

How Irregular is WH in Situ in Indonesian?

Peter Cole, Gabriella Hermon, and Yassir Nasanius Tjung

University of Delaware and Max Planck Institute for Evolutionary Anthropology

Peter Cole: pcole@udel.edu

Department of Linguistics
University of Delaware
46 East Delaware Avenue
Newark, Delaware
U.S.A. 19716-2551

Gabriella Hermon: gaby@udel.edu

Department of Linguistics
University of Delaware
46 East Delaware Avenue
Newark, Delaware
U.S.A. 19716-2551

Yassir Tjung: yns@udel.edu

University of Delaware
46 East Delaware Avenue
Newark, Delaware
U.S.A. 19716-2551

Abstract

How Irregular is WH in Situ in Indonesian?

Contemporary approaches to Generative syntax lead to the expectation that WH-in-situ would be subject to few distributional restrictions, but, a series of complex constraints apply to in-situ WH in subject position in Standard Indonesian. We argue that these distributions do not follow from universal grammar per se, but, rather from a constraint on the relationship between syntax and information structure.

We then turn to Colloquial Jakarta Indonesian, a dialect similar to Standard Indonesian with regard to grammatical restrictions on WH-in-situ, but lacking the constraint on the relationship between syntax and information structure found in Standard Indonesian. We contend that the seeming differences between the grammars of Standard Indonesian and Jakarta Indonesian do not reflect differences in grammar but rather in how the dialects relate formal grammar and pragmatics.

How Irregular is WH in Situ in Indonesian?*

Peter Cole, Gabriella Hermon, and Yassir Nasanius Tjung

University of Delaware and MPI EVA Leipzig and Jakarta Field Station

1 Introduction

Contemporary approaches to Generative syntax (e.g. Principles and Parameter Theory (Chomsky 1981, 1986) and Minimalism (Chomsky 1995, 1998, 1999)) would lead to the expectation that WH in situ sentences like (1) in Standard Indonesian (SI) would be subject to few restrictions and constraints:

(1) Siti mau apa?

Siti want what

'What does Siti want?'

In (1), an information question, the SVO word order of the question mirrors that of a declarative sentence:

(2) Siti mau buku ini.

Siti want book this

'Siti wants this book.'

Recent work on the analysis of WH in situ (e.g. Cole and Hermon 1995, 1998, 2000 inter alia) points to the conclusion that in situ WH forms are in situ throughout the derivation. Thus, it would not be expected that in situ WH forms would be subject to special constraints and restrictions on their distribution (e.g. constraints on movement). But, as we will show in section 2, a series of complex and seemingly arbitrary constraints apply to in situ WH subjects in Indonesian. We will argue that the distributions observed do not follow from any known principles of universal grammar, and, indeed, seem mutually contradictory. Thus, they constitute a challenge to one of the fundamental working hypotheses of Generative Grammar, that seeming complexity in the syntax of languages will be reducible to "noise" caused by the interaction of simple, universal syntactic principles.

To put the matter differently, current Generative approaches like the Minimalist Program take it as given that languages do not differ in having different syntactic rules and constraints. Rather, seeming differences in the syntax of languages are expected to evaporate once the languages are examined at an appropriate level of abstraction. In contrast, typological linguists and others of an empirical bent (e.g. Comrie 1989 and many others) believe that there is considerably greater diversity among languages than is generally acknowledged by Generative linguists, and that languages can, indeed, differ in a variety of ways including the constraints and restrictions on the in situ appearance of question words.

