Australian National University

CLEFTS AND ANTI-SUPERIORITY IN MOKEN[1]

Kenneth Baclawski Jr.
University of California, Berkeley
<kbaclawski@berkeley.edu>

Peter Jenks
University of California, Berkeley
<jenks@berkeley.edu>

Abstract
We describe an extraction asymmetry in Moken that presents apparent Anti-Superiority effects. We then show that this asymmetry is not rooted in Superiority at all. Evidence from island effects is used to demonstrate that the left-dislocation of *wh*-phrases is not the result of *wh*-movement as standardly conceived. Furthermore, the same Anti-Superiority effect obtains for non-*wh*-phrases and clefts. At the same time, standard Superiority effects in Moken do arise in certain environments. These observations lead to the conclusion that Anti-Superiority effects in Moken are not counterexamples to the universality of Superiority, but instead arise due to a constraint on crossed dependencies between arguments and non-argument positions.
Keywords: Moken, constituent question, cleft, superiority
ISO 639-3: mwt, cjm

1 Introduction

The Moken language[2] (Austronesian: Thailand, Burma) displays an extraction asymmetry that, on the surface, is plainly an Anti-Superiority effect. The principle of Superiority was proposed by Chomsky (1973) to account for the requirement that subjects be fronted before objects in multiple *wh*-questions. Languages such as Bulgarian require that all *wh*-pronouns be fronted (1), and modern descendants of Superiority have been invoked to account for the requirement that subjects precede objects in this configuration (Richards 2001). However, when multiple *wh*-pronouns are fronted in Moken, the pattern is precisely the reverse (2) of the Bulgarian one.

(1) a. *Koj kogo običa?*   (Bulgarian)
    who whom loves
    'Who loves whom?' (Bošković: 11a)

b.   **Kogo   koj   običa?*
         whom   who   loves
         ('Who loves whom?') (Bošković: 11b)

(2)   a.   **aca:=la: ano:=la: mane*   (Moken)
           who=Q   what=Q ask
           ('Who asked what?')

       b.   *ano:=la: aca:=la: mane*
           what=Q who=Q ask
           'Who asked what?'

Since modern syntactic theory captures Superiority effects like (1) and other similar cases with a hard-wired formal procedure called Agree (e.g. Miyagawa 2009), true counterexamples to Superiority could have profound consequences to syntactic theory.

Further investigation into the Moken asymmetry, however, reveals that it does not seem to be an intervention effect specific to *wh*-movement. Left-dislocated *wh*-phrases do not obey island constraints, implying that they do not undergo *wh*-movement. In addition, we will show that the same extraction asymmetry also applies to the extraction of non-*wh*-phrases and clefts (3).

(3)   a.   **olan   kəlan   nə   mətok*
           snake   monkey COP   bite
           'It's the monkey that the snake bit.'

       b.   *kəlan   olan   nə   mətok*
           monkey snake   COP   bite
           'It's the snake that bit the monkey.'

We will conclude that the apparent Anti-Superiority effects in Moken are actually due to a constraint on crossed dependencies headed by non-argument positions. Along the way, we provide the first description of Moken *wh*-questions and clefts as well as a basic analysis of these constructions.

This paper will examine the following questions: Do Moken questions involve movement? (**Section 2**); What is the structure of Moken clefts? (**Section 3**); and How can we account for the systematic extraction asymmetry in *wh*-questions and clefts? (**Section 4**). Finally, we will show that a Superiority effect does actually obtain, but only in the presence of an overt *wh*-cleft.

1.1 The Moken language

Moken is an Austronesian language spoken on Mainland Southeast Asia primarily in the Mergui Archipelago. It is an endangered language (cf. UNESCO 2010) spoken by fewer than 10,000 people, and the situation may be even more dire than previous estimates imply (Say Bay 2015). The name "Moken" refers to what is considered a dialect continuum, including Moklen, which is spoken in the northern regions of the archipelago in Burma (Larish 2005). Data for this paper comes from the second author's fieldwork on the Moken spoken in the southern end of the archipelago in Thailand, off the coast of Ranong from 2005 to 2007.

Much of the linguistic literature on Moken concerns its place within the Austronesian family (e.g. Larish 1999, Pittayaporn 2005). Similar to the Chamic languages (Thurgood 1999), Moken has undergone intense contact with Mainland Southeast Asian languages, rendering it difficult to assess what is an innovation and what is a contact effect. Theoretical analyses of Moken are largely restricted to the work of Larish on the Moklen variety. Since Moklen is not considered mutually intelligible with the Moken of Thailand, this data will only be alluded to for comparative reference when possible.

Unlike insular Austronesian languages, Moken lacks initial voice morphology, likely due to a contact-induced shift to final stress, resulting the degradation of phonological material in non-final syllables. In fact, Moken is now 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. Because of this, arguments for syntactic analyses must be based on configurational facts.

The basic, unmarked word order is given in (4a). In the elicitation context (i.e. a sentence offered out of context), SOV, OSV, and VSO orderings are acceptable, perhaps due to scrambling and/or discourse configuration (4b–d). OSV order will become especially important later on in the discussion of the Anti-Superiority effect. Subject-final orderings, OVS and VOS result in the unintended gloss, 'The bone is eating a dog'.

(4)   a.   *oj*   *ŋam*   *kəlan*
             dog   eat   bone
             'The dog is eating a bone.'
       b.   *oj*   *kəlan*   *ŋam*
       c.   *kəlan*   *oj*   *ŋam*
       d.   *ŋam*   *oj*   *kəlan*
       e.   **kəlan*   *ŋam*   *oj*
       f.   **ŋam*   *kəlan*   *oj*

Subject-final orderings are possible, however, under what is likely a marked information structure, such as a right-dislocated anti-topic (5). The precise information structural status of these forms is outside the scope of this paper. Suffice it to say that Moken appears to be strongly discourse configurational.[3]

(5)   *makaːw*   *toŋ*   *acaː=laː*   *biːŋ*
       tell   with   who=Q   2SG
       'Who are you talking about?'

