Clefts and Wh-questions in Moken
Peter Jenks, Harvard University
January, 2008

This paper examines cleft structures in Moken, an Western Malayo-Polynesian language.¹ The basic questions which this paper examines are: (1) Do Moken questions involve movement? (2) What is the structure of Moken clefts? and (3) How can we account for systematic similarities in extractions asymmetries between wh-questions, multiple wh-questions, and wh-cLEFTs? I argue that a shared restriction on extraction of subjects between what is likely a cleft and wh-questions points towards a cleftlike structure for questions, as is found in many Austronesian languages.

Moken has a number of properties which render it perplexing to structural analysis. A shift to final stress under the influence of Mainland Southeast Asian languages is likely responsible for systematic loss of the initial voice morphology found in most Austronesian languages. Additionally, Moken in its current state features relatively flexible SVO word order compared to its non-Austronesian counterparts in Southeast Asia. An ancillary goal of this paper is to begin exploring some of the structural roots of word order flexibility, exploiting configurational properties which are likely to be present in Moken by analogy with other Austronesian languages.

In the next section I will introduce the wh-question data and point to analytic problems which that data presents. Following that section, I examine what appears to be a cleft-construction in Moken associated with a particular lexical item which I argue is a copula. In the second part of that section, I present evidence for a cleft construction by demonstrating that the construction is biclausal. In the final section, I examine a basic restriction on extracting subjects from the embedded clause in both clefts, clefted wh-constructions, and multiple wh-questions. This restriction does not hold for objects, which can be extracted. I present an analysis of these facts consistent with an analysis of Moken as having clefts.

1. Basic properties of wh-questions

¹ Thanks to the generous hospitality of the Moken on Hlao Island and Elephant Island in Ranong Province, Thailand, and to See, my primary consultant. Special thanks also to Maria Polinsky, without whom this paper would never have existed in its current form, and also to helpful suggestions from Cedrick Boeckx and Eric Potsdam. Errors are mine.
Moken is a morphologically isolating language, with relatively few functional
formatives and a total lack of tense inflection or subject agreement on verbs, a total lack of
number and gender markers on nouns, and no case to boot. As such, our only arguments for
syntactic analyses come from purely configurational data, which is ample.

Moken wh-phrases all license a wh-particle -la: ‘Why’ is formed compositionally as
‘do-what’ (1f), as is common in many languages (e.g. Thai).

(1)  a. aca:-la:  nam  co:n
     who-WHQ  eat  rice
     ‘Who ate (rice)?’

b. som?qo:  nam  ano:-la:
     NAME  eat  what-WHQ
     ‘What did Som’oo eat?’

c. som?qo:  ni?en  bita:-la:
     NAME  play  where-WHQ
     ‘Where is Som’oo playing?’

d. som?qo:  ni?en  jipa:-la:
     NAME  play  how-WHQ
     ‘How is Som’oo playing?’

e. som?qo:  ni?en  kotan  ope?e-n-la:
     NAME  play  forest  when-WHQ
     ‘When did Som’oo play in the forest?’

f. canat  ni?en  toŋ  som?qo:  bo?-ano:-la:
     child  play  with  NAME  do-what-WHQ
     ‘Why did the children play with Som’oo?’

The wh-particle is actually optional in all of the examples above, and I have found no reported
effects on interpretation. I assume that it’s presence is the realization of some question feature
which is syntactically present regardless of whether the particle is realized overtly. Such a
move is justified by the simple fact that the interpretation of questions is unaffected by the absence of the particle.

While the wh-particle la: can appear immediately after the wh-constituent that licenses it, it can also appear sentence finally. The following data also demonstrate that it is optional in both positions, or, alternately, can appear in both:

\[(2)\]
\[
a. \text{aca-} \text{la:} \quad [\text{nam con}] \\
\text{who-} \text{WHQ} \quad \text{eat} \quad \text{rice}
\]
\[
b. \text{aca:} \quad [\text{nam con}] \quad \text{la:} \\
\text{who} \quad \text{eat} \quad \text{rice} \quad \text{WHQ}
\]
\[
c. \text{aca-} \text{la:} \quad [\text{nam con}] \quad \text{la:} \\
\text{who- LA:} \quad \text{eat} \quad \text{rice} \quad \text{WHQ}
\]
\[
d. \text{aca:} \quad [\text{nam con}] \\
\text{who} \quad \text{eat} \quad \text{rice} \\
(a-d) \text{‘Who’s eating (rice)?’}
\]

