

Friday, June 25, 1999

Clefted Questions in Malay *

Peter Cole, Gabriella Hermon and Norhaida Aman

To appear in Malay/Indonesian Linguistics (working title), David Gil and James

Collins (eds), London: Curzon Press

1. Introduction

The Malay language is blessed with a variety of forms for asking information (WH) questions. In addition to WH in situ questions like (1),

(1) Ali membeli apa?

Ali buy what

'What did Ali buy?'

there exist in Educated Informal Malay (EM) two types of questions in which the question word appears at the beginning of the clause over which it has scope:¹

(2) Apa Ali beli?

what Ali buy

'What did Ali buy?'

(3) Apa yang Ali beli?

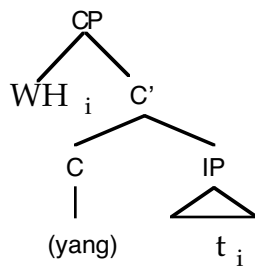
what that Ali buy

'What did Ali buy?'

This paper addresses the question of what the correct analysis is for questions like (3).

We will consider two potential analyses of (3). According to the first analysis, which we shall refer to as the Matrix Complementizer Hypothesis (MCH), (2) and (3) are both derived from the same structure:

(4)



According to the Null Complementizer Hypothesis, in both (2) and (3), the question word *apa* has moved from its base generated position within IP to Spec of CP. The only difference between (2) and (3) is that in (2) the complementizer is null while in (3) it is filled by *yang*. Thus, according to the Null Complementizer Hypothesis, Malay WH questions like (3) differ from those found in European languages like English only in that Malay permits the complementizer position in main clauses to be filled while English does not.

The second analysis, which we shall refer to as the Headless Relative Clause Hypothesis (HRCH), claims that while (2) does, in fact, have a structure like that of English WH questions, sentences like (3) have a very different structure. According to the Headless Relative Clause Hypothesis, questions like (3) are nominal or null copula sentences like (5):

(5) Ali guru.

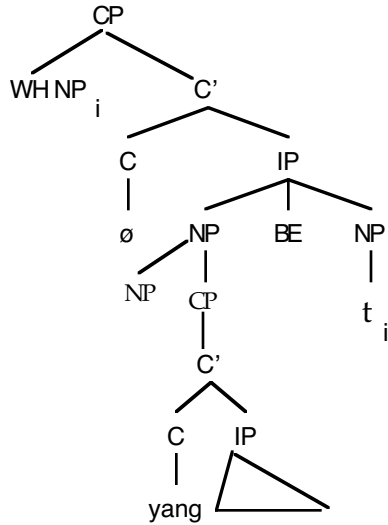
Ali teacher

'Ali is a teacher.'

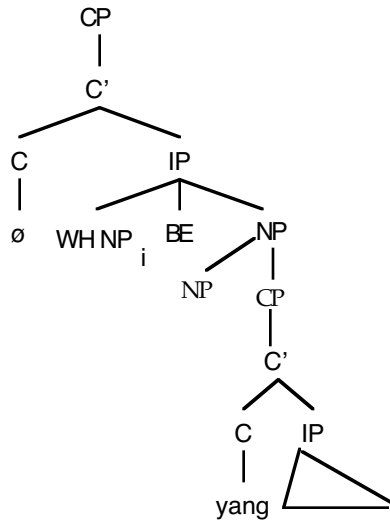
We assume that (5) is comprised of an IP or small clause containing two NPs, possibly related by an abstract verb 'be'. We shall, for convenience, refer to sentences like (5) as "null copula sentences".² According to the Headless Relative Clause analysis, (3) consists of two NPs, a headless relative clause [yang Ali beli]

'that which Ali bought' and *apa* 'what', which are related in the same way *Ali* and *guru* are related in (5), either as, roughly, (6) or (7).

(6)



(7)



Later in the paper we shall opt for structure (6),³ but for the moment we shall ignore the empirical differences between (6) and (7) (each of which have been proposed by different authors in the literature on Malay syntax), and will concentrate on the differences between the MCH and the HRCH.⁴ To the best of our knowledge, there has never been a systematic investigation of the arguments distinguishing between these hypotheses. In the discussion which follows, we shall present a number of arguments that support the HRCH over the MCH for EM. Thus, we argue that, at least for EM, the apparent similarity between (2) and (3) is illusory, and that EM WH questions employing *yang* have a structure quite different from those without *yang*, and also quite different from WH questions in English and other European languages.

Upon establishing the superiority of the the HRCH over the the MCH for EM, we will turn to the question of whether (6) or (7) is to be preferred. We will contend that an argument put forward by Kader (1976, 1980) provides reason to believe that (6) is the correct structure. We shall then turn to the issue of whether movement is involved in the derivation of questions like (3), and will argue against the claims of Martohardjono that a non-movement analysis is to be preferred. Finally, we will turn to the status of *yang* questions in one variety of colloquial Malay, that spoken in Singapore (CSM). We shall argue that a variety of syntactic changes in progress are conspiring to make (4) a possible structure for *yang* questions in that dialect. We leave open for further research whether the changes occurring in CSM are also taking place in the colloquial Malay of other regions.

2. Restrictions on the Distribution of Yang

Our first argument will be based on seemingly mysterious restrictions on the distribution of *yang* in questions. For many speakers, questions without *yang* can be formed using the full range of question words:

(8) a. Apa dibiaki Ali?

 what was fixed Ali

'What was fixed by Ali?'

b. Apa kau nak makan?

 what you want eat

'What do you want to eat?'

(9) a. Siapa nampak kau?

who see you

'Who saw you?'

b. Siapa kau nampak?

who you see

'Who do you see?'

(10) Di mana kau tinggal?

at where you stay

'Where do you live?'

(11) Ke mana kau pergi?

to where you go

'Where did you go?'

(12) Bagaimana Ali baiki kereta itu?

How Ali fix car that

'How did Ali fix that car?'

(13) Macam mana awak buat itu?

How you do that

'How did you do that?'

(14) Kenapa Ali dipecat?

Why Ali was fired

'Why is it that Ali was fired?'

In contrast, when the question is formed with *yang* as in (3), only questions with *apa* and *siapa* are completely well-formed:

(8') a. Apa yang dibaiki Ali?

what that was fixed Ali

'What was fixed by Ali?'

b. Apa yang kau nak makan?

what that you want eat

'What do you want to eat?'

(9') a. Siapa yang nampak kau?

who that see you

'Who saw you?'

b. Siapa yang kau nampak?

who that you see

'Who do you see?'

(10') ?? Di mana yang kau tinggal?

at where that you stay

'Where do you live?'

(11') ?? Ke mana yang kau pergi?

to where that you go

'Where are you going?'

