The Amazing Morphology of Thai

Sompong WITAYASAKPAN

Department of Thai, Chiang Mai University

1. Introduction

It is generally assumed that compound words are formed by combining words together. For example, a noun can be combined with any lexical category forming compound nouns, as in (1).

1   a. Noun + Noun: ปากกา /pàak-kaa/[1]

(beak/mouth-crow)   "a pen"

  b. Noun + Adjective: ของเก่า /khɔ̌ɔŋ-kàu/

(thing-old)   "an antique"

  c. Noun + Verb: คนดู /khon-duu/

(person-look)   "an onlooker"

  d. Noun + Preposition: ละครใน /lakhoon-nai/

(play-in/inside)   "a court's play"

The same process can be done with other lexical categories, namely, a verb, an adjective, and a preposition[2]. The difference is the productivity of each lexical category. The open class categories like a noun, a verb and adjective seem to be more productive, while, the closed class as a preposition is less productive[3]. Generally, this type of compounds is called primary compounds[4].

There is another type of compound which has been assumed that it is formed in the same way as that in (1), that is, a mere combination of words. However, it can be noticed that this type of compound differs from (1) in that it is similar to phrase or sentence construction.   Furthermore, the interpretation is straightforward.   Noticeably, those compounds in (1) are traditionally called exocentric, while the following compounds in (2) below endocentric, compounds.

(2) a. คนขับรถ /khon-khàp-rót/
         (person-drive-car)   "a car-driver"
     b. ไก่ย่าง /kài-yâaŋ/
         (chicken-BBQed)   "a BBQed chicken"
     c. หม้อหุงข้าว /mɔ̂ɔ-hǔŋ-khâau/
         (pot-cook-rice)   "a rice-cooker"
     d. สถาบันสอนภาษา /sathǎaban-sɔ̌ɔn-phaasǎa/
         (institute-teach-language) "a language institute"
     e. วันไหว้ครู /wan-wâai-khruu/
         (day-pay respect to-teacher) "the Teacher's Day"

     This type of compounds is similar to that of English
which is usually called synthetic compounds[5]. The question to be
addressed here is how these lexical categories are combinded.
Is it a mere combination of lexical items or is there any rule or
principle governing this process of these lexical combinations at
all?   And most of all, how can those compounds in (2) be
distinguished from sentences in (3) ? Are they stemmed from the
same source or they are merely accidentally identical?

     (3) a. คนขับรถ /khon khàp rót/ "a person drives a car".
         b. ไก่ย่าง(อยู่) /kài yâaŋ (yuu)/ "a chicken is being BBQed"
     cf. คนย่าง /khon-yâaŋ/ and คน(กำลัง)ย่าง /khon (kamlaŋ) yâaŋ/
         "BBQ person" and "a person is BBQing (something)", but
         unlikely "a BBQed man, of course".
         c. หม้อหุงข้าว(อยู่)   /mɔ̂ɔ hǔŋ khâau (yuu)/ "a pot is
         cooking rice".
         d. สถาบันสอนภาษา /sathǎaban sɔ̌ɔn phaasǎa/ "an institute
         teaches languages".
         e. *วันไหว้ครู /*wan wâai khruu/ *"the day is paying
         respect to teachers". But ไหว้ครูวันนี้ /wâai khruu wan
         nii/ "to pay respect to teachers on the day" is fine

It is argued in this paper that in fact compounds in
(2) and sentences in (3) are underlyingly related. They are
derived from the same origin. This can be observed by looking at
their grammatical relations or functions. This evidence leads to
the rejection of the assumption that compounding is a mere
combination of words.

2. The Theory of Grammatical Relations

First of all, let us assume that a word, a phrase, a
sentence, construction must have its head. A word, for example
can be a head itself or a nonhead constituent. In a word like นัก
ศึกษา /nák-sìksǎa/ "student", it can be argued that นัก- /nák/, an
affix is a head of this word construction, following Williams
(1981), Selkirk (1982), but following Witayasakpan (1990), the
head is ศึกษา /sìksǎa/. The head of a compound in English is
almost always on the right hand side[6], for example, country road,
chicken wing, university lecturer, where road, wing, and lecture
are heads of these compound constructions. Whereas in Thai the
situation is the opposite. The heads of the following compound
constructions are on the lefthand side: ถนนหลวง /thanǒn-lǔaŋ/
"higway" ปีกไก่ /pìik-kài/ "chicken wing", and อาจารย์มหาวิทยาลัย
/ʔaajaan-mahǎawíthayaalai/ "university lecture". Let us assume
that these head and nonhead constituents are grammatically
related. and call it the head-complement relation.