One restriction on the iteration of the wh-question particle illustrated in (2c) is that sentences are marked where the particle occurs twice in a row, with no intervening material:

\[(3)\]
\[
\text{olan} \quad \text{mətok} \quad \text{ano-la:-(?la:)} \\
\text{snake} \quad \text{bite} \quad \text{what-WHQ}
\]

‘What did the snake bite?’

There are a number of possible explanations for (3), among them some trivial phonological restriction on a function word repeating. The crucial piece of evidence would involve an adverb or other adjunct intervening between the two — a piece of evidence which is unavailable, unfortunately (xDATAx). The following data will frequently omit the parentheses indicating the optionality of the wh-question particle for expository purposes; to my knowledge, the marker is always optional.

The immediate question which arises in response to (2) is where the particle is when it appears sentence finally. There is good evidence to suggest that it is outside of VP, in some
higher clausal domain. The first clear piece of evidence is that the second particle is high is in (4), which demonstrates that the particle cannot attach to the verb:

(4)  *aca:-la:  nam-la:kəla:
      who- WHQ        eat-WHQ         monkey

‘Who eats monkeys?’

This is only significant in light of the fact that particles appearing inside the verbal domain, such as the perfective particle \( ka? \), can appear either after the verb or at the end of the VP:

(5)  a. pito:  nam  con-ka?
     NAME  eat       rice-PRF

b. pito:  nam-ka?  con
     NAME  eat-PRF  rice

‘Peter ate (rice).’

Regardless of one’s interpretation, the fact that \( la: \) cannot appear inside the VP dissociated from its host indicates that it must be higher in the clause structure, regardless.

Stronger evidence for the position of \( la: \) can be found in its interaction with adverbs. While \( la: \) can alternate in position with time adverbs (5), which are assumed to be relatively high in the structure, it cannot appear inside of ‘low’-adverbs such as manner adverbs (6). The sentences are shown with assumed structures, illustrating the point being made about the height of the position of \( la: \):

(5)  aca-la:  [\( vp\)mələn  oy]  \{la:\}  aloy-duy  \{la:\}
      who    hit  dog  WHQ    day-this  WHQ

‘Who hit a dog today?’

(6)  aca-la:  [\( vp\)mələn  oy]  \{*la:\}  liŋ]  \{la:\}
      who    hit  dog  WHQ    hard  WHQ

2 It is not inconceivable that there are semantic subtleties between the two positions of the aspect particle \( ka? \), as is the case, for example, with Chinese \( le \), which also occurs either post verbally or after VP but with a perfective vs. resultative aspect, respectively.
‘Who hit a dog hard?’

While the sentences above demonstrate the particle being licensed by a subject wh-phrase, the same facts obtain for wh-words in object position when it is not ‘in situ’ (xDATAx).

More evidence for the high position of the wh-question particle comes from verbal elements which we might want to assume are higher than VP itself. Thus, laː cannot occur inside of final modals (7), which are traditionally assumed to dominate VP (and in terms of semantics must dominate V for reasons of scope) and sentence final aspect particles (8), which are likewise assumed to be higher than VP in structural syntax:

(7) aca-la: [VP[VP bo-kan (*laː)] moy-ha]-laː)
who-Q do-work WHQ can-neg-WHQ
‘Who can’t do work?’

(8) aca-la: [[VP nam con (*la)] haree]-laː)
who-Q eat rice LAː negprf- LAː
‘Who hasn’t eaten yet?’

The structure of the clauses in (7) in (8) is unclear but immaterial to the point being made: under any theory which would assume that syntactic scope reflects semantic scope in the ordering of heads in the verbal domain, laː must be structurally high, somewhere in the “left” periphery.

