

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 resolution¹ 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 which 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 are reviewed 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 which 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 respect to features of pronominal and anaphor² as below:

	-anaphor	+anaphor
-pronominal	wh-trace(variable)	NP-trace
+pronominal	pro	PRO

¹The 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.

²These features are also used to categorize overt NPs into reflexives and reciprocals, pronominals, and names or r-expressions,.

A trace is normally analyzed as a result of move-alpha³. An NP-trace is left when an NP is moved from one A-position⁴ 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: John₁ seems *t*₁ to be nice

(b) wh-trace: What books₁ do you like *t*₁? (Cook 1988:163)

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 1988:38)

'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 1988:38). This fact reflects one parameter of universal grammar, which is 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: John₁ wants *PRO*₁ to go

It is time *PRO* to go (Cook 1988:164)

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 below 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 sub-theory that explains anaphoric relations between NPs in a sentence. The three principles of binding theory are:

A. An anaphor (+anaphor) is bound⁵ in its governing category⁶

B. A pronominal (+pronominal) is free in its governing category

C. An R-expression (-anaphor,-pronominal) is free

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

(a). John₁ likes himself₁

(b). John₁ likes him_{*1,2}

(c). John₁ believes that Peter₂ likes him_{1,3}

(d). *John₁ believes that Mary₂ likes himself₁

(e). He₁ thinks that John_{*1,2} 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

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

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

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

⁶"a governing category [of a] is a maximal projection containing both a subject and a lexical category governing a" (Chomsky 1986a: 169)

must be free. Thus, it cannot be bound to 'John'. It will have a different index from 'John'. In (c), the principle-B prohibits 'him' to be 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 the principle-A. In (e), 'John' is not bound to 'he' because the principle-C prohibits 'John', which is an R-expression, to be 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 suggesting 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) Who₁ do you think [John likes t₁]?

(b) That report which₁ I filed e₁ without PRO reading e₁

(Lasnik & Uriagereka 1988:78)

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 subadjacency principle⁷, '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

⁷Subadjacency 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.

because the first trace does not c-command it. Rather, it is a variable because it is locally A'-bound by 'which'⁸.

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) John₁ asked PRO₁ to go

(b) John₁ asked Peter₂ PRO₂ to go

(c) It is time PRO to go (Cook 1988:162)

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

Status of ECs, whether they are NP-traces, variables, pros or PROs, are 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 θ -role." (Chomsky 1982:35)

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-drop language like Chinese and Thai, an additional criteria 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 that a lexical item cannot be present is a PRO. An EC which 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 are 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.

⁸This structure may be 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 1991:324)

3 Empties categories in Thai

Thai is a pro-drop language. Thai 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 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, subadjacency principle will be violated. Her example is provided below:

'waan ní chán hěn [NP nák-khǎn [S' thîi
day this I see writer THAT
[S1 Nit bòk Noy [S' wǎa [S2 Dang kam-lan] ?àn
Nit tell Noy COMP Dang PROG read
[NP nǎŋ-sǔu [S' thîi [S3 EC wí-can EC]
book THAT criticize

(a) 'Today I saw the writer_i that Nit told Noy that Dang was reading the book_j that (he)_i criticized EC_j'

(b) 'Today I saw the writer_i that Nit told Noy that Dang was reading the book_j that EC_j criticized (him)_i'

(Hoonchamlong 1991:187)

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, subadjacency 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. Using of resumptive pronouns in English is very marginal. They are used in a sentence in which a gap cannot take place because of subadjacency 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 relative clause can be succeeded by predication rule (see Law 1991, Ch5). 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 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 topicalized sentence is not a trace in s-structure because the movement would violate 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 pronoun.

According to Hoonchamlong, topicalization in Thai does not involve move-alpha. If an EC is a trace resulting from movement, it would violate subjacency principle. In addition, topicalization does not exhibit Strong Cross-over effect⁹. In example below, subjacency principle will be violated if EC is a trace. (The relation between 'kháw' or EC and the antecedent TOP crosses two bounding nodes, S1 and NP1¹⁰.)

[TOP Suda ná][S1 chán dây-yin [NP1 khàaw [S' wâa [S2 John pân
Suda TM I hear news COMP John just

phaa kháw/EC pay roon-phà-yaa-baan mûaa-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 1991:93)

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

[TOP khay_i ná] [S1 kháw_i khít wâa [S2 EC_i chà-ná]]
who TM he think COMP win

'Who, he thought that won?' (Hoonchamlong 1991:198)

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 predication rule. (It is bound by the topic NP at LF)

⁹Strong Cross-over is a phenomena where "one of a pair of coreferential expressions crosses over another via Wh-movement" (Lasnik&Uriagaraka 1988:41) 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_i does he_i think Mary likes t_i

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

3.3 Serial verb constructions

Serial verb constructions are "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". (Baker 1989:513) 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 which 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 verbs¹¹ 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 1989:516)