Like words, phrases also have their own heads. It has
been argued that phrasal construction are normally endocentric[7].
A noun phrase, for example, has a noun as its head, followed by
its complement(s). The same is true for a verb phrase, adjective
phrase, and preposition phrase. Each constituent of a phrase is
grammatically related as well. And the relation is that of head
and its complement(s). In formal grammar the head of a sentence
could be either a verb or an inflection (INFL) element[8]. Other
constituents of a sentence can be considered as complement(s) of
a head[9]. Thus a sentence construction can also be viewed as a
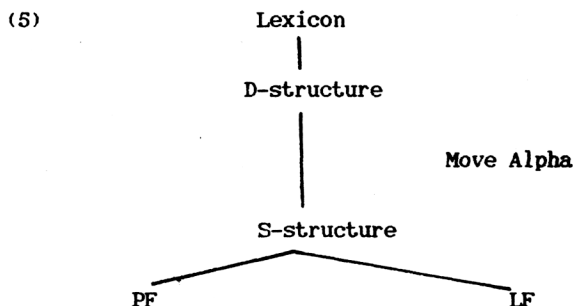combination of a head and its complement(s), both of which are
grammatically related.

In short, every construction in a language must have a head as its basic, and optionally with one or more complements. The occurrence of any complement must be subcategorized or licensed by the head.

3. Grammar and Subcomponents of Grammar

Following Chomsky (1981, 1986), let us assume that the grammar consists of various subcomponents as follows:

(4) (i)   lexicon
     (ii) syntax
               (a) categorial component
               (b) transformational component
     (iii) Phonetic Form (PF) component
     (iv) Logical Form (LF) component

The lexicon consists of lexical items, irregular phrases or sentences and other irregularities[10]. The lexicon and the categorial component constitute the base, which generates D-structures. These D-structures are mapped onto S-structures by the rule of Move Alpha (Affect-Alpha), which constitutes the transformational component. The S-structure then are assigned PF and LF representations by PF- and LF-,components, respectively. The grammatical processes can be illustrated as follows.

(5)   Lexicon
                          |
                    D-structure
                          |
                          | Move Alpha
                          |
                    S-structure
                 /   \
            PF   LF

In order to obtain grammatical results, the processes are governed by subsystems of principles of the grammar, which consist of, for example, X-bar theory, theta theory, government theory, and Case theory[11].

4. The Theory of Phrase Formation

Equipped with the grammatical thoery roughly outlined above, let us first consider lexical properties before the derivation of phrase structure.

4.1 Lexical Properties.

Lexical properties leads to the elimination of phrase structure rules in that each lexical item is the head of a phrase. And it is also specified as to lexical category, noun, verb, adjective, or preposition. Furthermore, it is provided with its complement (argument) structure or subcategorization frames.A complement structure consists of a number of arguments a lexical item can take -- obligatory or optional, external and/or internal arguments. For example, the verb กิน /kin/ "to eat" has the following lexical information:

(6) กิน /kin/ "to eat"   : verb
                           : ⟨Agent,   Patient⟩
                             External   Internal
                             Subject   Object

Normally, optional complements are not necessarily listed. The lexical information tells us that กิน /kin/ "to eat" required an Agent which will be realized as a Subject and a Patient an Object. An Agent and a Patient (or a Subject and an Object) are grammatically related to the verb กิน /kin/ in that they are subcategorized by the verb which is the head of the construction. The order of the element is specified by language specific. Theoretically, it is assumed, according to theta theory, that the head assigned theta (semantic) roles to its complements at D-structure and Case at S-structure, according to

Case theory. Thus, every element in a syntactic construction is grammatically related. There is no independent element allowed. The grammatical relations can be shown as in (7). After lexical insertion, and Move Alpha operation, (8) is the grammatical result. Both complements are assigned theta roles by their head.