Before we draw conclusions as to the significance of a high question particle in Moken, we need to consider further data which demonstrates that Moken wh-phrase are not actually in-situ. Thus, in (9) we see that the subject wh-phrase can appear in a number of positions which are ordinarily unacceptable for subjects (xDATAx):

(9) a. aca-la: [VP layam lelen toŋ-ŋan] -laː)
who-WHQ douse candle with-hand WHQ

b. [layam] aca-la: [VP lelen toŋ-ŋan] -laː)
douse who-WHQ candle with-hand WHQ

V-SWh-Obl

c. [VP layam lelen] aca-la: [toŋ-ŋan] -laː)
douse candle who-WHQ with-hand WHQ

V-O-SWh-Obl
d.  
\[ [\text{VP layam lelen toŋ-ŋan}] \text{ aca-la:} \quad \text{V-O-Obl-SWh} \]
\[ \text{douse candle with-hand who-WHQ} \]
(a-d) ‘Who put out the candle with his hand?’

The brackets and labels in (9) represent a suggestion as to how the data could be derived under without abandoning the notion of the VP as a constituent, which we just saw strong evidence for. Basically, the brackets represent the possibility that the wh-phrase is in some VP-external position, and that the various word orders are derived by scrambling elements out of VP or VP itself, stranding others to the left of the wh-phrase.

The following examples, with object wh-phrases, perhaps make the point more clearly. (10a) could not be derived via any process besides some sort of long distance dependency between the wh-phrase and its canonical position as the complement of the verb.

(10)  
a.  
\[ \text{ano-la: } [\text{apoŋ [layam toŋ-ŋan ka?] (-la:)} \]
\[ \text{what-WHQ father douse with-hand PRF WHQ} \]

b.  
\[ [\text{VP apoŋ layam}] \text{ ano-la: [toŋ-ŋan ka?] (-la:)} \]
\[ \text{father douse what-WHQ with-hand PRF WHQ} \]

c.  
\[ [\text{apoŋ layam toŋ-ŋan ka?}] \text{ ano-la: } \]
\[ \text{father douse with-hand PRF what-WHQ} \]
(a-c) ‘What did dad put out with his hand?’

(10) clearly demonstrates that the position of \( la? \) is outside the aspectual \( ka? \). (10c) additionally shows that the wh-word itself is outside of the VP. (10b), on the other hand, is a bit of an analytic puzzle if wh-phrases are assumed to move, and in fact seems to indicate that wh-phrases only move optionally.

In summary, Moken wh-phrases have the following properties:

(11)  
i. Wh-phrases license a wh-particle.
ii. The particle can optionally appear either after the wh-constituent, as a discontinuous final particle, or both.
iii. The position of the wh-particle when detached from the wh-word is high in the structure, presumably in CP.
iv. Wh-phrases can appear either \textit{in situ} or higher in the clause.
Together, these generalizations seem to point towards an analysis of Moken as exhibiting optional \textit{wh}-movement, licensed by some higher \textit{wh}-operator. This operator seems to somehow originate in the \textit{wh}-phrase, a fact wholly compatible with recent analyses of \textit{wh}-movement (e.g. Cable 2007), though it could also be base generated in a higher position and license the in-situ constituent (Tsai 1996). In the following section I will argue that Moken actually does not exhibit \textit{wh}-movement of any sort, and instead seems to exhibit the grammar of clefts in Moken, an unsurprising fact in light of the fact that Moken is an Austronesian language, which have been argued to express \textit{wh}-questions as clefts or pseudoclefts in a number of languages (Oda 2002, Potsdam 2006).

2. Backwards Superiority and the Cleft Connection

One property which has been linked to \textit{wh}-movement is superiority (Chomsky 1973). In English, when there are two \textit{wh}-phrases in a sentence, \textit{wh}-movement is assumed to apply to the constituent which is superior (see reference for relevant definitions). For our purposes, the relevant fact is that subjects are superior to objects. The Superiority Condition thus explains why (12a) is acceptable while (12b) is ill-formed:

\begin{enumerate}
\item[(12)]
\begin{enumerate}
\item a. Who bought what?
\item b. *What did who buy?
\end{enumerate}
\end{enumerate}

Superiority effects, then, demonstrate a clear link with \textit{wh}-movement which has been shown to be restrict in both \textit{wh}-movement and \textit{wh}-in situ languages (Huang 1982, Lasnik & Saito 1992). If the variability in word order in (9) and (10) were due to movement, we would expect Moken \textit{wh}-phrases to exhibit superiority effects.

