

# Referent Resolution for Zero Pronouns in Thai

Wirote Aroonmanakun

Chulalongkorn University

## 1. INTRODUCTION

Resolving zero pronouns is a major problem in developing a natural language understanding (NLU) system for Thai. Since subject and object pronouns in Thai can be omitted from a sentence, an NLU system must be capable of identifying the missing subjects or objects in the sentence. This process of identifying referents for zero pronouns, which is a part of referent resolution<sup>1</sup> process, is the concern of this paper.

Basically, I assume that referent resolution for zero pronouns can be done at two levels, the sentence level and the discourse level. Some zero pronouns can be resolved on the basis of sentence-grammar principles. These principles are implemented as a part of syntactic/semantic parser. As for zero pronouns that cannot be resolved by a sentence grammar, discourse principles will be used.

The sentence grammar that is adopted in this paper is that of government and binding theory (Chomsky, 1981, 1982, 1986a, 1986b). In this framework, zero pronouns are analyzed as empty categories. An overview of empty categories and related principles is given in section 2. In section 3, zero pronouns in different syntactic structures in Thai and the domain in which the government and binding theory is applicable are discussed. Section 4 is concerned with the centering theory that is adopted as the basic discourse principle for resolving zero pronouns at the discourse level.

## 2. GOVERNMENT AND BINDING THEORY

This section begins with an overview of empty categories in government and binding theory. Then, principles that relate to the process of identifying antecedents for empty categories are discussed in the following order: binding theory, bounding theory, and control theory. It concludes with the process of identifying empty categories and their antecedents.

### 2.1 Empty Categories

In government and binding theory, zero pronouns are analyzed as empty categories (ECs). An EC is considered a gap in the s-structure. A sentence contains an EC whenever it does not have a lexical item in a position that is assigned a theta-role. ECs are categorized into four types: *wh*-trace (variable), NP-trace, *pro*, and *PRO*, with

---

<sup>1</sup>The term “referent” used in this paper refers to discourse referents (Karttunen, 1976), not referents in the world. A discourse referent is an entity that is evoked from the discourse context.

respect to features of pronominal and anaphor<sup>2</sup> as below:

	-anaphor	+anaphor
-pronominal	<i>wh</i> -trace(variable)	NP-trace
+pronominal	pro	PRO

A trace is normally analyzed as a result of move-alpha.<sup>3</sup> An NP-trace is left when an NP is moved from one A-position<sup>4</sup> to another A-position, while a *wh*-trace is left when an argument is moved from an A-position to an A'-position. Examples of NP and *wh*-traces are shown below.

(a) NP-trace: John<sub>1</sub> seems *t*<sub>I</sub> to be nice

(b) *wh*-trace: What books<sub>1</sub> do you like *t*<sub>I</sub>? (Cook, 1988, p. 163)

In (a), *John* is moved to the subject position, which is an A-position, while in (b), *what books* is moved to the specifier of CP, which is an A'-position. When an argument is moved, a chain of movement is created. A trace and its antecedent are coindexed within the chain by the movement.

(c) pro: Italian: Sono il tricheco

English: \*(I) am the walrus (Cook, 1988, p. 38)

The EC "pro" is established from the fact that some languages such as Italian and Spanish can have null subjects in declarative sentences, while other languages like English cannot (Cook, 1988, p. 38). This fact reflects one parameter of universal grammar, which is that a language can be either pro-drop or non-pro-drop. A language in which "pro" exists is called a pro-drop language.

(d) PRO: John<sub>1</sub> wants PRO<sub>I</sub> to go

It is time PRO to go (Cook, 1988, p.164)

Since English is a non-pro-drop language and an English sentence must have a subject (as a result of extended projection principle (Chomsky, 1982)), the embedded S in the example above must contain another kind of EC, which is not "pro" or trace, but PRO. PRO can be either A-bound or A-free. PRO in the first example is bound to *John*, while PRO in the second example is free.

## 2.2 Binding Theory

Binding theory (Chomsky, 1982,1986a) is a subtheory that explains anaphoric relations between NPs in a sentence. The three principles of binding theory are:

- A. An anaphor (+anaphor) is bound<sup>5</sup> in its governing category.<sup>6</sup>
- B. A pronominal (+pronominal) is free in its governing category.
- C. An R-expression (-anaphor,-pronominal) is free.

<sup>2</sup>These features are also used to categorize overt NPs into reflexives, reciprocals, pronominals, and r-expressions.

<sup>3</sup>Move-alpha is a syntactic process that moves a constituent to another place. The movement is restricted by other principles, such as the subjacency principle.