In Moken narratives, SVO ordering is predominant. Because of radical pro-drop (cf. Huang 1984) and frequent serial verb constructions, however, sentences often resemble long strings of verbs, as in (6).

(6)   *malak*   *chuwat*   *bɔ*   *salɔːj*   //   *taloŋ*   *malak*   *bɔ*   *salɔːj*
       wring   cloth   make   dry   help   wring   make   dry
       '(She's) wringing the cloth to dry (it).
       (He comes to) help wring (the cloth) to make (it) dry.' [Narration-7_21: p.9]
       CONTEXT: Two friends are sitting together doing chores.

Nevertheless, it is worth noting that of the six narratives recorded in our fieldwork, there are zero credible instances of a non-(S)V(O) utterance. While this may call into question the word order permutations accepted in (4) above, we assert that our consultants provided strong grammaticality judgments in elicitation and that narration typically exhibits low information load per sentence, at least in this narrative tradition.

In the following sections, we will lay out basic properties of *wh*-questions, then an analysis of clefts, and finally the apparent Anti-Superiority effects.

2 Basic properties of *wh*-questions

Moken *wh*-phrases are typically in-situ and optionally license a *wh*-particle *laː* (cf. Cable 2010). These facts are displayed in (7a–f); note the optionality of *laː* throughout. We have not yet been able to detect any

---

[3]   Larish (1999) even attests OVS ordering for Moklen, apparently derived via object topicalization. This example is
       repeated in (i). It is unclear why the verb precedes the subject here, though; it seems likely that there is more to this
       sentence than just topicalization.
       (i) *kuupʰut*   *duk*   *ʔɔlaːŋ*   *hẽʔ*   *namaʔ*   (Moklen)
            husk   put   they   on   flag
            'As for coconut husks, they [Moklen people] put (husks) on (the) flag [pole(s)].' (Larish 1999: p.239)

meaning difference between the forms with and without this *wh*-particle. Larish (2005) notes a similar distribution with Moklen *lay* and its allomorphs *may* and *ŋay*, which also optionally mark in-situ *wh*-phrases.

(7) a. *aca:(=la:)*[4]   *ɲam*   *co:n*
          who(=Q)   eat   rice
          'Who ate rice?'

     b. *somʔo:* *ɲam*   *ano:(=la:)*
          Som'oo   eat   what(=Q)
          'What did Som'oo eat?'

     c. *somʔo:* *niʔen* *bita:(=la:)*
          Som'oo   play   where(=Q)
          'Where is Som'oo playing?'

     d. *somʔo:* *niʔen* *jipa:(=la:)*
          Som'oo   play   how(=Q)
          'How is Som'oo playing?'

     e. *somʔo:* *niʔen* *kotan*   *ope:n(=la:)*
          Som'oo   play   forest   when(=Q)
          'When did Som'oo play in the forest?'

     f. *canat* *niʔen* *toŋ*   *somʔo:* *bo*   *ano:(=la:)*
          child   play   with   Som'oo   do   what(=Q)
          'Why did the children play with Som'oo?'

Narratives confirm the tendency of *wh*-phrases to remain in-situ and the optionality of *la:*. In these two utterances from the same narrative, *la:* marks the *wh*-phrase (8a), or is absent (8b); in all cases, the *wh*-phrases are in-situ. Note that *la:* also appears sentence-finally in (8a). The following section discusses the distribution of *la:* in more detail.

(8) a. *eŋ*   *bɔ*   *ano:=la:* *ta*   *ana:t* *cuj=la:*
          you   do   what=Q   PART   child   1SG=Q
          'What are you doing to my child?' [Narration-7_20: p.26]

     b. *eŋ*   *buno*   *lamat*
          you   do.what   inside
          'What are you doing inside?' [Narration-7_20: p.53]

2.1 The wh-particle *la:*

This particle examines the syntactic properties of the *wh*-particle *la:*. We will conclude that *la:* is likely licensed by a high syntactic head. It may appear after *wh*-phrases or sentence finally. The examples in (9)

---

4   Throughout this paper, we use parentheses in examples to indicate optionality.

demonstrate that *la:* is optional and may appear multiple times in a sentence (cf. (8a) above for a naturally occurring example).[5]

(9)   a.   *aca:=la: ɲam   con*
               who=Q eat   rice
               'Who ate rice?'

        b.   *aca: ɲam   con=la:*

        c.   *aca:=la: ɲam   con=la:*

        d.   *aca: ɲam   con*

Evidence that *la:* occupies a position high in the syntactic structure comes from its interaction with adverbs. While *la:* can alternate in position with temporal adverbs (10a), which are assumed to be relatively high in the structure (in the sense of Cinque 1999), it cannot appear inside of "low"-adverbs such as manner adverbs (10b).

(10)   a.   *aca=la: mələn   oj{ }   aloj   duj{=la:}*[6]
              who=Q hit   dog   day   this=Q
              'Who hit a dog today?'

        b.   *aca=la: mələn   oj{*}   li:ŋ{=la:}*
              who=Q hit   dog   hard=Q
              'Who hit a dog hard?'

More evidence for the high position of the *wh*-particle comes from verbal elements, which we might want to assume are higher than VP itself. For instance, *la:* cannot occur inside of final modals (11a), which are traditionally assumed to dominate VP (and in terms of semantics must dominate V for reasons of scope) and sentence final aspect particles (11b), which are likewise assumed to be higher than VP in structural syntax.