(12') ?? Bagaimana yang Ali baiki kereta itu?

how that Ali fix car that

'How did Ali fix that car?'

(13') ?? Macam mana yang awak buat itu?

how that you do that

'How did you do that?'

(14') ?? Kenapa yang Ali dipecat?

Why that Ali was fired

'Why is it that Ali was fired?'

Speakers vary with regard to their absolute judgements regarding (8')-(14'), but there is general agreement that while (8')-(9') are perfect, (10')-(14') are of diminished grammaticality, ranging from odd for some speakers to ungrammatical for others.

Let us consider the predictions of the MCH and the HRCH regarding (8')-(14'). According to the MCH, (8') and (9') differ from (10')-(14') only with respect to whether the WH word is an NP or some other grammatical category (prepositional phrase or adverb). There is no reason to expect that the grammatical category of the element filling the Spec of CP would affect whether it is grammatical to fill the complementizer position. Thus, the MCH provides no explanation for the differences in well formedness between (8')-(9') and (10')-(14').

Turning now to the HRCH, according to this analysis, *yang* is not the main clause complementizer but rather the complementizer introducing a headless relative clause.⁵ Thus, according to the HRCH, the restrictions on (8')-(14') should follow from the restrictions on which positions can be relativized in headless relative clauses. Speakers who find (10')-(14') illformed, should also find sentences like (15)-(19) illformed to roughly the same degree, a prediction which is substantiated by the facts:

(15) ?? Yang aku tinggal (ialah) di K.L.⁶

that I stay is at K.L.

'The place that I live is in K.L.'

(16) ?? Yang aku pergi (ialah) ke K.L.

that I go is to K.L.

'Where I am going is to K.L.'

(17) ?? Yang Ali baiki kereta itu (ialah) dengan alatnya.

that Ali fix car that is with tool-his

'The way that Ali fixed the car is with his tool.'

(18) ?? Yang Ali baiki kereta itu (ialah) dengan hati-hati.

that Ali fix car that is with care

'The way that Ali fixed the car is with care.'

(19) ?? Yang Ali dipecat (ialah) kerana dia cuai

that Ali was fired is because he careless

'Why Ali was fired is because he was careless.'

The ungrammaticality of (15)-(19) is in sharp contrast to the grammaticality of (20)-(21):

(20) Yang aku makan nasi goreng (-lah).

that I eat rice fried focus

'Fried rice is what I am eating.'

(21) Yang kau nampak Siti (-lah).

that you see Siti focus

'Siti is who you see.'

The generalization is that NP arguments can be relativized by headless *yang* relative clauses, but adverbial adjuncts like 'location' (15), 'direction' (16), 'instrument' (17), 'manner' (18) and 'reason' (19) cannot.

Turning now to (8')-(14'), according to the HRCH, the sentences of structure of (8')-(9') are based on headless relative clauses like (20)-(21), and, therefore, well-formed. Sentences (10')-(14'), however, are postulated to be based on headless relative clauses like those in (15)-(19), and, hence, are expected to be less than fully grammatical. Thus, the HRCH correctly predicts the illformedness of (10')-(14') without the need for additional stipulation.

3. Possible Replies to Questions

We will turn now to a consideration of the possible replies to WH questions with and without *yang*. The examples of (22)-(23) show the possible replies to (8)-(9), questions formed without *yang*:

(22) 'What' Question without *yang*

a. Apa kau nak makan?

what you want eat

'What do you want to eat?'

b. Aku nak makan ayam buah keluak.

I want eat chicken fruit keluak

'I want to eat chicken with keluak nuts.'

c. @Yang aku nak makan ayam buah keluak.⁷

that I want eat chicken nut keluak

'What I want to eat is chicken with keluak nuts.'

(23) 'Who' Question without *yang*

- a. Siapa kau nampak?
who you see
'Who do you see?'
- b. Aku nampak Fatimah.
I see Fatimah
'I see Fatimah.'
- c. @Yang aku nampak Fatimah.
that I see Fatimah
'The one that I see is Fatimah.'

While the appearance of a headless relative clause is infelicitous in replies to questions without *yang*, they are felicitous when *yang* appears in the question:

(24) 'What' Question with *yang*

- a. Apa yang kau nak makan?
what that you want eat
'What do you want to eat? What is the thing that you want to eat?'
- b. Aku nak makan ayam buah keluak.
I want eat chicken fruit keluak
'I want to eat chicken with keluak nuts.'
- c. Yang aku nak makan ayam buah keluak.
that I want eat chicken fruit keluak
'What I want to eat is chicken with keluak nuts.'

(25) 'Who' Question with *yang*

a. Siapa yang kau nampak?

who that you see

'Who do you see? Who is the one that you see?'

b. Aku nampak Fatimah.

I see Fatimah

'I see Fatimah.'

c. Yang aku nampak Fatimah.

that I see Fatimah

'The one that I see is Fatimah.'

To summarize, questions with *yang* can receive replies with headless relative clauses, but those without *yang* cannot.

Let us compare the predictions of the HRCH and the MCH regarding (22)-(25). According to the MCH, sentences with and without *yang* have the same structure except for the occurrence or nonoccurrence of an optional morpheme. Thus, it would be expected that answers with and without *yang* should be possible in reply to question without *yang*, just as in English the absence of a complementizer in a question does not prevent its use in an answer:

(26) a. Who does Mary think Tom saw?

b. She thinks he saw Anne.

c. She thinks that he saw Anne.

Thus, the MCH fails to predict the ungrammaticality of the (c) sentences of (22)-(24).

In contrast, the the HRCH provides an explanation for the illformedness of answers with *yang* in (22)-(23) and their wellformedness in (24)-(25). If

questions like those of (24a)-(25a) have the structure [headless relative clause NP], it is unsurprising that headless relative clauses can be used in answers. The infelicity of *yang* in the replies to (22a)-(23a) is also predicted since answers are expected to preserve the overall structure of the question asked. Thus, the the HRCH (but not the MCH) provides an explanation for the otherwise inexplicable differences in the sets of possible replies to questions with and without *yang*.⁸

4. The Distribution of Yang and Bahawa

We shall now turn to an argument due to Kader (1976, 1980), based on the distribution of the complementizer *yang* versus that of the complementizer *bahawa*. Both *yang* and *bahawa* can be used to introduce complement clauses:

(27) Ali memberitahu saya tadi yang Fatimah sakit semalam.

Ali told me just now that Fatimah was sick yesterday

'Ali told me earlier that Fatimah was sick yesterday.'

(28) Ali memberitahu saya tadi bahawa Fatimah sakit semalam.

Ali told me just now that Fatimah was sick yesterday

'Ali told me earlier that Fatimah was sick yesterday.'

