

# The Transfer of Anaphors in Translation

HSIN-HSI CHEN

National Taiwan University, Republic of China

## Abstract

This paper adopts Government-Binding (GB) Theory and some Chinese-specific conditions to explain the various uses of anaphors such as the reflexive, pronoun, PRO, trace, pro and variable in Mandarin Chinese. The transfer of each anaphor is specified for Chinese-English and English-Chinese machine translation systems. We identify the situations in which the coreferential relationship between an anaphor and its antecedent is useful for transfer. Chinese reflexive is translated according to its function and its antecedent. In some situations, the interpretation of pronouns in English is ambiguous at sentence level, but its Chinese counterpart has only one reading. One-to-many interpretation in Chinese-English translation may be disambiguated by the readers. However, many-to-one interpretation in English-Chinese translation cannot be treated without pronoun resolution. PRO has direct translation between English and Chinese. For translating topic-comment structures and ba-constructions, the moved constituent is taken back from its landing site to its empty site (trace), and then the structural transfer is done. For relative clauses and bei-constructions, the position occupied by trace in English is still left empty in Chinese. Because variable does not appear in English, the other English anaphors are used during the transfer of Chinese pro and variable. The translation of them should involve the discourse knowledge.

## 1. Introduction

Different languages use different anaphors to denote their referents in the sentences. For improving the translation quality, to explain the various functions of anaphors is indispensable for machine translation system design (Kay, 1986; Noord *et al.*, 1989; Wada, 1990; Nakaiwa *et al.*, 1992). For example, the Chinese reflexive '自己' (ziji), an overt anaphor, may be translated into a possessive pronoun (see ex1) or a reflexive pronoun (see ex2) in English. Its number and gender also depend on the antecedent. The first selects the third singular masculine pronoun, and the second uses the third plural pronoun.

- (cx1) 約翰愛自己的太太。  
John loves his wife.

- (cx2) 他們對自己沒信心。  
They don't have confidence in themselves.

Another apparent phenomenon is that Chinese pronouns do not have case markers. The Chinese word '他' may be used as subject form (he, see ex3) or object form (him, see ex4). Which form is selected affects the lexical transfer.

**Correspondence.** Hsin-Hsi Chen, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan 10764, Republic of China. Internet: 'hh\_chen@nlg.csie.ntu.edu.tw' or 'hh\_chen@csmn.csie.ntu.edu.tw'

Literary and Linguistic Computing, Vol. 7, No. 4, 1992

- (cx3) 他看見那個人。  
He saw that man.

- (cx4) 那個人看見他。  
That man saw him.

Besides the overt anaphors, empty anaphors are also used in natural language statements. Based on the freedom of their uses, natural languages may be classified into two types—cool and hot (Ross, 1982). Chinese and English are typical cool and hot languages, respectively (Huang, 1984). The following shows their differences:

- (cx5)  $e_i$  來了。  
\*  $e_i$  came. / He came.
- (cx6) 約翰看見了  $e_i$ 。  
\* John saw  $e_i$ . / John saw her.
- (cx7)  $e_i$  看見了  $e_j$ 。  
\*  $e_i$  saw  $e_j$ . / He saw her.
- (cx8) 約翰<sub>i</sub>說  $e_{(i,j)}$  看見了瑪琳。  
\* John<sub>i</sub> said that  $e_{(i,j)}$  saw Mary. /  
John<sub>i</sub> said that he<sub>(i,j)</sub> saw Mary.
- (cx9) 約翰說瑪琳看見了  $e_i$ 。  
\* John said that Mary saw  $e_i$ . /  
John said that Mary saw him.
- (cx10) 約翰<sub>i</sub>設法  $e_i$  來。  
John<sub>i</sub> tried  $e_i$  to come.
- (cx11) 我告訴過約翰他們要這些書。  
I have told John that they want these books.
- (cx12) [<sub>NP</sub> 約翰]<sub>i</sub>, 我告訴過  $e_i$   
[<sub>S</sub> 他們要這些書]。  
(cx13) 我告訴過 [<sub>NP</sub> 約翰]  
[<sub>S</sub> [<sub>NP</sub> 這些書]<sub>j</sub> 他們要  $e_j$ ]。  
(cx14) [<sub>NP</sub> 約翰]<sub>i</sub>, 我告訴過  $e_i$   
[<sub>S</sub> [<sub>NP</sub> 這些書]<sub>j</sub> 他們要  $e_j$ ]。  
(cx15) [<sub>NP</sub> 約翰]<sub>i</sub>, [<sub>NP</sub> 這些書]<sub>j</sub>,  
我告訴過  $e_i$  [<sub>S</sub> 他們要  $e_j$ ]。  
(cx16) 約翰看見  $e_i$  的人<sub>i</sub> 來了。  
The man<sub>i</sub> who John met  $e_i$  came.

