Mon Khmer Word Order from a Crosslinguistic Perspective

Matthew S. Dryer  
SUNY Buffalo

Mon-Khmer languages overwhelmingly exhibit word order properties that can be described as head-initial. If we ask why we find this strong tendency for consistent head-initial order among Mon-Khmer languages, a popular answer is that languages tend to be consistently head-initial or consistently head-final, and Mon-Khmer languages are simply instances of the former type. However, as I have argued elsewhere (Dryer 1988, 1991, 1992), the assumption that languages tend to be consistently head-initial or head-final is not true, that the order of a number of kinds of modifiers with respect to their heads does not exhibit any crosslinguistic correlation with the order of verb and object. I will discuss alternative explanations for the pattern we find in Mon-Khmer languages.¹

The table in (1) summarizes properties that can be said to characterize languages which are consistently head-final or consistently head-initial.

(1)

<table>
<thead>
<tr>
<th></th>
<th>Head-final</th>
<th>Head-initial</th>
</tr>
</thead>
<tbody>
<tr>
<td>verb</td>
<td>occurs last, following subject, object, adpositional phrases, adjuncts, adverbs</td>
<td>verb precedes object, adpositional phrases, adjuncts, adverbs, though not necessarily the subject</td>
</tr>
<tr>
<td>postpositions</td>
<td></td>
<td>prepositions</td>
</tr>
<tr>
<td>noun</td>
<td>occurs last in noun phrase</td>
<td>noun occurs first in noun phrase</td>
</tr>
<tr>
<td>modifiers</td>
<td>in general precede the modified element</td>
<td>modifiers in general follow the modified element</td>
</tr>
</tbody>
</table>

The examples in (2) (from Jacob 1968) illustrate how Khmer conforms to the general characteristics of head-initial languages.

(2) Khmer

a. VO

\[ ?o:púk thiú: tũ: \]

father make cupboard
‘Father is making a cupboard’ (p. 262)

b. Verb - Nominal Adjunct
   po:nu tu sa:la:-riŋ (p. 262)
   younger.sibling go school
   ‘(My) younger brother/sister is going to school’

c. Verb - Verbal Adjunct
   vi:tu tυu rahas
   3SG go quick
   ‘He goes quickly’ (p. 79)

d. Noun-Adjective
   phṭe:ah to: c
   house small
   ‘a small house’ (p. 60)

e. Noun-Genitive (i.e. noun possessor)
   phṭe:ah ta:
   house old.man
   ‘the old man’s house’ (p. 263)

f. Noun-Possessor (i.e. pronoun possessor)
   ðo:puk khjom
   father 1SG
   ‘my father’ (p. 60)

g. Noun-Numeral
   phṭe:ah πi:r
   house two
   ‘two houses’ (p. 62)

h. Noun-Demonstrative
   bɔntup niŋ
   room this
   ‘this room / these rooms’ (p. 64)

i. Noun-Interrogative
   phṭe:ah ra:
   house which
   ‘which house’ (p. 266)

j. Adjective-Intensifier
   thmvu nas
   new very
   ‘very new’
As shown by Dryer (1992), a number of the word order characteristics illustrated for Khmer in (2) are indeed typical of VO languages. These include the characteristics listed in (3).

(3) Typical VO characteristics (Greenberg 1963, Dryer 1992)

Noun-Genitive
Noun-Relative Clause
Prepositions
Verb-PP
Verb-Adverb
Adjective-Marker-Standard
etc.

However, some of the characteristics of Khmer illustrated in (2) are not typical of VO languages, despite their involving head-initial order. These include those listed in (4).

(4) Typical Mon-Khmer but NOT typical VO characteristics

Noun-Adjective
Noun-Demonstrative
Adjective-Intensifier

Below, I will illustrate each of these characteristics from other Mon-Khmer languages, and present data from a database containing data for over 750 languages showing that these are not typical VO characteristics.

The examples in (5) illustrate the noun-adjective order from a number of Mon-Khmer languages.