The following sentences in Moken demonstrate that Moken does not seem to exhibit superiority effects, but in fact shows the converse from what one might expect:

\begin{enumerate}
\item[(13)]
\begin{enumerate}
\item a. ano:-la: aca:-la: mane?
\begin{tabular}{lll}
what-WHQ & who-WHQ & ask
\end{tabular}
\item b. aca:-la: mane? ano:-la:
\begin{tabular}{lll}
what-WHQ & ask & who-WHQ
\end{tabular}
(a-b) ‘What did who ask?’
\end{enumerate}
\end{enumerate}
The sentences in (13) represent the opposite pattern as we saw for English, with structurally lower arguments, a wh-object, able to occur to the left of the wh-subject. One conclusion which could be drawn from these facts is simply that Moken does not actually have wh-movement in the sense that we are familiar with from European languages, but forms questions with some kind of cleft structure.

The clearest evidence for this is a similar restriction on extraction between (13) and from putative cleft structures, whose structure we will examine in the following section:

(14) *Object NP scrambling out of a subject /nə/-cleft*

a. [olanₐₐ nə] [ECₐₐ matok kəla]  
snake COP bite monkey  
‘It’s the snake that bit the monkey.’

b. kəlaₐₐ [olan nə] [ECₐₐ mətok kəla]  
monkey snake COP bite  
‘It’s the snake that bit the monkey.’

(15) *Subject NP scrambling out of an object /nə/-cleft*

a. aloy-bubut [kəlaₐₐ nə] [olan mətok ECₐₐ]  
day-yester monkey COP snake bite  
‘It’s the monkey that the snake bit yesterday.’

b. *olanₐₐ aloy-bubut [kəlaₐₐ nə] [ətan mətok ECₐₐ]  
snake day-yester monkey COP bite  
‘The monkey it’s the snake that bit yesterday.’

---

3 A possible reaction to these data would be to assert that (13a) might be an echo question, which show obviated superiority effects in English. I have no evidence that (13a) is not an echo question — such subtleties are often difficult to determine on the field — however, the unacceptability of (13c) would go unexplained under this account, as would its relationship to the data from cleft constructions. Another reaction which seems incorrect would be to claim simply that the Superiority Condition does not hold in Moken, or is ‘parameterized’. Besides being theoretically problematic this approach, too, ignores the asymmetry in (13) and its relationship to clefts.
The basic ingredient of the cleft under consideration is the copular element \( nə \). Arguments of the clause to the right can be fronted to the left of the copula. This is shown in (14a) for subjects and in (15a) for objects. I represent their ordinary position in the lower clause with the label EC for empty category. In (14b) and (14c) we see the effects of attempting to dislocate the remaining lower argument to the front of the clause. In (14b), we see that the object can occur before the clefted subject. However, as (15b) shows, embedded subjects cannot be preposed before clefted objects.

The cleft structure in (14-15) can be seen as showing essentially the same restriction on ordering as the multiple wh-questions in (13a-c). The basic schema of this restriction is shown below:

\[(16)\]

a. *Multiple Wh-questions*

(i) \( \text{Obj-wh} \quad \text{Sbj-wh} \quad \text{V} \)

(ii) \( \ast \text{Sbj-wh} \quad \text{Obj-wh} \quad \text{V} \)

b. *Clefts*

(i) \( \text{Obj-NP} \quad \text{Sbj-NP} \quad nə? \quad \text{V} \)

(ii) \( \ast \text{Sbj-NP} \quad \text{Obj-NP} \quad nə? \quad \text{V} \)

In Section 4, I will argue that despite a lot of other positive evidence, wh-questions in Moken are in fact clefts, which, if true, may show that languages may retain specific syntactic constructions historically even in the face of significant syntactic and morphological change elsewhere in the language.

Before we can explore what that might mean, we need to see that the sentences with \( nə? \) are actually clefts of some sort. In the following section, I demonstrate that the construction in (14-15) above is a copula from a number of data which show, among other things, the status of \( nə? \) as a copula, the biclausality of the cleft-construction, and also some basic comparative evidence from Malagasy.