In this paper, we shall examine the complex conditions on WH in situ in Indonesian. We shall argue, on the one hand, that these conditions provide support for the typologist's view that there is considerable diversity in the syntactic patterns manifested by languages. But, on the other hand, we will also argue that the diversity in this case is not internal to the syntax (taking "syntax" in the narrow and technical sense usual in Minimalist works), but rather derives from the interaction of purely syntactic requirements with those of pragmatics (specifically, information structure). We conclude that both the Generative and the typological positions are correct (though incomplete), and that, surprisingly, these positions are considerably more compatible than might be expected. While the typological approach is correct in claiming considerable diversity among languages in the rules determining the distribution of words in sentences ("syntax" in the less technical sense), the Generative claim, that, the core syntactic principles are universal also seems to be correct. That is, the distribution of WH in situ in Indonesian will be seen to follow from the interaction between a simple, constrained and (potentially) universal formal syntax and language specific conditions on the relationship between syntax and information structure.

This paper will be organized as follows. We shall first examine the restrictions on WH in situ in Standard Indonesian, and will show that, as far as we can tell, the restrictions are not explained by the principles of universal grammar, but, rather by a constraint on the relationship between syntax and information structure. We will then turn to Colloquial Jakarta Indonesian, a

variety heard not only on the streets of Jakarta but also on Indonesian radio and television. We shall show that CJI is similar to SI with regard to grammatical restrictions on WH in situ, but this variety of Indonesian lacks the constraint on the relationship between syntax and information structure found in SI.

2 The Distribution of WH Argument Words in SI

In this section we lay out in some detail the distribution of WH argument words in three types of sentences, (i) verbal sentences, (ii) sentences with the complementizer *yang*, and (iii) nominal sentences. Our purpose is to present a clear picture of the constraints on WH in situ.

2.1 Questioning in Verbal Sentences

As seen in (3), WH in situ is grammatical in direct object position in both root and embedded clauses in SI.

(3) a. **WH in Direct Object Position**

Siti akan membeli apa?

Siti FUT MEN-buy what

'What will Siti buy?'

b. Saya tidak tahu Siti akan membeli apa?

1SG NEG know Siti FUT MEN-buy what

'I don't know what Siti will buy.'

The examples in (4) show that question words in SI can also appear in situ in a variety of other positions in verbal sentences.

(4) a. **WH as Possessor**

Ani sudah mengetik laporan siapa?

Ani PAST MEN-type report-AN who

'Whose report did Ani type?'

b. **WH as Indirect Object**

Ayah harus mengirim siapa uang setiap bulan?

Father must MEN-send-I who money every month

'Who must Father send money to every month?'

c. **WH as Object of Preposition**

Kamu akan membayar hutangmu dengan apa?

2SG FUT MEN-pay debt-2SG with what

'How will you pay your debt?'

Although WH in situ is grammatical in a wide variety of contexts, WH in situ in the subject position of a verbal sentence in SI unexpectedly renders the sentence ungrammatical.

(5) a. *Siapa akan menjadi wasit pertandingan itu?

who FUT MEN-become referee PER-match-AN that

'Who will be the referee of that match?'

- b. *Boleh saya bertanya [siapa akan menjadi wasit pertandingan itu]?
 may 1SG BER-ask who FUT MEN-become referee PER-match-AN that
 'May I ask who the referee of that match will be?'
- c. *Apa membuatmu demikian gembira hari ini?
 what MEN-make-2SG so happy day this
 'What makes you so happy today?'
- d. *Saya heran [apa membuatmu demikian gembira hari ini].
 1SG surprised what MEN-make-2SG so happy day this
 'I wonder what makes you so happy today.'

The examples in (5) involve active sentences, those marked by the prefix *MEN-* on the verbs. The following examples show that the constraint against WH in situ in subject position in SI also applies to the canonical passive sentences, those marked by the prefix *DI-* on the verb, and to the bare passive sentences, passives in which the verb appears without a voice prefix.

- (6) a. *Siapa dirampok tadi malam?
 who DI-rob just.now night
 'Who was robbed last night?'
- b. *Boleh saya bertanya [siapa dirampok tadi malam]?
 may 1SG BER-ask who DI-rob just.now night
 'May I ask who was robbed last night?'
- (7) a. *Apa harus kamu masak untuk dia hari ini?
 what must 2SG cook for 3SG day this
 'What must be cooked by you for him today?/What must you cook for him today?'
- b. *Saya ingin tahu [apa harus kamu masak untuk dia hari ini]?
 1SG want know what must 2SG cook for 3SG day this
 'I want to know what must be cooked by you for him today./I want to know what you must cook for him today.'