<sup>4</sup>"A-positions—positions which may in principle be filled by arguments laid down in lexical entries..." (Cook, 1988, p. 113). On the other hand, A'-positions refer to non A-positions.

<sup>5</sup>A binds B if A c-commands B, and they are coindexed. Binding theory refers to only A-binding. It means that A and B must be in an A-position.

<sup>6</sup>"...a governing category [of a] is a maximal projection containing both a subject and a lexical category governing a" (Chomsky, 1986a, p. 169).

The coindexing in sentences below can be explained by these principles.

- (a). John<sub>1</sub> likes himself<sub>1</sub>
- (b). John<sub>1</sub> likes him\*<sub>1,2</sub>
- (c). John<sub>1</sub> believes that Peter<sub>2</sub> likes him<sub>1,3</sub>
- (d). \*John<sub>1</sub> believes that Mary<sub>2</sub> likes himself<sub>1</sub>
- (e). He<sub>1</sub> thinks that John\*<sub>1,2</sub> is lazy.

In (a), since *himself* is +anaphor, it must be bound in its governing category. Thus, *himself* must have the same index as *John*. On the other hand, since *him* is +pronominal, it must be free. Thus, it cannot be bound to *John*. It will have a different index from *John*. In (c), principle-B prohibits *him* from being coindexed with *Peter*, but does not exclude the coindexing between *him* and *John*, since *John* is not in the governing category of *him*. In (d), *himself* is not bound within its governing category, which is the embedded sentence. Thus, this sentence is ungrammatical because it violates principle-A. In (e), *John* is not bound to *he* because principle-C prohibits *John*, which is an R-expression, from being bound in any category.

The binding theory applies not only to overt NPs but also to covert NPs, or ECs. Thus, we can conclude the following facts about ECs:

- As a result of being +anaphor, an NP-trace must observe principle-A. Thus, it is bound in the governing category.
- As a result of being +pronominal, a “pro” observes principle-B. It is free in the governing category.
- Since a variable is both –pronominal and –anaphor, it observes principle-C. Thus, it is free in all governing categories.
- Since PRO is both +anaphor and +pronominal, it should observe both principle-A and principle-B. But it is impossible for PRO to be free and bound in the governing category at the same time. However, the contradiction does not really occur because PRO is ungoverned (Chomsky, 1986a). Therefore, it does not have any governing category.

In sum, the binding theory provide us the coindexation between NP-traces and their antecedents. It does not directly explain the coindexation of other ECs. What it does is suggest what cannot be coindexed with “pro” and *wh*-trace.

### 2.3 Bounding Theory

While binding theory explains a coindexation between arguments in A position (A-binding), bounding theory deals with A'-binding in which an argument in A-position is bound to an argument in A'-position. Bounding theory relates to only one type of ECs, variables or *wh*-traces. It explains coindexation between variables and their antecedents, and the sequence of *wh*-movements.

- (a) Who<sub>1</sub> do you think [ John likes t<sub>1</sub> ]?
- (b) That report which<sub>1</sub> I filed e<sub>1</sub> without PRO reading e<sub>1</sub>

(Lasnik & Uriagereka, 1988, p.78)

Coindexation between *who* and *wh*-trace in (a) is an example of A'-binding resulting from *wh*-movements. In (a), A'-binding is generated by application of *wh*-

movements. To observe the subjacency principle<sup>7</sup>, *who* is moved to the specifier of the embedding clause first, then, it is moved to the specifier of the main clause. *Who* and *wh*-trace get the same index as a result of the movements.

Even though A'-binding is usually a result of *wh*-movement, some variables are not directly related to *wh*-movements. An EC is considered a variable whenever it is locally A'-bound. This is shown in the example of "parasitic gaps" in (b). In (b), the first EC is a result of *wh*-movement and is determined as a *wh*-trace. The last EC is not related to *wh*-movement because the first trace does not c-command it. Rather, it is a variable because it is locally A'-bound by *which*.<sup>8</sup>

The bounding theory explains the coindexation between variables or *wh*-traces and their antecedents. If there is a movement, coindexing is a direct result of the movement. Otherwise, coindexing is determined by A'-binding.

## 2.4 Control Theory

Control theory (Chomsky, 1981, 1986a) is a sub-theory that determines antecedents for PROs. PROs can be either obligatory PROs or arbitrary PROs. Obligatory PROs are bound in a sentence, while arbitrary PROs are free. Control theory assigns antecedents for obligatory PROs, which can be either subject control or object control.

(a) John<sub>1</sub> asked PRO<sub>1</sub> to go.

