An Explanation For Inconsistent Word Order Typologies in Some Southeast Asian Languages

Saranya Savetamalya
Chiangmai University

Lawrence A Reid
University of Hawaii

1. INTRODUCTION

In the conclusion to Starosta's treatise on Lexicase, he states, "Areas in which lexicase has a promising but as yet unrealized potential to make significant contributions to linguistic knowledge include ... language typology ..." (1988, p. 258). In this paper we would like to begin to explore some problems in the word order typology of Southeast Asian (SEA) languages, using lexicase as the theoretical framework within which to seek explanations. We will step into one of the well-trodden areas of linguistic structure in Southeast Asia, that of quantifier constructions, an area in which Professor Vichin Panupong (1970, pp. 56-66) led the way by providing, from a structuralist point of view, the earliest insightful description for Thai.

It is our purpose to show that some of the apparent anomalies in the typological characteristics of quantifier constructions are resolved when they are analyzed within the constraints of lexicase. Along the way we shall digress into the structure of noun phrases containing lexical items translated as adjectives in certain Philippine languages, which we will show to be typologically identical in many respects to noun phrase structures of some mainland SEA languages containing quantifiers.

We begin by reviewing the word order typology of SEA languages as it has been discussed in the literature, paying special attention to those constructions which do not conform to the expected typology. We then discuss the analysis of some of these constructions within lexicase, showing that the analysis forced upon us by the theory in fact reveals the languages to be typologically more consistent than previous analyses have implied.

It must be emphasized here that the objective of this paper is to divorce ourselves from a semantically based characterization of notions such as subject and object, as well as semantic definitions of lexical categories, and to confine ourselves to morpho-syntactic characterizations. Word order typology is after-all an attempt to characterize recurring patterns of a syntactic nature in language, and one can only expect to get meaningful results when the terms that are used are not based on the intuitions of native English
speakers, but are carefully formulated within the constraints of a single grammatical theory.

Finally we suggest that the analyses provided here have implications for the historical reconstruction of earlier stages of the syntax of these languages.

2. GENERAL CHARACTERISTICS OF WORD ORDER TYPOLOGY IN SEA LANGUAGES

The general characteristics of the word order typology of a number of languages in SEA were included in the often-cited, pioneering work of Greenberg (1966), in which he related the position of the verb (V) in relation to subject (S) and object (O) with other aspects of word order, drawing from this a number of supposed universals of word order. Thai, Khmer, Vietnamese, and Malay were cited as examples of SVO languages which have prepositions and show NG (Noun-Genitive) and NA (Noun-Adjective) word orders. Tagalog and other Philippine languages were cited as examples of VSO languages which also have prepositions and NG word order, but which have the opposite order, AN, for noun phrases containing adjectives. Other features, such as the position of auxiliary verbs, relative clauses, and numerals were also shown to be implied by the position of the verb in relation to the subject and object.

Subsequent work by Vennemann (1975, 1976), building on insights presented in Lehmann (1973), showed that what is relevant in determining the word order typology of a language was not the relative positions of subject, object, and verb per se, but the relative positions of operator in relation to operand, or attribute to head. In each of the languages cited above, the object follows the verb, putting them into the class of languages in which operators typically follow their operands. Thus in Thai, Khmer, and Indonesian (as shown in ex. 1-12), head nouns are typically followed by demonstratives, adjectives, genitive constructions, and relative clauses, all of which modify in some sense the reference of their head noun and appear to be attributive to it. Philippine languages are also typically operand-operator languages, with heads preceding their modifiers.

Thai
1. bāan jā
   house big

2. rōŋtjan nī
   school this

3. nāŋšī khōŋ phōm
   book of me

'big house'
'this school'
'my book'

1 Other linguists have also contributed to the discussion of word order typology, including Comrie (1981), and Stampe & Donegan (1983).
4. bán thỉ chăn plăuk
    house that I build
    'the house that I built'

Khmer (Jacob, 1968)^2
5. 'ptĕoh thom
    house big
    'big house'

6. ckae nihn
    dog this
    'this dog'

7. sówphiiw rebah khnom
    book of me
    'my book'

8. seh dael ceh khmar
    student who knows Khmer
    'a student who knows Khmer'

Indonesian (MacDonald & Soenjono, 1967)
9. rumah besar
    house big
    'big house'

10. buku ini
    book this
    'this book'

11. rumah saja
    house my
    'my house'

12. jang pergi
    the goes
    'the one who goes'

Burmese, on the other hand is an example of a SEA language with the opposite
typology. As a Sino-Tibetan language it is typically SOV, therefore in Vennemann's
terms it is an example of an operator-operand language. Attributes such as adjectives,
determiners, numerals, genitive constructions, and relative clauses precede their head
nouns, as in ex. 13-17.