As seen in the examples (5) - (7) above, the subject position **cannot** be filled by a WH word in verbal sentences.

2.2 Questioning in Sentences with *Yang*

The ill-formed sentences in (5) - (7) become grammatical if the complementizer *yang* follows the WH word in SI. Notice that the sentences in (5) - (7) are, on the surface, very similar to those in the following examples.

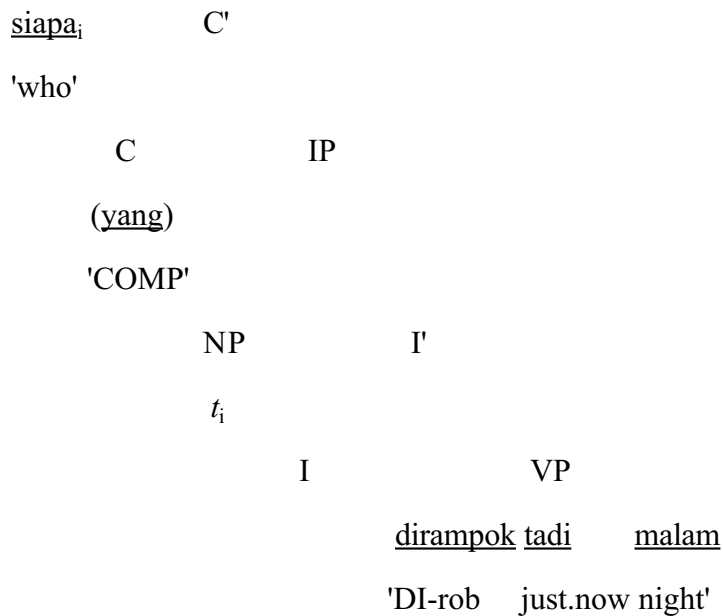
- (8) a. **Siapa yang** akan menjadi wasit pertandingan itu?
who COMP FUT MEN-become referee PER-match-AN that
'Who will be the referee of that match?'
- b. Boleh saya bertanya **siapa yang** akan menjadi wasit pertandingan itu?
may 1SG BER-ask who COMP FUT MEN-become referee PER-match-AN
that
'May I ask who the referee of that match will be?'
- c. **Apa yang** membuatmu demikian gembira hari ini?
what COMP MEN-make-2SG so happy day this
'What makes you so happy today?'
- d. Saya heran **apa yang** membuatmu demikian gembira hari ini.
1SG surprised what COMP MEN-make-2SG so happy day this
'I wonder what makes you so happy today?'

- (9) a. Siapa yang dirampok tadi malam?
 who COMP DI-rob just.now night
 'Who was robbed last night?'
- b. Boleh saya bertanya siapa yang dirampok tadi malam?
 may 1SG BER-ask who COMP DI-rob just.now night
 'May I ask who was robbed last night?'
- (10) a. Apa yang harus kamu masak untuk dia hari ini?
 what COMP must 2SG cook for 3SG day this
 'What must be cooked by you for him today?'
- b. Saya ingin tahu apa yang harus kamu masak untuk dia hari ini.
 1SG want know what COMP must 2SG cook for 3SG day this
 'I want to know what must be cooked by you for him today.'

Even though the surface form of the sentences in (8) - (10) is very similar to that in (5) - (7), *yang* questions are grammatical with the WH word in sentence initial position in SI. This appears surprising, because it might appear that (5) - (7) differ from (8) - (10) only with respect to whether the matrix complementizer position is filled:

- (11) a. *Siapa dirampok tadi malam?
 who DI-rob just.now night
 'Who was robbed last night?'
- b. Siapa yang dirampok tadi malam?
 who COMP DI-rob just.now night
 'Who was robbed last night?'