Sentences (27) and (28) differ significantly in register, but both are well formed (in their appropriate registers).

In the case of relative clauses, however, in all registers the modifying clause can be introduced with *yang* but not with *bahawa*:

(29) perempuan $\left\{ \begin{array}{l} yang \\ *bahawa \end{array} \right\}$ telah diberi hadiah

woman $\left\{ \begin{array}{l} that \\ that \end{array} \right\}$ already was given gift

'the woman that was already given a gift'

(30) buku $\left\{ \begin{array}{l} \textit{yang} \\ \textit{*bahawa} \end{array} \right\}$ John beli

book $\left\{ \begin{array}{l} \textit{that} \\ \textit{that} \end{array} \right\}$ John bought

'the book that John bought'

The same pattern seen in headed relative clauses is found in headless relative clauses:

(31) $\left\{ \begin{array}{l} \textit{yang} \\ \textit{*bahawa} \end{array} \right\}$ Fatimah suka

$\left\{ \begin{array}{l} \textit{that} \\ \textit{that} \end{array} \right\}$ Fatimah likes

'what Fatimah likes'

(32) $\left\{ \begin{array}{l} \textit{yang} \\ \textit{*bahawa} \end{array} \right\}$ tinggal di Singapura

$\left\{ \begin{array}{l} \textit{that} \\ \textit{that} \end{array} \right\}$ stays in Singapore

'the one that lives in Singapore'

To summarize, either the complementizer *bahawa* or the complementizer *yang* can be used to introduce complement clauses, but only *yang* can be used to introduce relative clauses, both headed and headless.

Let us turn now to the predictions regarding choice of complementizers made by the two hypotheses under consideration. The MCH claims that Malay lacks the prohibition found in English blocking the presence of complementizers in main clauses. Thus, the MCH predicts that Malay main clauses will show the same range of complementizers found in complement clauses. Specifically, both *yang* and *bahawa* should be possible in questions.

In contrast, the HRCH claims that the complementizer found in *yang* questions is the complementizer introducing a headless relative clause. Since only *yang* and not *bahawa* can be used to introduce a relative clause, the HRCH predicts that only *yang* will be grammatical in forming WH questions like (3).

The predictions of the the HRCH are, in fact, correct. As (33) and (34) show, only *yang* can be used in the formation of WH questions:

(33) Siapa $\left\{ \begin{array}{l} \textit{yang} \\ \textit{*bahawa} \end{array} \right\}$ kau nampak?

who $\left\{ \begin{array}{l} \textit{that} \\ \textit{that} \end{array} \right\}$ you see

'Who did you see?'

(34) Semalam apa $\left\{ \begin{array}{l} \textit{yang} \\ \textit{*bahawa} \end{array} \right\}$ Ali makan?

Yesterday what $\left\{ \begin{array}{l} \textit{that} \\ \textit{that} \end{array} \right\}$ Ali eat

'What did Ali eat yesterday?'

Thus, the ungrammaticality of *bahawa* in the formation of WH questions provides an additional argument in favor of the HRCH over the MCH.

5. The Status of WH Questions Without Yang

We have presented three arguments that WH questions with *yang* like (3) derive from a headless relative clause structure like (6) or (7), rather than from structure like (4), in which *yang* is an optional matrix clause complementizer. At this stage, we would like to consider briefly whether WH questions without *yang* could be derived from a headless relative clause structure, thereby permitting a unified analysis of WH questions with and without *yang*. More concretely, can questions like (2) (repeated)

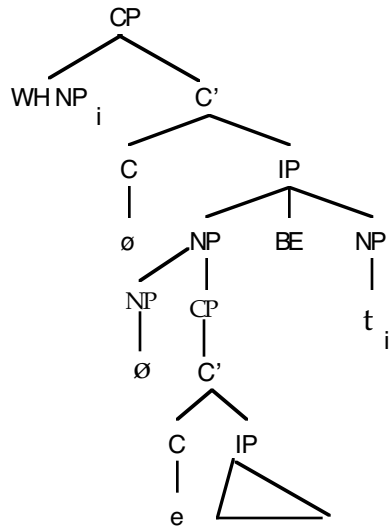
(2) Apa Ali beli?

what Ali buy

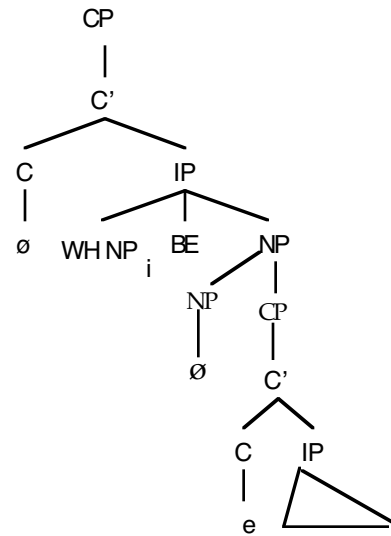
'What did Ali buy?'

be derived from (6') or (7')?⁹

(6')



(7')



Structures (6')-(7') differ from (6)-(7) in that the complementizer of the headless relative clause is *yang* in (6)-(7) but null in (6')-(7'). That is, the distinction between (6)-(7) and (6')-(7') reduces to the claim that the complementizer *yang* is optional in headless relative clauses.

This claim, however, is incorrect, as can be seen from the ungrammaticality of (36):

(35) [_{NP} Yang Ali baca semalam] ialah buku itu.

That Ali read yesterday is book that

'What Ali read yesterday is that book.'

(36) *_{[NP} Ali baca semalam] ialah buku itu.

Ali read yesterday is book that

'What Ali read yesterday is that book.'

While headless relative clauses like (35), which are introduced by the complementizer *yang*, are well-formed, those in which *yang* is missing are ungrammatical. Thus, we conclude that WH questions like (2), in which *yang* does not occur, cannot be derived from headless relative clause structures like (6')-(7').

6. The Position of Question Words in Yang Questions

We shall now return to WH questions with *yang*, and will examine whether the question word originates in object position, as in (6), or subject position, as in (7). The argument which we shall provide is an elaboration of that of Kader (1976, 1980), for whom this issue was a central concern. While we believe that the argument put forward by Kader remains convincing today, we would like to point out that the central judgment on which it is based is subject to variation: It appears to hold for Formal Malay (FM), but not for Educated Informal Malay (EM). We shall return to the issue of dialectal variation later.

Kader notes the existence of a restriction on the distribution of the interrogative focus particle *kah*. Oversimplifying slightly, this particle, like the affirmative focus particle *lah*, can appear on elements in the predicate, but not on the subject:

(37) Fatimah kata [Siti membeli [**buku itukah**] semalam]?