Chinese demonstrates the maximal freedom for the uses of empty anaphors. It allows the empty anaphor to refer to some element that is mentioned in the context. Sentences (ex5)–(ex9) are such examples. In English, if we do not place some overt pronoun at the empty position, the sentence is unacceptable. The second English sentences in (ex5)–(ex9) specify the use of a pronoun. The empty anaphor in (ex10) is co-indexed with a nominal element in the same sentence. Sentences (ex11)–(ex15) show that Chinese is a highly

© Oxford University Press 1992

flexible language. The same meaning is represented in many different patterns. They correspond to the same English sentence 'I have told John that they want these books'. Sentence (ex16) specifies the directions of relativization in Chinese and in English are different. The former is rightward, and the latter leftward. The other constituents that may produce this type of empty anaphors in Chinese are '把' (ba)-construction, '被' (bei)-construction, *etc.* For the correct transfer, it is necessary to determine the relationship between the moved constituent and trace during Chinese-English translation.

This paper will consider several types of anaphors in Mandarin Chinese such as reflexive, pronoun, PRO, trace, pro and variable, and discuss their transfers in English-Chinese and Chinese-English machine translation systems. We will identify the situations in which the coreferential relationship between an anaphor and its antecedent is useful for transfer, and give analyses of some special anaphors.

## 2. Overt Anaphors

### 2.1 Reflexives

Chinese reflexives have two functions: anaphoric and intensifying (Li and Thompson, 1981; Tang, 1989). Anaphoric reflexive occurs in argument positions, and intensifying reflexive appears in non-argument positions. The following shows the transfer rules for each function:

1. anaphoric ziji ('自己')
  - (a) as a subject of an NP  
 (pronoun) ziji + de + N  $\longleftrightarrow$   
 possessive pronoun + own + N  
 For example, 約翰愛(他)自己的太太。  
 (John loves his own wife.)
  - (b) as a direct object, an indirect object or an oblique object  
 (pronoun) ziji  $\longleftrightarrow$  reflexive  
 For example, 瑪琍在責備(她)自己。  
 (Mary is blaming herself.)
2. intensifying ziji ('自己')
  - (a) between the subject NP and the predicate  
 Subject + ziji + predicate  $\longleftrightarrow$   
 Subject + reflexive + predicate /  
 Subject + predicate + by reflexive  
 For example, 他們自己會補衣服。  
 (They themselves can mend clothes/  
 They can mend clothes by themselves.)
  - (b) between the auxiliary verb and the main verb  
 auxiliary verb + ziji + VP  $\longleftrightarrow$   
 auxiliary verb + VP + by reflexive  
 For example, 瑪琍會自己回家。  
 (Mary will go home by herself.)
3. intensifying and anaphoric ziji ('自己') in a sentence  
 Neglect the intensifying ziji in the translation.  
 For example, 他們會自己照顧自己。  
 (They will take care of themselves.)

The above rules show that the translations of reflexives from English to Chinese are easier than those transla-

tions from Chinese to English. In the former translation, lexicon gives the corresponding Chinese words. However, the latter translation must select suitable possessive pronouns or reflexive pronouns. A reflexive resolution algorithm is needed to find the coreferential relationship between the Chinese reflexive and its antecedent. Binding Principle A (Sells, 1985:68) and some Chinese parameters such as subject-orientation condition and sub-command condition (Tang, 1989) are used. Binding Principle A suggests the candidates for the anaphor. The Chinese specific conditions try to find such an NP in the subject position as top level as possible. Algorithm 1 shows the resolution of Chinese reflexives.

---

#### Algorithm 1. Reflexive Resolution Algorithm.

**Input:** An annotated parse tree.

**Output:** The same index for Chinese reflexive and its referent.

**Method:**

1. Traverse the parse tree to find out the reflexive  
 If the reflexive functions as an anaphoric element, go to step (2). Otherwise, stop.
  - 2: [Binding Principle A and subject-orientation Condition]  
 Find the subject that c-commands the reflexive.  
 [sub-command Condition]
    - (a) The subject is a pronoun or a simple noun.  
 Check its semantic feature. If it is an animate noun, then we co-index it with the reflexive, else the sentence is semantic-illegal.
    - (b) The subject is a complex noun phrase.  
 Check the semantic feature of the head noun first. If it meets the condition, then we co-index the head noun with the reflexive. Otherwise, go to step (2) to check the associative phrase or modifying phrase
- 