(12) CP



If structure (12) is correct, it is mysterious why (5) - (7) are ungrammatical but (8) - (10) are well-formed. We have argued, however, in Cole and Hermon (to appear) that the structure of WH questions with *yang* is quite different from that of questions without *yang* and that (12) is incorrect. This question is discussed extensively below.

2.3 Questioning in Nominal Sentences

Let us now turn from verbal sentences to nominal sentences, those which have two NPs as their immediate constituents.

- (13) a. [IP [NP **Siapa**] [NP Presiden RI keempat]]?
 who president RI KE-four
 'Who is the fourth president of RI (Republic of Indonesia)'

- b. [[NP **Apa**] [NP jawabanmu atas pertanyaannya]]?
 what reply-AN-2SG for PER-ask-AN-NYA
 'What is your reply to his question?'

Unlike verbal sentences, in nominal sentences the questioning of subjects without *yang* appears to be grammatical, as shown in (13). Furthermore, *yang*, unexpectedly, is not allowed to follow the WH words in nominal sentences:

- (14) a. ***Siapa yang** Presiden RI keempat?
 who COMP president RI KE-four
 'Who is the fourth president of RI (Republic of Indonesia)?'
- b. ***Apa yang** jawabanmu atas pertanyaannya?
 what COMP reply-AN-2SG for PER-ask-AN-NYA
 'What is your reply to his question?'

The sentences of (14) show that, unlike verbal sentences (8) - (10), *yang* cannot follow WH words in nominal sentences.

3 Why WH in situ in Subject Position is Ungrammatical in SI

The distribution described in the previous section appears to be internally inconsistent. WH in subject position is banned in verbal sentences; however, it is allowed in nominal sentences. Furthermore, the presence of the complementizer, *yang*, appears to repair the ungrammaticality of subject questions in verbal sentences, while making ungrammatical otherwise well-formed subject questions in nominal sentences. In this section, we shall first show that a seemingly likely grammatical explanation does not appear to provide a solution to this puzzle. We shall then show that the seemingly contradictory restrictions on WH in situ in subject position in SI turn out not to be contradictory at all. The solution to this puzzle will be built upon three crucial claims: (a) In SI there is a requirement of parallelism between syntactic structure and information structure; (b) although *yang* sentences like (8) - (10) appear to be verbal sentences, a closer examination reveals that they are, in fact, nominal sentences; and (c) the grammaticality of apparent subject questions in nominal sentences like (8) - (10) and (13) is due to a general rule that optionally

moves focused predicates to initial position, and such sentences are not instances of either WH movement or of WH in situ in subject position.

3.1 An ECP Account of the Constraints on in Situ WH Subjects

It might be proposed that the constraints on WH in situ in SI are *that t* facts and are due to the effects of the ECP (or to whatever principles in current grammatical theory turn out to account for the range of facts accounted for by the ECP in the Principles and Parameters framework).

