Finding Subject in Thai

Kitima Indrambarya[1]

1. Introduction

Just like a few other Southeast Asian languages, the Thai language exhibits an areal feature known as pronominal-dropping. Thai is a pro-drop language and hence allows a subject and an object to be missing within context. This characteristic not only poses problems for learners of Thai as a foreign language just as Hartman and Hudak (2005) noted:

"...*only the skilled student or native speaker is able to figure out from the topic and context who the centers of action and focus are with any degree of certainty.*"

but it also poses problems for natural language processing. Let's look at the following example:

(1) หน่อไม้ฝรั่งที่**พบเห็น**อยู่ทั่ว ๆ ไป **มี** ทั้งชนิดหน่อสีขาวซึ่ง**ใช้** สำหรับแปรรูป
มี **ปลูก** กันมากที่จังหวัดสุพรรณบุรี และชนิดหน่อสีเขียว ซึ่ง**ใช้** **รับประทาน**สด
มี **ปลูก** กันมาก ที่จังหวัดนครปฐม กาญจนบุรี นนทบุรี และนครราชสีมา[2]

The above excerpt contains only one sentence. It is easy to find that หน่อไม้ฝรั่ง is the topic and focus of the discussion. Yet, it is not easy for a non-native speaker of Thai nor a computer machine to identify the missing subjects of each verb in this sentence. There are a number of verbs in this sentence, namely พบเห็น ใช้ ใช้ แปรรูป มี มี มี ปลูก ปลูก, all of which are shown in bold letters in the above example. How then can we find the subject for each finite verb? And if a subject is missing, how can we find the actor or the performer of such finite verb? These are not simple questions and analysis of sentence structure is needed to sort them all out.

Within the Lexicase Dependency Grammatical Framework, this paper aims at finding 1) the subjects of finite verbs within a complicated sentence; and 2) an interpretation

of the missing Nominative of a finite verb. The data used are written documents from various online sources. It is hoped that the Dependency approach will shed some light and provide guidance for learners of Thai language as a foreign as well those involved in natural language processing.

This paper is divided into four sections. The first section gives an overview of the problem. The second section provides an overview of the Lexicase Dependency Grammar related to the point under discussion. The third section explains how to identify a remote subject and to make an interpretation concerning the missing subject. The last section presents a conclusion.
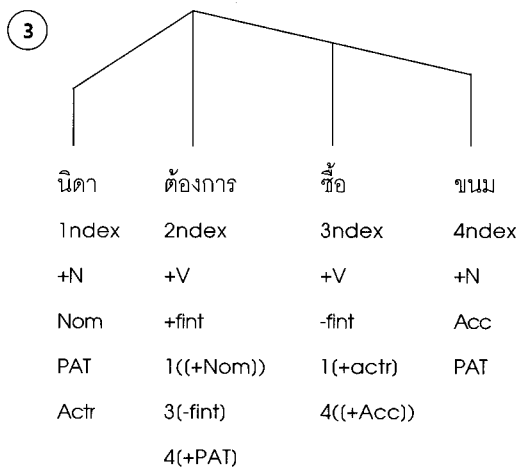
2. Grammatical Framework

This study analyzes Thai sentence structures using the Lexicase Dependency Gram- mar developed by Stanley Starosta of the University of Hawaii. In this section, I will discuss two relevant points. They are finite vs. non-finite verbs, on the one hand, and Nominative vs. subject on the other.

2.1 Finite vs. Non-finite Verbs

Lexicase uses binary features to indicate syntactic and semantic information on each word. A verb can be either finite **[+fint]** or non-finite **[–fint]**. A finite verb refers to a verb which requires, as its dependent sister complement, a Nominative actant. This Nomina- tive may be optional contextually for 'subject pro-drop' languages. The parentheses on the contextual feature **[?([+Nom])]** distinguishes a subject in pro-drop languages from that in non-pro-drop languages such as French and English which is shown by **[?[+Nom]]** (Kitima Indrambarya, 1994: 34). A non-finite verb, on the other hand, refers to a verb which does not have a Nominative of its own. The actor or performer of the action can be retrieved through a Control Chaining Rule. In other words, the contextual feature **[?([+Nom])]** on the finite verb is used to distinguish a finite verb from a non-finite counterpart in Thai. It says the verb looks for a Nominative of a certain index which can be left out contextually. An example is shown in 2.