### 2.2 Pronominal Anaphor

Unlike English pronominal anaphors, Chinese pronouns do not have case markers. The Chinese word '他' ('她') may be used as subject form (he/she) or object form (him/her) or possessive form (his/her) in serial noun constructions like '他爸爸' (his father). Which translation is correct depends on its position in the sentence (see ex3 and ex4). Does it mean the pronoun resolution is not necessary for machine translation? Compare the constraints of the interpretations of pronouns on English and Chinese. 'A pronoun may not c-command its antecedent' is a sufficient condition for English (Reinhart, 1976), whereas it is necessary but insufficient for Chinese (Huang, 1982). Teng (1985) formulated a condition on pronominal anaphora in Chinese: A pronoun may not maximal-c-command its antecedent. The definition of maximal-c-command is shown as follows:

A maximal-c-commands B if and only if

1. A c-commands B, or
2. The minimal maximal NP dominating A and not immediately dominated by another NP c-commands B, or
3. The first maximal projection dominating A c-commands B.

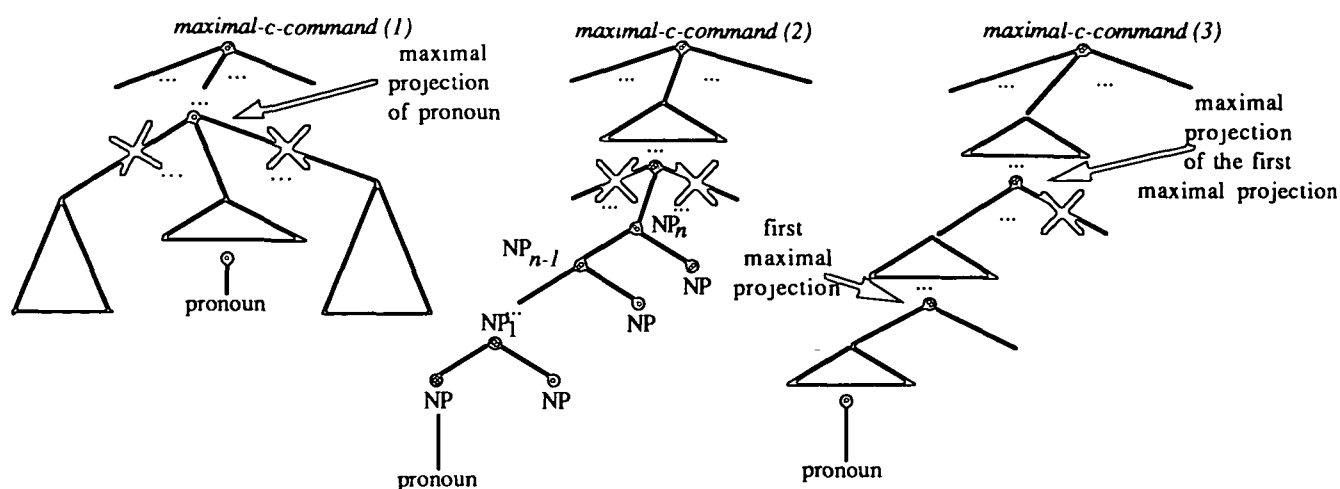


Fig. 1 Condition on pronominal anaphor in Chinese

Besides Binding Principle B, the condition also excludes the possible referent of a pronominal anaphor. Figure 1 demonstrates this concept. The element(s) in the maximal-c-command domain are deleted. One should note that the interpretation of Chinese pronoun is more restrictive than that of the English one on the basis of this Chinese-specific condition (Table 1). Assume the English and the Chinese structures are not changed during translation. The second and the third rows of this table tell us there are a one-to-many interpretation and a many-to-one interpretation for Chinese-English and English-Chinese translations at sentence level, respectively. Sentences (ex17) and (ex18) depict the difference. For Chinese-English translation, i.e. one-to-many interpretation, two approaches may be adopted. The first one is to use pronoun resolution from discourse level to identify the antecedent, e.g. Jack in (ex19). The second one is to use direct translation, i.e. (ex18), and to let the readers themselves disambiguate its use by discourse knowledge. For English-Chinese translation, i.e. many-to-one interpretation, the pronoun resolution is necessary. The translation may be either (ex17) or (ex20), but not both.