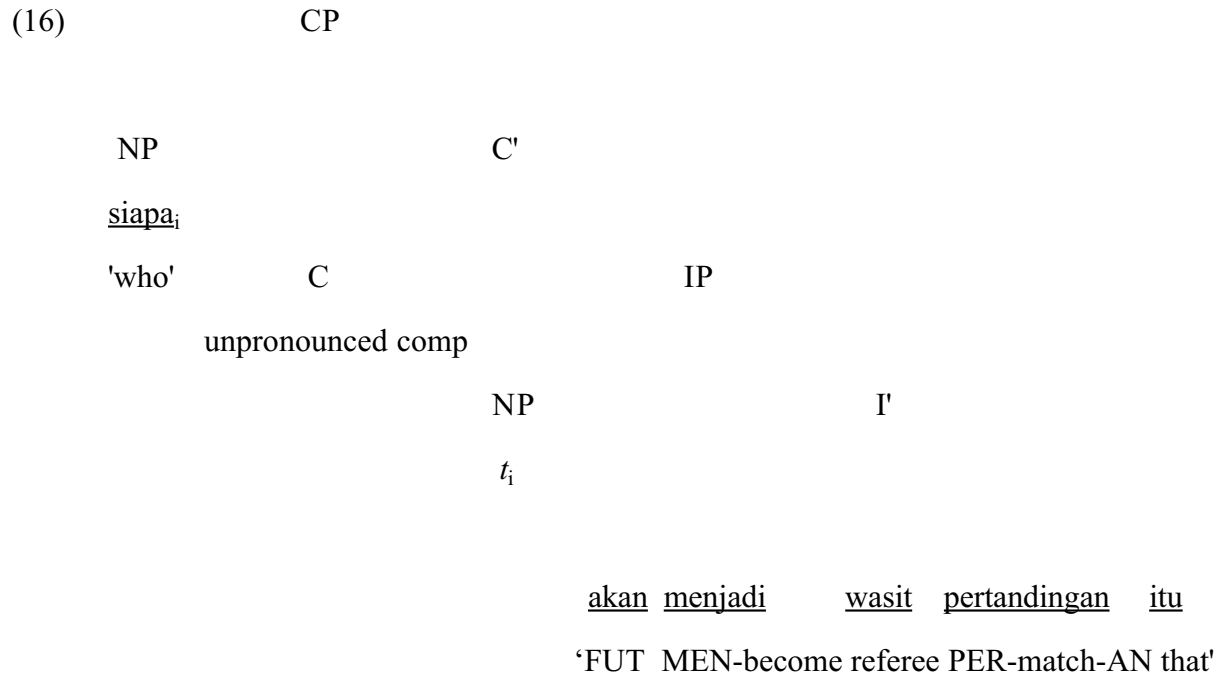
Under such an analysis, it would be claimed that in a sentence like (5a), repeated as (15) below,

(15) *Siapa akan menjadi wasit pertandingan itu?

who FUT MEN-become referee PER-match-AN that

'Who will be the referee of that match?'

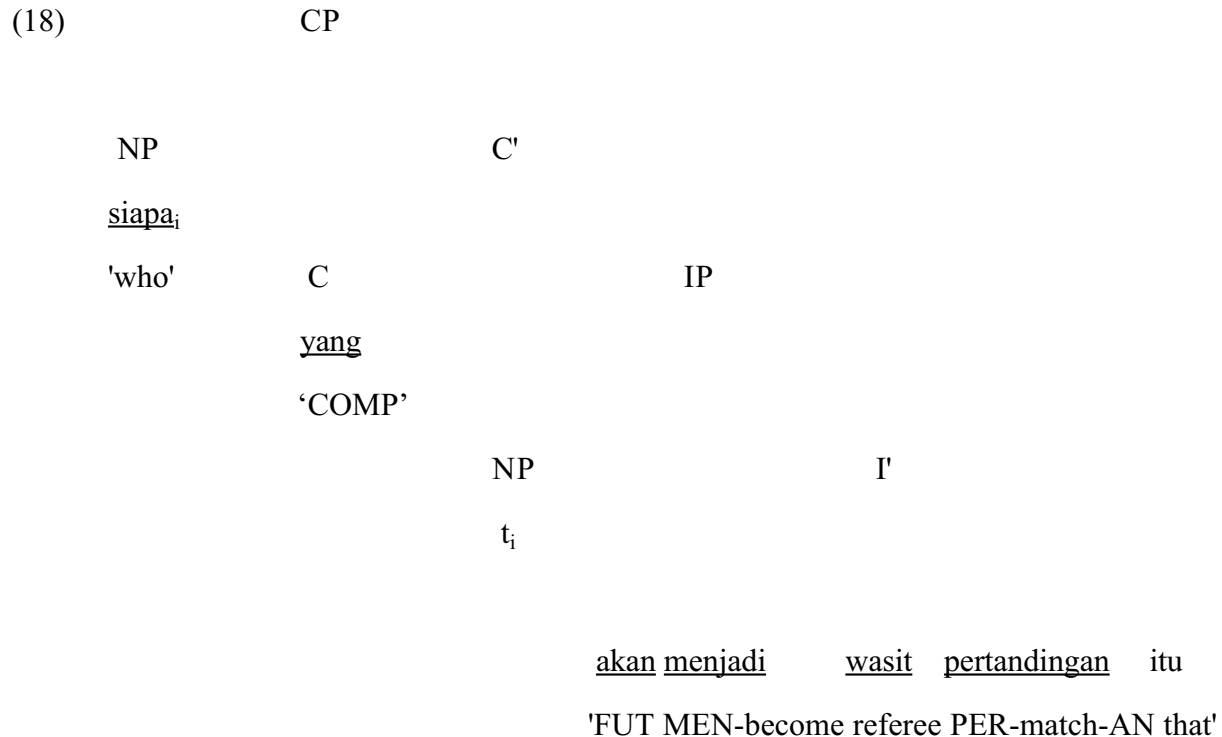
the question word *siapa* is extracted from the IP at LF. The extraction, however, is blocked because a phonologically null complementizer occurs in the matrix comp position, resulting in the ill-formed sequence *complementizer t*.