(7)   กิน /kin/, V,   <NP1,   NP2>
                      <Agent, Patient>

(8)   คน กิน ข้าว   /khon kin khâau/
      <Agent   Patient>
      Subject   Object

This accords with Chomsky's (1982) Projection Principle which states that representation at each syntactic level (i.e., LF, and D- and S-structures) are projected from the lexicon in that they observe the subcategorization properties of lexical items.
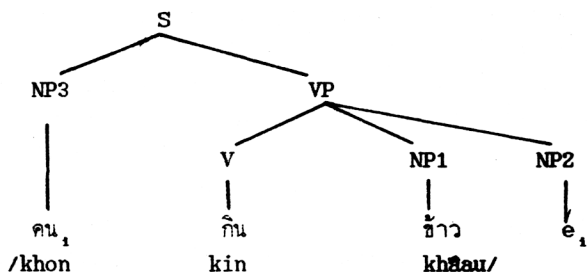
4.2 Theta Theory and the Directionality of Theta Role Assignment

Now let us see how theta roles are assigned and in which direction. Theta theory requires that the assignment must be done under sisterhood and is constrained by the theta criterion which says that each argument bears one and only one theta role, and each theta roles is assigned to one and only one arguemt (Chomsky 1982:36). In the SVO or head-initial language the directionality of theta role assignment is rightward. This can be shown in (9).

(9)

VP

V   θ   NP1   NP2
          <Patient> <Agent>
     θ

กิน   ข้าว   คน
/kin   khâau   khon/

The Move Alpha operation yields the following structure.

(10)

```
                          S
              ┌───────────┴────────┐
            NP3                    VP
             │          ┌──────────┼──────────┐
             │          V         NP1        NP2
             │          │          │          │
            คน₁        กิน        ข้าว         e₁
           /khon       kin      khâau/
```

Under this grammatical theory, it is obvious that words are related. The relation is innate in the lexicon and is projected at the base. This relation is grammatical and semantic. The interpretation follows without any other interpretive rules. Notice also that this syntactic process is absolutely productive.

In the following sections let us consider Thai compounds as given in (2) in particular. We will see if the generation of these compounds bears any similarities to that of phrases at all.

5. Compounding in Thai

5.1 Identifying Grammatical Relations

To begin with, let us consider the grammatical relations of lexical items in each compound. First, คนขับรถ /khon-khap-rot/ shows that it has the same grammatical relations as in a sentence คนขับรถ /khon khàp rót/. That is, the verb ขับ /khàp/ needs two agrument, one an Agent and the other a Patient. Both arguments satisfy the requirement of the subcategorization frame of the head in the compound and in the sentence. Does this identicality occur by accident? It is unlikely to be so. Following the theory of grammar outlined above, it is certain that these lexical and phrasal constructions must be related underlyingly.

Second, the thematic relations in ไก่ย่าง /kài yaaŋ/ are
the same as those of a sentence (คน)ย่างไก่ /(khon-) yâaŋ kài/,
where ไก่ /kài/,occurring either preverbally or postverbally, is
the Patient or the object of the verb ย่าง /yâaŋ/. Both
constructions imply the Agent which is not spelled out
phonetically[12]. The status of this empty category[13] can, of
course, satisfy the requirement of the head. The difference in
positions does not alter the theta relation between the head and
its complement.

Third, หม้อ /mɔ̂ɔ/ in หม้อหุงข้าว /mɔ̂ɔ-hǔŋ-khâau/ is not the
Agent argument of the verb. It bears the Instrumental relation
to the head. That is, "someone cooks rice with หม้อ /mɔ̂ɔ/". It
could also be assumed that หม้อ /mɔ̂ɔ/ originates postverbally,
according to the directionality of theta role assignment. In หม้อ
หุงข้าวอยู่ /mɔ̂ɔ hǔŋ khâau yùu/ "the pot is cooking rice" can
considered as deriving from the same base.

Fourth, the same can be argued for สถาบันสอนภาษา
/sathǎaban-sɔ̌ɔn-phaasǎa/ where สถาบัน /sathǎaban/ can never be an
Agent, but a Locative. And, fifth, วัน /wan/ in วันไหว้ครู
/wan-wâai-khruu/ cannot function as an Agent either.