- (ex17) 他<sub>i</sub> 的 媽媽 喜歡 約翰<sub>j</sub> 。
- (ex18) His<sub>(i,j)</sub> mother likes John<sub>j</sub> .
- (ex19) Jack's<sub>i</sub> mother likes John<sub>j</sub> .
- (ex20) 約翰<sub>j</sub> 的 媽媽 喜歡 約翰<sub>j</sub> 。

GB Theory and Chinese specific condition only tell us what it is not rather than what it is. Algorithm 2 depicts pronoun resolution at sentence level. It is just a weak method, so that we cannot find the coreferential relationship without a context model.

#### Algorithm 2. Pronoun Resolution Algorithm

**Input:** An annotated parse tree.

**Output:** A list of candidates that can serve as a referent of the pronominal anaphor.

#### Method:

1. Traverse the parse tree to find the occurrence of the pronominal anaphor. If it does not exist, stop. Otherwise, go to next step.
2. [Binding Principle B]  
Find the first NP (or S) upward on the path where a pronoun is found. Remove the nominal element dominated by such a category because it c-commands the pronoun.
3. [Condition on Pronominal Anaphor in Chinese]
  - (a) [maximal-c-command 1]  
Find the first VP, PP, or S' that dominates the pronoun. Cut the other branch(es) of the dominating node.
  - (b) [maximal-c-command 2]  
Skip the intermediate successive NP(s), stop at the topmost NP and then do the same job as (3a).
  - (c) [maximal-c-command 3]  
Find the first NP, VP, PP, or S' that dominates

Table 1 Coreferential relationships in Chinese and in English

coreferential relationship maximal-c-command \ language	Chinese	English
(1) is satisfied	A and B cannot be co-indexed	A and B cannot be co-indexed
(2) is satisfied	A and B cannot be co-indexed	A and B may be co-indexed
(3) is satisfied	A and B cannot be co-indexed	A and B may be co-indexed
(1), (2) and (3) are not satisfied	A and B may be co-indexed	A and B may be co-indexed

the pronoun. Do the operation again. Cut the other branch(es) of the dominating node. The only exception is that we stop at the first VP if we traverse upward from the right branch.

4. Traverse the remaining parse tree and look for the nominal elements whose semantic features (e.g. type, gender, number, etc.) coincide with the pronoun. Those elements are put into the candidate list.

### 3. Empty Anaphors

#### 3.1 Identification of Empty Anaphors

There are four types of empty anaphors in Mandarin Chinese: PRO, trace, pro and variable. The antecedents of them appear at different referential levels (sentence or context). PRO and trace belong to sentence level, variable context level, and pro either level. Empty anaphors are different from overt ones in that they have no semantic contents. To determine their existences and identify their types is an indispensable step during transfer. The following proposes an algorithm to determine the uses of the empty anaphors in Chinese. We use the subcategory of verb to explain whether there are empty anaphors in the input sentence. If there exists, apply the steps shown below in sequence:

1. If the matrix verb belongs to the types (Li, 1988) shown below, then the empty anaphor is PRO.

- (a) Control Verbs.

For example, "企圖" (attempt), "打算" (intend), "計劃" (plan), "開始" (begin), "準備" (prepare), "允許" (permit), "鼓勵" (encourage), etc.

- (b) Clause Union Verbs.

For example, "設法" (try), "努力" (strive), "勸" (persuade), "逼" (force), "請" (invite), "要求" (ask), "讓" (allow), "使" (cause), "派" (send), "叫" (tell), etc.

- (c) Exceptional Case Marking Verbs.

For example, "要" (want), "喜歡" (like), "討厭" (hate), etc.

The following general rule explains which constituent is referred by this type of empty anaphors:

*If the verb subcategorizes an object, then the object is the antecedent. Otherwise, the subject is co-indexed with PRO.* For example, the verb "要" (want) may have an object or not. The index assignment obeys the above rule (see ex21 and ex22). The verb "設法" (try) has only one use (see ex10). It also satisfies the rule

(ex21) 我<sub>i</sub> 要 e<sub>j</sub> 去學校。

(I<sub>i</sub> want e<sub>j</sub> to go to school.)

(ex22) 我要他<sub>i</sub> e<sub>j</sub> 去學校。

(I want him<sub>i</sub> e<sub>j</sub> to go to school.)

2. If an empty anaphor is generated by topicalization or relativization under C-Command and Subjacency conditions, then it is a trace. Sentences (ex11)–(ex16) meet

these two movement constraints. The ba-constructions and the bei-constructions are treated in a similar way.