(5) a. Vietnamese

nhà nhở
house small
‘a small house’ (Thompson 1965: 221)

b. Palaung

rpyā kariär
girl beautiful
‘the beautiful girl’ (Milne 1921: 38)

c. Khmu?

kôn láʔ
man good
The data in (6) below illustrates the distribution of AN versus NA order among VO languages from my database. The data is organized in the following way (cf. Dryer 1989a, 1992). First, I have classified the languages into genetic groups, which I call genera, which are roughly comparable to the subfamilies of Indo-European, and the numbers in (6) represent the number of genera containing languages of each type. In addition, I group these genera into six large geographic areas and count the number of genera within each area.² For example, the ‘5’ in the upper lefthand corner of (6) indicates that there are 5 genera in
Africa containing languages in my database which are VO and AN, while the ‘32’ below it indicates that there are 32 genera in Africa containing languages in my database which are VO and NA. The larger figure for each area is enclosed in a box.

\[
\begin{array}{ccccccc}
\text{Area} & \text{Afr} & \text{Eur} & \text{SEA} & \text{A-NG} & \text{NAm} & \text{SAm} & \text{Total} \\
\text{VO&AN} & 6 & 4 & 5 & 17 & 3 & 40 \\
\text{VO&NA} & 32 & 12 & 3 & 7 & 5 & 63
\end{array}
\]

NA order is often thought to be typical of VO languages, but the data in (6) shows that both orders are common, and that AN order is more common than NA order in 3 of the 6 areas. The order VO&NA is somewhat more common overall (by 63 genera to 40), but this difference is completely attributable to the large number of VO&NA languages in Africa: 32, or more than half of the 63 VO&NA genera, are from Africa. Outside of Africa, VO&AN is slightly more common, by 35 genera to 31.

Now one might argue that NA order is still somewhat more common than AN order among VO languages and that this suggests that there is a weak correlation between VO order and NA order. However, it turns out that this weak preference for NA order is also found among OV languages, as demonstrated by the data in (7).

\[
\begin{array}{ccccccc}
\text{Area} & \text{Afr} & \text{Eur} & \text{SEA} & \text{A-NG} & \text{NAm} & \text{SAm} & \text{Total} \\
\text{OV&AN} & 7 & 28 & 2 & 5 & 9 & 7 & 58 \\
\text{OV&NA} & 24 & 4 & 5 & 17 & 15 & 19 & 86
\end{array}
\]

The data in (7) show that in 5 areas out of 6, OV&NA order is more common than OV&AN, that the latter order is more common only in Eurasia.

In short, while NA order may be slightly more common among VO languages, it is also somewhat more common among OV languages. Hence, the occurrence of NA order among Mon-Khmer languages cannot be attributed to a general tendency for VO languages to be NA.

The situation with the order of noun and demonstrative is somewhat similar. First, the examples in (8) illustrate how the normal order among Mon-Khmer languages is for the demonstrative to follow the noun.
(8) Noun-Demonstrative

a. Vietnamese

nhà này
building this
‘this building’ (Thompson 1965: 191)

b. Palaung

jùär inân
basket this
‘this basket’ (Milne 1921: 47)

c. Khmu?

kxn nê? ki:
child little this
‘this little child’ (Premsrirat 1987: 31)

d. Katu

tariq gamak achek
buffalo big that
‘that big buffalo’ (Costello 1969: 34)

e. Stiêng

côc nêy
place that
‘that place’ (Miller 1976: 43)

f. Chrau

ier heq
chicken this
‘this chicken’ (Thomas 1971: 139)

g. Sre

?aso nc
dog that
‘that dog’ (Manley 1972: 156)

h. Sedang

hái tá
day that
‘that day’ (Smith 1979: 78)
i. Mon

kwan te?
village that
‘that village’ (Bauer 1982: 321)

One Mon-Khmer language which is an exception to this pattern is Khasi, as illustrated in (9).

(9) Khasi: Dem-“Article”-Noun

utra briew
MASC.That MASC person
‘that man’ (Nagaraja 1985: 13)

I assume that this is related to the geographical location of Khasi, in northeast India, outside the area in which NDem order is common and in an area in which most other languages are DemN.