(b) John<sub>1</sub> asked Peter<sub>2</sub> PRO<sub>2</sub> to go.

(c) It is time PRO to go.

(Cook, 1988, p. 162)

Obligatory PROs in the infinitive clauses in (a) and (b) are bound to *John* and *Peter* respectively. On the other hand, PRO in (c) is arbitrary and free.

## 2.5 Determining ECs and Their Antecedents

The status of ECs, whether they are NP-traces, variables, pros, or PROs, is functionally determined by their roles in the sentence. "An EC is a variable if it is in an A-position and is locally A'-bound. An EC in an A-position that is not a variable is an anaphor. Note that if not a variable, a pronoun is either free or locally A'-bound by an antecedent with an independent q-role" (Chomsky, 1982, p. 35).

Thus, if an EC is A'-bound by an element in a non-theta-position, and observes the locality condition (subjacency principle), it is a variable. If an EC is A'-bound by an element in a non-theta-position, and observes the subjacency principle, it is an NP-trace. If an EC receives independent theta-role, it can be either PRO or pro. Since English is not a pro-drop language, the only possible category is PRO. But in a pro-

<sup>7</sup>The subjacency principle limits the distance of movement so that an argument cannot move across more than one bounding node. Bounding nodes may vary in different languages. Bounding nodes for English are IP and NP.

<sup>8</sup>This structure maybe analyzed in another way such that the last EC is A'-bound by a null operator. The last EC is a variable because it is locally A'-bound by a null operator. (It is still licensed by the first trace)

The paper [O1 [that you filed t1

[O2 [without [PRO reading t2]]]]]

(Law, 1991, p. 324)

drop language like Chinese and Thai, an additional criterion is needed to distinguish between PRO and *pro*. Since PRO is ungoverned, it cannot receive a case. Thus, an overt NP cannot occur in the same position as PRO because it must be governed to receive a case. Therefore, an EC that occurs in the position where a lexical item cannot be present is a PRO. An EC that is not a trace or a PRO is a *pro*.

Coindexation between ECs and their antecedents can be determined by the principles discussed above. Coindexing between NP-traces, *wh*-traces and their antecedents is created directly by the movement. For a variable that does not involve move-alpha, coindexing is a result of A'-binding in the s-structure. It will get the same index as its binder. For obligatory PROs, their antecedents are determined by the control theory. Obligatory PROs will receive the same index as their antecedent. On the other hand, arbitrary PROs and *pros* cannot be assigned antecedents directly by these principles. The principle-B of the binding theory provides us only what are not antecedents of *pros*, not what are antecedents of *pros*.

### 3. EMPTY CATEGORIES IN THAI

Thai is a *pro*-drop language. It can have a null subject in declarative sentences. According to Pingkarawat (1989) and Hoonchamlong (1991), Thai can have both subject “*pro*” and object “*pro*.” In brief, Pingkarawat argued against Huang (1984), who proposed that object “*pro*” does not exist in any language. Pingkarawat's argument is strengthened by the analysis of Hoonchamlong, who provided evidence from topicalization and relative clauses to show that Thai can have object *pros*.

In this section, analyses of different syntactic structures in Thai will be reviewed. Determining statuses of ECs and their antecedents in different constructions will be discussed.

#### 3.1 Relative Clauses

An EC in relative clauses in Thai is analyzed as a null resumptive pronoun rather than a variable (Hoonchamlong, 1991). Since an EC does not observe the subadjacency principle, it is not a variable at the s-structure. Rather, Thai can be categorized as a language having a parameter of *wh*-in-situ in relativization (see Demirdache, 1991). In this view, an EC is a variable at the level of LF. Coindexation between an EC and the head noun, then, can be succeeded by a rule of predication.

Unlike English, an EC in Thai relative clauses cannot be analyzed as a variable resulting from *wh*-movement. As pointed by Hoonchamlong, if an EC is a result of move-alpha, the subadjacency principle will be violated. Her example is provided below:

'wan nǐ chán hěn [NP nák-khǎn [S' thǐ  
 day this I see writer THAT  
 [S1 nít bòk nòy [S' wâa [S2 dæŋ kam-laŋ ?àan  
 Nit tell Noy COMP Daeng PROG read  
 [NP nǎŋ-sǔu [S' thǐ [S3 EC wǐ-can EC  
 book THAT criticize

- (a) 'Today I saw the writer<sub>i</sub> that Nit told Noy that Daeng was reading the book<sub>j</sub> that (he)<sub>i</sub> criticized EC<sub>j</sub>.'
- (b) 'Today I saw the writer<sub>i</sub> that Nit told Noy that Daeng was reading the book<sub>j</sub> that EC<sub>j</sub> criticized (him)<sub>i</sub>.' (Hoonchamlong. 1991, p. 187)

In addition, resumptive pronouns in Thai can alternate quite freely with gaps in relative clauses. Thus, it is possible to view an EC as a null resumptive pronoun. Since resumptive pronouns in relative clauses are base-generated and not related to move-alpha at s-structure, subjacency is not relevant in relative clauses. This analysis corresponds to Demirdache's proposal (1991), who proposes that a language can have a parameter of *wh*-in-situ not only in question-formation but also in relativization.