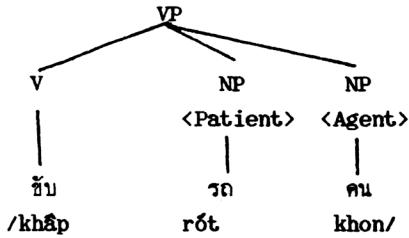## 5.2 The Theory of Compounding

Now let us consider a theory of compounding which
stipulates that a compound and a phrase share the same base or
D-structure. It is then not surprising to see that a compound
structure and phrase structure are similar or even identical in
many respects. The base looks like this:

(11)   XP
                         /   / | \
                        /   / | \
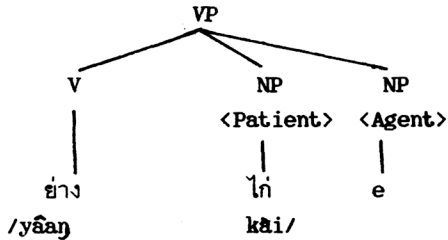   X   YP   ZP   ...
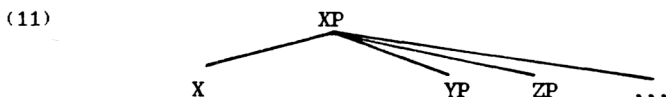
where X stands for any lexical category which always
functions as a head of the phrase XP, whereas YP, ZP
and ... stand for arguments in the subcategorization
frame required obligatorily or optionally by the head.

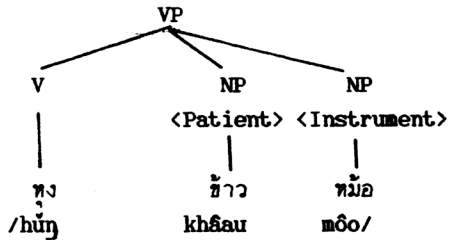Apply the template to our compounds in (2). What we obtain is the following:
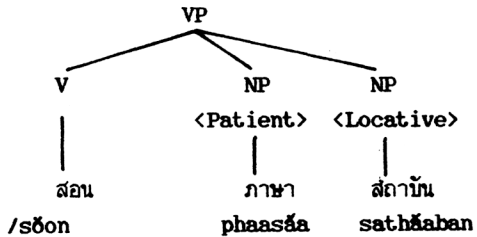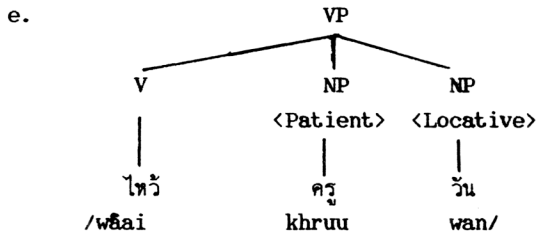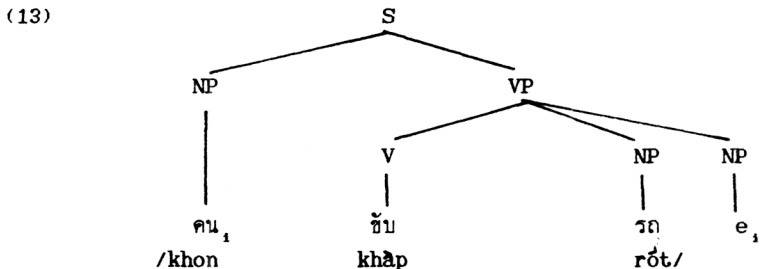
(12) a.

```
              VP
        /      |      \
       V       NP      NP
       |   <Patient> <Agent>
       |       |        |
      ชับ      รถ       คน
     /khâp     rót     khon/
```

b.

```
              VP
        /      |      \
       V       NP      NP
       |   <Patient> <Agent>
       |       |        |
     ย่าง     ไก่        e
    /yâaŋ      kài/
```

c.

```
              VP
        /      |      \
       V       NP          NP
       |   <Patient>  <Instrument>
       |       |           |
      หุง     ข้าว        หม้อ
     /hǔŋ     khâau       môo/
```

d.

```
              VP
        /      |      \
       V       NP          NP
       |   <Patient>   <Locative>
       |       |           |
      สอน    ภาษา        สถาบัน
     /sɔ̌ɔn   phaasǎa     sathǎaban
```

e.

```
                        VP
        ┌────────────────┼────────────┐
        V               NP            NP
        │           <Patient>     <Locative>
        │               │             │
      ไหว้             ครู            วัน
     /wǎai           khruu          wan/
```

If the second NP moves to the empty node under S, the construction will then become a sentence construction as shown below.