### 3. The structure of the \( nə? \)-cleft

This section of the paper deals with the distribution of the copula \( nə? \) and some evidence which suggests certain structures for the cleft which it forms. We will begin by examining where it occurs as a simple copula, then proceeding to note some restrictions on its complement in the cleft structure and also evidence that that structure is biclausal.
3.1 nəʔ as Copula

The copular use of nəʔ is actually a number of uses which can presumably be tied together with a formal semantic analysis. The verb has clear differences from be in English, some affinity to predicational copula in Romance (e.g. Spanish *estar*) and is nearly identical to Thai and Chinese *have* copula (*mīī* and *yōu*, respectively). As this is the case, nəʔ is defined by two basic but related uses.

The first common use of nəʔ is as the verb ‘have.’ The following examples are representative:

(17) thiː omak cuy nə cəpoʔ-thəwa dʒiːluy
    at home 1SG have ten-two CLF
    ‘I have 12 children at home’

(18) cuy nəʔ huŋ haʔ laluy
    1SG have money NEG AT.ALL
    ‘I don’t have any money at all.’

The second use of nəʔ is in presentational constructions. The examples below are the simplest type of example:

(19) a. cəm nəʔ kaa
    rice COP YNQ
    ‘Is there rice?’

    b. cəm haʔ || nəʔ khanom
       rice NEG COP bread
    ‘There’s no rice. There’s bread.’

(20) a. nəʔ kujaːn
    COP rain

    b. kujaːn nəʔ
       rain COP
    (a-b) ‘It’s raining’ (lit: ‘There’s rain.’)
As the above sentences show, the word order relative to *nəʔ* is relatively variable. Below we suggest some possible meaning correlates to the different orders.

Other uses of *nəʔ* are less straightforward but generally can be considered presentational constructions. The examples in (21) and (22) represent instances of what appears to be a constituent separated by the copula:

(21) batuy *nəʔ* toŋ kati:
stone COP at chair
‘The stone is by the chair.’

(22) Moken *nəʔ* thəwe dilay
COP two CLF
‘There are two Moken’

The prepositional phrase in (21) can either occur as an adverb or as an adnominal expression when it occurs in other contexts. Here, we find it modifying the subject of the copular construction, which I interpreted as definite in the context it was uttered. The fact that the precopular expression is definite is unsurprising given the role of *nəʔ* in presentational constructions, which usually require indefinites (or other ‘weak’ pronouns in the sense of Milsark 1977). The construction in (22) is also attested in Chinese but with the predicative/equative *shi*. The fact that *nəʔ* is also being used in this context is evidence that its distribution may extend past the equivalent ‘have’ words in Chinese and Thai.

A final case where *nəʔ* appears in copular constructions is in the presentational time phrases below:

(23) a. *nəʔ* aʔ-aloy [cuy lɔlɔ.adə’]
COP INDF-day I body.big
‘There will be a day when I am fat.’

b. *nəʔ* aʔ-aloy [canaat gayuŋ]
COP INDF-day child tall
‘There will be a day when the child is tall.’

---

4 This fact and analysis makes the prediction that the opposite order of constituents in (21) would be interpreted with ‘stone’ as indefinite, data which I do not have.
The sentences in (23) include a simple indefinite time phrase, predicated by an embedded clause. These examples seem to relatively straightforward instances of the presentational use of nəʔ.

There are also important restrictions on the use of nəʔ. The first is that it cannot take an adjective in its complement, as adjectives are stative verbs:

(24) cuy (*nə) lələ ade’
1sg COP body big
‘I’m big.’

These facts reflect the fact that adjectives in Moken are stative verbs, as in Thai, which shows the same restriction:

(25) phôm (*pen) tua jaj (Thai)
1sg (*COP) body big
‘I’m big.’

While (24) is understandably bad in light of (25), It is hard to see why (24) could not be interpreted as a cleft, with the subject appearing preverbally. At this point I have no answer to this question.