(11)   a.   *aca=la: bo   ka:n{*} moj   ha{=la:}*
              who=Q do   work   can   NEG=Q
              'Who can't do work?'

        b.   *aca=la: ɲam   con{*}   haɾe:{=la:}*
              who=Q eat   rice   NEG.PRF=Q
              'Who hasn't eaten yet?'

The structure of the clauses in (11) 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, *la:* must be structurally high, somewhere in the left periphery.

---

[5]   One restriction on *la:* is that it may not appear twice in a row, such as sentence-finally after a sentence-final *wh*-phrase (ii). There are a number of possible explanations for this restriction, such as a trivial phonological restriction on a function word repeating.

(ii)   *olan   mətok ano:=la:(*[??]*=la:)*
       snake   bite   what=Q(=Q)
       'What did the snake bite?'

[6]   Throughout this paper, we will use curly brackets to indicate sets of possible orders. In (10a), *la:* may appear in either of the bracketed positions. In (10b), the former bracketed position is ungrammatical, but the latter is acceptable.

2.2 Word order and movement

Despite the general tendency of *wh*-phrases to remain in-situ, they may in fact appear in a number of positions, in parallel with the word orders attested above in (4–5). Subject *wh*-phrases, as in (12b–d) may appear lower in the clause.

(12)   a.   *aca:=la: layam   lelen   toŋ   ŋan*
            who=Q   douse   candle   with   hand
            'Who doused the candle with (their) hand?'
      b.   *layam   aca:=la: lelen   toŋ   ŋan*
      c.   *layam   lelen   aca:=la: toŋ   ŋan*
      d.   *layam   lelen   toŋ   ŋan   aca:=la:*

The same is true for *wh*-objects, which may likewise be fronted. In (13a), there clearly must be some sort of dependency between the *wh*-phrase and the canonical object position after the verb. Note the fact that the subject *wh*-phrase appears after the perfect aspect marker *ka* in (13c), which seems to imply that it has been right-dislocated outside the clause.

(13)   a.   *ano:=la: apoŋ   layam   toŋ   ŋan   ka*
            what=Q   father   douse   with   hand   PRF
            'What did the father douse with (his) hand?'
      b.   *apoŋ   layam   ano:=la: toŋ   ŋan   ka*
      c.   *apoŋ   layam   toŋ   ŋan   ka   ano:=la:*

These word order facts lead to several hypotheses. For one, Moken could be seen as a *wh*-in-situ language with optional *wh*-movement (cf. Denham 2000 on Babine-Witsuwit'en), especially to explain sentences like (12d) and (13a). It could also be seen as a language with prevalent scrambling effects. Alternately, non-canonical positions of *wh*-phrases could be ascribed to information structure positions, like right- and left-dislocated topics. For this paper, we will focus on *wh*-phrases extracted to sentence-initial position (as in 13a). The data will suggest that these sentence-initial *wh*-phrases do not undergo movement of any kind, and instead must be base-generated. We will note that the data here do not necessarily bear on any of the other possible positions seen in (12–13).

Preliminary evidence against *wh*-movement comes from island restrictions. Moken *wh*-phrases violate several island restrictions, which is generally taken to be diagnostic of base-generation as opposed to syntactic movement. First, subject and object *wh*-phrases may be fronted out of adjunct clauses (14a–b).

(14)   a.   *{aca:=la:}   enaw   lakaw   malaj   kana: { }   məlak   kəbaŋ*
            {who=Q}   Naw   go   out   because   like   boat
            'Who did Naw leave because (they) like boats?'

      b.   *{ano:=la:}   enaw   lakaw   malaj   kana: ethip   məlak   { }*
            {what=Q}   Naw   go   out   because Thip   like
            'What did Naw leave because Thip likes?'

Second, *wh*-subjects and objects may also be fronted out of complex noun phrases, as in 'a story about X being afraid of Y', as shown in (15a–b).

(15)   a.   *{aca:=la:}   enaw   məkaw   jalan   { }   lakot   kəjuj*
            {who=Q}   Naw   tell   story   be.afraid.of shark
            'Who did Naw tell a story (about) being afraid of sharks?'

b.   *{ano:=la:}*   *enaw*   *məkaw*   *jalan*   *pito*   *lakot*   *{ }*
           {what=Q}   Naw   tell   story   Peter   be.afraid.of
           'What did Naw tell a story (about) Peter being afraid of?'

It should be noted that island violations are not specific to *wh*-phrases. For example, in (16), a non-*wh*-phrase undergoes the same island-violating movement out of an adjunct clause.

(16)   *{ethip}*   *enaw*   *lakaw*   *malaj*   *kana:*   *{ }*   *məlak*   *kəbaŋ*
       {Thip}   Naw   go   out   because   like   boat
       'Thip, Naw went out because (he) likes boats.'

A full exploration of the word order capacity of *wh*-phrases in Moken is beyond the scope of this paper. Suffice it to say that sentence-initial *wh*-phrases do not obey usual movement restrictions. This implies that *wh*-movement is not the best analysis for the present phenomenon. While this data could be explained as syntactic movement with null resumptive pronouns (cf. Boeckx 2003), it will become clear that base-generation is a preferable analysis in the following sections, due to the apparent Anti-Superiority effect shared by fronted *wh*-phrases and clefts.

3 Properties of clefts

In this section, we will turn away from *wh*-phrases to cleft constructions with the presentational copula *nə*. Moken has a copula derived from a verb meaning 'have' (reminiscent of the 'acquire' grammaticalization path described in Enfield 2003). Two basic examples of *nə* as a verb meaning 'have' are given in (17).