In Demirdache's view, for some languages, resumptive pronouns are in-situ at s-structure, and move at LF. But, for some languages, *wh*-movements always apply in relative clauses at s-structure, not at LF. In this analysis, Thai would be in the first group, while English would be in the second group. The use of resumptive pronouns in English is very marginal. They are used in a sentence in which a gap occur because of subjacency violation (Georgopoulos, 1991). Following this analysis, a resumptive pronoun in Thai would move to the Spec of CP and leave a trace at LF. An EC in relative clauses, which is a null resumptive pronoun, will be a variable at LF.

Coindexation between the head noun and an EC in a relative clause can be succeeded by the predication rule (see Law, 1991, Chap. 5). According to Browning (1987), relative operators must move to the Spec of CP to satisfy the licensing condition for subject-predicate relation. Since an EC in relative clauses in Thai is bound to an empty operator at LF (from *wh*-movement), it will have the same index as the operator. And the empty operator gets the same index as the head noun by the predication rule. Therefore, the EC will receive the same index as the head noun.

### 3.2 Topicalization

Topicalization can be analyzed in a similar way to relative clauses (see Hoonchamlong, 1991). An EC in a topicalized sentence is not a trace in s-structure because the movement would violate the subjacency principle. Rather, the EC is a covert pronoun, or pro. In this view, topicalization shares the same structure and distribution as left-dislocation. Both constructions are analyzed as base-generated. No movement is involved at s-structure. The difference between them is a result of the difference between covert and overt pronouns.

According to Hoonchamlong, topicalization in Thai does not involve move-alpha. If an EC is a trace resulting from movement, it would violate the subjacency principle. In addition, topicalization does not exhibit Strong Cross-over effect.<sup>9</sup> In the example below, the subjacency principle will be violated if EC is a trace. (The relation between

<sup>9</sup>Strong Cross-over is a phenomenon in which "one of a pair of coreferential expressions crosses over another via Wh-movement" (Lasnik&Uriagaraka, 1988, p. 41). A chain between *Who* and *wh*-trace, in the sentence below, crosses over *he*. The sentence below is ungrammatical because *wh*-trace is also A-bound by *he* (violation of principle-C).

\*Who<sub>1</sub> does he<sub>1</sub> think Mary likes t<sub>1</sub>

*kháw* or EC and the antecedent TOP crosses two bounding nodes, S1 and NP1.)<sup>10</sup>

[TOP sudaa ná][S1 chǎn dâ-yin [NP1 khâaw [S' wâa [S2 cɔɔn p̄n̄  
 Suda TM I hear news COMP John just  
 phaa khǎw/EC pay rooŋ-phà-yaa-baan m̄aa-cháaw ní]]]]  
 take s/he go hospital morning this  
 ‘Suda, I heard the news that John just took her/EC to the hospital this morning.’  
 (Hoonchamlong, 1991, p. 93)

The example below indicates that ECs in topicalized sentences do not exhibit Strong Cross-over effect. The EC in the example below can be A-bound by *kháw*. Thus, it cannot be a variable. (If it is a variable, it will violate principle-C.)

[TOP khra-yi ná] [S1 kháw<sub>i</sub> khít wâa [S2 EC<sub>i</sub> chà-ná ]]  
 who TM he think COMP win  
 ‘Who, he thought that won?’ (Hoonchamlong, 1991, p. 198)

Since Thai topicalization is analyzed as a left dislocation structure, coindexation between an EC and the topic NP can be done by the predication rule in the same way as relativization. An EC in topic construction would receive the same index as the topic-NP at LF by the predication rule. (It is bound by the topic NP at LF.)

### 3.3 Serial Verb Constructions

Serial verb constructions are (Baker, 1989, p. 513) “constructions in which a sequence of verbs appears in what seems to be a single clause. Usually there is only one tense/aspect specification for the whole chain of verbs.” Serial verb constructions usually contain fewer overt NPs than the number of arguments subcategorized by all verbs in the construction. Missing arguments can be analyzed at least in two ways. They may be analyzed as ECs that are coindexed with overt NPs, or they may be analyzed as sharing arguments.