There are, however, a number of arguments against such an analysis. First, while it provides an account for (5) - (7), it provides no explanation for the remaining data in section 2: It does not explain why sentences like (8a), repeated below as (17), should be grammatical.

- (17) **Siapa yang** akan menjadi wasit pertandingan itu?
 who COMP FUT MEN-become referee PER-match-AN that
 'Who will be the referee of that match?'

Under the *complementizer t* analysis, (17) would have the same structure as (15) except that that the complementizer in (17) is pronounced as *yang*:



Thus, it would be expected that sentences like (17) would be ungrammatical. They are, however, well-formed, suggesting that the ECP approach could not be correct.

Secondly, the *that t* analysis would not explain why in nominal sentences WH in situ in subject position is well formed **without** the *yang* complementizer, but ungrammatical when *yang* appears (the reverse of the pattern seen in verbal sentences).

Thirdly, there has been considerable work showing that WH in situ in Indonesian does not involve movement at any level. For example, in Saddy (1991, 1992), as well as Cole and Hermon (1995, 1998, 2000), it is shown that in relativization in Indonesian the movement of the relative operator out of a Ross Island (Ross, 1967) is ungrammatical. In contrast, WH in situ inside Ross islands is well-formed. This suggests that WH in situ in Indonesian does not involve covert movement. But if WH in situ in Indonesian does not involve movement, the *that t* account could not be correct, since the structure of sentences like (15) would not be (16), but rather like (19), in which the WH word is never extracted from the IP:



C

IP

e

siapa akan menjadi wasit pertandingan itu

'who FUT MEN-become referee PER-match-AN that'

Thus, while a *that t* account of the facts in section 2 may seem initially promising, such an account does not appear to be the correct explanation for the complex distribution of WH in situ in Indonesian. In the absence of alternative grammatical accounts for these facts, we shall now turn to what we consider a more promising approach to the problem.

3.2 The Parallelism Hypothesis (PH)

Turning to information structure, all languages provide formal devices to distinguish old information from new information, topic from focus, etc. (Lambrecht 1994, Van Valin and LaPolla 1997, and many others). There is, in all languages, a high correlation between information theoretic constructs like topic and focus, and grammatical constructs like subject and predicate. For instance, the prototypical subject in English is also the topic, while the prototypical predicate is the focus. We contend that in SI, the crosslinguistically prototypical situation has become the rule: In SI, the subject **must** be the topic and cannot be the focus of the sentence. That is, SI requires a strict parallelism between topic and subject, a parallelism not required in such languages as English.