The data in (10) from my database shows that both orders of demonstrative and noun are common among VO languages. Although the overall pattern is one whereby both orders are about equally common (54 versus 56), again the large number of VO&NDem languages in Africa is distorting and DemN order is actually more common outside of Africa (by 50 genera to 22).

(10) Afr Eur SEA A-NG NAm SAm Total
VO&DemN 4 8 7 6 20 9 54
VO&NDem 34 1 12 0 7 2 56

The examples in (11) illustrate how what I call intensifiers, words modifying adjectives and indicating degree, typically follow the adjective in Mon-Khmer languages.

(11) Adjective-Intensifier

a. Khmer

thmyu nas
new very
‘very new’ (Jacobs 1968)
b. Sre
   mhar nənj
   fast very
   'very fast' (Manley 1972: 219)

c. Sedang
   lən tələi
   beautiful exceptionally
   'exceptionally beautiful' (Smith 1979: 128)

d. Mon
   khoh dən
   good quite
   'quite good' (Bauer 1982: 387)

For this pair of elements, the preference for head-initial order is less strong among Mon-Khmer languages. The examples in (22) to (26) illustrate how a number of Mon-Khmer languages exhibit some words of this sort following the adjective but others preceding. In the (a) examples, the intensifier precedes the adjective while in the (b) examples, it follows.

(12) Vietnamese
   a. ếń vui
      very happy
      'very happy' (Thompson 1965: 222)

   b. dép ləm
      beautiful very
      'very beautiful' (Thompson 1965: 272)

(13) Palaung
   a. gōp ūr
      very tired
      'very tired' (Milne 1921: 105)

   b. dāng gāt
      large very
      'very large' (Milne 1921: 104)
Despite the number of languages in which both orders of intensifier and adjective are found, I am not aware of any Mon-Khmer language in which the normal order is Intens-Adj. Among languages in which one order is dominant, the order Adj-Intens appears to be the dominant one.

Once again, crosslinguistic data shows that there is no crosslinguistic tendency for the intensifier to follow the adjective in VO languages. The data in (17) shows that both orders are about equally common among VO languages.

(14) Khmu?

a. lāk pŋkà
   really shy
   ‘really shy’ (Premsrirat 1987: 42)

b. kòn lā? phó:t
   too.much good too.much
   ‘too good’ (Premsrirat 1987: 68)

(15) Katu

a. long gamak
   very big
   ‘very big’ (Costello 1969: 31)

b. gamak pablōng
   big very
   ‘very big’ (Costello 1969: 31)

(16) Chau

a. mo’yāh māq
   very big
   ‘very big’ (Thomas 1971: 110)

b. māq trōq
   big extremely
   ‘extremely big’ (Thomas 1971: 110)

We can summarize what has been demonstrated to this point as follows. Mon-Khmer languages tend to be consistently
head-initial, as reflected by the common orders NA, NDem, and AdjIntens. A common assumption is that these characteristics are simply typical of VO languages, as reflections of head-initial order. I have shown, however, that these characteristics are by no means typical of VO languages, so that their occurrence in Mon-Khmer languages cannot simply be explained in terms of a supposed tendency for VO languages to be head-initial. In the remainder of this paper, I will examine some alternative hypotheses for why Mon-Khmer languages might exhibit consistent head-initial order.

Even though I have demonstrated that VO languages do not exhibit a general tendency to be head-initial, one might suggest that the VO languages which are not consistently head-initial are languages which were once OV and have retained certain head-final characteristics. Such an explanation is plausible for languages like Finnish, for example, which have demonstrably changed their order from OV to VO while still retaining a number of characteristics from the time when the language was OV. One might argue that the VO order in Mon-Khmer is very old and that Mon-Khmer languages are in some sense “purer” VO languages, without OV ancestry. The problem with this explanation is that there are a number of families or areas in which adjectives or demonstratives commonly precede the noun or intensifiers precede the adjective and in which an explanation in terms of OV ancestry seems unlikely. This includes the Austronesian family, where there is no evidence of OV ancestry and where there is considerable variation in the order of adjective and noun, sufficient variation that the prenominal position of the adjective in many languages must be relatively recent. In addition, there are two areas in the New World in which AN order and DemN order are quite common in which VO is the normal order as part of an areal phenomenon, namely Meso-America and the Pacific Northwest. The common AN order in these areas does not seem attributable to any OV ancestry.