Burmese (Cornyn, 1944)
13. kāundé lú
    good person
    'good person'

14. dī lūha
    this person
    'this person'

^2 All Khmer, Indonesian, and Burmese examples are given using the transcription of the
sources.
15. nā mani? 'five minutes'  
    five minute
16. dī châundēga yēihā 'the water of this stream'  
    this stream water
17. bamá mahouttè lì 'a person who is not Burmese'  
    Burmese not-is person

Hawkins (1983) provides a greatly expanded sample of languages and language types in terms of their word order typologies including information from some 357 languages, a large number of which are SEA languages. After providing an extensive review of the contributions of Greenberg, Lehmann and Vennemann to the discovery of word order universals, Hawkins, like Vennemann, rejects the Greenbergian characterization of the relative positions of S, V, and O as the main explanatory principle of universal word order typologies, opting to follow Vennemann's operator-operator order as the major explanatory principle. Hawkins states, "The modifier-head principle is claimed to be a valid cross-categorial generalization about language. Like all major generalizations, it exemplifies a number of phenomena under a higher regularity: The categories N, V, Adp [Adposition], and Adj are assigned the common status 'head' within their respective phrasal categories, and all other constituents within those are assigned the status 'modifiers of the head.'" (1983, p. 292). Hawkins further delineates and describes a number of other principles which he believes account for the variety of word order types found universally.

It has long been recognized that simply by knowing the position of the object with reference to the verb in a sentence, especially if the language is SVO, it is almost impossible to predict with assurance the relative orders of other constituents in that language. Many such languages have word order patterns that are apparently not consistent with the operand-operator order implied by the position of the object following the verb. SEA languages are no exception, as the data presented below will show.

Although both Thai and Khmer have structures containing a quantifier which seem to conform to the expected head-modifier (noun-numeral) order of these languages, as in ex. 18-19, in other quantifier constructions in Thai, the numeral always precedes any classifier which specifies a quantified occurrence, time, distance, or measurement noun (ex. 20-22). In this construction the apparent order, modifier-head, is contrary to the expected typology. However in other types of quantified noun phrase, the numeral-classifier sequence itself follows the quantified noun (ex. 23-24), and is therefore in the appropriate position typologically.
Thai
18. bāan nǐŋ³
    house one

Khmer (Sak-Humphrey, 1994)
19. trey pii [dael c?oen nuh kgnom gnam haoey]
    fish two which cook those I eat already
'Those two fish which are cooked, I ate already.'

Thai
20. sāam khrāŋ
    three time

21. sāam thii
    three occurrence

22. hāa méet
    five meters

23. nāŋsūi sāam lèm
    books three volume

24. mææw hòk tua
    cat six body

Similarly, in Khmer, numerals precede classifiers in apparent violation of the predominant head-modifier order elsewhere in the language, as in ex. 25-28, although, as in Thai, a quantified noun is followed by the numeral-classifier sequence (ex. 29-30), matching the expected typology.

Khmer (Sak-Humphrey, 1994)
25. pram lau
    five dozen

26. dap snet
    ten bunch

27. samsep dolor
    thirty dollar

28. moy maong
    one hour

³ That nǐŋ in post-nominal position in Thai is not a numeral at all but an indefinite determiner is discussed in section 3 below.
29. pteh moy knang  'one house'
    house one  classifier

30. kou pram kpal  'five cows'
    cow five  classifier

In Vietnamese, a typically head-modifier language, with adjectives, demonstratives and relative clauses appearing after the nouns they modify (Nguyen 1990, p. 57), numerals can appear either before or after. They can follow their head noun (the typologically appropriate order) in cases where the numeral is interpreted as an ordinal (ex. 31), but cardinal numbers, as in Thai and Khmer, precede temporal nouns (ex. 32-33). Quantifier plus classifier (clsf) expressions may follow the nouns they quantify (as in Thai and Khmer) (ex. 34), but that they usually precede them (ex. 35-36), is probably the result of the influence of Chinese.