Fatimah say Siti bought book that-Q yesterday

'Did Fatimah say that Siti bought **that book** yesterday?' (Boldface indicates focus)

(38) *Fatimah kata [[**Sitikah**] membeli buku itu semalam]?¹⁰

Fatimah say Siti-Q bought book that yesterday

'Did Fatimah say that **Siti** bought that book yesterday?' (Boldface indicates focus)

(39) Fatimah kata [Siti membeli [**apakah**] semalam]?

Fatimah say Siti bought what-Q yesterday

'What did Fatimah say that Siti bought yesterday?'

(40) *Fatimah kata [[**siapakah**] membeli buku itu semalam]?

Fatimah say who-Q bought book that yesterday

'Who did Fatimah say bought that book yesterday?'

In (37) and (39) the interrogative particle is associated with the direct object, and the sentence is well-formed. In contrast, in (38) and (40), *kah* is associated with the subject, and the resulting sentence is ill-formed.

This pattern holds not only for verbal sentences but for null copula sentences as well:¹¹

(41) Ali rasa [Rahman **seorang yang baikkah**]?

Ali feel Rahman one-man that good-Q

'Does Ali feel that Rahman is **a good person**?'

(42) *Ali rasa [[**Rahmankah**] seorang yang baik]?

Ali feel Rahman-Q one-man that good

'Does Ali feel that **Rahman** is a good person?'

As in verbal sentences, *kah* can be associated with the predicate but not the subject.

Let us now consider the predictions made by (6) and (7) with regard to the occurrence of *kah* in questions like (3). If (6) is the structure underlying (3), it is predicted that *kah* can appear on the question word since the question word originates in predicate position. In contrast, if (7) is the structure underlying (3), it would be expected that the occurrence of *kah* would be illformed since the question word originates in subject position. In fact, *kah* is well-formed on questions with *yang*,

(43) Apakah yang Ali beli?

what-Q that Ali buy

'What did Ali buy?'

(44) Siapakah yang datang?

who-Q that came

'Who came?'

Thus, as Kader showed, the grammaticality of *kah* on *yang* questions is persuasive evidence that the question word originates in predicate position, as in (6), rather than in subject position, as in (7).

7. Evidence for Movement

We have argued so far that the structure of WH questions with *yang* is like (6), rather than like (7) or (4). We shall next turn to the implications of adopting structure (6) for the issue of whether Malay WH questions like (2) and (3) involve movement. We shall also examine whether the movement in question is movement of the WH word itself or of an operator.

We shall argue that WH movement occurs in both (2) and (3), but that in (3) the movement of the question word is strictly local while the phonologically null operator internal to the headless relative clause can potentially move successive cyclically. Our position is to be contrasted with two other analyses, those of Cheng (1992) and Martohardjono (1993). Both Cheng and Martohardjono contend that Indonesian (and, hence, presumably, Malay as well) lack WH movement, and that the WH word is in situ at S-structure (i.e. structure (7)). While Cheng does not discuss OP movement, it would appear to be compatible with her analysis. Martohardjono, however, explicitly denies that Indonesian has operator movement.

We shall turn first to the evidence for movement of the WH word to Spec CP in WH questions with *yang*. We saw in the previous section that in WH questions with *yang*, the question word originates in predicate position. Thus, movement to pre-complementizer position (i.e. Spec CP) is necessary in order to explain the surface word order. Viewed from this perspective, the argument for (6) is also an argument that the question word in (3) is not in situ at S-structure, but, rather, has undergone movement to Spec CP, pace Cheng (1992) and Martohardjono (1993).

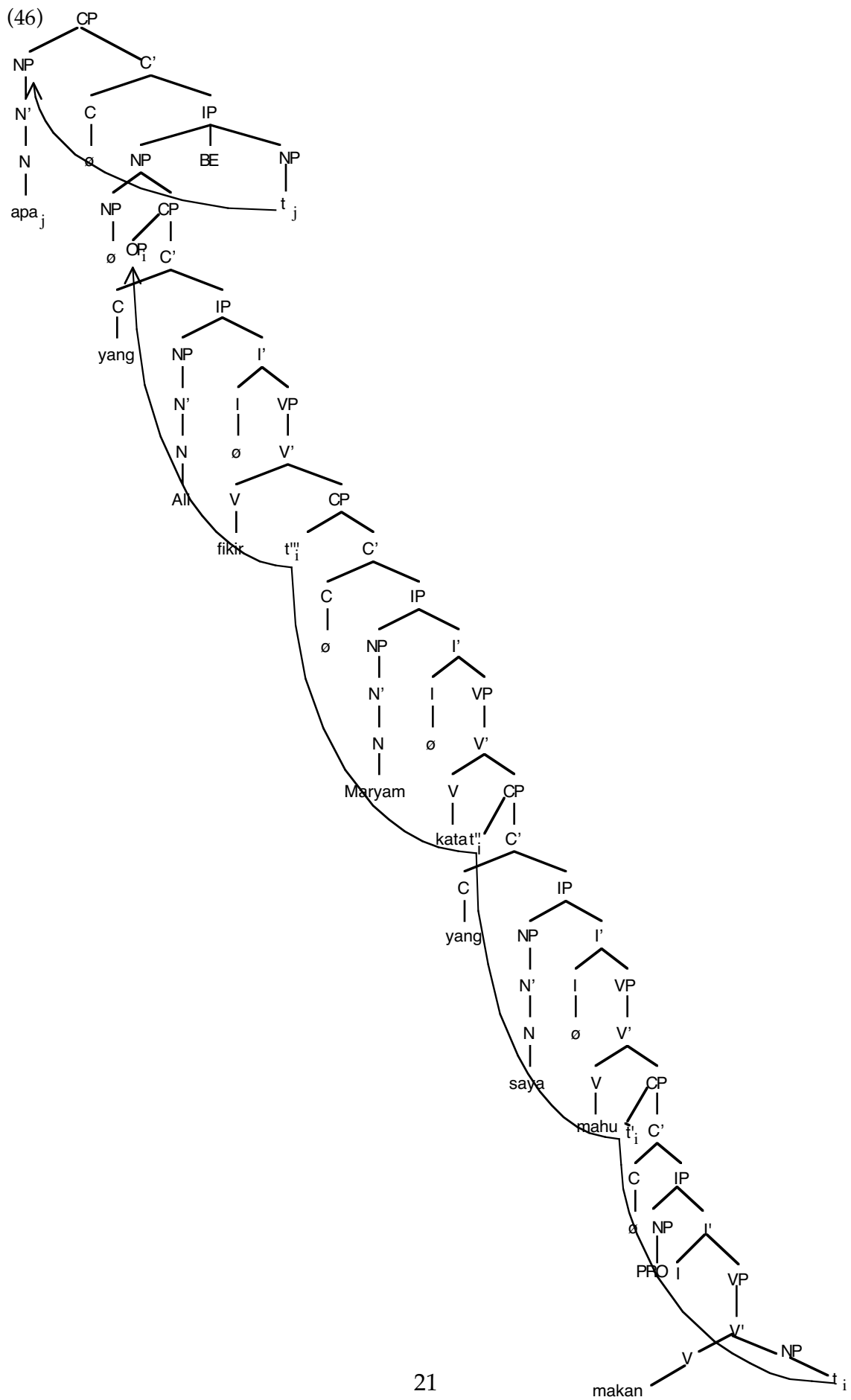
We next turn to the question of whether apparent long distance WH movement in questions with *yang* is to be viewed in our analysis as long distance movement of the WH or of a phonologically null operator. Consider the derivation of questions with *yang* like (45), questions that appear on first examination to involve long distance WH movement:

(45) Apa yang Ali pikir Maryam kata yang saya mahu makan?

what that Ali think Miriam say that I want eat

'What does Ali think Miriam says that I want to eat? What is the thing that Ali thinks Miriam says that I want to eat?'

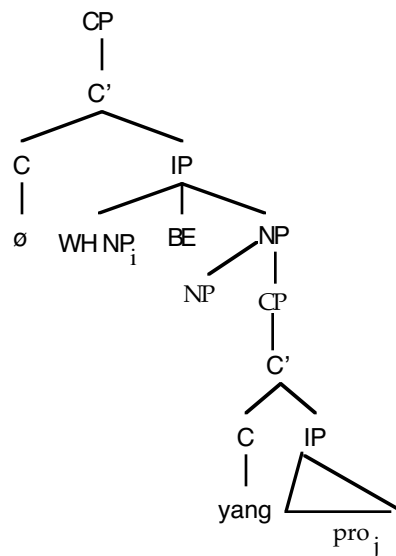
It would appear on superficial examination that (45) involves interclausal WH movement. However, as tree (46) shows, according to the HRCH the movement of the WH word *apa* is local. The element which moves interclausally is the phonologically null OP internal to the headless relative clause:



In (46) OP moves from the object position of the lowest clause (object of *makan* 'eat') successive cyclically to the specifier position of the headless relative clause. In contrast, the question word *apa* has only moved locally from the matrix predicate to Spec CP.

This position is to be contrasted with that of Martohardjono, who claims not only that sentences like (3) do not involve WH movement, but that there is also no operator movement in the derivation of these sentences. According to Martohardjono, the empty category corresponding to the original site of the relativized NP is an instance of *pro* rather than a trace of A'-movement of a relative operator:

(47)



The main reason to believe that questions like (3) do in fact involve operator movement is that *yang* questions, like questions without *yang* are subject to island conditions similar to those found in English. This is described in detail in Cole and Hermon (1995) and (1998). We show there that question both

with and without *yang* are barred when the relativized NP originates in a relative clause, sentential subject, adjunct, or WH island. The situation is described similarly by Rogayah (1995) for formal Malay and by Saddy (1991 and 1993) for Indonesian. Examples adapted from Cole and Hermon (1998) are provided below:

(48) Extraction from Relative Clause

*Siapa_i yang Ali suka perempuan_j yang e_i fikir e_i tinggal di Singapura?

who that Ali like woman that think live in Singapore

'Who does Ali like the woman that thinks lives in Singapore?'

(49) Extraction from Sentential Subject

*Siapa_i yang [yang e_i mengahwini Fatimah] mengecewakan ibu Fatimah.

who that that married Fatimah upset mother Fatimah

'Who did that married Fatimah upset Fatimah's mother?'

(50) Extraction from Adjunct

*Siapa_i yang Ali dipecat kerana e_i seorang pencuri?

who that Ali was fired because a thief

'Who was Ali fired because was a thief?'

(51) Extraction from WH Islands

*Siapa_i yang awak agak sama ada e_i ternampak Ali?

who that you wonder whether saw Ali

'Who do you wonder whether saw Ali?'

Martohardjono explains the existence of island violations similar to (48)-(51) as due to a version of Huang's (1989) Generalized Control Rule rather than to subjacency. According to Martohardjono's version of Generalized Control, pro must be identified by being coindexed with the closest c-commanding NP.

For example, on the assumption that *e* is pro in (48), Martohardjono's version of Generalized Control would require that *e* be coindexed with *perempuan* 'woman', not with *siapa* 'who', thereby predicting the ungrammaticality of (48).

While Martohardjono's version of Generalized Control does correctly predict many of the island facts, it fails to block extraction from subject position inside sentential subjects as in (49). *Siapa* is the closest c-commanding NP to *e*. Thus, Martohardjono's analysis predicts that (49) should be well-formed. Note that the ungrammaticality of (49) could not be due to any general interpretive constraint against questions inside sentential subjects since WH in situ in sentential subjects is well-formed:

(52) Yang siapa mengahwini Fatimah mengecewakan ibunya?

that who married Fatimah disappointed mother her

'Who did that married Fatimah dissappoint her mother? That who married Fatimah disappointed her mother?'

Not only does Martohardjono's use of Generalized Control make incorrect predictions regarding islands, it makes incorrect predictions regarding the possibility of relativization or questioning from deeply embedded positions. Since Martohardjono's analysis requires that the antecedent for pro be the closest c-commanding NP, relative clauses like those in (53)-(54) would be ruled out:

(53) Saya membaca [_{NP} buku_i [_{CP} yang [Fatimah kata [*e*_i dibeli *e*_i

I read book that Fatimah say was bought

oleh Ali tahun lepas]]]]

by Ali year last

'I read a book that Fatimah said was bought by Ali last year.'

(54) Ali mengahwini [_{NP} perempuan_i [_{CP} yang ibunya fikir

Ali married woman that mother-his thinks
 e_i tidak tahu memasak kue lapis]]

not know cook cake "lapis"

'Ali married a woman that his mother thinks does not know how to make "lapis" cake.'

According to Martohardjono's Generalized Control analysis, both (53) and (54) should be ungrammatical since the closest c-commanding NP to e in (53) is *Fatimah* and not *buku* 'book'. Similarly, the closest c-commanding NP to e in (54) is *ibunya* 'his mother' and not *perempuan* 'woman'.

Similar examples involving questions rather than relative clauses are given in (55)-(56):

(55) Buku apa_i yang Fatimah kata [e_i dibeli e_i
book what that Fatimah say was bought
oleh Ali tahun lepas]?
by Ali year last

'What book did Fatimah say was bought by Ali last year.'