A final construction in which nəʔ is prohibited is in predicational or equative constructions. In these situations, another copula, borrowed from Thai is used:

(26) oləŋ mokaw ɲa pĩn Moken
person say 3SG COP
‘He said he is Moken’

The fact that Moken has borrowed a copula from Thai is surprising but may have a natural historical explanation: as Moken underwent a significant shift from an Austronesian language and towards an SVO isolating language, holes in its grammar may have begun to appear. When those gaps appear, Moken would naturally borrow from a language in which it is in heavy contact and which has a similar structure, at least on the surface, to that of Moken — in this case, Thai. It is also possible to imagine the opposite scenario, that contact with Thai somehow induced the speakers to borrow their copula, but this seems intuitively unlikely.
### 3.2 noʔ and Clefting

As we have seen, an argument can be dislocated from a clause and appear to the left of noʔ. Both subjects and direct objects can appear in this construction,

(27) **Subject cleft**

olanₐ₉ nə [ECₐ₉ mətok kəla]

snake COP bite monkey

‘It’s the snake that bit the monkey.’

(28) **Object cleft**

aloy-bubut kəlaₐ₉ nə [olan mətok ECₐ₉]

day-yester monkey COP snake bite

‘It’s the monkey that the snake bit yesterday.’

(29) **Wh-subject cleft**

anoₐ₉-la nəʔ [ECₐ₉ mətok kəla]

what-Q COP bite monkey

‘What is it that bit the monkey?’

(30) **Wh-object cleft**

anoₐ₉-la nəʔ [olan mətok ECₐ₉]

what-Q COP snake bite

‘What is it that the snake bit?’

At this point, the only evidence that (27-30) are clefts at all involves the fact that they include a copula. A strong piece of evidence for the status of (27-30) as a cleft would involve demonstrating that the structure is biclausal, or that there is some sort of boundary in the lower clause.

One such piece of evidence come from scope facts. While negation in can usually take either wide or narrow scope with respect to the subject in Moken, quantifiers in the cleft construction cannot, instead

(31) taphuŋ bo-kan moy-haʔ
everybody do-work can-not

‘Everybody can’t do work.’

All > Neg, Neg > All
The fact that inverse scope is blocked when negation appears in the clause to the right of the cleft is clear evidence for a clause boundary, indicating that the structure as a whole is biclausal.

Another more intricate piece of evidence comes from quantifier float. In normal monoclausal sentences, subject quantifiers can appear sentence finally:

(33) oy nemaʔ bulat nam kalăn
dog five CLF eat bones
‘Five dogs are eating bones.’

(34) oy nam kalăn nemaʔ bulat
dog eat bones five CLF
‘(Five) dogs are here eating (five) bones.’

The classifier bulat is a general classifier used for inanimate objects and animals. In (34) it can refer to either the subject or the object. The reading where it refers to the subject exemplifies quantifier float. Thai exhibits a similar phenomenon (Jenks 2004). Quantifier float can also be seen within the embedded verb in a cleft:

(35) a. binay nাʔ [olan thəwa-bulat mətōk]
   woman COP snake two-CLF bite
   b. binay nাʔ [olan mətōk thəwa-bulat]
   woman COP snake bite two-CLF
   (a-b) ‘It’s a woman that the two snakes bit.’

The classifier in (35) could not be mapped to ‘woman’, as there is a separate classifier for humans.

Now we have seen that quantifiers can float around objects, so a natural question would be whether the quantifiers of clefted subjects can similarly float to the end of the sentence.
What we find is that while these quantifiers can float to the right of the copula (36b), they cannot float to the end of the sentence (36c):

(36)  

a. binay thəwa-jiluy nəʔ [olan mətok]
     woman two-clf COP snake bite

b. binay nəʔ thəwa-jiluy [olan mətok]
     woman COP two-clf:person snake bite

c. *binay nəʔ [olan mətok] thəwa-jiluy
     woman COP snake bite two-clf:person

     ‘There are the two women that the snake bit.’

It is admittedly difficult to interpret these facts without an in-depth study of quantifier float, which is yet to be undertaken, but the difference between (36) and (35b) is striking. Similar facts obtain from English show that extraposition is marked in clefts, demonstrating that it must be clausebound:

(37)  

a. The monkey died that you hated so much.

b. It’s the monkey that you hated so much that died.

c. ?? It’s the monkey that died that you hated so much.