(17)   a.   *thi:*   *omak*   *cuj*   *nə*   *cəpo*   *thəwa*   *dʒilu:j*
           at   home   1SG   have   10   2   CLF
           'I have 12 children at home.'

       b.   *cuj*   *nə*   *huŋ*   *ha*   *ləluj*
           1SG   have   money   NEG   AT.ALL
           'I don't have (any) money at all.'

The second use of nə is in presentational constructions. Some examples are given below in (18–19).

(18)   a.   *co:n*   *nə*   *ka:*
           rice   cop   Y/N.Q
           'Is there rice?'

       b.   *co:n*   *ha*   *//*   *nə*   *khanom*
           rice   NEG   COP   bread
           '(There's) no rice; there's bread.'

(19)   a.   *nə*   *kuɟa:n*
           COP   rain
           'It's raining.' [lit: 'There's rain.']

       b.   *kuɟa:n*   *nə*

The copula *nə* exhibits an apparent definiteness effect. Pre-copular nouns tend to be definite, as in (20a), while post-copular nouns indefinite (20b; cf. 4a). While the exact structure of these sentences is unclear, this effect is unsurprising given the cross-linguistic interplay between presentational constructions and definiteness ('strong' and 'weak' pronouns in the sense of Milsark 1977).

(20) a. *batuj* *nə* *toŋ* *kati:*
      stone COP at chair
      'The stone is by the chair.'

      b. *nə* *oj* *ŋam* *kəlan*
      COP dog eat bone
      'A dog is eating the bone.'

There are several restrictions on the distribution of the copula *nə*. It may not take an adjective in its complement, as adjectives are stative verbs (21a). It is also prohibited in predicational and equative constructions as in (21b); instead, a copula borrowed from Thai is used.[7]

(21) a. *cuj* *(*nə)* *lɔlɔ* *adɛ*
      1sg COP body big
      'I'm big.'

      b. *olaŋ* *mokaw* *ɲa* *(pɪn/*nə) moken*
      person say 3SG COP Moken
      'He said he is Moken.'

With the presentational copula use in mind, we will argue that *nə* can also mark clefts. The argumentation will center around establishing that sentences like (22a–b) are biclausal.

(22) a. *olan* *nə* *mətok* *kəla*
      snake COP bite monkey
      'It's the snake that bit the monkey.'

      b. *aloj bubut* *kəla* *nə* *olan* *mətok*
      yesterday monkey COP snake bite
      'It's the monkey that the snake bit yesterday.'

It is not entirely straightforward to establish biclausality in Moken clefts for several reasons. For one, Moken lacks verbal morphology, precluding many biclausality tests in the literature (e.g. Potsdam 2006). Moken clefts also appear to be "small", in the sense that *nə* licenses few, if any aspect or modality markers. This is illustrated in (23a) with the ungrammaticality of the progressive aspect marker *kaloj*, which is conversely licensed by the raising verb *dəni* (23b).[8]

---

7 Larish (1999: 219–220) does not mention a cognate form for Moklen. Instead, copular sentences are either unmarked or marked with the same Thai borrowing *pɪn*. Larish (1999: 238) does describe *ne:* as a "topicalizer", apparently equating it with the Thai topicalizer *nâ*. Moklen *ne:* is claimed to mark human animate topics. While Moken *nə* is likely cognate, it does not seem to have anything to do with animacy, based on the examples in this section (e.g. 20). We will avoid making statements about the specific information structural status of Moken clefts, though it seems likely it marks some kind of focus, given copular clefts in other languages like English.

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

(a) It can't take voice markers, such as the negative imperative circumfix *na... la*;

(iv) a. **na=dəni=la* *ŋabut* *kotan*
      NEG=INCEPT=IMP run forest
      ('Don't be about to be running in the forest.')

      b. *na=ŋabut=la* *kotan*
      NEG=run=IMP forest
      'Don't run in the forest.'

(23)　　a.　　*{*nə}　cana:t　{*}　　kaloj　labut*
　　　　　　　COP　　child　　　　　　PROG　run
　　　　　　　'The child is running.' (*'It is the child who is running.')

　　　　b.　　*{dəni}　cana:t　{ }　　kaloj　labut*
　　　　　　　INCEPT　child　　　　　PROG　run
　　　　　　　'The child is about to be running.'

Nevertheless, we will present two pieces of evidence for the biclausality of Moken cleft sentences from scope and quantifier float. First, scope in cleft constructions does not behave as expected in a monoclausal construction. For example, negation can normally take narrow or wide scope with respect to the subject. In (24a), note that two readings are possible. By contrast, the wide scope reading is impossible with a *nə*-cleft, as in (24b). 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.

(24)　　a.　　*taphuaŋ　　bo　　ka:n　moj　ha*　　　　　　$\forall > \neg, \neg > \forall$
　　　　　　　everybody　do　　work　can　NEG
　　　　　　　'Everybody can't do work.' / 'Not everybody can do work.'

　　　　b.　　*taphuaŋ　　nə　　bo　　ka:n　moj　ha*　　　$\forall > \neg, * \neg > \forall$
　　　　　　　everybody　COP　do　　work　can　NEG
　　　　　　　'It's everybody that can't do work.' (*'It's not everybody who can do work.')

The second, more intricate, piece of evidence comes from quantifier float effects. In monoclausal sentences, subject quantifiers may appear sentence-finally. For instance, note the possible positions of the numeral and classifier *nema bulat* in (25). When the numeral and classifier are sentence-final, it is ambiguous whether it modifies *oj* 'dog' or *kəlan* 'bone', as both nouns take the same classifier. Thai exhibits a similar phenomenon (Jenks 2013).

(25)　　*oj　　ɲam　　kəlan　nema　bulat*
　　　　　dog　eat　　bone　5　　　CLF
　　　　　'Five dogs are eating bones.' / 'The dogs are eating five bones.'