(13)

```
                        S
            ┌───────────────────┴────────────┐
            NP                               VP
            │            ┌──────────────┬─────┴──────┐
            │            V             NP           NP
            │            │             │            │
           คน₁          ขับ           รถ            e₁
         /khon         khàp          rót/
```

This type of movement is pervasive in the syntax. Grammatically, this movement is required by Case Theory which applies at S-structure (Chomsky 1986)

Under our theory of compounding, the movement is required as well, but not by Case theory. The movement is constrained by what is called the "Head Movement Constraint" proposed by Travis (1984:41). The constraint states that an X^o may only move into Y^o which properly governs[14] it. In other word, to form a lexical compound, a lexical item has to move to the head of the construction which properly governs it. The movement can be illustrated as follows.

(14)
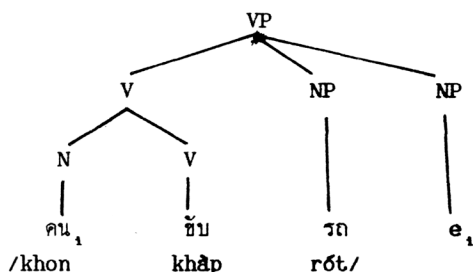
The moved element is coindexed with its trace or an empty category. The empty category is licensed by various principles, including the Empty Category Principle[15].

The moved element moved to adjoin to the the left side of the head entailing a complex verb. However, this does not yield grammatical structure as expected. Yet, another principle has to apply. The principle is the Percolation Principle[16] which postulates that the category features of the left hand side of the head which now functions as the head of the construction will percolate to the upper node. The result is that the V node in every upper node will turn into an N node, as shown below.

(15)
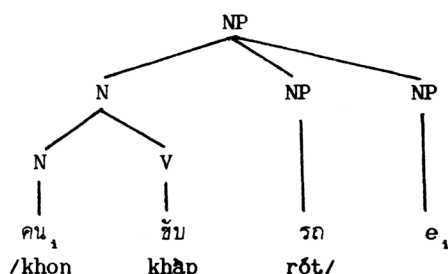
Consequently, we obtain a complex N node which consists of a noun and a verb, where the noun is the head of the construction. And we call this a compound.

Another movement can apply to move a Patient รถ /rót/ to adjoin to the verb that governs it, following again the Head Movement Constraint. This will yield the following structure in (16) which is also a wellformed compound.

(16)

คน₁   ขับ   รถ_t   e_t   e₁
/khon   khàp   rót/
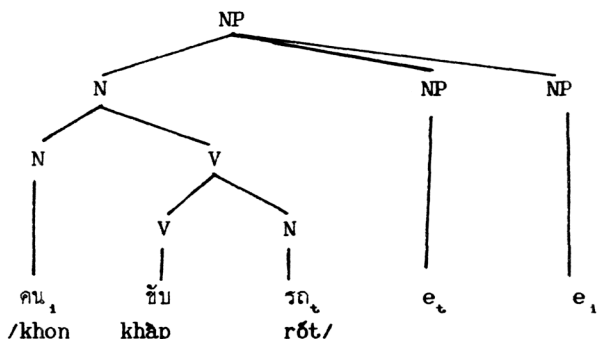
The same process can be applied to other examples, namely, ไก่ย่าง /kài-yâaŋ/, หม้อหุงข้าว /mɔ̂ɔ-hǔŋ-khâau/, สถาบันสอนภาษา /sathǎaban-sɔ̌ɔn-phaasǎa/, and วันไหว้ครู /wan-wâai-khruu/. Here, we will not show the derivation of these compounds.

It is clear now that a compound and a sentence are derived from the same base by the same principles but these principles apply at difference level. Movement can apply at phrasal level, thus yielding a sentence, if apply at lexical level, thus a compound. In short, compounds and phrases make use of the same set of syntactic principles. The identicality in semantic interprettion is then expected.

6. Theoretical Implications

The theory of compounding proposed above bears some theoretical implications. First, it predicts that the process is productive. Second, it shows that semantic interpretations of these compounds are transparent. And third, if there is any semantic irregularity, it is because of other principles getting involved, for example, pragmatic principles. Phrase or word formation processes have nothing to do with semantic interpretation.