[CP Kofi [IP [VP [V' [V naki] [NP Amba] [V' [V Kiri]]]]]]

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

(a) [CP khǎw [IP[VP[V'[V wâat] [NP rûup] [V'[V khǎi]]]]]]

 he paint picture sell
'He paints a picture and sells it'

(b) [CP khǎw [IP[VP[V'[V chái] [NP mîit] [V'[V tât][NP n-aa]]]]]]

 he use knife cut meat
'He uses a knife to cut meat'

(c) [CP Dum [IP [VP [V' [V tii] [NP ŋuu] [V' [V tai]]]]]]

 Dum hit snake die
'Dum hits a snake dead'

(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. 'Dum' receives only one theta-role from 'hit' while 'snake' receives two theta-roles from 'hit' and from 'die'.

(d) [CP khǎw [IP[VP[V'[V' [V1 kin] [NP1 khâw]] [V'[V2 ?im]]]]]]

 he eat rice full
'He ate and became full'

Examples (a)-(c) above indicate object sharing in serial verb constructions. However, the serial verb constructions do not necessary 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

¹¹This 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 θ -role as long as all its θ -roles are assigned to the same structural position." (Baker 1989:521)

sentence type is different from the structure above. 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 above analysis, serial verb constructions maybe 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. Example of Sranan language in the section above, when analyzed in this way, may look like this:

1. Kofi [VP hit Amba_i [V' kill pro_i]]

2. Kofi [VP hit Amba_i [XP O_i [VP kill t_i]]] (baker 1989:518)

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). Example below indicates 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 'khăw' because 'yàak' is a subject-control verb.

(a) khăw 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 another word, 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 already 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 and Sidner 1986). A discourse segment is a group of utterances which are locally coherent. The centering theory (Grosz et.al. 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 et.al. 1993). An entity that is in focus usually contain less information in itself. It is normally realized as a pronoun or a zero pronoun (see Gundel et.al 1993, Givon 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 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_i in a discourse segment U_1, \dots, U_m :

1. There is precisely one Cb.
2. Every element of $Cf(U_i)$ must be realized¹² in U_i .
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. 1990:2)

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 $Cb(U_i) \triangleleft Cp(U_i)$

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

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

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 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 the next section). 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 other Cf, 'Mike', is realized as a pronoun.

U1) John_i wanted to go for a ride yesterday

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

U2) He_i called up Mike_j

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

U3) He_j was annoyed by John_i's call.

4.2 Centering in Thai

Since zero pronouns contains less information than pronouns, zero pronouns are assumed here as more focused than pronouns. In this paper, only zero pronouns are considered, and assumed as the basis form for the centering rule. The difference between zero pronouns and

¹²"An utterance U (of some phrase, not necessarily a full clause), realize c if c is an element of the situation described by U, or c is the semantic interpretation of some subpart of U " (Walker et.al 1990)

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 (Givon 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 are 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 U_1 has only one entity, Dang is the Cb and is the only member of Cf list. Thus, Dang will be the Cb of U_2 . Then, Dang would be the antecedent of zero pronoun since it is the only possible referent. Since Dang is the highest Cf realized in U_2 , it would be the Cb of U_3 . There are two possibilities of ordering Cfs, Cf1 and Cf2. But Cf1 is preferred to Cf2 because Cf1 represents continuing state while Cf2 represents a retention state. Thus, the centering theory predicts that the zero subject in U_3 would refer to Dang while the zero object would refer to Dum.

- (a) U_1 : Dang pay paa-tîi mûaa-waan
 Dang go party yesterday
 'Dang went to a party yesterday'
 Cb: Dang
 Cf: {Dang}
- U_2 : [Z] dâi rûu-càk kâp Dum
 ASP. meet with Dum
 '(He) met Dum'
 Cb: Dang
 Cf: {Dang, Dum}
- U_3 : [Z] kô læy chuaan [Z] pay duu năj
 CONJ then invite go see movie
 '(He) invited (Dum) to go to a movie'
 Cb: Dang
 Cf1: {Dang, Dum} C
 Cf2: {Dum, Dang} R

However, centering cannot eliminate all ambiguities in a discourse. Ambiguity may arise when the first utterance contains more than one entity. For example, U_1 in the example below contains two entities, Dang and Dum. Either one of them can be the Cb of U_2 because U_2 contains only one entity. Thus, the zero pronoun in U_2 can be interpreted either as Dang or Dum.

- (b) U_1 : Dang maa cə Dum
 Dang come meet Dum

'Dang met with Dum'
 Cb: ?
 Cf: {Dang, Dum}
 U2: tɔɔn-thîi [Z] kam-laŋ dɔɔn lɛn
 while PROG. walk ASP.
 'while (he) was walking'
 Cb1: Dang
 Cf1: {Dang}
 Cb2: Dum
 Cf2: {Dum}

In addition, the centering theory is still in the developing stage. It needs more researches done 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)=Dang, Cf11 is preferred to Cf12. If Cb(U2)=Dum, Cf21 is preferred to Cf22. However, Cf21 is unlikely to be acceptable. Zero subject in U3 should refer to Dang rather than Dum. 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 Dang regardless of the ambiguity in U2¹³. And U3 can have only one preferred interpretation. This suggests that subordination may not be analyzed as an individual utterance.