For the next examples, note that the classifier for *binaj* 'woman' is *ɉiluj*, while the classifier for *olan* 'snake' is *bulat*. In a cleft construction, quantifiers that float to the end of the sentence may only modify the subject of the embedded verb, not the subject of the cleft (26).

(26)　　*binaj　nə　　olan　mətok　thəwa　(bulat/*ɉiluj)*
　　　　　woman　COP　snake　bite　　2　　　CLF
　　　　　'It's the woman that two snakes bit.' (*'It's two women that the snake bit.')

Instead, quantifiers that modify subjects of clefts may only float to the edge of the cleft itself. This constitutes strong evidence that clefts are clauses, at least for the purposes of quantifier float.

(27)　　*binaj　{thəwa　ɉiluj}　nə　　{ }　　olan　mətok*
　　　　　woman　2　　　CLF　　COP　　　　snake　bite

(b) It can occur with inanimate subjects.

(v)　　*{dəni}　lase:　{ }　　pəloh*
　　　　{INCEPT}　book　　　　fall
　　　　'The book is about to fall.'

'It's two women that the snake bit.'

It is admittedly difficult to interpret these facts without an in-depth study of quantifier float in Moken, which is yet to be undertaken, but the difference between (26) and (27) is striking. All in all, we will assert that Moken has a cleft construction marked by the copula *nə*, along the lines of English it-clefts. In the next section, we will show that clefts and *wh*-phrases both display apparent Anti-Superiority effects.

4 Superiority and Anti-Superiority

As mentioned in the introduction, multiply fronted Moken *wh*-phrases exhibit apparent Anti-Superiority effects. We use the term *fronted* here and below as a descriptive term that is not taken to imply *wh*-movement to this position. After briefly reviewing the standard analysis of Superiority, we offer several arguments why these restrictions would not be expected to surface in multiply fronted *wh*-questions in Moken. In addition, we show that Superiority effects do surface in Moken *nə*-clefts under certain circumstances. We then propose that the apparent Anti-Superiority effects in Moken are attributable to a constraint on crossed A-bar dependencies.

Superiority was proposed by Chomsky (1973) as a constraint on *wh*-movement. In English, when there are two *wh*-phrases in a sentence, *wh*-movement is assumed to apply to the constituent which is structurally superior. While for our purposes it will suffice to assume that subjects are superior to objects, the notion of superiority is ultimately reducible to a more basic structural relationship like c-command. Superiority, thus, explains why (28b) is ill-formed.

(28)   a.   *Who bought what?*
        b.   **What did who buy?*

For languages with multiple *wh*-movement, like Bulgarian, *wh*-movement is taken to occur first with the superior *wh*-phrase; then the next *wh*-phrase "tucks in" beneath it in the tree, to its right, resulting in the SOV order seen in (29b) (Richards 2001).

(29)   a.   *Koj   kogo   običa?*   (Bulgarian)
                  who   whom   loves
                  'Who loves whom?' (Bošković 2002: 11a)

        b.   **Kogo   koj   običa?*
                  whom   who   loves
                  ('Who loves whom?') (Bošković 2002: 11b)

Crucially, in both English and Bulgarian, *wh*-movement in constituent questions is obligatory, and in Bulgarian, multiple *wh*-movement is obligatory. *Wh*-movement in both languages is detectable via the standard empirical diagnostics for movement such as island phenomena and crossover effects.

To be clear, we will be using the term "Anti-Superiority" to refer to the reverse grammaticality effect: ungrammaticality in a sentence like (29a) and grammaticality in those like (29b).[9] Such effects have been reported in languages like Japanese, Korean, and Turkish (e.g. Watanabe 1992; Jeong 2003; İşsever 2009). All of these cases involve *wh*-objects and *wh*-adjuncts. True Anti-Superiority effects involving core arguments are vanishingly rare.

---

[9]   It is known that Superiority effects do not obtain in all languages and all constructions. D-linked *wh*-phrases, sentences with more than two *wh*-phrases, and echo questions all can avoid Superiority violations in languages like English. Other languages like Serbo-Croatian display a lack of such effects in sentences like (29a–b). Some linguists (e.g. Bošković 2002; Fanselow 2004) use the term "Anti-Superiority" to describe just this: a lack of expected Superiority effects. The term has also even been used to describe the pattern seen in Bulgarian multiple *wh*-fronting (Bošković 1997).

On the surface, Moken seems to present a particularly strong challenge to Superiority: it exhibits Anti-Superiority effects with core arguments. In multiple constituent questions, OSV order is grammatical, but the SOV order predicted by Bulgarian is strictly and consistently ungrammatical (30b).

(30)　　a.　　　*ano:=la: aca:=la: mane*
　　　　　　　　what=Q who=Q ask
　　　　　　　　'What did who ask?'

　　　　b.　　　**aca:=la:*　　　　*ano:=la: mane*
　　　　　　　　who=Q　　　　　what=Q ask
　　　　　　　　('Who asked what?')

A bona fide case of Anti-Superiority would falsify a major prediction of contemporary syntactic theory, which is that a movement triggering head will always target the closest available constituent. This prediction arises because Superiority and similar intervention are derived as a special instance of the general operation Agree, which is built into the formal system (Chomsky 2000).[10]

However, there are several reasons to doubt that the multiply fronted *wh*-questions in (30) have the same syntactic status as their Bulgarian counterparts in (29). The first, most obvious reason to doubt this equivalence is that Moken is not an obligatory *wh*-movement language. We will assume that Moken *wh*-phrases are fronted due to information-structural requirements such as topic or focus marking. The relevant data is not available to understand exactly what motivates the fronting of *wh*-phrases in Moken. Baclawski (2015), however, proposes that left-dislocated phrases, including *wh*-phrases in Eastern Cham (Austronesian: Vietnam) are licensed by discourse anaphora accessibility in the sense of López (2009). We assume that similar discourse factors could be at work in Moken, a related language with very similar phenomena.