#### 3.3.1 Object Sharing

According to Baker (1989), serial verb constructions may not contain an EC. Rather, overt NPs in the sentence are shared by different verbs. For example, *Amba* in the sentence below is an internal argument of both *naki* and *kiri*. It receives the theta-role theme from both verbs,<sup>11</sup> while *Kofi* receives the theta-role Agent from both verbs. The structure of this sentence is represented below.

Sranan: Kofi naki Amba kiri  
 Kori hit Amba kill  
 ‘Kofi struck Amba dead.’ (Baker, 1989, p. 516)  
 [CP Kofi [IP [VP [V' [V naki] [NP Amba] [V' [V Kiri]]]]]]

<sup>10</sup>It is assumed here that Thai has the same bounding nodes as English, NP and S.

<sup>11</sup>This analysis is possible on the modification of theta-theory so that an argument can receive more than one theta-role in certain conditions: “...most current versions [of theta-theory] allow an argument to receive more than one  $\theta$ -role as long as all its  $\theta$ -roles are assigned to the same structural position” (Baker, 1989, p. 521).

Serial verb constructions in Thai can be analyzed in a similar way. Examples of a sentence in which the second verb is a transitive verb are shown in (a) and (b).

(a) [CP *kháw* [IP [VP [V' [V *wâat*] [NP *rûup*] [V' [V *khăay*]] ]]]]

he paint picture sell

'He paints a picture and sells it.'

(b) [CP *kháw* [IP [VP [V' [V *cháy*] [NP *míit*] [V' [V *tàt*] [NP *núa*]] ]]]]

he use knife cut meat

'He uses a knife to cut meat.'

(c) [CP *dam* [IP [VP [V' [V *tii*] [NP *ηuu*] [V' [V *taay*]] ]]]]

Dam strike snake die

'Dam strikes a snake dead.'

Example (b) has the same structure as (a), except that it subcategorizes for one more NP. In (a), *picture* receives the theta-role theme from both *paint* and *sell*. And both verbs assign the external theta-role Agent to *he*. In (b), *knife* receives the theta-role theme from *use* and the theta-role instrument from *cut*, while *meat* gets the theta-role theme from *cut*. And both verbs assign the theta-role agent to *he*. Unlike (a) and (b), (c) is an example in which the second verb is an intransitive verb. *Dam* receives only one theta-role from *hit* while *snake* receives two theta-roles from *hit* and from *die*.

(d) [CP *kháw* [IP [VP [V' [V' [V1 *kin*] [NP1 *khâaw*]] [V' [V2 *?im*]] ]]]]

he eat rice full

'He ate and became full.'

Examples (a)–(c) indicate object sharing in serial verb constructions. However, the serial verb constructions do not necessarily involve object sharing. For example, serial verbs in (d) do not share the same object. Rather, they share only the same subject. The structure of this sentence type is different from the structure above it. In this structure, V2 cannot assign a theta-role to NP1 because NP1 is not a sister node of V2.

### 3.3.2 ECs in Serial Verb Constructions

Contrary to the foregoing analysis, serial verb constructions may be analyzed in the way that ECs are in the structure. These ECs are coreferential with arguments in the structure. In this analysis, an EC is obligatorily coindexed with an argument in the sentence. The Sranan example in 3.3.1, when analyzed in this way, may look like this:

(1) Kofi [VP hit *Amba*<sub>i</sub> [V' kill *pro*<sub>i</sub>]]

(2) Kofi [VP hit *Amba*<sub>i</sub> [XP O<sub>i</sub> [VP kill *t*<sub>i</sub>]]] (Baker, 1989. p. 518)

The missing object of the second verb can be a *pro* that is coindexed with the object *Amba* as in (1), or it can be a variable resulting from movement of a null operator as in (2). This analysis is similar to the analysis of complement clauses in Pingkarawat (1989). The example below shows that the EC in the complement clause is a PRO, since that position is not governed and cannot be assigned case. (An overt NP cannot occur in that position.) It is coindexed with *kháw* because *yàak* is a subject-control verb.

(a) *kháw yàak* [S PRO *kin khâaw*]

he want eat rice

'He would like to eat.'



Which analysis is suitable for serial verb constructions in Thai is not discussed here. Whatever the analysis is, it seems to be that coindexation in serial verb construction is always constant, and does not depend on the discourse. In other words, coreferent in serial verb constructions can be resolved by principles available within the government and binding theory.

