Bureau
MOEA	The amended Taiwan Energy Policy, produced by the Energy Commission of the Ministry of Economic Affairs (MOEA)…	Ministry of Economic Affairs

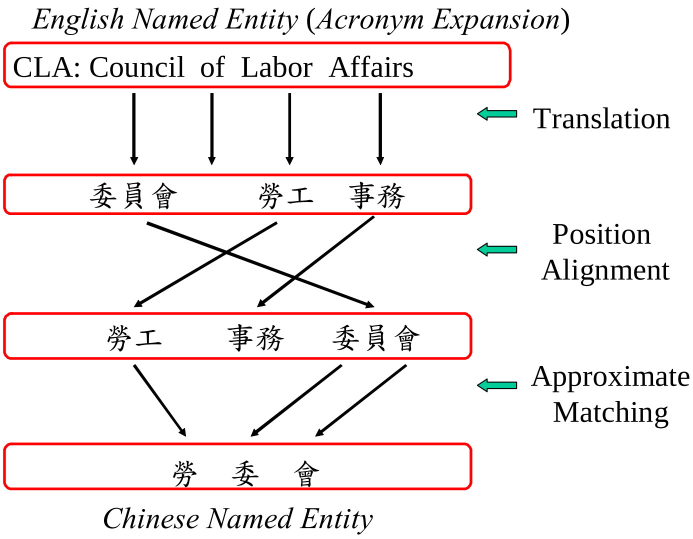

Figure 3.6 Alignment of the NE pair (CLA, 勞委會).