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\(^1\) 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

\(^1\)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 as below:

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

A trace is normally analyzed as a result of move-alpha. An NP-trace is left when an NP is moved from one A-position 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 seems t₁ to be nice
(b) wh-trace: What books do you like t₁? (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: John₁ wants PRO₁ 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 in its governing category.
B. A pronominal (+pronominal) is free in its governing category.
C. An R-expression (–anaphor, –pronominal) is free.

---

2 These features are also used to categorize overt NPs into reflexives, reciprocals, pronominals, and r-expressions.
3 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.
4 “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.
5 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.
6 “…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). \( \text{John}_1 \) likes himself_1
(b). \( \text{John}_1 \) likes him_1,2
(c). \( \text{John}_1 \) believes that \( \text{Peter}_2 \) likes him_1,3
(d). \( \text{John}_1 \) believes that \( \text{Mary}_2 \) likes himself_1
(e). He_1 thinks that \( \text{John}_1 \) is lazy.

In (a), since \textit{himself} is \(+\)anaphor, it must be bound in its governing category. Thus, \textit{himself} must have the same index as \textit{John}. On the other hand, since \textit{him} is \(+\)pronominal, it must be free. Thus, it cannot be bound to \textit{John}. It will have a different index from \textit{John}. In (c), principle-B prohibits \textit{him} from being coindexed with \textit{Peter}, but does not exclude the coindexing between \textit{him} and \textit{John}, since \textit{John} is not in the governing category of \textit{him}. In (d), \textit{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), \textit{John} is not bound to \textit{he} because principle-C prohibits \textit{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 provides 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 \textit{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 \textit{wh}-traces. It explains coindexation between variables and their antecedents, and the sequence of \textit{wh}-movements.

(a). \( \text{Who}_1 \) do you think [ \text{John likes t}_1 ]?
(b). That report which_1 I filed e_1 without PRO reading e_1

(Lasnik & Uriagereka, 1988, p.78)

Coindexation between \textit{who} and \textit{wh}-trace in (a) is an example of A'-binding resulting from \textit{wh}-movements. In (a), A'-binding is generated by application of \textit{wh}-
movements. To observe the subjacency principle\(^7\), *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*.\(^8\)

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) John\(_1\) asked PRO\(_1\) to go.
(b) John\(_1\) asked Peter\(_2\) PRO\(_2\) to go.
(c) It is time PRO to go. \(\text{(Cook, 1988, p. 162)}\)

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-

---

\(^7\)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.

\(^8\)This 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]]]]] \(\text{(Law, 1991, p. 324)}\)