3. If the matrix verb takes finite clauses (Li, 1988), then the empty anaphor is pro. For example, "同意" (consent), "告訴" (tell), "控告" (accuse), "答應" (promise), etc. The determination of the antecedent is pragmatic-controlled. Sentence (ex8) shows two possible interpretations at different levels.

4. If the sentence is well-formed, it is a variable. This case is irregular. Under zero-topic postulation, it is co-indexed with a zero-topic that was mentioned before. For example, given the sentence

(ex23) 約翰看見瑪琍了嗎?  
(Did John see Mary?)

sentences (ex6) and (ex7) are two possible answers. In another question:

(ex24) 比爾看見誰了?  
(Who did Bill see?)

sentence (ex8) "約翰<sub>i</sub> 說 e<sub>j</sub> 看見了瑪琍。" (John<sub>i</sub> said that he<sub>j</sub> saw Mary) is a possible answer. The empty anaphor refers to "比爾" (Bill) in the context. We may use feature structures collected from the context to identify the coreferential relationship.

#### 3.2 Empty Anaphors and Bilingual Translation

The identification of various empty anaphors is helpful to the lexical transfer and the structural transfer. The following shows the key ideas:

1. PRO

According to X' theory (Sells, 1985), we can formulate subcategorization frame(s) for each verb. This frame tells us whether there is a PRO, and which constituent it refers to if it does exist. Thus, it is simple for both English–Chinese and Chinese–English translations. The verb "要" (want) in sentences (ex17) and (ex18) is an example. It has two subcategorization frames:

[NP<sub>i</sub> "要" (want) PRO<sub>i</sub> VP] and  
[NP "要" (want) NP<sub>i</sub> PRO<sub>i</sub> VP].

2. trace

The generation of *trace* is considered to be a movement transformation in GB Theory. A sentence in English may have one-to-many mapping to many different structures in Chinese. The correspondence of English sentence (ex11) to Chinese sentences (ex11)–(ex15) is such a mapping. English–Chinese translation is simple because we can select the direct mapping, i.e., (ex11). Chinese–English translation is not trivial. We must apply movement transformation first to put the moved constituent back to its original position, and then do the structural transfer, which is a reversal operation of the former English–Chinese translation. The empty position is filled by an overt constituent in the topic-comment structures. Sentence (ex25) shows the similar transfer of the ba-constructions. The anaphoric relationship between the moved constituent and trace is resolved by a Government-Binding parser for Mandarin Chinese (Chen, 1990).

- (ex25) 我今天把三本書<sub>i</sub>賣<sub>e<sub>i</sub></sub>了。  $\xrightarrow{\text{move-}\alpha}$   
 我今天賣三本書了。  $\xrightarrow{\text{structural transfer}}$   
 I sold three books today.

For relative clauses and bei-constructions, the position occupied by trace in English is still left empty in Chinese (see ex16 and ex26).

- (ex26) 錢包<sub>i</sub>被偷<sub>e<sub>i</sub></sub>了。  $\xrightarrow{\text{structural transfer}}$   
 The wallet<sub>i</sub> was stolen <sub>e<sub>i</sub></sub>.

### 3. pro

Consider Chinese-English translation. The counter part of pro in English may be PRO (see ex27) or personal pronoun (see ex28). The interpretation is ambiguous at the sentence level. Which pronoun is selected depends on discourse knowledge.

- (ex27) 我<sub>i</sub>答應<sub>pro<sub>i</sub></sub>不再遲到。  $\xrightarrow{\text{transfer}}$   
 I<sub>i</sub> promise PRO<sub>i</sub> not to be late again.

- (ex28) 約翰<sub>i</sub>告訴瑪瑞<sub>j</sub> <sub>pro</sub> 明天不能去了。

$\xrightarrow{\text{transfer}}$   
 John<sub>i</sub> told mary<sub>j</sub> that <sub>pro</sub> cannot go tomorrow.  
 (where <sub>pro</sub> = he<sub>i</sub>, she<sub>j</sub>, or they<sub>(i+j)</sub>)

### 4. variable

In this interpretation, the variable refers to a nominal element in the context. Therefore, a pronoun is necessary during Chinese-English translation. How to select a suitable pronoun (he/him, she/her, they/them, it, ...) is still a research problem.

## 4. A Statistical Analysis of Zero and Pronominal Anaphors

### 4.1 Methodology

According to the above discussions, the transfer of some anaphors needs to identify the coreferential relationship at the discourse level. However, the resolution of anaphora depends on various kinds of knowledge (Hirst, 1981). This section adopts a statistical approach to analyse the behaviours of zero anaphors (pro and variable) and pronominal anaphors, and touches on their contributions to machine translations. A bilingual corpus which is selected from the Bilingual Sections of *Reader's Digest* in Chinese version is used. Each anaphor in the corpus is marked a tag of the form:

(index1, type, environment, syntactic role, semantic case, index2).