The second reason to doubt that Moken multiple *wh*-fronting has the same status as *wh*-fronting in Bulgarian is that the optional fronting of *wh*-phrases in Moken does not seem subject do ordinary locality constraints on *wh*-movement such as extraction from adjunct clauses (Section 2.2). Given this observation, at least some cases of the putative fronting operation in Moken can be attributed to the base-generation of a topic or focus phrase which binds a null pronoun in an argument position. We have seen that Moken allows subject and object pro-drop (cf. 6), so there is no empirical obstacle to positing null pronouns in this configuration.[11] Base-generation of the fronted element along with binding was also proposed by Pesetsky (1987) as an explanation for why English D-linked *wh*-phrases (e.g. *which girl*) do not respect Superiority; an analogous account is even more plausible for Moken given the general availability of null pronouns.

Third, and most importantly, Superiority effects show up in Moken *nə*-clefts when we restrict our attention to the fronting of a single *wh*-element. Specifically, while a subject *wh*-pronoun can be clefted when it has a co-argument object *wh*-pronoun (31a), an object *wh*-pronoun cannot be clefted in this configuration (31b):

(31)　　a.　　　*aca:=la: nə　　　makaw　ano:=la:*
　　　　　　　　who=Q COP　　　say　　what=Q
　　　　　　　　'Who is it that said　　　'What is it that who asked?'

---

[10]　A possible reaction to these data would be to assert that (30a) might be an echo question, which show obviated Superiority effects in English (cf. Comorovski 1996). I have no evidence that sentences like (30a) are not echo questions – such subtleties are often difficult to determine on the field – however, the unacceptability of (30b) would go unexplained under this account. 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 (30) and its relationship to clefts, as shown in the remainder of this section.

[11]　Nothing in our proposal hinges on the existence of null pronouns, however. As long as a theory countenances the distinction between (a theoretical analogue of) *wh*-movement and (a theoretical analogue of) binding from a non-argument position, it could accommodate the differences between Bulgarian and Moken we observe.

These data follow if Moken *nə*-clefts involve obligatory movement of a focused constituent.[12] It is an accidental gap in our data that none of the examples of fronting out of islands in Section 2.2 involved an overt *nə*-cleft. A movement-based explanation of the contrast in (31) thus predicts that fronting into a *nə*-cleft should respect island restrictions. Although we have no concrete evidence in support of this view, we leave it as a prediction to be confirmed or falsified by future work on the language.

In summary, there are three reasons why it would be a mistake to treat Moken examples like (30) as genuine Superiority violations. The first is that Moken is not an obligatory *wh*-movement language, while English and Bulgarian are. The second is that not all fronted constituents in Moken are moved to that position given that they do not respect the island-restrictions typical of *wh*-movement. And third, we have seen that in the restricted context of *nə*-clefts, Moken does in fact exhibit Superiority effects.

What is needed, then, is a constraint that can account for the putative Anti-Superiority effects in Moken illustrated in (30). The first clue for what that constraint might be is offered by completing the *ne*-cleft paradigm from (31) with multiply fronted *wh*-phrases. When a subject *wh*-phrase is overtly clefted, the object may be fronted over it (31c). However, the object *wh*-cleft is impossible even of the subject *wh*-pronoun appears to its left (31d):

(31) c. *ano:=la: aca:=la: nə   makaw*
          what=Q who=Q   COP   say
          'Who is it that said what?'

      d. **aca:=la:*   *ano:=la: nə   mane*
          who=Q   what=Q COP   ask

These examples are identical to the Anti-Superiority effects illustrated in (30). This observation sharpens the question about the contrast in (31c–d). We can now ask how the structure of multiply fronted wh-questions, clefts, and topics differs from the structure of the examples in (31a–b) with a single fronted constituent.

5 A constraint on crossed dependencies

In this section we propose that Anti-Superiority in Moken arise due to a constraint on crossed dependencies (cf. Pesetsky 1982's Path Containment Condition). This seems to be a general feature of fronted constituents in Moken, regardless of whether they have moved to that position.

First, we take the presence of argumental noun phrases either to the left of a cleft or in fronted positions more generally to be related to an empty category within the clause, either a trace or a silent pronominal element. Regardless of whether it is generated by movement, we assume that this relationship between a fronted constituent and an empty category involves a semantic dependency involving variable binding, as in many instances of cross-clausal anaphora (e.g. Grodzinsky & Reinhart 1993). The diagram below represents variable binding relationships for the two examples in (31c–d). Subscripts on the relevant expressions represent the dependency and *x* the clause-internal empty category:

(31') c'. $ano{:}{=}la{:}_2$ $aca{:}{=}la{:}_1$ nə $[x_1$ makaw $x_2]$

       d'. $*aca{:}{=}la{:}_1$ $ano{:}{=}la{:}_2$ nə $[x_1$ mane $x_2]$

The well-formed multiple dependency in (31c) – our putative instance of Anti-Superiority – is a nested dependency. In this configuration, dependencies are resolved by virtue of the occurrence of the coindexed empty category in the opposite order that they are established, e.g. 2, 1, 1, 2. In contrast, the ungrammatical (31d) involves crossed dependencies, where the variables appear in the same order that they are established, e.g. 1, 2, 1, 2.