(56) Siapa yang Maryam fikir
who that Miriam think
[e_i tidak tahu memasak kue lapis]?
not know cook cake "lapis"

'Who did Miriam think doesn't know how to cook "lapis" cake?'

In (55), according to Martohardjono's analysis, e should be coindexed with *Fatimah*, while in fact it is coindexed with *siapa*. Similarly, in (56), her analysis predicts that e will be coindexed with *Maryam*, while in reality it is coindexed with *siapa*.

Martohardjono does in fact provide one example similar to (53)-(56), which she does not discuss in detail:

(57) Siapa_i yang kau-kira e_i dilihat (oleh) Siti?

who that you think was seen by Siti

'Who do you think was seen by Siti?'

Martohardjono does not discuss the problem sentences like (57) pose for her analysis, but we infer from the hyphenation of *kau* 'you' and *kira* 'think' that Martohardjono believes that *kau* is cliticized to the verb, and, hence, does not constitute a potential c-commanding antecedent for *pro*. Whatever the case may be with regard to *kau*, *Fatimah* and *Maryam* in (55)-(56) are full NPs. Thus, there is no possibility that they are cliticized to the verb and, as a result, they constitute counter examples to Martohardjono's analysis. Thus, we conclude that Martohardjono's alternative to a movement analysis of islandhood is not viable.

We have argued so far that in WH questions with *yang* there is local movement of the WH nominal to matrix Spec CP and potentially long distance movement of OP to the Spec CP of the headless relative clause. We will conclude this section by showing that questions without *yang* involve potentially long distance movement of the WH word itself. There are two arguments for our position. First, we showed in Section 5 that WH questions employing *siapa* and *apa* without *yang* cannot be derived from headless relative clauses in which the complementizer *yang* has been omitted since *yang* is obligatory in headless relative clauses. This leaves movement to Spec CP of the WH words themselves as the only alternative derivation.

Secondly, there is no potentially plausible relative clause source for adjunct WH questions like those in (58)-(59).

(58) *Bila_i Maryam fikir [yang Ali akan datang ke sini t_i].*

when Mariam think that Ali will come to here

'When does Miriam think that Ali will come here?'

(59) *Kenapa_i Siti kata [yang Fatimah beli ikan itu t_i]?*

why Siti say that Fatimah bought fish that

'Why did Siti say that Fatimah bought that fish?'

We showed in Section 2 that structures of the form

(60) $\left[\text{headless relative clause BE } \left\{ \begin{array}{l} PP \\ AdvP \end{array} \right\} \right]$

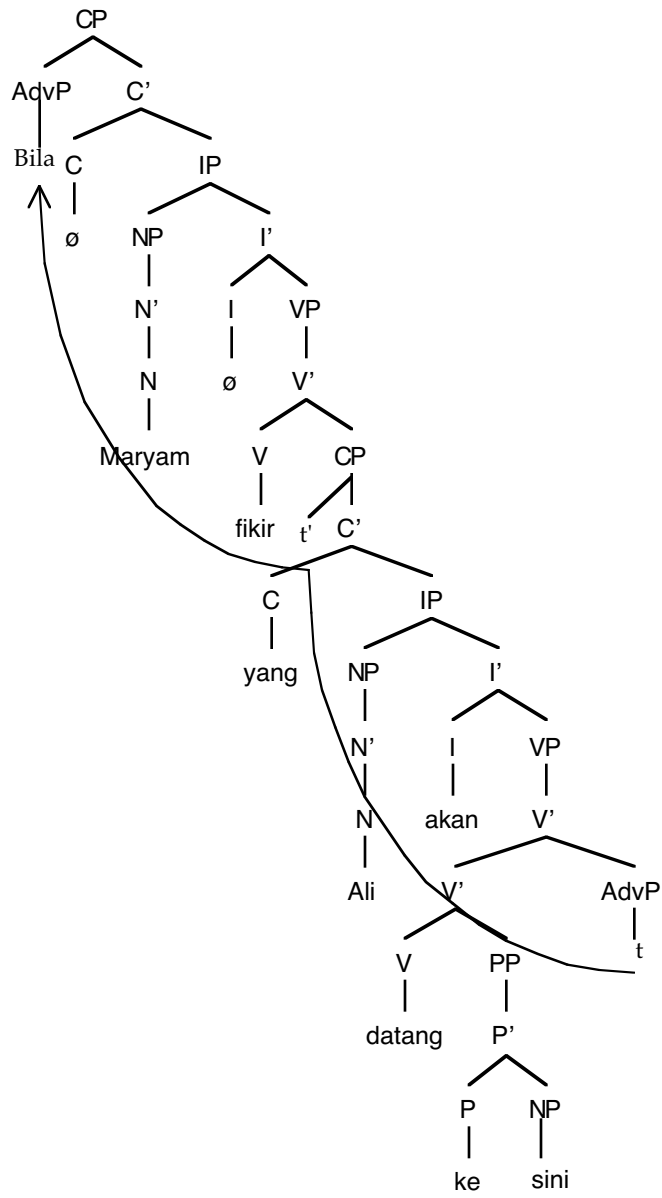
are illformed (due to the fact that headless relativization of NPs is wellformed but not of PPs and AdvPs). Thus, well-formed questions like those of (61)-(62) cannot be derived from a source like (61):

(61) $\left[\left[\begin{array}{l} +WH \\ \left\{ \begin{array}{l} +Adv \\ +prep \end{array} \right\} \end{array} \right] \left[\text{headless relative clause BE } [t] \right] \right]$

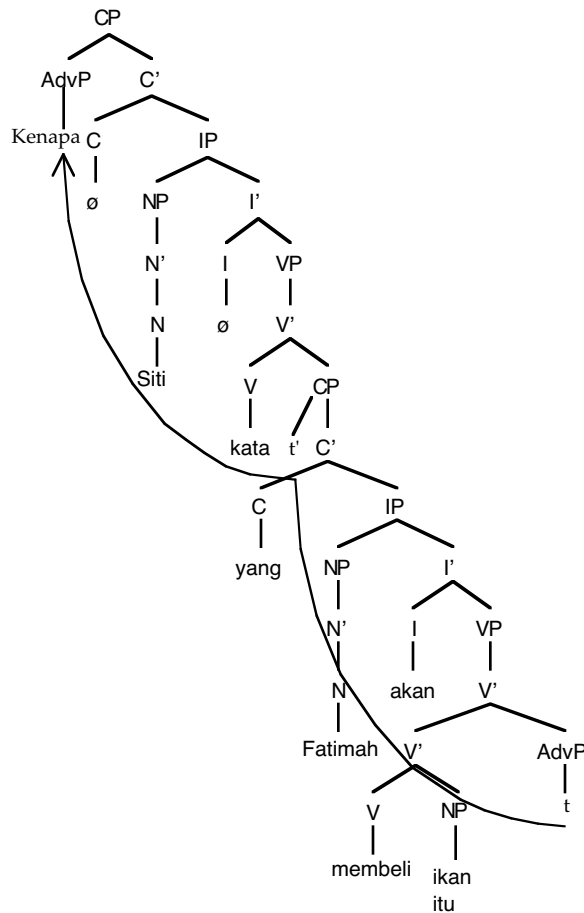
But questions like (58)-(59) are found in all varieties of Malay including Indonesian. Thus, there is reason to believe that interclausal movement of the WH word occurs across dialects.