6.1 Productivity

The theory of compounding, like any syntactic theories, is very productive in that it can generate compounds of this type indefinitely. Thus, we have an indefinite number of compounds, and new compounds can be formed all the time. Like phrases and sentences, compounds of this type have not to be memorized, thus do not exist in the lexicon. We form compounds in the same way that we construct phrases and sentences. Phrases and sentences are never listed either. Here are some examples.

(17)   a. รถลากไม้ /rót-lâak-máai/
                    "vehicle-pull-wood"
         b. พนักงานเก็บเงิน /phanákŋaan-kèp-ŋen/
                    "employee-collect-money"
         c. สภารักษาความสงบเรียบร้อย /saphaa-ráksǎa-
            khwaamsanòprîapróoi/   "council-keep-peace"

6.2 Semantic Transparency

There is no need to apply any rules of interpretation to these compounds. Anyone who knows Thai understands the meanings of them. Again, like phrases and sentences, the meanings of these compounds are straightforward and transparent, since the meaning follows the interpretation of theta roles and grammatical relations between a head and its complement(s). Consider examples in (17) above, it is obvious that the interpretation is straigtforward.

6.3 Semantic irregularities

It has been assumed for quite some time that compounds differs from phrases in that compounds have special meanings. In traditional grammars, a word หางเสือ /hǎaŋ-sǐa/ "tail-tiger" which means "a rudder" is a compound, but not "tiger tail". Our theory predicts that หางเสือ "tail-tiger" which means "tiger tail" is a regular compound which is productive and semantically transparent, whereas the one that means "a rudder" has semantic

irregularities that have to be learned. Notice that this type of compounds has to be listed and learned not only by foriegners but by native speakers as well. It is traditionally called exocentric compounds. We would rather call it idiom words or idiom compounds. This type of semantic irregularities occurs in phrases and sentences too. And we call them idioms or expressions. All these phrases, and compounds have to be listed in the lexicon since they cannot be learned from rules or principles. Unlike productive phrases and compounds, they is no need to list them in the the lexicon.

7. Conclusion

We have seen, so far, that the theory of compounding presented here is part of a grammar. It utilizes existing syntactic principles without adding specific rules or principles to account for its generation. The essence of this theory is that syntactic principles if apply at phrasal level, yield phrases and sentences, but if at lexical level compounds. The theory predicts that compound formation, like phrase formation, is productive and semantically transparent. Thus, no interpretive rules are required. Semantic irregularities of compounds exist, but this existence is comparable to idioms and expressions for phrases or sentences. All of these, either words, compounds, phrases, or sentences have to be learned, thus, listed.

[1]It is the author's intention to use a very broad phonemic transcription since the focus is not on the sounds but the words, especially how words are ordered in a compound.

[2]For the definition and more examples of Thai compound words, the reader is referred to อนงค์ เลียงอุบล (2525) พระยาอุปกิตศิล ปสาร (2517) กำชัย ทองหล่อ (2515) สุธิวงศ์ พงศ์ไพบูลย์ (2513) Fasold (1968) among others.

[3]See Emonds (1985), Lieber (1980, 1983), and Selkirk (1982).

[4]For further details on the classification of compounds, see Jespersen (1942), Bloomfield (1933), and Allen (1978).

[5]For example, churcher-goer, book-reading, football-player, etc. It has been noticed since Maetzner (1874) that grammatical relations in these compoundsare the same as those in sentences like: someone goes to church, someone is reading books, and someone plays football.

[6]See "the righthand head rule" in Williams (1981), Selkirk (1982), for example.

[7]See Fabb (1984), Emonds (1985), Williams (1980), among others. สำหรับภาษาไทย โปรดดู วิจินตน์ ภาณุพงศ์ (2522) นววรรณ พันธุเมธา (2525).

[8]For Chomsky (1981, 1982, 1986), INFL is the head of sentence construction.

[9]There are two types of complements, one is obligatory, the other optional. Complements of a verb sometimes are called arguments.

[10]Di Scuillo and Williams (1987:3) describe the lexicon as:"If conceived as the set of listemes, the lexicon is incredibly boring by its very nature. It contains objects of no single specifiable type (words, VPs, morphemes, perhaps intonation patterns, and so on), and those objectss that it does not contain there because they fail to conform to interesting law. The lexicon is like a prison -- it contains only the lawless, and the only thing that its inmates have in common is lawlessness."