We propose that this basic difference between nested and crossed dependencies is the actual explanation of the contrast in (31c–d). Crucially, the ungrammatical configuration in (31d) will be the same regardless of

---

[12]   More explicitly, overt *wh*-clefts would include a syntactic head with both a [*u*wh] and EPP feature.

whether the fronted constituent and its clause-internal correlate are generated by movement or by base generation. Thus, we can extend this explanation to cover all fronted constituents in Moken regardless of whether movement is involved.

Consider the case of the object-subject-verb sequence in (32a, repeated from 3a).[13] This sentence involves a subject *nə*-cleft. Fronting of the object to the left of the clefted subject is grammatical, because this configuration involves a nested dependency.

(32)   a.   *{kəla}*   *olan*   *nə*   *mətok*   *{ }*
                monkey snake COP   bite
                'It's the snake that bit the monkey.'

        b. $kəla_j$ [$olan_i$ $nə$]   [x_i $mətok$ x_j]

By contrast, the presence of an object cleft (33a) creates a restriction: the subject may not be fronted ahead of the cleft. A parallel schematization is given in (33b). Again, the difference between (32) and (33) seems to be that the two dependencies are nested in the former, but crossed in the latter. This implies that clefts and left-dislocated phrases obey a constraint against crossed dependencies.

(33)   a.   *{*}*   *kəla*   *nə*   *{olan}* *mətok*
                  monkey COP   snake bite
                'It's the monkey that the snake bit.'

        b.   *$olan_i$ [$kəla_j$ $nə$]   [x_i $mətok$ x_j]]

Due to the absence of island effects in these kinds of structures, we can safely conclude
that the fronted NP's above are base-generated in their surface positions. Thus, the constraint on crossed dependencies crucially must constrain these kind of base-generated dependencies between noun phrases in A-bar positions while still allowing multiple *wh*-movement as seen in Bulgarian.

Here, then, we can see what crucial difference between Bulgarian and Moken may be: while Bulgarian superiority violations (29) are driven by a recursive movement operation driven by a single head (C_{wh}) requiring movement to its specifier, the Moken structures above involve two distinct non-argument (A-bar) positions at the left edge of the clause which bind variables inside that clause. It seems clear that only the latter structures could evoke crossed dependencies:

(34)   a.   [wh_i wh_j C_{wh} [$t_{wh\text{-}i}$ … $t_{wh\text{-}j}$]]
        b.   *[NP_i [NP_j [x_i … x_j]]

As left-dislocated *wh*-phrases in Moken are also coindexed with an element in the embedded clause, we can extend the constraint on crossed dependencies in (33b) to these examples. This observation also extends to the cases observed by Pesetsky (1982), where examples in violation of his Path Containment Condition involve distinct non-argument positions.[14]

The following example provides further evidence that fronted noun phrases in Moken occupy distinct positions. An adverb can intervene between the two positions, and the constraint on crossed dependencies still holds. It is worth noting that these examples also illustrate that the effect holds when there is one *wh*-phrase and one non-*wh*-phrase.

---

[13]   We tentatively hypothesize that this fronting is a topicalizing left-dislocation, but confirmation would require future information structure-based research.

[14]   Cf. Pesetsky (1982)'s examples (29–30), repeated here as (vi–vii):
    (vi)   What subject$_i$ do you know [$_{S'}$ who$_j$ [$_S$ PRO to talk to t$_j$ about t$_i$]]
    (vii) *Who$_j$ do you know [$_{S'}$ what subject$_i$ [$_S$ PRO to talk to t$_j$ about t$_i$]]

(35) a. *{kəla}*   *aloj bubut*   *ano:=la: nə*   *mətok*   *{ }*
           monkey   yesterday   what=Q COP   bite
           'What is it that bit the monkey?'

      b. *{*}*   *aloj bubut*   *ano:=la: nə*   *{olan}*   *mətok*
                       yesterday   what=Q COP   snake   bite
           ('What is it that the snake bit yesterday?')

      c. *{ano:=la:}*   *aloj bubut*   *olan*   *nə*   *mətok*   *{ }*
           what=Q   yesterday   snake   COP   bite
           'It's the snake that bit what yesterday?'

      d. *{*}*   *aloj bubut*   *kəla*   *nə*   *{ano:=la:}*   *mətok*
                        yesterday   monkey   COP   what=Q   bite
           'It is the monkey that what bit yesterday?'

The generalization from this data is that for *wh*-phrases and clefts, multiple dependencies may be nested (35a), but they may not cross (35b). A full formalization of this effect would require further research, particularly into the information structural characteristics of clefts and extraction. Regardless, it is clear that *wh*-movement is not the root of this phenomenon. Hence, Superiority effects are not predicted.

5 Conclusion

To summarize, we will not conclude that Moken presents true Anti-Superiority effects; it does not challenge any theory rooted in Superiority. Instead, left-dislocated *wh*-phrases do not seem to undergo *wh*-movement proper. Furthermore, *wh*-phrases and clefts both display the same effect. We assert that Moken seems to have a constraint against crossed dependencies, which applies to left-dislocated *wh*-phrases, as with clefts. This assumes that left-dislocated *wh*-phrases and the pivots of clefts must bind an element in the embedded clause. Superiority effects do not obtain, because there is no *wh*-movement, just base-generation. Overt *wh*-clefts, however, may be possible landing sites for *wh*-movement, as Superiority effects do obtain in this one case. These results are schematized in (37).

(37) a. $[DP_j [[DP_i COP] [x_i V x_j]]]$ **"Anti-Superiority" with clefts**
      b. $*[DP_j [[DP_i COP] [x_j V x_i]]]$

      c. $[wh_j [wh_i [x_i V x_j]]]$ **"Anti-Superiority" with *wh*-phrases**
      d. $*[wh_j [wh_i [x_j V x_i]]]$

      e. $[wh_i COP] [x_i V wh_j]$ **Superiority effect with overt *wh*-clefts**
      f. $*[wh_j COP] [wh_i V x_j]$

These results generally uphold the theory of Superiority and encourage further study of word order and information structure in Moken and typologically similar languages. In particular, a more precise understanding of extraction and clefting would enrich our understanding of why this Anti-Superiority effect is apparent. For now, we have asserted that it is not a challenge to standing syntactic theory, as there is architecture that could explain its existence.