Each item is defined in the following:

#### 1. index1 and index2

The indices index1 and index2 denote an anaphor and its antecedent respectively.

#### 2. anaphor types

The referring expressions are classified into several types: PRO, zero, pronoun, proper noun, definite NP,

indefinite NP, bare NP, demonstrative NP, quantifier NP, quantifier, etc.

### 3. syntactic environments

Since the environment of an anaphor may influence its anaphoric expression, the factor of environment is taken into consideration. Example (ex29) signals causality.

- (ex29) He is loved by all, because he is honest.  
 因為他<sub>i</sub>很誠實，所以<sub>e<sub>i</sub></sub>受人喜愛。

Some optional rules are set: coordination, subordination, predicate construction, verb category, etc.

### 4. syntactic roles

The specification of an anaphor containing its grammatical function or syntactic role is based on the categories: subject, object, prepositional object, etc.

### 5. semantic cases

The semantic case of each anaphor in the sentence which it happens is given. There are several kinds of semantic cases to be used: agent, patient, source, goal, etc.

A paragraph selected from *Reader's Digest* (1991) is tagged below as an example:

"Albie<sup>^1</sup>,proper,\_,MS,AG,\_,^ arrives early in the morning<sup>^2</sup>,DNP,\_,MPO,TI,\_,^, ^3,PRO,parti,C1S,AG,1<sup>^</sup> kicking up swirls<sup>^4</sup>,BP,\_,C1O1,\_,\_,^ of dust<sup>^5</sup>,BP,\_,C1PO,TH,\_,^ with his<sup>^6</sup>,POSSA,\_,C1PO,\_,3<sup>^</sup> pickup truck<sup>^7</sup>,ProN-NP,\_,C1PO,IN,\_,^ He<sup>^8</sup>,ProN,\_,MS,AG,6<sup>^</sup> unloads his<sup>^9</sup>,POSSA,\_,MO1,\_,\_,8<sup>^</sup> tools<sup>^10</sup>,ProN-NP,\_,MO1,TH,\_,^ and<sup>^11</sup>,zero,AND,C1S,TH,9<sup>^</sup> starts<sup>^12</sup>,PRO,T2,C2S,AG,11<sup>^</sup> to work."

"艾爾比<sup>^1</sup>,proper,\_,MS,AG,\_,^一大清早<sup>^2</sup>,INP,\_,MAD,TI,\_,^就來到，他的<sup>^3</sup>,POSSA,\_,C1S,\_,1<sup>^</sup>小貨車<sup>^4</sup>,ProN-NP,\_,C1S,AG,\_,^揚起一陣塵土<sup>^5</sup>,QNP,\_,C1O1,TH,\_,^。他<sup>^6</sup>,ProN,\_,MS,AG,3<sup>^</sup>把工具<sup>^7</sup>,BP,BA,MPO,TH,\_,^從車上<sup>^8</sup>,LOC,\_,MPO,SO,4<sup>^</sup>拿下來，<sup>^9</sup>,zero,AND,C1S,AG,6<sup>^</sup>開始<sup>^10</sup>,PRO,T2,C2S,AG,9<sup>^</sup>工作。"

### 4.2 Results

After the bilingual texts are tagged, computers are used to summarize the rules. Table 2 presents some statistics for four texts. The coreferential relationship may occur in the same clause, previous clause, same sentence (across more than one clause in a sentence), previous sentence, same paragraph (across more than one sentence in a paragraph), or previous paragraph. Only zero anaphor (ZA) and pronominal anaphor (PA) are discussed. Both ZAs and PAs in English favour a position whenever the antecedent occurs in a short distance. The majority of occurrences of ZAs centre on the previous clause. The difference between them is mainly reflected in that ZAs are preferred with the same sentence; the previous sentence position favours personal pronouns, but at a somewhat higher rate (40 per cent). In a word, pronouns mostly occur in the previous clause of a sentence or the previous sentence. However, possessive pronouns (e.g. *his* hat) occur mostly in the same clause. The referential distance between a possessive pronoun and its antecedent is very short. ZAs in Chinese tend to occur in the previous clause, but they