We, therefore, conclude that (58) would have a derivation roughly like that of (62) and (59) like that of (63)

(62)



(63)



To summarize, we have argued in this section that both WH questions without *yang* like (2) and those with *yang* like (3) involve movement of WH to Spec of CP. The two question types differ in that in *yang*-less questions only the WH element moves, in a potentially long distance fashion, while in questions with *yang* the movement of the WH word is local, and the element which undergoes movement is the null operator internal to the headless relative clause.

8. Dialectal Variation

We have shown so far that in EM WH questions like (3) are derived from a headless relative clause structure like (6) rather than from a structure like (4). We

would like to turn now to a number of ways in which CSM differs from EM and assess the significance of these differences for the analysis of WH questions in that dialect. There are three differences between CSM and EM that are relevant to the analysis of *yang* questions:

1) In CSM the complementizer *bahawa* has been replaced by *yang* and Ø (true also of EM but not of FM).

2) CSM questions with the structure of (64)

(64)

$$\left[\left[\begin{array}{l} +WH \\ \left\{ \begin{array}{l} +Adv \\ +prep \end{array} \right\} \end{array} \right] \left[\text{headless relative clause BE } [t] \right] \right]$$

like those of (10'')-(14'') are grammatical:

(10'') *Di Mana Yang* in CSM

Di mana yang kau tinggal?

at where that you stay

'Where do you live?'

(11'') *Ke Mana Yang* in CSM

Ke mana yang kau pergi?

to where that you go

'Where are you going?'

(12'') *Bagaimana Yang* in CSM

Bagaimana yang Ali baiki kereta itu?

how that Ali fix car that

'How did Ali fix that car?'

(13") *Macam Mana Yang* in CSM

Macam mana yang awak buat itu?

how that you do that

'How did you do that?'

(14") *Kenapa Yang* in CSM

Kenapa yang Ali dipecat?

Why that Ali was fired

'Why is it that Ali was fired?'

but headless relative clauses in which the relativized element occurs in positions analogous to that of the question words in (10")-(14") are illformed.

3) The constraint against the placement of the focus marker *kah* on subjects described in Section 6 does not hold for null copula sentences in CSM. Thus, while the distinction between (39) and (40) (repeated)

(39) Fatimah kata [Siti membeli [**apakah**] semalam]?

Fatimah say Siti bought what-Q yesterday

'What did Fatimah say that Siti bought yesterday?'

(40) *Fatimah kata [[**siapakah**] membeli buku itu semalam]?

Fatimah say who-Q bought book that yesterday

is found in both FM and CSM, that seen in (41)-(42) holds in FM but not in CSM:

(41') *Kah* on Predicate of Null Copula Sentence in CSM

Ali rasa [Rahman **seorang yang baikkah**]?

Ali feel Rahman one-man that good-Q

'Does Ali feel that Rahman is a **good person**?'

(42') *Kah* on Subject of Null Copula Sentence in CSM

Ali rasa [[**Rahmankah**] seorang yang baik]?

Ali feel Rahman-Q one-man that good

'Does Ali feel that **Rahman** is a good person?'

We assume that EM represents the state of affairs from which Malay is changing while that in CSM represents the direction of change. The fact that these three changes occur together would be explained by the following scenario:

- 1) The complementizer *bahawa* is lost, and *yang* takes on the functions previously filled by *bahawa*.
- 2) As a result of the loss of the *bahawa-yang* distinction, the fact that *bahawa* does not occur in place of *yang* in *yang* questions like (3) no longer constitutes evidence for the HRCH and against the MCH. Thus, speakers assume that two distinct structures are available for *yang* questions, (4) and (6)/(7).
- 3) Because structure (4) has become available for *yang* questions, the construction is extended to prepositional phrases and adverbial WH questions.

Thus, the loss of *bahawa* leads, ultimately, to the adoption of a new structure for *yang* questions, that of (4).

This scenario, while speculative, consists of a plausible series of changes, and explains the differences between CSM and EM. Children's acquisition of Malay provides some empirical support for this scenario. There is evidence that in CSM *yang* questions are no longer all derived from headless relative clauses. Aman (1999) and Hermon and Aman (1999) describe the acquisition of simple and complex questions in CSM. Naturalistic data from two young children ages 2;8 – 3;2 were collected. Children show both in situ and fully moved questions, but notably lack *yang* wh-questions.

At the same time, these very young children already have some headless relative clauses using *yang* (even though the distribution of these relative clauses is quite limited), and one child produces *yang* with a focused constituent. Neither the children nor their caretakers use *yang* as a complementizer in structures other than relative clauses, and while mothers use *yang* as a relative clause complementizer in both headless and headed relative clauses, these CSM speakers do not use *yang* questions in their spontaneous speech.

In addition to the collection of naturalistic data, older children (mean age 4;2) were tested on their knowledge of *yang* questions using an elicited imitation task (in which children are asked to repeat sentences containing the syntactic structure under investigation). According to this methodology, speakers will often change structures not in their grammars to those that are. When they were presented with *yang* questions like (3), most children converted these questions into *yang*-less moved questions like (2). Our results were corroborated by researchers in Malaysia (Tye 1999), who found that in naturalistic data from children of this age (age 4;6) collected in Kuala Lumpur, Malaysia, *yang* questions are missing. At the same time, the Malaysian children use *yang* in headless relative clauses quite frequently, as exemplified below:

(65) Ah nanti malam adik dia yang masak

AH later night younger.sibling YANG cook

'Later at night, the younger sibling is the one who will cook.'

(66) Awaklah yang basuh

you-LAH YANG wash

'You are the one to wash it.'

We assume that young children are primarily acquiring the colloquial form of the language. The evidence described shows that in the colloquial language the knowledge of relative clauses with *yang* does not entail the knowledge of *yang* questions. Thus, contrary to the situation in EM, it appears that *yang* relative clauses and *yang* questions are no longer two instances of the same construction in Colloquial Malay. This corroborates partially the scenario described earlier in this section.