A further hypothesis would be that there is a subtype of OV languages which are consistently head-final and a subtype of VO languages which are consistently head-initial, that both subtypes belong to a general type that involves crosscategorial consistency in head position in contrast to languages that do not. But while there may indeed be such a subtype for OV languages, illustrated by Japanese, closer examination of Mon-Khmer languages shows that they are often not entirely head-initial, that
many of them exhibit sporadic exceptions to the trend towards head-initial order, so that in fact we cannot say that Mon-Khmer languages are consistently head-initial without exception. Consider, for example, the order of numeral and noun. While some Mon-Khmer languages are NNum, as in (18), others place the numeral before the noun, as in (19).

(18) Noun-Numeral
a. Khmu?
   kxeŋ pàːr kɔn
   child two CLSFR
   ‘two children’ (Premšriat 1987: 34)

b. Khmer
   phèːh pàːr
   house two
   ‘two houses’ (Jacob 1968: 62)

c. Mon
   nɔm chuŋ pɔn nɔm
   CLSFR tree four CLSFR
   ‘four trees’ (Bauer 1982: 359)

(19) Numeral-Noun
a. Vietnamese
   bà cái nhà
   three CLSFR house
   ‘three houses’ (Thompson 1965: 198)

b. Katu
   pe (panong) anuŋ
   three (CLSFR) dog
   ‘three dogs’ (Costello 1969: 26)

c. Chrau
   dʊ vanɔnɔŋ gəpu
   one CLSFR buffalo
   ‘one buffalo’ (Thomas 1971: 133)
Other languages, such as Palaung and Stiêng, illustrated in (20) and (21), employ two patterns with numerals, one without a classifier, in which the numeral precedes the noun, and a second with a classifier in which the numeral plus classifier follows the noun.

(20) Palaung: Num+Noun or Noun+Num+Clsfr
   a. ḁr rè b. bāŋ u tō
two village horse one CLSFR
   ‘two villages’ ‘one horse’ (Milne 1921: 57)

(21) Stiêng: Num+Noun or Noun+Num+Clsfr
   a. pēy sām b. nē bāar ḋc
three year house two CLSFR
   ‘three years’ ‘two houses’ (Miller 1976: 31, 32)

Now one might argue that in constructions with a classifier, the classifier is really the head and the noun is itself a modifier, unlike the analysis commonly assumed for European languages. However, it seems unlikely that in some Mon-Khmer languages the classifier is head while in others the noun is head. If we claim that the classifier is consistently the head, then the languages in (18), in which the numeral plus classifier follows the noun would not be exhibiting a head-initial construction. Furthermore, if we restrict attention to the order of numeral and classifier, which is apparently always Num + Classifier in Mon-Khmer languages, we have a further problem, because the numeral is most plausibly analysed as a modifier of the classifier, in which case we have a further example of an order which is not head-initial.
A further example of a type of modifier that commonly precedes the noun is plural words (cf. Dryer 1989b), separate words indicating the plurality of the noun phrase, as in (22).

(22) a. Vietnamese

những chó
PLURAL dog
‘dogs’ (Thompson 1965: 179)

b. Stiêng

bol ổi au
PLUR morning here
‘these mornings’ (Miller 1976: 34)

c. Sedang

vai đrôh
PLUR girl
‘the girls’ (Smith 1979: 92)

Such plural words typically precede the noun in Mon-Khmer languages, again contrary to the general trend towards consistent head-initial order.