**Vietnamese** (Nguyen, 1990)
31. tháng ba  'March, i.e., third month'
    month three

32. hai năm  'two years'
    two year

33. ba tần  'three weeks'
    three week

34. tiêu vài chú  'a few wood cutters'
    wood.cutters few clsf

35. vài chú tiêu  'a few wood cutters'
    few clsf wood.cutters

36. hai cây nến  'two candles'
    two clsf candle

Greenberg (1975, p. 37) noted that the doubly anomalous order, numeral-classifier-noun, commonly found in Vietnamese (ex. 35-36 above), is also found in some of the Tai languages of southern China and Vietnam (ex. 37).

**Tai Dam** (Strecker, 1990, p. 26)
37. sųŋ fin faa  'two pieces of cloth'
    two  clsf cloth
In Indonesian, a quantified noun phrase always has the numeral in the initial, pre-head position, counter to the expected typology, regardless of whether a classifier is present or not\(^4\), as in ex. 38-41.

**Indonesian** (MacDonald & Soenjono, 1967)

38. empat kursi  
    four chairs

39. tiga hari  
    three days

40. dua mahasiswa  
    two student

41. se orang mahasiswa  
    oneclf student
    'one student'

Almost all of the more than one hundred Philippine languages are verb-initial, and typically have operand-operator word order. But in most, if not all of these languages, numerals typically precede the nouns that they enumerate, as in ex. 42-44, again in apparent violation of the expected word order.

**Hokano**

42. maya nga aso  
    one linker dog
    'one dog'

43. dua nga aldaw  
    four linker day
    'four days'

44. lima a balay  
    five linker houses
    'five houses'

It is not just the VO languages of SEA that show anomalies with respect to the word order of quantifier phrases. Burmese, an OV language, in addition to phrases such as ex. 15 above in which the expected modifier-head word order appears, has examples such as ex. 45-47, where the opposite order is found.

**Burmese** (Cornyn, 1944)

45. mêmimh nhayau?  
    women two
    'two women'

\(^4\) According to MacDonald and Soenjono (1967, pp. 133-135) the use of classifiers in Indonesian is becoming rare. In modern times, the number of classifiers is reduced to only three, i.e., *orang* 'human being', *ekor* 'tail', and *buah* 'fruit'.

7
46. myín láizé hóres forty

47. máin táyá miles hundred

Attempts to explain anomalous word order patterns, and competing word order patterns in specific languages have been made by many linguists. Hawkins (1983, p. 242) summarizes the various mechanisms that have been proposed. Among others, N. Smith (1981) discusses language contact as a motivating force for change. Li & Thompson (1974) suggest that word order change in Chinese has been the result of a change in the grammar by which a verb developed into a case marker. Parker (1980) and Stockwell (1977) stress the importance of analogy as the basis for word order change. Hawkins (1983, p. 242) states, "It seems eminently reasonable that there will be a multiplicity of change-inducing factors operating upon different languages, and even upon one and the same language--just as there are numerous factors that contribute to the explanation for synchronic universals. The proposed explanations are therefore complementary rather than conflicting, and in all likelihood represent only a fraction of the total causes underlying word order change."

It is our contention that trying to understand the processes of change can best be achieved only when the descriptions of the relevant languages have been made within a common theoretical framework, and specifically one that is highly constrained. Lexicase is such a theory, and descriptive studies of a considerable number of languages in SEA using this theory have been made, providing a base for a clearer understanding of the problems involved, and a principled method for seeking solutions to these problems.

3. EXPLANATIONS FOR SOME APPARENTLY INCONSISTENT DATA

Lexicase is a variety of dependency grammar primarily concerned with the discovery and description of the relationships which exist between and among pairs of words in construction with each other in a language. According to Starosta (1988, p. 104), "a lexicase representation can be viewed as a network of dependencies obtaining between (actual or virtual) pairs of lexical items in a sentence." The concepts of 'head' and 'attribute', therefore, have vital roles to play in the theory, and are formally defined.