#### 4. REFERENT RESOLUTION AT THE DISCOURSE LEVEL

The government and binding theory discussed above provides us some principles for referent resolution at the sentence level. As we have seen, some zero pronouns, those which are categorized as traces and obligatory PROs, can be resolved by some principles in the grammar. But some zero pronouns, those which are categorized as pros and arbitrary PROs, cannot be resolved by any principle in the theory. Their antecedents are identified at the discourse level. In this section, the centering theory is adopted as the basis of discourse principle for resolving these zero pronouns.

##### 4.1 The Centering Theory

A discourse can be analyzed as a structure of discourse segments (Grosz & Sidner, 1986). A discourse segment is a group of utterances that are locally coherent. The centering theory (Grosz, Joshi, & Weinstein, 1983, 1986) is a computational model that accounts for the local coherence in a discourse segment. The analysis is based on the discovery that different NP forms signify different cognitive status of discourse entities (see Gundel, Hedberg, & Zacharski, 1993). An entity that is in focus usually contains less information in itself. It is normally realized as a pronoun or a zero pronoun (see Gundel et al., 1993; Givón, 1983). Thus, a referent of a pronoun or a zero pronoun can be identified from salience of discourse entities. In other words, if we can keep track of discourse entities that are in focus, we should be able to identify the referent of a pronoun or a zero pronoun. The process of keeping track of salient entities is generally called focusing.

Centering is one of the focusing mechanisms. It exhibits coherence in a discourse segment in terms of centers. Centers are discourse entities that serve to link utterances in a segment. It is assumed in the theory that an utterance contains two kinds of centers: backward looking center (Cb) and forward looking center (Cf). An utterance can have many Cfs, but it can have only one Cb. Cfs are ordered according to discourse salience. One of the Cfs would be the Cb of the utterance. The highest rank of Cf would be a preferred Cb (Cp) of the next utterance. Constraints and rules of the centering theory are stated below:

Constraints:

For each  $U_j$  in a discourse segment  $U_1, \dots, U_m$ :

1. There is precisely one Cb.
2. Every element of  $Cf(U_j)$  must be realized<sup>12</sup> in  $U_j$ .

---

<sup>12</sup>An utterance  $U$  (of some phrase, not necessarily a full clause), realizes  $c$  if  $c$  is an element of the situation described by  $U$ , or  $c$  is the semantic interpretation of some subpart of  $U$  (Walker, Iida, & Cote, 1990):

3. The center,  $Cb(U_i)$ , is the highest-ranked element of  $Cf(U_{i-1})$  that is realized in  $U_i$ .

Rules:

For each  $U_i$  in a discourse segment  $U_1, \dots, U_m$ :

1. If some element of  $Cf(U_{i-1})$  is realized as a pronoun in  $U_i$ , then so is  $Cb(U_i)$ .
2. Transition states are ordered. CONTINUING is preferred to RETAINING is preferred to SHIFTING-1 is preferred to SHIFTING.

(Walker et al., 1990, p. 2)

Transition states are determined from realization of Cbs as below:

Continuing:  $Cb(U_i) = Cb(U_{i-1})$  and  $Cb(U_i) = Cp(U_i)$

Retaining:  $Cb(U_i) = Cb(U_{i-1})$  and  $Cb(U_i) \lt \! \! \! \lt Cp(U_i)$

Shifting-1:  $Cb(U_i) \lt \! \! \! \lt Cb(U_{i-1})$  and  $Cb(U_i) = Cp(U_i)$

Shifting:  $Cb(U_i) \lt \! \! \! \lt Cb(U_{i-1})$  and  $Cb(U_i) \lt \! \! \! \lt Cp(U_i)$

The centering rule states that if an utterance contains one or more pronouns, one of them must be the Cb. The Cb of an utterance is determined from the highest rank of the previous utterance's Cfs that are realized in the current utterance. Ranks of Cfs are determined from syntactic properties and preferred order of transition state (see example (a) in section 4.2). These constraints and rules are used as the basis for identifying referents for pronouns or zero pronouns (see 4.2). The centering rules can explain why the following discourse (from Grosz et al., 1986) is unacceptable. U3 is unacceptable because it violates the first centering rule. Since *John* is the highest Cfs (of U2) realized in U3, it must be the Cb. But it is not realized as a pronoun while the other Cf, *Mike*, is realized as a pronoun.