**Table 2** The referential distance between an anaphor and its antecedent in English (E) and in Chinese (C)

distance → type ↓	same clause	previous clause	same sentence	previous sentence	same paragraph	previous paragraph
zero anaphor (E)	1 4.35%	22 95.65%				
personal pronoun (E)	6 7.89%	23 30.26%	7 9.21%	30 39.47%	3 3.95%	7 9.21%
possessive pronoun (E)	26 56.52%	8 17.39%	3 6.52%	7 15.22%	1 2.17%	1 2.17%
zero anaphor (C)		80 72.73%	14 12.73%	13 11.82%	2 1.82%	1 0.91%
personal pronoun (C)	5 7.81%	19 29.69%	7 10.94%	26 40.63%	1 1.56%	6 9.38%
possessive pronoun (C)	3 23.08%	2 15.38%	3 23.08%	4 30.77%		1 7.69%

are also distributed in the same sentence or the previous sentence. Chinese ZAs outnumber English ones. The reason may be that pronouns are not repeatable in Chinese, while they are in English. Instead of repeating overt pronouns, zero pronouns should be used. With respect to pronouns, the distribution of PAs in Chinese is similar to those in English. Yet the number of possessive pronouns is lower than that of English ones. One possible reason is that Chinese uses a lot of bare NP instead of possessive NP. As a result, there are a lot of absent genitive pronouns in Chinese discourse.

Our quantitative measures show that given the choice of subject (S) vs. object (O) or prepositional object (PO) as antecedent of a ZA or a PA, the subject is most preferred both in English and in Chinese as illustrated in Table 3, no matter what its semantic role. We can get the following hierarchy of the syntactic role of an antecedent: S>O>PO.

**Table 3** The syntactic roles of antecedent in English (E) and in Chinese (C)

Type	S	O	PO
ZA (E)	21 (91.30%)	2 (8.70%)	
PA (E)	80 (65.57%)	24 (19.67%)	18 (14.75%)
ZA (C)	90 (81.82%)	15 (13.64%)	5 (4.55%)
PA (C)	57 (74.03%)	12 (15.58%)	8 (10.39%)

In closer analysis, Table 4 shows the relationship among environment, and the syntactic roles of ZA and its antecedents. English ZAs occur in the coordinate

**Table 4** The environment, and the syntactic roles of ZA and its antecedent

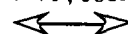
ZA ↓	antecedent →	S	O	PO
Coord (E)	S	21		
	O		2	
Coord (C)	S	63	4	2
	O		3	
Subord (C)	S	8	1	1
	O	1	2	
SV (C)	S	7	1	
Bei (C)	PO	1		
irregular (C)	S	7	2	1
	O	3	2	1

constructions. The anaphor and its referent have the same syntactic role. As for Chinese, ZAs may appear in much more sentence types. Besides the regular constructions such as coordinate constructions, subordinate constructions, serial verb constructions and bei-constructions, there are some irregular types as mentioned in Section 3.1. The frequency of the former is larger than that of the latter, i.e., 6:1, and the conjunctions are at a higher rates (near 80%). Summing up, the referential distance, the syntactic environment and the syntactic roles are important in referent-tracking. We may consider the sentence types first, and then try the subject, the object and oblique positions in sequence within the referential boundary.

Under the statistic analysis, two more transfer rules are formulated for English-Chinese and English-Chinese translation:

1. Coordination

- (a)  $NP_i + VP, \text{coordinate} + ZA_i + VP$  (Chinese)

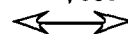


$NP_i + VP, \text{coordinate} + ZA_i + VP$  (English)

e.g. 艾爾比給棚子漆上綠色的油漆，讓它吹乾。

(Albie painted the shed green and let it dry.)

- (b)  $NP_i + VP, \text{coordinate} + ZA_i - NP + VP$  (Chinese)



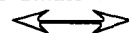
$NP_i + VP, \text{coordinate} + PA_i - NP + VP$  (English)

e.g. 瑪麗很高，而且腳很修長。

(Mary is tall, and her legs are long.)

2. Subordination

- (a) subordinate +  $ZA_i + VP, NP_i + VP$  (Chinese)



subordinate +  $PA_i + VP, NP_i + VP$  (English)

e.g. 雖然已七十歲了，他還像以前那樣的強壯。

(He is as strong as ever, although he is seventy.)

- (b) subordinate +  $ZA_i - NP + VP, NP_i + VP$

(Chinese)



subordinate +  $PA_i - NP + VP, NP_i + VP$  (English)

e.g. 雖然外表看起來很冷酷，他內心還是一位仁慈的男人。

(Though his appearance looks cold, he is a kind man at heart.)

The phenomena as (1b) and (2b) cause problems in machine translation because components not overtly indicated in the source language (i.e. Chinese) become compulsory elements in the target language (i.e. English).