(37c) is marked if it is interpreted as synonymous to (37b). So while the analysis of quantifier float in Moken is unclear, the fact that quantifiers cannot float to the end of the embedded clause demonstrates a structural difference between non- and clefted sentences.

A final observation about the nəʔ-cleft is that its complement seems ‘small,’ meaning that it doesn’t not seem license the full range of TMA markers. While this avenue certainly begs for further exploration, the available data is suggestive:

(38)  

(*nəʔ) canaat (*nəʔ) kaloy labut
     COP child COP PROG run

     ‘The child is running’ (without the particle)
Unlike $nəʔ$, raising verbs such as $dəni$ can take a complement with $kaloy$:

(39) a. dəni canaat kaloy labut
close child PROG run
‘The child is about to be running.’

b. canaat dəni kaloy labut
child close PROG run
‘The child is about to be running.’

Thus, the complement of $nəʔ$ does not appear to be a full CP, as is usually the case for raising verbs, but rather some smaller piece of phrase structure. More evidence is needed to determining an analysis of these facts.

In summary, this section has laid out a number of facts. First, $nəʔ$ is a copula which participates in basic existential constructions. Second, sentences formed with $nə$ are biclausal. Third, the complement of $nə$ is smaller than normal clausal complements.

The following two pieces of data suggest that $dəni$ is a raising verb:

(i) _It can't take voice markers, e.g. the negative imperative circumfix na-la:_

*na-dəni-la ŋabut kotan
NEG-close-IMP run forest

'Don’t run into the forest'

(ii) _It can occur with inanimate subjects:_

a. dəni lasee pəloհ
close book fall

b. lasee dəni pəloհ
book close fall
(a-b) ‘The book is about to fall.’

c. *lasee pəloհ dəni
book fall close
4. A Constraint on Fronting in Clefts

As we saw in Section 2, Moken exhibits systematic restrictions on whether subjects can raise across fronted objects which are present both in cleft constructions and in multiple questions:

(40) a. Multiple Wh-questions (repeated from 16)
   (i) Obj-wh    Sbj-wh    V
   (ii) *Sbj-wh   Obj-wh    V

b. Clefts with NP
   (i) Obj-NP    Sbj-NP    nəʔ  V
   (ii) *Sbj-NP  Obj-NP    nəʔ  V

Recall that these facts seemed exceptional because they seemed to demonstrate the opposite effect from what we would expect by analogy with Superiority effects in languages with Wh-movement.

Interestingly, identical effects hold only if either argument is a wh-phrase:

(41) Subject WhP-cleft with extracted NP
    kəla    aloy bubut  anoː-laːm  nəʔ  [ ECm  matok  kəla ]
    monkey    yesterday    what-WHQ    COP    bit
    ‘What is it that bit the monkey?’

(42) Object WhP-cleft with extracted NP
    *olan    aloy-bubut  ano_m-la  nəʔ  [ Ṽám  matok  ECm ]
    snake    yesterday    what    COP    bite

(43) Subject NP-cleft with extracted WhP
    anoː-la:  aloy-bubut  olan_m  nəʔ  [ ECm  matok  anoː-la ]
    what-Q    yesterday    snake    COP    bite
    ‘It is the snake that bit what yesterday?’

(44) Object NP-cleft with extracted WhP
    *anoː-la:  aloy-bubut  [kəla_m  nəʔ]  [ anoː-la  matok  ECm ]
    what    yesterday    monkey    foc    bite
These results are summarized below:

(45)  
\[ \text{a. Wh-cleft with extracted NP} \]
\[
(\text{i}) \quad \text{Obj-NP} \quad \text{Sbj-wh} \quad \text{noʔ} \quad \text{V} \\
(\text{ii}) \quad * \text{Sbj-NP} \quad \text{Obj-wh} \quad \text{noʔ} \quad \text{V}
\]

\[ \text{b. NP-cleft with extracted WhP} \]
\[
(\text{i}) \quad \text{Obj-Wh} \quad \text{Sbj-NP} \quad \text{noʔ} \quad \text{V} \\
(\text{ii}) \quad * \text{Sbj-Wh} \quad \text{Obj-NP} \quad \text{noʔ} \quad \text{V}
\]

To this point, the data in (45) represent the only instances of wh-phrases appearing in an overtly clefted structure. They suggest one final data set which we have not seen yet, namely, two wh-phrases in an overtly clefted sentence.