And in various Mon-Khmer languages there are further sporadic instances of other sorts of prenominal modifiers. For example, in Mon, the quantifier meaning ‘every’ precedes the noun (in contrast to numerals which follow the noun), as in (23).

(23) Mon: NNum but QuantN

rôh ɲoa
every day
‘every day’ (Bauer 1982: 343)

In (24) is a summary chart of the Mon-Khmer languages in my database. In most of these languages, we find at least one instance in which a modifier precedes the word it modifies, thus showing that completely consistent head-initial order is apparently the exception.
(24) Summary chart

<table>
<thead>
<tr>
<th></th>
<th>Gen</th>
<th>Adj</th>
<th>Dem</th>
<th>Num</th>
<th>Art</th>
<th>Plur</th>
<th>Intens</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khasi</td>
<td>NG</td>
<td>NA</td>
<td>DemN</td>
<td>NumN</td>
<td>ArtN</td>
<td>PIN</td>
<td>AdjInt/IntAdj</td>
<td>NOrd QuantN/IntN</td>
</tr>
<tr>
<td>Palaung</td>
<td>NG</td>
<td>NA</td>
<td>NDem</td>
<td>NumN/Num</td>
<td></td>
<td></td>
<td>IntAdj/AdjInt</td>
<td></td>
</tr>
<tr>
<td>Khmu</td>
<td>NG</td>
<td>NA</td>
<td>NDem</td>
<td>NNum</td>
<td></td>
<td></td>
<td>IntAdj/AdjInt</td>
<td></td>
</tr>
<tr>
<td>Vietnamese</td>
<td>NG</td>
<td>NA</td>
<td>NDem</td>
<td>NumN</td>
<td>DefN</td>
<td>PIN</td>
<td>IntAdj/AdjInt</td>
<td></td>
</tr>
<tr>
<td>Kau</td>
<td>NG/gn</td>
<td>NA</td>
<td>NDem</td>
<td>NumN/num</td>
<td></td>
<td></td>
<td>IntAdj/AdjInt</td>
<td></td>
</tr>
<tr>
<td>Siêng</td>
<td>NG</td>
<td>NA</td>
<td>NDem</td>
<td>NumN/Num/Num</td>
<td>PIN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chau</td>
<td>NG</td>
<td>NA</td>
<td>NDem</td>
<td>NumN</td>
<td></td>
<td></td>
<td>IntAdj/AdjInt</td>
<td></td>
</tr>
<tr>
<td>Sre</td>
<td>NG</td>
<td>NA</td>
<td>NDem</td>
<td>NumN</td>
<td></td>
<td></td>
<td>AdjInt</td>
<td></td>
</tr>
<tr>
<td>Bao</td>
<td>NG</td>
<td>NA</td>
<td>NDem</td>
<td>NNum</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sedang</td>
<td>NG</td>
<td>NA</td>
<td>NDem</td>
<td>NumN</td>
<td>PIN</td>
<td>AdjInt</td>
<td>NOrd QuantN/QuantN</td>
<td></td>
</tr>
<tr>
<td>Khmer</td>
<td>NG</td>
<td>NA</td>
<td>NDem</td>
<td>NNum</td>
<td>NArt</td>
<td></td>
<td>AdjInt</td>
<td>NInt</td>
</tr>
<tr>
<td>Mon</td>
<td>NG</td>
<td>NA</td>
<td>NDem</td>
<td>NNum</td>
<td>NDef</td>
<td>NPI</td>
<td>AdjInt</td>
<td>QuantN</td>
</tr>
</tbody>
</table>

This suggests that while there may be a general trend towards head-initial order in Mon-Khmer, we cannot say that they are consistently head-initial, thus casting doubt on the idea that the instances in which they are head-initial can be explained in terms of some general rule of head-initial order.