U1) John<sub>i</sub> wanted to go for a ride yesterday

$Cf(U1) = \{John\}$

U2) He<sub>i</sub> called up Mike<sub>j</sub>

$Cb(U2) = John, Cf(U2) = \{John, Mike\}$

U3) He<sub>j</sub> was annoyed by John<sub>i</sub>'s call.

## 4.2 Centering in Thai

Since zero pronouns contains less information than pronouns, zero pronouns are assumed here to be more focused than pronouns. In this paper, only zero pronouns are considered, and assumed as the basic form for the centering rule. The difference between zero pronouns and pronouns in the centering, if any, is not discussed in this paper. The centering rule for Thai can be stated as below:

Rules:

For each  $U_i$  in a discourse segment  $U_1, \dots, U_m$ :

1. If some element of  $Cf(U_{i-1})$  is realized as a zero pronoun in  $U_i$ , then so is  $Cb(U_i)$ .

2. Transition states are ordered. CONTINUING is preferred to RETAINING is preferred to SHIFTING-1 is preferred to SHIFTING.

One of the major issues in applying the centering algorithm to the resolution of zero pronouns in Thai is to determine the order of Cfs. Following the analysis that subject NPs in many languages are more prominent than other NPs (Givón, 1983), it is

assumed here that subject NPs in Thai also have a higher rank than object NPs. However, ranking of NPs in other positions is not discussed here. Further research is needed to determine the order of Cfs with respect to other syntactic positions.

Referent resolution for zero pronouns, especially for pros and arbitrary PROs, can be resolved by applying the centering algorithm. For example, zero pronouns in the following discourse can be resolved by applying the centering theory. Since U1 has only one entity, Daeng is the Cb and is the only member of Cf list. Thus, Daeng will be the Cb of U2. Then, Daeng would be the antecedent of the zero pronoun since it is the only possible referent. Since Daeng is the highest Cf realized in U2, it would be the Cb of U3. There are two possibilities of ordering Cfs, Cf1 and Cf2. But Cf1 is preferred to Cf2 because Cf1 represents a continuing state, while Cf2 represents a retention state. Thus, the centering theory predicts that the zero subject in U3 would refer to Daeng, while the zero object would refer to Dam.

- (a) U1: dæŋ pay paa-tii m̂a-waan  
 Daeng go party yesterday  
 'Daeng went to a party yesterday.'  
 Cb: Daeng  
 Cf: {Daeng}
- U2: [Z] d̂ay r̂u-ĉak k̂ap dam  
 ASP. meet with Dam  
 '(He) met Dam'  
 Cb: Daeng  
 Cf: {Daeng, Dam}
- U3: [Z] k̂o l̂æy chuan [Z] pay duu n̂aŋ  
 CONJ then invite go see movie  
 '(He) invited (Dam) to go to a movie.'  
 Cb: Daeng  
 Cf1: {Daeng, Dam} C  
 Cf2: {Dam, Daeng} R

However, centering cannot eliminate all ambiguities in a discourse. Ambiguity may arise when the first utterance contains more than one entity. For example, U1 in the example below contains two entities, Daeng and Dam. Either one of them can be the Cb of U2 because U2 contains only one entity. Thus, the zero pronoun in U2 can be interpreted either as Daeng or Dam.

- (b) U1: dæŋ maa ĉə dam  
 Daeng come meet Dam  
 'Daeng met with Dam.'  
 Cb: ?  
 Cf: {Daeng, Dam}
- U2: t̂oŋ-tĥi [Z] kam-l̂aŋ d̂əŋ l̂ēn  
 while PROG. walk ASP.  
 'While (he) was walking'  
 Cb1: Daeng  
 Cf1: {Daeng}  
 Cb2: Dam  
 Cf2: {Dam}

In addition, the centering theory is still in the developing stage. It needs more research on different discourse genres to strengthen its explanation power. For example, subordination is normally assumed to behave like a separate utterance in the centering theory (see Walker et al., 1990; Kameyama, 1985). But this claim is unlikely to hold in Thai. The example below indicates that the subordinate clause does not behave like a separate utterance, but rather like a part of the main clause (U1). Since U2 is ambiguous (as shown above), U3 can be interpreted in two ways. If Cb(U2)=Daeng, Cf11 is preferred to Cf12. If Cb(U2) = Dam, Cf21 is preferred to Cf22. However, Cf21 is unlikely to be acceptable. The zero subject in U3 should refer to Daeng rather than Dam. On the other hand, if we analyze U2 as a part of U1, U3 will be the next utterance of U1. In this view, the Cb of U3 can be only Daeng regardless of the ambiguity in U2.<sup>13</sup> And U3 can have only one preferred interpretation. This suggests that subordination may not be analyzed as an individual utterance.