(2)

แพร	พิมพ์	รายงาน
1ndex	2ndex	3ndex
+N	+V	+N
Nom	+trns	Acc
AGT	+fint	PAT
Actr	1((+Nom))	
	3((+Acc))	
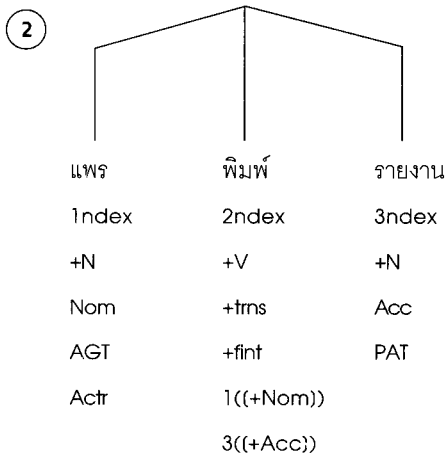
In Lexicase, verbs are the most important parts of the sentence. Each verb must be marked in its lexical entry for what is required as its complement. For example, the transitive verb พิมพ์ is finite because it requires a Nominative noun phrase **(Nom)** and an Accusative noun phrase **(Acc)** as its complement. This is shown with an implied feature **[?([+Nom])]** and **[?([+Acc])]**. Once the index number is filled as **[1([+Nom])]**, we know that แพร is a Nominative of พิมพ์. Similarly, **[3([+Acc])]** on the verb พิมพ์ indicates that รายงาน is an Accusative of พิมพ์. Now let's consider sentence 3.

(3)

นิดา	ต้องการ	ซื้อ	ขนม
1ndex	2ndex	3ndex	4ndex
+N	+V	+V	+N
Nom	+fint	-fint	Acc
PAT	1((+Nom))	1(+actr)	PAT
Actr	3(-fint)	4((+Acc))	
	4(+PAT)		

In this sentence, the verb ต้องการ is finite as it takes the noun phrase นิดา as subject. This is shown by the feature **[1([+Nom])]** on the verb. It also requires a non-finite verb complement shown by the implied feature **[3[-fint]]**. On the other hand, the verb ซื้อ is non-finite because it does not take a Nominative dependent. Instead, the actor (performer of the action) of ซื้อ can be retrieved by the Control Chaining Rule which states that an actor of ซื้อ must be co-referential with the Patient of the higher verb ต้องการ. Since Nida is an Accusative patient[3] of the regent[4] transitve verb ต้องการ, it is then the actor of the verb ซื้อ.

2.2. Nominative vs. Subject

Starosta (2001:4) points out that the terminology 'subject' is not universal. He states the drawbacks in maintaining the term 'subject' as follows:

(1) there is no linguistic justification for a category of 'subject'; (2) all generalizations statable in terms of 'subject' can be formulated at least as well using lexicase case forms, case relations, or the actor macrorole; and (3) maintaining a category of subject in syntactic descriptions repeatedly obstructs the positing of generalizations that could be easily stated if it weren't there to get in the way. (Starosta, 2001: 4)

Starosta then proposes that the term 'subjects' be replaced with Nominatives. In lexicase, a Nominative, identified with the feature **[+Nom]**, can account for both ergative and accusative language[5]. For Accusative languages such as Thai, the feature **[+Nom]** (Nominative) is syntactically required by a finite verb. It is used to distinguish a finite verb from a non-finite counterpart. Hence, the term Nominative **(Nom)** is used here to identify a subject of finite verbs.

---

[3] *In Lexicase, every sentence needs to have a Patient. Lexicase identifies a Patient syntactically. In accusative languages, it is a Nominative of an intransitive verb and an Accusative of a transitive verb.*

[4] *The term 'regent' refers to a sister head.*

[5] *An Accusative language refers to one that marks the direct object of transitive verbs with an Accusative case, distinguishing them from the subject of both transitive and intransitive verbs, which carries a Nominative case. An Ergative language, on the other hand, is one that treats the subject of transitive verbs distinctly from the subject of intransitive verbs and the object of transitive verbs.*

3. Looking for the Nominative

As an SVO language, one can find a Nominative in a thai sentence by looking leftwards of the finite verb. Yet, it is not always easy to find a Nominative for some Thai finite verbs. This section will discuss two points: an overt Nominative and interpretation of a covert Nominative.

3.1 Looking for an overt Nominative

As a pro-drop language, Thai allows both subject and object to be missing within context. In conversation, interlocutors can understand each other through pragmatic context. However, Nominatives within complex sentence structures in written texts such as those found in written documents and news articles might be hard to recognize.

As an SVO language, Thai generally has a Nominative immediately to the left of a finite verb in an affirmative sentence. In a more complex sentence, the Nominative could be remote. Within a sentence, a noun phrase on the left of the finite verb could function as a Nominative on the condition that it is not a Nominative or an Accusative of another verb or of a preposition. Let's consider our first example.