Further research should also investigate whether these effects contribute to the historical development of Moken from its Austronesian roots. If it could be shown, for example, that Moken questions are formed with covert clefts, this could represent a subtle Austronesian retention, as *wh*-clefts are common in the family, particularly in verb-initial languages like Malagasy (cf. Oda 2002; Potsdam 2006; Aldridge 2013).

References

Aldridge, Edith. 2013. Wh-clefts and verb-initial word order in Austronesian languages. In, Katharina Hartmann and Tonjes Veenstra (eds.), *Cleft structures*, 71–96. Amsterdam/Philadelphia: John Benjamins.

Baclawski Jr., Kenneth. 2015. Topic, focus, and *wh*-phrases in Cham and Moken. (Paper presented at the Information Structure in Spoken Language Corpora 2 (ISSLAC2) Workshop, Paris, France, 2 December 2015.)

Boeckx, Cedric. 2003. *Islands and chains: Resumption as stranding.* Amsterdam/Philadelphia: John Benjamins.

Bošković, Željko. 2002. On multiple wh-fronting. *Linguistic Inquiry* 33. 351–383.

Cable, Seth. 2010. The grammar of Q: Q-particles, wh-movement, and pied-piping. New York: Oxford University Press.

Chomsky, Noam. 1973. Conditions on transformations. In Stephen Anderson and Paul Kiparsky (eds.), *A festschrift for Morris Halle*, 232–286. New York: Holt, Rinehart and Winston.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In Roger Martin, David Michaels and Juan Uriagereka eds., *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, 89–155. Cambridge, Mass: MIT Press

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. New York: Oxford University Press.

Comorovski, Ileana. 1996. *Interrogative phrases and the syntax-semantics interface*. Dordrecht/Boston: Kluwer Academic Publishers.

Denham, Kristin. 2000. Optional wh-movement in Babine-Witsuwit'en. *Natural Language & Linguistic Theory* 18. 199–251.

Enfield, Nicholas J. 2003. Linguistic epidemiology: semantics and grammar of language contact in mainland Southeast Asia. London: Routledge Curzon.

Fanselow, Gisbert. 2004. The MLC and derivational economy. In, Arthur Stepanov, Gisbert Fanselow, Ralf Vogel (eds.), *Minimality effects in syntax*, 73-124. Berlin/New York: Mouton de Gruyter.

Grodzinsky, Yosef & Tanya Reinhart. 1993. The innateness of binding and coreference. *Linguistic Inquiry* 24. 69–101.

Huang, James C.-T. 1984. On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15. 531–574.

İşsever, Selçuk. 2009. A syntactic account of *wh*-in-situ in Turkish. In, Sıla Ay (ed.), *Essays on Turkish Linguistics: Proceedings of the 14th International Conference on Turkish Linguistics*, 103–112. Weisbaden: Harrassowitz Verlag.

Jenks, Peter. 2013. Quantifier float, focus, and scope in Thai. *Proceedings of BLS 39.* http://ling.auf.net/lingbuzz/001960

Jeong, Youngmi. 2003. Deriving Anti-Superiority effects: Multiple *wh*-questions in Japanese and Korean. In, Cedric Boeckx and Kleanthes K. Grohmann (eds.), *Multiple wh-fronting*, 131–140. Amsterdam/Philadelphia: John Benjamins.

Larish, Michael D. 1999. The position of Moken and Moklen within the Austronesian language family. University of Hawai'i. (Doctoral dissertation.)

Larish, Michael D. 2005. Moken and Moklen. In, K. Alexander Adelaar and Nikolaus Himmelmann (eds.), The Austronesian Languages of Asia and Madagascar, 513–533. London: Routledge.

Milsark, Gary. 1977. Peculiarities of the existential construction in English. *Linguistic Analysis* 3. 1–29.

Miyagawa, Shigeru. 2009. Why agree? Why move?: Unifying agreement-based and discourse-configurational languages. Cambridge, MA: MIT Press.

Oda, Kenji. 2002. Wh-questions in V-initial languages. University of Toronto. (Master's thesis.)

Pittayaporn, Pittayawat. 2005. Moken as a mainland Southeast Asian language. In Anthony Grant and Paul Sidwell (eds.), *Chamic and beyond: Studies in mainland Austronesian languages*, 198–210. Canberra: *Pacific Linguistics* 569.

Pesetsky, David. 1982. Paths and categories. MIT. (Doctoral Dissertation.)

Pesetsky, David. 1987. Wh-in-situ: Movement and Unselective Binding. In, Eric Reuland and Alice G. B. ter Meulen (eds.), *The Representation of (In)definiteness*, 98–129. Cambridge, MA: MIT Press.

Potsdam, Eric. 2006. More concealed pseudoclefts and the clausal typing hypothesis. *Lingua* 116. 2154–2182.

Richards, Norvin. 2001. *Movement in language*. New York: Oxford University Press.

Say Bay, Naw. 2015. The process of preserving Moken from extinction. (Paper presented at SoLE-4, Payap University, 26 May 2015.)

Thurgood, Graham. 1999. From Ancient Cham to modern dialects: Two thousand years of language contact and change. Honolulu: University of Hawai'i Press.

Watanabe, Akira. 1992. Subjacency and S-Structure movement of wh-in-situ. *Journal of East Asian Linguistics* 1.3. 255–291.