Of key importance to the analysis of the apparently inconsistent data discussed in the previous section is the determination of the nature of the dependency in the cited examples, i.e., which item is the head of its construction and which is an attribute. A lexical item such as a numeral, or one which would be translated into English as an adjective in construction with a noun, has traditionally been considered to be an attribute. Thus Philippine languages which typically have noun phrases of the kind given in ex. 48-49, in which what appears to be an adjective can either precede or follow a noun, have
been described as having variable word order. Thus Hawkins (1983, p. 339) characterizes Tagalog (following Schachter and Otanes 1972) as being AN/NA, but other Philippine languages, such as Cebuano, Hiligaynon, and Tagabili are cited as being only AN, whereas in all of these languages, either order occurs.

Tagalog
48. ang bahay na maliit
   the house lkr little
   'the little house'
   (Lit. 'the house which is a little one')
49. ang maliit na bahay
   the little lkr house
   'the little house'
   (Lit. 'the little one which is a house')

Starosta makes explicit (1988, p. 105) that "the head of a construction can be defined as the indispensable representative of that construction". If one examines the indispensibility of items in each of the above examples, one discovers that it is only the first constituent following the Determiner ang that is obligatory, and is therefore the head, while the rest of the phrase is optional. Since these are both noun phrases, marked as such by the Determiner as well as by the fact that they may occur in any position in the language, such as subject, which requires a noun phrase, the head of each of these constructions is necessarily a noun, as shown in Fig. 1a. and 1b. It should be noted that the form of the lexicase representation displays the dependency relationships within a construction. Vertical lines mark the heads of constructions, slanting lines mark the dependents of heads.\(^5\)

Figure 1a.    Figure 1b.

\[\begin{array}{c}
\text{ang} \\
+ \text{Det} \\
\text{ang bahay 'the house'} \\
\text{the house}
\end{array}
\]

\[\begin{array}{c}
\text{ang} \\
+ \text{Det} \\
\text{ang maliit 'the little one'} \\
\text{the little}
\end{array}
\]

The modifying constituents in ex. 48-49, introduced by the prepositions na both have the form of a relative clause and occur in the typologically expected position following their head nouns. All relative clauses in Tagalog form exocentric constructions with a preceding preposition na (or its phonologically determined alternate post-clitic form -ng), commonly referred to in the literature as a 'linker' or 'ligature'. A relative

\(^5\) Only those features of the lexical items that are directly relevant to the discussion here are included. A full lexicase representation of the examples is far beyond the scope of this paper.
clause consists of the predicate head of a clause, either nominal or verbal, with its non-subject dependents. The missing subject of such a clause is ultimately co-referential with the head of the noun phrase, as illustrated below. Compare the structures of Fig. 2a. and 2b., and 3a. and 3b.

**Figure 2a.**

```
+Det  +N
ang bahay  
      +P
na maliit  
      +N
        +prdc
```

ang bahay na maliit 'the house which is a little one'

**Figure 2b.**

```
+Det  +N
ang bahay  
      +P
na binili  
      +V
        +prdc
niya  
      +N
```

ang bahay na binili niya 'the house that he bought'

**Figure 3a.**

```
+Det  +N
ang maliit  
      +P
na bahay  
      +N
        +prdc
```

ang maliit na bahay 'the little one which is a house'
In most of the work on word order typology and language universals, the presence of adjectives in the languages referred to is simply assumed. Forms that translate into English as adjectives have been treated as though they are adjectives also in the source language. Yet a number of linguists have questioned the validity of adjective as a syntactic word class in various SEA (and other) languages. Savetamalya (1989, p. 76) states, "Stative verbs such as siyaf in Thai translate as adjectives, or they function like adjectives in other languages, but they are not adjectives in Thai." Hudak (1990, p. 420) comes to the same conclusion. Nguyen (1990, p. 63) says the same for Vietnamese.

The word class Adj (adjective) is defined in lexicase as "the head of an adjective phrase, an endocentric, non-predicational attribute of a noun" (Starosta 1988, p. 51, italics added). Given analyses such as the above, it is clear that Tagalog, and Philippine languages which have similar structures, probably do not have a class of adjectives either, since the lexical items translatable in English as adjectives are all predicational. They probably constitute a subclass of verbs. As such, they are typologically consistent, forming noun phrases having a head-attribute word order.