(1) หน่อไม้ฝรั่งที่**พบเห็น**อยู่ทั่ว ๆ ไป **มี** ทั้งชนิดหน่อสีขาว**ซึ่ง ใช้** สำหรับแปรรูป
มี **ปลูก** กันมากที่จังหวัดสุพรรณบุรี และชนิดหน่อสีเขียว **ซึ่ง ใช้** รับประทานสด
มี **ปลูก** กันมากที่จังหวัดนครปฐม กาญจนบุรี นนทบุรี และนครราชสีมา

The above excerpt contains only one sentence. Here, we find as many as ten verbs in this sentence but only six of them are finite verbs. Four of these verbs have their subject on their left as illustrated below.

[หน่อไม้ฝรั่ง   [ที่**พบเห็น**อยู่ทั่ว ๆ ไป]   **มี** ทั้งชนิดหน่อสีขาว
[ซึ่ง**ใช้** สำหรับแปรรูป]
และชนิดหน่อสีเขียว
[ซึ่ง**ใช้** รับประทานสด]]

The compound verb พบเห็น finds the first noun phrase on its left, ที่, as its nominative. The verbs **ใช้** and **ใช้** have ซึ่ง and ซึ่ง, immediately on their left as Nominatives, respectively. Unlike other verbs mentioned, the verb มี is not adjacent to any NP. The first

available noun phrase on its left is หน่อไม้ฝรั่ง which is its regent noun phrase. This noun phrase is the Nominative of มี$_1$. We will discuss how to find the reference for the missing Nominative of มี$_2$ and มี$_3$ in the next section.
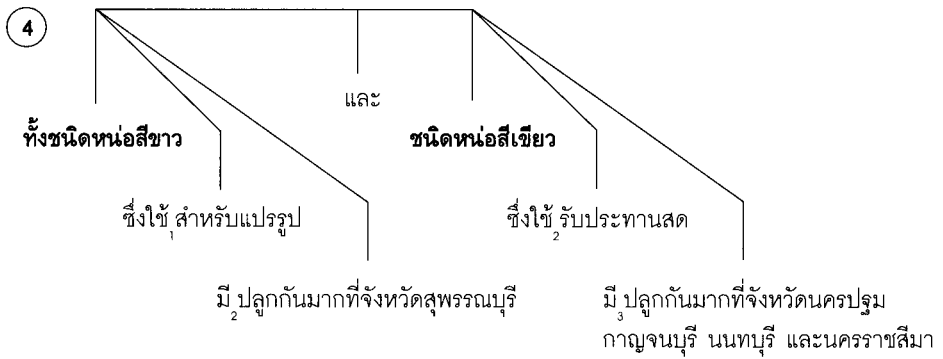
3.2. The Interpretation of covert Nominative

While some finite verbs may have their Nominative present, others may exhibit the pro-drop characteristics and leave the subject position missing. We still need to look for the interpretation of the covert Nominative through pragmatic context. Sometimes, the interpretation could be retrieved from a regent noun phrases. In other cases, the interpretation must be retrieved from an external noun phrase outside a sentence boundary. We will explore both of these cases in this section. Let us now turn to what remains problematic in sentence 1.

(1) หน่อไม้ฝรั่งที่**พบเห็น**อยู่ทั่ว ๆ ไป **มี**$_1$ ทั้งชนิดหน่อสีขาวซึ่ง **ใช้**$_1$ สำหรับ**แปรรูป**

 มี$_2$ **ปลูก**$_1$ กันมากที่จังหวัดสุพรรณบุรี และชนิดหน่อสีเขียว ซึ่ง **ใช้**$_2$ **รับประทาน**สด

 มี$_3$ **ปลูก**$_2$ กันมากที่จังหวัดนครปฐม กาญจนบุรี นนทบุรี และนครราชสีมา

Here the verbs มี$_2$ and มี$_3$ do not seem to have any overt subject in its own clause. The question remains whether these verbs are finite verbs with missing Nominatives (a characteristic of a pro-drop language) or they are simply non-finite verbs. Generally, space in the Thai writing signals a pause in thought or a break between sentences. However, as a Thai non-finite verb never appears as head of an independent clause, I maintain then that มี$_2$ and มี$_3$ are finite verbs with missing Nominatives.