- (c) U1: dæŋ maa cəə dam  
 Daeng come meet Dam  
 'Daeng met with Dam.'  
 Cb: ?  
 Cf: {Daeng, Dam}
- U2 tɔ̀n-thii [Z] kam-ləŋ dən lən  
 while PROG. walk ASP.  
 'While (he) was walking'  
 Cb1: Daeng  
 Cf1: {Daeng}  
 Cb2: Dam  
 Cf2: {Dam}
- U3: [Z] kə ləy chuan [Z] pay duu nəŋ  
 CONJ then invite go see movie  
 '(He), then, invited (him) to go to a movie.'  
 Cb1: Daeng  
 Cf11: {Daeng, Dam} C  
 Cf12: {Dam, Daeng} R  
 Cb2: Dam  
 Cf21: {Dam, Daeng} C  
 Cf22: {Daeng, Dam} R

## 5. CONCLUSION

In this paper, I provide an overview of referent resolution for zero pronouns, with an emphasis on Thai. I assume that the resolution can be done at two levels: the sentence level and the discourse level. The resolution at the sentence level can be implemented on the basis of principles of a sentence grammar, which is in accordance with the government and binding theory. Zero pronouns that cannot be resolved by the government and binding theory are resolved on the basis of discourse principles. The

<sup>13</sup>Both *Daeng* and *Dam* are realized as zero pronouns. But *Daeng* has a higher rank than *Dam*. Thus, *Daeng* must be the Cb of U3. And Cf11 is preferred to Cf12.

centering theory is the discourse principle used in this paper. Zero pronouns are resolved by keeping track of discourse-salient entities. The referents of the zero pronouns are expected to be the most focused entity, or the (backward) center of an utterance. The theory has been used in pronoun resolution in many languages, such as English, Italian, and Japanese. It is shown in this paper that the theory is also applicable in resolving zero pronouns in Thai texts. However, since the theory was developed on the basis of constructed discourses, further research based on naturally occurring discourses is therefore needed, especially on complex sentences. Such research will strengthen the centering theory.

## REFERENCES

- Baker, M. C. (1989). Object sharing and projection in serial verb constructions. *Linguistic Inquiry*, 20, 513–553.
- Browning, M. (1987). *Null operator constructions*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. Cambridge, Mass.: MIT Press.
- Chomsky, N. (1986a). *Knowledge of language: Its nature, origin and use*. New York: Praeger.
- Chomsky, N. (1986b). *Barriers*. Cambridge, Mass.: MIT Press.
- Cook, V.J. (1988). *Chomsky's universal grammar: An introduction*. Cambridge, Mass.: Basil Blackwell.
- Demirdache, H. (1991). *Resumptive chains in restrictive relatives, appositives and dislocation structures*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge.
- Georgopoulos, C. (1991). *Syntactic variables: Resumptive pronouns and A' binding in Palauan*. Dordrecht: Kluwer Academic Publishers.
- Givón, T. (Ed.). (1983). Introduction. In *Topic continuity in discourse: A Quantitative cross-language study*. Amsterdam: John Benjamin.
- Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12, 175–204.
- Grosz, B.J., Joshi, A.K., & Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, 44–50.
- Grosz, B.J., Joshi, A.K., & Weinstein, S. (1986). *Towards a computational theory of discourse interpretation*. Unpublished manuscript.
- Gundel, J.K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274–307.
- Hoonchamlong, Y. (1991). *Some issues in Thai anaphora: A government and binding approach*. Unpublished doctoral dissertation, University of Wisconsin, Madison.
- Huang, C.-T.J. (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry*, 15, 531–574.
- Kameyama, M. (1985). *Zero anaphora: The case of Japanese*. Unpublished doctoral dissertation, Stanford University, Stanford.

- Karttunen, L. (1976). Discourse referents. In J. D. McCawley (Ed.), *Syntax and semantics 7: Notes from the linguistic underground* (pp. 363–385). New York: Academic Press.
- Lasnik, H., & Uriagereka, J. (1988). *A course in GB syntax: Lectures on binding and empty categories*. Cambridge: MIT Press.
- Law, P. S. (1991). *Effects of head movement on theories of subjacency and proper government*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge..
- Pingkarawat, N. (1989). *Empty noun phrases and the theory of control, with special reference to Thai*. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.
- Walker, M, Iida, M., & Cote, S.. (1990). Centering in Japanese discourse. *COLING-90: Papers presented to the 13th International Conference on Computational Linguistics, 1*, 1-6.