(c) U1: Dang maa cəə Dum
 Dang come meet Dum
 'Dang met with Dum'
 Cb: ?
 Cf: {Dang, Dum}
 U2: tɔɔn-thîi [Z] kam-laŋ dɔɔn lɛn
 while PROG. walk ASP.
 'while (he) was walking'
 Cb1: Dang
 Cf1: {Dang}
 Cb2: Dum
 Cf2: {Dum}
 U3: [Z] kô ləəy chuaan [Z] pay duu năŋ
 CONJ then invite go see movie
 '(He), then, invited (him) to go to a movie'
 Cb1: Dang
 Cf11: {Dang, Dum} C
 Cf12: {Dum, Dang} R
 Cb2: Dum
 Cf21: {Dum, Dang} C

¹³Both 'Dang' and 'Dum' are realized as zero pronouns. But 'Dang' has a higher rank than 'Dum'. Thus, 'Dang' must be the Cb of U3. And Cf11 is preferred to Cf12.

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 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 pronouns resolution in many languages, such as English, Italian, Japanese, etc. It is shown in this paper that the theory is also applicable in resolving zero pronouns in Thai texts. However, since the theory is developed on the basis of constructed discourses, further researches based on naturally occurring discourses is therefore needed, especially on complex sentences. Such researches will strengthen the centering theory.

REFERENCES

- Baker, Mark C. (1989). *Object Sharing and Projection in Serial Verb Constructions*. Linguistic Inquiry, Vol.20, No.4, 513-553.
- Browning, M. (1987). *Null Operator Constructions*, Ph.D. Dissertation, MIT.
- Chomsky, Noam. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, Noam. (1982). *Some concepts and Consequences of the Theory of Government and Binding*. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. (1986a). *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.
- Chomsky, Noam. (1986b). *Barrier*. Cambridge, Mass.: MIT Press.
- Cook, V.J. (1988). *Chomsky's Universal Grammar: An Introduction*. Cambridge, Mass.: Basil Blackwell.
- Demirdache, Hamida. (1991). *Resumptive Chains in Restrictive Relatives, Appositives and Dislocation Structures*. Ph.D. Dissertation, MIT.
- Georgopoulos, Carol. (1991). *Syntactic Variables: Resumptive Pronouns and A' Binding in Palauan*. Dordrecht: Kluwer Academic Publishers.
- Givon, Talmy. (1983). ed., *Topic Continuity in Discourse: A Quantitative Cross-language Study*. Amsterdam: John Benjamin.
- Grosz, Barbara J. and Candance L. Sidner. (1986). Attention, Intentions, and the Structure of Discourse. In *Computational Linguistics*, Vol.12, No.3, 175-204.
- Grosz, B.J., A.K. Joshi, and S. Weinstens. (1983). Providing a Unified Account of Definite Noun Phrases in Discourse. In *Proceedings of ACL*, 44-50.
- Grosz, B.J., A.K. Joshi, and S. Weinstens. (1986). *Towards a Computational Theory of Discourse Interpretation*. Unpublished Manuscript.
- Gundel, J.K., Nancy Hedberg, and Ron Zacharski. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, Vol. 69(2), 274-307.

- Hoonchamlong, Yuphaphann. (1991). *Some Issues in Thai Anaphora: A Government and Binding Approach*. Ph.D. Dissertation, University of Wisconsin-Madison.
- Huang, C.-T.J. (1984). On the Distribution and Reference of Empty Pronouns. In *Linguistic Inquiry*, 15, 531-574.
- Kameyama, Megumi. (1985). *Zero Anaphora: The Case of Japanese*. Ph.D. Dissertation. Stanford University
- Karttunen, Lauri. (1976). Discourse Referents. In James D. McCawley, ed., *Syntax and Semantic 7: Notes from the Linguistic Underground*. New York: Academic Press. 363-385.
- Lasnik, Howard and Uriagereka, Juan. (1988). *A Course in GB Syntax: Lectures on Binding and Empty Categories*. Cambridge: The MIT Press.
- Law, Paul S. (1991). *Effects of Head Movement on Theories of Subjacency and Proper Government*. Ph.D. Dissertation, MIT.
- Pingkarawat, Namtip. (1989). *Empty Noun Phrases and the Theory of Control, with Special Reference to Thai*. Ph.D. Dissertation, University of Illinois at Urbana-Champaign.
- Walker, Marilyn., Masayo Iida, and Sharon Cote. (1990). Centering in Japanese Discourse. In *COLING-90*, Vol.I, 1-6.