Let us now consider the status of numerals in Philippine languages. Lexicase assumes a limited set of eight word classes with which a language forms constructions: noun (N), verb (V), adjective (Adj), adverb (Adv), determiner (Det), pre- or postposition (P), conjunction (cnj), and sentence particle (SPart). A language may have fewer than these, but it will not have more. The theory does not allow the possibility of a class of Numerals (or of Classifiers) distinct from the above set. The theory then forces us to decide to which class such lexical items belong. In Philippine languages, as in most languages, numerals typically precede the nouns that they enumerate, as illustrated in ex. 50.

**Tagalog**

50. kinuha niya yong apat na mangga  'He got those four mangoes.'
got he that four 1kr mango
But as with other adjective-like words occurring immediately following a
determiner, the numeral is the obligatory constituent, it is the head of its construction, and
must therefore be a noun. The modifying element in the phrase is not the numeral, but
the optional relative clause which follows it, see ex. 51.

51. kinuha niya yong apat 'He got those four.'
got he that four

Fig. 4 displays the structure of ex. 50, and illustrates again the typological
regularity of the word order, with a head noun followed by its attribute.

Figure 4.

kinuha
+V

niya
+N

yong +Det

apat
+N

na
+P

mangga
+N

+prdc

kinuha niya yong apat na mangga 'He got the four which are mangoes.'
got he that four mango

Although Tagalog has been shown to be strongly right branching, with relative
clauses following their head nouns, a glance at the tree diagrams will show that not all
nominal attributes follow their heads. Determiners invariably precede their heads.
Historically though, these can also be shown to have developed from the sequence of a
demonstrative noun followed by a relative clause. The form yong, for example, can be
shown to have developed from the demonstrative iyan + -ng, the latter being the form of
the preposition na following a vowel or an alveolar nasal (which it replaces), as shown in
Fig. 5. The two patterns both occur in Tagalog today.
The conditions which brought about this structural change are not clear. Certainly phonological reduction and semantic bleaching were involved, but they resulted in a word order which is contrary to what is expected. Interestingly, exactly the same changes have taken place, apparently independently, in many Philippine languages, including Ilokano and Bontok. Other Philippine languages, such as Ivatan and Isinai, have developed sets of post-nominal determiners in a similar manner, that is, by reducing a sequence of relative noun plus demonstrative to a postclitic determiner.

We shall now return to the problem of apparently inconsistent word orders in quantifier constructions, discussed in section 2 above. An examination of numeral classifier constructions in Khmer as discussed by Sak-Humphrey (1994) is instructive. In a revealing lexicase analysis, she has shown that at least in that language, numerals must be considered to be nouns. She provides data which shows that Khmer numerals function in every respect as nouns. They may occur without dependents as the heads of subject and object noun phrases. They may take as dependents a demonstrative noun, or a relative clause. Ex. 53-54 illustrate some of these characteristics. The lexicase representations which follow the examples are Sak-Humphrey's.

**Khmer** (Sak-Humphrey, 1994)

52. moy nh tvoe pi chhoe
   one   this  made of wood

'This one is made of wood.'
Figure 6.

```
moy  nih  tvoe  pi  chhoe
  +N    +Det +V    +P    +N
```

moy nih tvoe pi chhoe
one this made of wood

'This one is made of wood.'

53. swaay nih pram lau
mango this five dozen

'These mangoes are five dozens.'

Figure 7.

```
swaay  pram  lau
  +N    +N    +N
   nih    +prdc  +prdc
   +Det
```

swaay nih pram lau
mango this five dozen

'These mangoes are five dozens.'