When we examine no-cLEFTs with two wh-phrases, we find that while subjects can be clefted and objects can raise out, as we saw for all of the other cleft sentences:

(46)  
\[ \text{a. aca-la: noʔ makaw ano-la:} \]
\[
\text{who COP said what}
\]

\[ \text{b. ano-la: aca-la: noʔ makaw} \]
\[
\text{what-Q who-Q COP said} \\
(\text{a-b) ‘Who is it that said what?’}
\]

\[ \text{c. Obj-NP Sbj-NP noʔ V} \]

Wh-objects cannot be clefted at all if there is a wh-subject, nor can the wh-subject appear to its left:

(47)  
\[ \text{a. * ano-la: noʔ aca-la: maneʔ} \]
\[
\text{what COP who ask}
\]

\[ \text{b. * aca-la: ano-la: noʔ maneʔ} \]
\[
\text{who what COP ask} \\
(\text{‘What is it that who asked?’}
\]
Why would the First, note that (47a) is essentially our missing superiority violation, directly contrasting with (46b), which is grammatical and represents the same word order modulo the position of naʔ. If we assume that the highest wh-phrase must appear as the clefted constituent, in response to the superior ity condition, then we can explain why (47a) is ill-formed, where the object moved first. Furthermore, if we assume that at least one wh-phrase must front, and that one must be the highest, we can explain why the object can stay in situ in (46a). The final claim is simply that the lower constituent can front once superiority has been satisfied.

The following well-formed and ill-formed derivations demonstrate this difference:

(48) **Well-formed derivation of (46a-b)**

- a. naʔ Sbj-wh V Obj-wh
- b. Sbj-wh naʔ Sbj-wh V Obj-wh
- c. Obj-wh Sbj-wh naʔ Sbj-wh V Obj-wh (optionally)

(49) **Ill-formed derivation of (47a-b)**

- a. naʔ Sbj-wh V Obj-wh
- b. Obj-wh naʔ Sbj-wh V Obj-wh *
- c. Sbj-wh Obj-wh naʔ Sbj-wh V Obj-wh *

Now compare these derivations, omitting the copula, to the original multiple question data from (13) which appear to have no cleft:

(50) a. ano-la: aca-la: mane? (cf. 48c or 49b) what-WHQ who-WHQ ask

- b. aca-la: mane? ano-la: (cf. 48b) what-WHQ ask who-WHQ
  (a-b) ‘What did who ask?’


It appears that the pattern we find in multiple wh-questions can be derived by essentially the same abstract procedure that we illustrated for clefts in (48-49). One hypothesis to explain this
non-surface true parallelism would be to posit a null copula for the multiple wh-questions in (50). In the case of the apparent superiority violation (50a), there is a possible structure that the speaker could posit in which that structure would be licensed, namely, where both wh-phrases have been fronted, thus, the sentence is grammatical. In the case of (50c), however, there is no possible structure or derivation in which it could have arisen, and thus, it is ruled out.

I take the collection of facts detailed above as evidence for the following claims:

(51)   a. Moken questions are formed by a cleft-like structure similar to the noʔ-cleft.
       b. Wh-phrases undergo obligatory clefting.
       c. The superior (c-commanding) wh-phrase must cleft first.
       d. NPs apart from wh-phrases can be similarly marked for clefting and then exhibit identical constraints.

The final claim in (51) attempts an explanation of the clefts with either two regular NPs or a combination of NPs and wh-phrases. In essence, it says that the NPs which are “destined” for movement do so because they exhibit essentially the same syntax as wh-phrases.

It is beyond the scope of the present to say more, and to do so at this point would likely be imprudent. A number of questions still loom: What is the feature which is triggering movement in the above examples, and where are their landing sites? How can any constituent be extracted from the complement of a cleft which has been shown to be a barrier for scope? What is the specific structure of the cleft? Further work is needed.

References