### 5. Concluding Remarks

High quality translation is an ultimate goal for machine translation systems. This paper identifies the differences of anaphors between Chinese and English, and addresses the importance of anaphors in machine translation. Table 5 lists the lexical transfer of anaphors between English and Chinese. Chinese reflexive is translated according to its function and its antecedent. In some situations, the interpretation of pronouns in English is ambiguous at sentence level, and its Chinese translation has only one reading. One-to-many interpretation in Chinese-English translation may be disambiguated by the readers. Many-to-one interpretation

in English-Chinese translation cannot be treated without pronoun resolution. For empty anaphors, we propose an algorithm to determine their uses. PRO has direct translation between Chinese and English, however, trace, pro and variable need further processing during Chinese-English translation. The moved constituent is taken back from its landing site to its empty site (trace). And then the structural transfer is done. Discourse knowledge should be involved to identify their antecedent during translating pro and variable. Posting-editing the translating results containing the anaphors whose antecedents appear in the discourse is impracticable because it takes much time for editors to understand the text. Automatic resolution of anaphors is indispensable for machine translation system design. Referential distance, syntactic roles and other knowledge are considered to select a suitable pronoun from context information to replace the empty anaphors.

**Table 5** The lexical transfer of anaphors in English-Chinese (E-C) and Chinese-English (C-E) translations

Chinese → English ↓	nominal anaphor	reflexive pronoun	personal pronoun	possessive pronoun	PRO	trace	pro	variable
nominal anaphor	E-C C-E					C-E	C-E(*)	C-E(*)
reflexive pronoun		E-C C-E						
personal pronoun	E-C		E-C C-E			C-E	E-C C-E	C-E
possessive pronoun	E-C	C-E (own)	C-E	E-C C-E				
PRO					E-C C-E		C-E	
trace						E-C C-E		
pro							E-C C-E	

\*: Nominal element is used for rhetorical reason, e.g. ‘他吃了 e’ is translated into ‘It ate the fish’ instead of ‘It ate it’.

### Acknowledgement

Research on this paper was partially supported by National Science Council grant NSC81-0408-E002-007.

### References

Chen, H.-H. (1990). A Logic-Based Government-Binding Parser for Mandarin Chinese, *Proceedings of 13th COLING*, 2: 48–59 Helsinki, Finland  
Hirst, G. (1981). *Anaphora in Natural Language Understanding A Survey*. Springer-Verlag, New York.  
Huang, C.-T. (1982). *Logical Relations in Chinese and the Theory of Grammar*. Ph.D. dissertation, Massachusetts Institute of Technology.

— (1984). On the Distribution and Reference of Empty Pronouns, *Lingustic Inquiry*, 15.4: 531–74.  
Kay, M. (1986). Forum on Machine Translation: Machine Translation Will Not Work. *Proceedings of 24th Annual Meeting of ACL*, 268. New York.  
Li, M -D. (1988). *Anaphoric Structures of Chinese*. Student Book Company, Taipei, Taiwan.  
Li, C N. and Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. University of California Press.  
Nakaiwa, H. and Ikehara, S. (1992). Zero Pronoun Resolution in a Japanese to English Machine Translation System Using Verbal Semantic Attributes. *Proceedings of Applied Natural Language Processing*, 201–8. Trento, Italy  
Noord, G V., Dorrepaal, J., and Arnold, D. (1989). An

- Approach to Sentence-Level Anaphora in Machine Translation *Proceedings of the European Chapter of ACL*, 299–307. Manchester, England.
- Reader's Digest* (1991) A Master's Touch, 52 6: 150–2. Bilingual Section.
- Reinhart, T. (1976). *The Syntactic Domain of Anaphora*. Ph D. dissertation, Massachusetts Institute of Technology.
- Ross, J. R. (1982) Pronoun Deleting Process in German. Paper presented at the Annual Meeting of the Linguistic Society of American, San Diego, California
- Sells, P. (1985). *Lectures on Contemporary Syntactic Theories*. Stanford, Center for the Study of Language and Information.
- Tang, C.-C. (1989) Chinese Reflexives, *Natural Language and Linguistic Theory*, 7: 93–121.
- Teng, T.-H. (1985) Pronominal Anaphora in Chinese Master thesis, Graduate Institute of English, National Taiwan Normal University, Taipei, Taiwan, Republic of China.
- Wada, H (1990). Discourse Processing in MT: Problems in Pronominal Translation. *Proceedings of 13th COLING*, 1. 73–5. Helsinki, Finland.