Given analyses such as these, the numeral constructions in Khmer can be seen to be consistent with the expected word order typology. They are noun phrases, just as in Philippine languages, with the numerals functioning as the nominal heads of their constructions. The analyses of the classifier constructions given in ex. 29-30 are shown here in Fig. 8-9, to further illustrate this point.
Sak-Humphrey's analysis of numeral classifiers as predicate nouns in Khmer corresponds precisely to the analysis of classifiers in Thai described in Savetamalya (1989, p. 155). Savetamalya is forced however to consider numerals as adjectives in Thai, because they only occur as dependents of nouns. They may never function as the sole constituent of a noun phrase (however see the discussion of Fig. 13 and 13' below), nor can they be modified by determiners or relative clauses, as numerals can in Khmer. The analysis of classifier constructions in Thai differs therefore from what we have seen for Khmer. The analysis of ex. 23 (taken from Savetamalya 1989, p. 155) and 24 are shown in Fig. 10-11.

Although it seems that in modern Thai numerals function as adjectives, there is evidence to suggest that at some earlier stage of the language they may have formed a class of nouns. Example 18 above shows the numeral นี่่ 'one' occurring following its head noun. However, only the cardinal numeral นี่่ 'one' may occur in this position, and
it functions as an indefinite determiner rather than as a numeral (Hudak 1990, p. 43), so that it must be analyzed as a Determiner, as in Fig. 12.

\[ \text{Figure 12.} \]

\[ \text{baan} \quad +\text{N} \]
\[ +\text{Det} \]

\[ \text{baan} \quad \text{nin} \quad \text{'a house'} \]

house one

In the same position, an ordinal number may also occur. Ordinal numbers in Thai have traditionally been considered to be compound nouns as illustrated in Figure 13. However, the structure of ordinal numbers consists of a sequence of what must be analyzed elsewhere in the language as a relator noun \( \text{thii} \) followed by a dependent nominal sister as in Figure 14. Historically, even if not synchronically, numbers must have been nouns in earlier stages of Thai.

---

6 Although today the form meaning 'one' is pronounced the same whether it occurs preceding a classifier or whether it occurs following a noun, at earlier stages of the language, they must have been pronounced differently. The former is written with an initial aspiration, the latter is not. Just as in English, the numeral 'one' became phonologically reduced to 'an' when it developed into a determiner, similarly in Thai the determiner is phonologically reduced from its numeral counterpart.

7 This is one of four homophonous \( \text{thii} \) forms, described in Savetamalya 1989. Relator nouns serve to carry the localistic feature of the noun phrase of which they are head. Since numerals do not themselves carry localistic features, they cannot directly carry a Locus case relation. Since relator nouns do carry this information, they serve to mark the case relation of the whole noun phrase.
Figure 13.

baan
+N
thīi-sāam
+Det

baan thīi-sāam "third house"
house third

Figure 14.

khon
+N
thīi
+N
+rtl
+letn
LOC

khon thīi bāan "person in a house"
person in house

This pattern strongly suggests that the origin of ordinal nouns in Thai was a structure in which the numeral was a noun preceded by the relator noun thīi, as in Figure 15.
Figure 15.

```
bāan  +N
     |  +N
     | +ritr
     +letn
     LOC

thīi  +N

sāam
+N
```

bāan thīi sāam  Lit. 'house in place three'
house in three

4. CONCLUSION

We have attempted to show that in Philippine languages (all Austronesian), Khmer (an Austroasiatic language), and at least historically in Thai (in the Tai-Kadai family), numerals must be analyzed as nouns and therefore function as heads of their own noun phrases. Whether similar analyses are justified for the data cited from the other languages given in section 2 is unknown at this point. Without access to native speakers, and with the published materials on the languages inadequate for this purpose, we must leave the question open.

The analysis that has been shown to be best for the Philippine languages as well as for Khmer, has the advantage of showing that these languages are typologically more consistent as far as their word order is concerned than previous descriptions have implied. The analyses also have implications for earlier stages of the languages, as we have suggested for Thai. We should note also that Greenberg attempted to determine which of the alternate word order patterns in which quantifiers are involved in Thai was the original order. He concluded (Greenberg 1975, p. 37?) that the Noun-Quantifier order of the type illustrated in ex. 13 was older, and that the reverse order was an innovation. It is probable, however, that if quantifiers were nouns, as our analysis suggests, they could have occurred either preceding or following other nouns, much as they do in Philippine languages and in Khmer, with those following quantified nouns acting as dependent sisters, and those preceding quantified nouns being their phrasal heads. There is no need to consider either order as somehow earlier or more basic.

REFERENCES