9. Conclusions

We have shown in this paper that WH questions employing overt WH movement in EM are derived from two very different structures. Those without the complementizer *yang* are similar to WH questions in English and other European languages: The WH word moves successive cyclically to scopal Spec CP. Those with *yang*, however, are derived from a very different structure in which the matrix clause is a null copula sentence, the subject of which is a headless relative clause.

There is suggestive evidence that the distinction between these two structures is being lost in Colloquial Singapore Malay. We propose that the loss of the distinction between the complementizers *yang* and *bahawa* has led to a reanalysis of *yang* questions as having a structure like (4) in which *yang* is a main clause complementizer rather than a complementizer introducing a headless relative clause.

Footnotes

* This research was supported in part by the National Science Foundation (grants SBR-9121167, SBR-9729519 and INT-9423291). We would like to thank David Gil and Mashudi Kader for their suggestions and comments on the ideas expressed in this paper.

¹ By "educated informal Malay" we mean the Malay spoken by educated speakers (e.g. graduates of secondary and tertiary school) when they are attempting to speak "correct" Malay. EIM differs from Formal Malay (FM) in that speakers of EIM often use grammatical constructions that are considered "incorrect" or "colloquial" in the more formal register. The register differs from Colloquial Malay (CM) in a variety of features including the choice of lexical items, the frequent use of the complementizer *yang* and the frequent use of WH fronting rather than WH in situ. The particular characteristics of EIM discussed here may be specific to the speech of Malays from Singapore and surrounding urban areas in Malaysia.

² Presumably the IP or small clause is contained within a CP. Evidence for this is the fact that when null copula sentences are themselves the complement of a higher verb, they can be introduced by a complementizer:

(i) Siti percaya [yang [Ali pencuri]].

Siti believes that Ali thief

'Siti believes that Ali is a thief.'

³ See Section 6.

⁴ Structure (6) was proposed by Kader (1976, 1980). His main argument for (6) is summarized in Section 6. Variants on (7) have been proposed by Alsagoff (1992), Cheng (1992) and Martohardjono (1993). See the discussion in Section 7.

⁵ The standard arguments that *that* in relative clauses in English is not a relative pronoun but rather a complementizer carry over to Malay. For instance, as is shown in (i),

- (i) a. the book about which you told me
- b. the book that you told me about
- c. *the book about that you told me

pied piping is possible with relative pronouns (ia), but not with *that* (ib-c). This is explained by the hypothesis that *which* is a relative pronoun but *that* is a complementizer. The preposition *about* can optionally move to clause initial position together with the relative pronoun (i.e. the preposition is optionally pied piped by the relative pronoun). The complementizer *that* cannot pied pipe the preposition because, unlike the relative pronoun, the complementizer is base generated in clause initial position.

Similar facts are found in Malay:

- (ii) a. buku yang kau bilang saya
 book that you tell me
 'the book that you told me about'
- b. *buku tentang yang kau bilang saya
 book about that you tell me
 'the book that you told me about'

Malay differs from English in that no relative pronoun option exists for subject and object relativization. Thus, there is no Malay equivalent to (ia). It follows from the ungrammaticality of pied piping in (iib) that *yang* is a base generated complementizer rather than a relative pronoun.

⁶ Speakers find (15)-(19) somewhat worse than similar sentences in which the adverbial or WH phrase appears before the headless relative clause, e.g. (18) is worse than (i):

(i) ?? Dengan hati-hati yang Ali baiki kereta itu.

with heart-heart that Ali fix car that

'With care is how Ali fixed the car.'

Since sentence with the word order given in (i) is also illformed, the reduced grammaticality of (15)-(19) cannot be attributed to the order of the AdvP/PP and the headless relative clause.

⁷ The symbol "@" indicates "unsuitable as an answer to to this question". In an appropriate context, the sentence is grammatical. See below.

⁸ The question remains why questions employing headless relative clauses can be answered employing a structure that does not employ a headless relative clause. Note, however, that the same pattern is observed in English:

(i) a. Who bought the book?

b. Jon bought it.

c. @John is the one who bought it.

(ii) a. Who is the one that bought the book?

b. Jon bought it.

c. John is the one that bought it.

⁹ We postpone the choice between (6) and (7) to a later section.

¹⁰ For many Singapore speakers of EIM, the asterisk overstates the degree of illformedness of the question. However, the same general pattern holds: *Kah* on subjects is degraded in comparison with *kah* on other constituents.

¹¹ This is a point at which speakers of colloquial Singapore Malay part company from speakers of Formal Malay. In Colloquial Singapore Malay the

appearance of *kah* on subjects in verbal sentences like (40)-(43) reduces significantly the grammaticality of the sentence. In the case of nominal null copula sentences like (44)-(45), however, the expected reduction in grammaticality does not occur. This dialect variation will be discussed below.

References

- Aman, Norhaida. 1999. How to Ask What in Malay: the Acquisition of Wh Questions in Singapore Malay. Ph.D. Dissertation, University of Delaware.
- Cheng, Lisa. 1991. *On the Typology of Wh-Questions*. PhD dissertation, MIT.
- Cole, Peter and Gabriella Hermon. 1998. The Typology of Wh-Movement: Wh Questions in Malay, *Syntax*: 1: 221-258.
- Cole, Peter and Gabriella Hermon. 1995. Is Wh-in-situ really In-situ - Evidence from Malay and Ancash Quechua, Proceedings from the 1994 West Coast Conference on Formal Linguistics, Stanford, CA, CSLI.
- Hermon, Gabriella & Norhaida Aman. 1999. When an Island is Not an Island: Long-Distance Questions in Singapore Malay. *Proceedings of the 23rd Annual Boston University Conference on Language Development*, Cascadilla Press.
- Huang, C.T. James. 1989. Pro-Drop in Chinese: A Generalized Control Theory, in Jaeggli, O. (ed.); Safir, K. (ed.) *The Null Subject Parameter*. Dordrecht : Kluwer.
- Kader, Mashudi. 1976 (1980). *The Syntax of Malay Interrogatives*, PhD. dissertation.

- Martohardjono, Gita. 1993. *Wh-Movement in the Acquisition of a Second Language: A Cross-Linguistics Study of 3 Languages with and without Overt Movement*. Doctoral Dissertation, Cornell University, Ithaca, NY.
- Rogayah, A. Razak. 1995. *The Syntax and Semantics of Quantification in Malay- A Government and Binding Approach*, PhD. dissertation.
- Saddy, Douglas. 1991. Wh Scope Mechanisms in Bahasa Indonesia, in L. Cheng and H. Demirdash (eds.), *More Papers on Wh-Movement*, MITWPL 15.
- Saddy, Douglas. 1992. A versus A-bar Movement and WH Fronting in Bahasa Indonesia, manuscript, University of Queensland, Australia.