[11]See Chomsky (1981, 1986) for details of this subsystems of principles and their application.

[12]For details, see Roeper (1984) who calls this phonetic null element an implicit argument.

[13]For the definition, the principle and the distribution of empty categories, the reader is referred to Chomsky (1986), Baker (1988), and Sproat (1985), and the references cited therein.

[14]Proper Government is defined as: ∅ properly governs & iff ∅ theta-governs, and Case-marks, or antecedent-governs & (Chomsky 1986:22).

[15]The Empty Category Pinciple (Chomsky 1982:7) states that traces must be properly governed.

[16]See Lieber (1983).

References

กำชัย ทองหล่อ. (2515) **หลักภาษาไทย.** บำรุงสาส์น.

นววรรณ พันธุเมธา. (2525) **ไวยากรณ์ไทย.** รุ่งเรืองสาส์นการพิมพ์.

วิจินตน์ ภาณุพงศ์. (2522) **โครงสร้างภาษาไทย:ระบบไวยากรณ์.** มหาวิทยาลัย
    รามคำแหง

สุธิวงศ์ พงศ์ไพบูลย์. (2513) **หลักภาษาไทย: เรียบเรียงตามหลักสูตรวิชาชุดครูมัธยม
    ตามระเบียบว่าด้วยการสอบวิชาชุดครู 2505 วิชาภาษาไทย.** ไทยวัฒนาพานิช.

อนงค์ เอียงอุบล. (2525) **คำประสม.** วิทยานิพนธ์อักษรศาสตรมหาบัณฑิต จุฬาฯ.

อุปกิตศิลปสาร, พระยา. (2517) **หลักภาษาไทย:อักขรวิธี วจีวิภาค วากยสัมพันธ์
    ฉันทลักษณ์.** ไทยวัฒนาพานิช.

Allen, Margaret R. (1978) **Morphological Investigations.** Ph.D.
    dissertation, University of Connecticut.

Baker, Mark. (1988) **Incorporation: A Theory of Grammatical
    Function Changing.** The University of Chicago Press.

Bloomfield, Leonard. (1933) **Language.** Holt, Rinehart and Winston.

Chomsky, Noam. (1981) **Lectures on Government and Binding.** Foris
    Publications, Dordrecht.

        (1982) **Some Concepts and Consequences of the
    Theory of Government and Binding.** MIT Press.

        (1986) **Knowledge of Language: Its Nature, Origin
    and Use.** Praeger.

Di Sciullo, Anna Maria and Edwin Williams. (1987) **On the
    Definition of Word.** The MIT Press.

Emonds, J. (1985) **A Unified Theory of Syntactic Categories.** Foris
    Publications, Dordrecht.

Fabb, Nigel. (1984) **Syntactic Affixation.** Ph.D. dissertation, MIT.

Fasold, Ralph William August. (1968) **Noun Compounding in Thai.**
    Ph.D. dissertation, University of Chicago.

Lieber, Rochelle. (1980) **On the Organization of the Lexicon.** Ph.
    D. dissertation, MIT.

Maetzner. (1878) **An English Grammar: Methodological, Analytical,
    and Historical.** trans. by Clair James Grece, Foresman
    and Company.

Roeper, Thomas. (1986) "Implicit Arguments and the Head-Complement Relation", **Linguistic Inquiry.** 18:267-310.

Selkirk, E. (1982) **The Syntax of Words.** Linguistic Inquiry Monograph 7, MIT Press, Cambridge, Massachusetts.

Sproat, Richard. (1985) **On Deriving the Lexicon.** Ph.D. dissertation, MIT.

Travis, Lisa. (1984) **Parameters and Effects of Word Order.** Ph.D. dissertation, MIT.

Williams, Edwin. (1981a) "Arguement Structure and Morphology", **The Linguistic Inquiry Review.** 1:81-114.

(1981b) "On the Notions 'Lexically related' and 'Head of a Word", **Linguistic Inquiry.** 12:245-274.

Witayasakpan, Sompong. (1990) **A Theory of Syntactic Compounding in English.** Ph.D. dissertation, University of Washington.