In this case, I propose that one can find the interpretation of a covert Nominative from a regent noun phrase if there is one. That is, the head noun phrase. In sentence 1, the verb phrase มี$_2$ ปลูก$_1$ กันมากที่จังหวัดสุพรรณบุรี modifies the noun phrase ชนิดหน่อสีขาว. The verb phrase มี$_3$ ปลูก$_2$ กันมากที่จังหวัดนครปฐม กาญจนบุรี นนทบุรี และนครราชสีมา modifies the noun phrase ชนิดหน่อสีเขียว. Hence, ชนิดหน่อสีขาว and ชนิดหน่อสีเขียว provide the reference for the missing subject of verbs มี$_2$ and มี$_3$. It is not possible, however, to take ชนิดหน่อสีขาว and ชนิดหน่อสีเขียว as Nominative of these มี$_2$ and มี$_3$. This is because the two noun phrases have already been taken as direct object complements of the regent verb มี$_1$. The tree diagram in 4. illustrates that the noun phrases ทั้งชนิดหน่อสีขาว and ชนิดหน่อสีเขียว are the regent noun phrases of the verb phrases headed by มี$_2$ and มี$_3$, respectively.

(4)

และ

ทั้งชนิดหน่อสีขาว   ชนิดหน่อสีเขียว

ซึ่งใช้$_1$สำหรับแปรรูป   ซึ่งใช้$_2$รับประทานสด

มี$_2$ปลูกกันมากที่จังหวัดสุพรรณบุรี   มี$_3$ปลูกกันมากที่จังหวัดนครปฐม
กาญจนบุรี นนทบุรี และนครราชสีมา

Question Pull Test (Hasagawa,1988: 71) can be used to confirm the interpretation of the missing Nominative through context. Consider the following example.

(5) Dao : ซื้อมาแล้วนะ
     Lek : ซื้ออะไร
     Dao : พริ้นเตอร์ไง
     Lek : แล้วใครซื้อล่ะ
     Dao : ดาว

In the above example, Dao told Lek that she has bought something. Without any subject and object in Thai, Lek got confused as to who did what and hence asked a few wh-questions to retrieve the information on subject and object of ซื้อ. Let's now apply this question pull test to *example 1*.

Speaker 1: หน่อไม้ชนิดใด มี$_2$ปลูกกันมากที่จังหวัดสุพรรณบุรี
Speaker 2: ชนิดหน่อสีขาว
Speaker 1: หน่อไม้ชนิดใด มี$_3$ปลูกกันมากที่จังหวัดนครปฐม กาญจนบุรี
Speaker 2: ชนิดหน่อสีเขียว

As shown above, this question pull test allows us to identify the interpretation of any missing the noun phrases. We can now confirm that the interpretation of the covert Nominatives of มี$_2$ and มี$_3$ are ชนิดสีขาว and ชนิดสีเขียว, respectively.

The following tree diagram in 6. provides the sentence structure of this sentence.
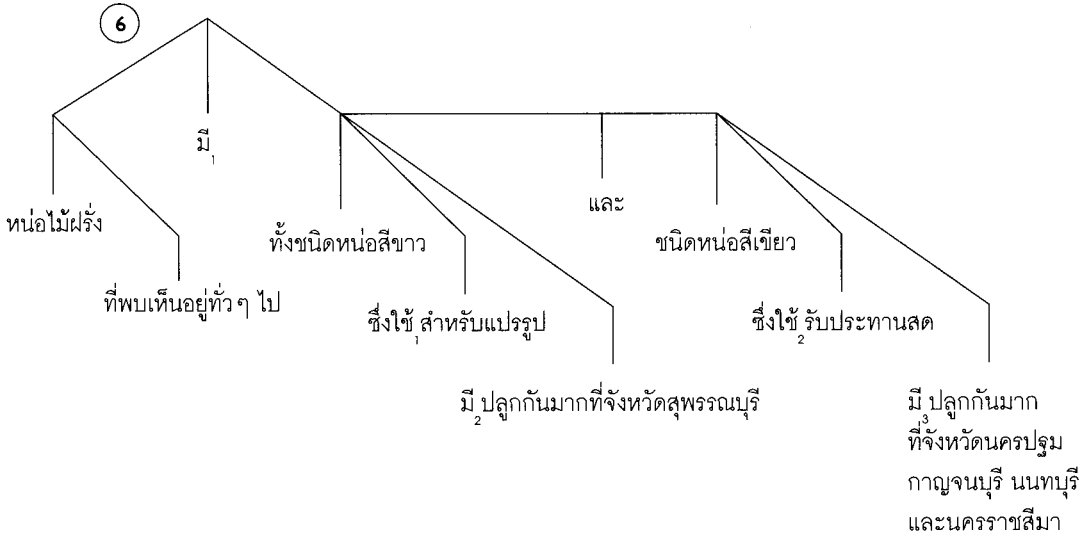
In this sentence, หน่อไม้ฝรั่ง is the Nominative of the main verb มี$_1$ which takes, as a complement, two noun phrases in a coordinate construction. These two noun phrases are ทั้งชนิดหน่อไม้สีขาว ซึ่งใช้$_1$ในการแปรรูป มีปลูกกันมากที่จัดหวัดสุพรรณบุรี and ชนิดหน่อสีเขียว ซึ่งใช้$_2$รับประทานสด มี$_3$ ปลูกกันมากที่จังวัดนครปฐม กาญจนบุรี นนทบุรี และนครราชสีมา