So why do the Mon-Khmer languages exhibit a stronger tendency towards head-initial languages than VO languages elsewhere in the world? The simplest explanation would be that these orders in MK languages are simply accidental, that their occurrence does not reflect a greater pattern, any more than a random win on a slot machine. While much effort in modern linguistics is to find appealing explanations for specific phenomena in different languages, the possibility that many phenomena have no real explanation should not be overlooked. According to this view, the fact that the adjective and the demonstrative generally follow the noun in Mon-Khmer languages, and that intensifiers so often follow the adjective are simply among the many arbitrary properties found in any language, and any attempt to reduce them to some more abstract principle may simply be misguided.
Notes

1 It should be noted that the pattern exhibited by Mon-Khmer languages is part of an overall pattern within Asia, in which Mon-Khmer languages exhibit head-initial characteristics, shared with Daic languages, and contrasting sharply with languages to the north and west, which exhibit head-final characteristics (e.g. Japanese, Korean, Mongolian, Turkic, Uralic, Indo-Iranian, Dravidian), with a belt of languages in between exhibiting mixed characteristics (Chinese, Tibeto-Burman, Miao-Yao). The properties of Mon-Khmer languages discussed here are thus part of a larger areal pattern that includes Daic languages, and one Mon-Khmer language, Khasi, spoken in northeast India, does not exhibit as clear a head-initial pattern. For this reason, the phenomenon being discussed is really an areal phenomenon rather than something specific to Mon-Khmer, though I will restrict attention here to Mon-Khmer.

2 The abbreviations for the six areas are: Afr = Africa, Eur = Eurasia, SEA = Southeast Asia and Oceania, A-NG = Australia-New Guinea, NAm = North America, and SAm = South America. SEAsia & Oceania includes the languages of Southeast Asia, Sino-Tibetan, and Austronesian languages. Eurasia contains the remaining languages of Europe and Asia.

3 Key to summary chart in (24):

<table>
<thead>
<tr>
<th>Gen</th>
<th>Genitive</th>
<th>Pl</th>
<th>Plural Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj</td>
<td>Adjective</td>
<td>Int</td>
<td>Intensifier</td>
</tr>
<tr>
<td>Dem</td>
<td>Demonstrative</td>
<td>Quant</td>
<td>Quantifier</td>
</tr>
<tr>
<td>Num</td>
<td>Numeral</td>
<td>Ord</td>
<td>Ordinal Numeral</td>
</tr>
<tr>
<td>Art</td>
<td>Article</td>
<td>Int</td>
<td>Interrogative Modifier</td>
</tr>
<tr>
<td>Def</td>
<td>Definite marker</td>
<td>Clsfr</td>
<td>Classifier</td>
</tr>
</tbody>
</table>

References


It should be noted that the pattern exhibited by Mon-Khmer languages is part of an overall pattern within Asia, in which Mon-Khmer languages exhibit head-initial characteristics, shared with Daic languages, and contrasting sharply with languages to the north and west, which exhibit head-final characteristics (e.g. Japanese, Korean, Mongolian, Turkic, Uralic, Indo-Iranian, Dravidian), with a belt of languages in between exhibiting mixed characteristics (Chinese, Tibeto-Burman, Miao-Yao). The properties of Mon-Khmer languages discussed here are thus part of a larger areal pattern that includes Daic languages, and one Mon-Khmer language, Khasi, spoken in northeast India, does not exhibit as clear a head-initial pattern. For this reason, the phenomenon being discussed is really an areal phenomenon rather than something specific to Mon-Khmer, though I will restrict attention here to Mon-Khmer.

The abbreviations for the six areas are: Afr = Africa, Eur = Eurasia, SEA = Southeast Asia and Oceania, A-NG = Australia-New Guinea, NAm = North America, and SAm = South America. SEAsia & Oceania includes the languages of Southeast Asia, Sino-Tibetan, and Austronesian languages. Eurasia contains the remaining languages of Europe and Asia.

Key to summary chart:

- **Gen**: Genitive
- **Adj**: Adjective
- **Dem**: Demonstrative
- **Num**: Numerical
- **Art**: Article
- **Def**: Definite marker
- **Pl**: Plural Word
- **Intens**: Intensifier
- **Quant**: Quantifier
- **Ord**: Ordinal Numeral
- **Int**: Interrogative Modifier
- **Clstr**: Classifier