As a topic language, Thai allows preposed topicalized noun phrases in both conversation and written text. A topic may be a noun phrase or a prepositional phrase. Consider the following sentence.

(7) สำหรับความลึกของการหยั่งราก ขึ้นอยู่กับ ความลึกของหน้าดิน

The prepositional phrase สำหรับความลึกของการหยั่งราก denotes a change in topic of discussion and focuses on the new information. It cannot be the subject of the compound verb ขึ้นอยู่กับ. Only a noun phrase, not a prepositional phrase, can function as a Nominative in Thai. In this sentence, the Nominative is covert. One must then look outside the sentence boundary to the left of the finite verb for the closest noun phrase to provide the interpretation of the missing Nominative. And the noun phrase closest to the finite verb ขึ้นอยู่กับ is ความลึกของการหยั่งราก. This noun phrase then provides the interpretation of the covert Nominative.

In other cases, the subject may not be present at all. This is shown in sentence 7. This sentence exhibits a true characteristic of pro-drop language in which a missing Nominative must be retrieved mainly through genuine context.

(8)   พบคนฉวยโอกาสปั่นพันธ์สบู่ดำสูงลิ่วถึงพันบาทต่อกิโลกรัม

Sentence 8 starts without an overt Nominative. By looking leftward, one cannot find any preceding noun phrase. An interpretation of the missing Nominative can only be made based on genuine pragmatic context. Possible interpretations for this sentence are, for example, officers, police, or even anyone at all. This type of missing Nominative would be the most difficult for language processing as well as learners of Thai as a foreign language.

4. Conclusion

In this paper I have discussed ways to identify a Nominative as well as find reference for a covert Nominative in Thai sentences. In a simple sentence, a Nominative generally appears immediately to the left of a finite verb. In a more complex construction in which a Nominative is missing, one needs to look for a regent noun phrase within the same sentence to find an interpretation of the Nominative. In a topicalized sentence, the closest noun phrase to the left of the finite verb can give the interpretation of the missing Nominative. As for sentences without any related noun phrases in previous sentences, the interpretation of missing Nominatives must be based solely on pragmatic context.

Bibliography

กรมส่งเสริมทางการเกษตร.   (โฮมเพจของกรมส่งเสริมการเกษตร). 2548.   (ออนไลน์).
เข้าถึงได้จาก: http://www.doae.go.th (วันที่เข้าถึง 11 กันยายน 2548).

การท่องเที่ยวแห่งประเทศไทย. 2548. **กิจกรรมท่องเที่ยว ดูนก**.   (ออนไลน์).   เข้าถึงได้จาก:
http://www.tat.or.th/thai/travelling_eco_tourism.php?travel_id=7
(วันที่เข้าถึง 11 กันยายน 2548).

สำนักข่าวไทย. (โฮมเพจของสำนักข่าวไทย).   2548.   (ออนไลน์).   เข้าถึงได้จาก:
http://tna.mcot.net (วันที่เข้าถึง 11 กันยายน 2548).

Hartmann, John and Hudak, Thomas John.   2005.   **A Profile of the Thai Language**.
(Online).   Available: http//www.seasite.niu.edu/thai/LLF/profile.htm (Accessed
September 11, 2005).

Hasagawa, Yoko.   1988.   "Question Pull: A Diagnostic Test for the Complement/Adjunct
Distinction in Japanese." **BLS** 13:66-77.

Kitima Indrambarya.   1994.   "The Subcategorization of Verbs in Thai."   Ph.D. Dissertation,
University of Hawaii.

Starosta, Stanley.   1988.   **The Case for Lexicase**.   New York: Pinter Publisher.

--------. 2001.   **Off the Subject**.   (Online).   Paper Presented at The Philological Society,
Cambridge, 09 June 2001.   Available: http://www2.hawaii.edu/~stanley/rt.html
(Accessed September 17, 2005).

Wikepedia: The Free Encyclopedia.   (Online).   Available: http://en.wikipedia.org/wiki/
Ergative-absolute_language (Accessed November 30, 2005).