Our argument for positing the Parallelism Hypothesis (PH) is based on Soemarmo's (1970) study of three types of sentential constructions in Indonesian and Javanese, Mashudi's (1976) work on the syntax of interrogatives in Formal Malay, and Poedjosoedarmo's (1977) work on thematization and information structure in Javanese. Turning first to Mashudi's work, he observed that in Formal Malay the most novel information within a sentence (the focus) can only occur in the predicate phrase; it cannot occur in, or be, the subject. His evidence for this claim is based on the distribution of the focus morpheme (FM) *-lah* and its interrogative counterpart

(IFM) *-kah*. A constituent which is followed by *-lah* or *-kah* must not be a subject, as shown in the Malay examples in (20) - (27) from Mashudi.

(20) ***Abulah** [_{VP} belajar di bilik itu tadi].

Abu-FM study in room that just.now

'It was Abu who was studying in that room just now.'

(21) ***Abukah** [_{VP} belajar di bilik itu tadi]?

Abu-IFM study in room that just.now

'Was it Abu who was studying in that room just now?'

(22) Abu [_{VP} belajar**lah** di bilik itu tadi].

Abu study-FM in room that just.now

'It was studying that Abu did in that room just now.'

(23) Abu [_{VP} belajar**kah** di bilik itu tadi]?

Abu study-IFM in room that just.now]

'Was it studying that Abu did in that room just now?'

(24) Abu [_{VP} belajar di bilik itulah tadi].

Abu study in room that-FM just.now

'It was in that room that Abu was studying just now.'

(25) Abu [_{VP} belajar di bilik itukah tadi]?

Abu study in room that-IFM just.now

'Was it in that room that Abu was studying just now?'

(26) Abu [_{VP} belajar di bilik itu tadilah].

Abu study in room that just.now-FM

'It was just now that Abu was studying in that room.'

(27) Abu [_{VP} belajar di bilik itu tadikah]?

Abu study in room that just.now-IFM

'Was it just now that Abu was studying in that room?'

As seen in (20) - (27) above, the focus morpheme *-lah* and its interrogative counterpart *-kah* can be attached freely to any constituent except the subject. This fact suggests that there exists a strict parallelism between syntactic and information structure. Similar proposals were made by Soemarmo (1970) for Indonesian and by Poedjosoedarmo (1977) for Javanese.

(28) **The Parallelism Hypothesis (PH)**

The focus or new information must occur in the predicate and the subject must be the old information (topic).

3.3 The PH and Verbal Questions

Let us now see how the PH accounts for the ungrammatical verbal sentences in (5) - (7). Consider the WH question in (5a), repeated as (29). The structure of (29) is given in (30).

(29) *Siapa akan menjadi wasit pertandingan itu?
 who FUT MEN-become referee PER-match-AN that
 Who is the referee of that match?'

(30) CP
 C'
 C IP
 NP [+topic] I'
siapa
 'who' I VP
akan V'
 'FUT'
 V NP
menjadi wasit pertandingan itu
 'MEN-become referee PER-match-AN that'

As seen in (29), the WH word *siapa* occurs in the spec of IP. According to the PH, the spec of IP (subject) is a position reserved for the topic, thus correctly predicting that (29) is ill-formed: All WH words have inherent focus and are therefore banned from occupying the spec of IP. The same restriction also applies to the WH questions in (5b, c, d) - (7): They are all ungrammatical because the WH words in these sentences occur in the spec of IP.

3.4 The PH and *Yang* Questions

Our analysis for *yang* questions is based on the claim that these questions are in fact nominal sentences, namely, [IP [NP] [NP]], rather than verbal sentences [IP [NP] [VP]]. This claim is argued for at length in Cole, Hermon, and Norhaida (to appear) with respect to *yang* questions in Malay. We shall claim that this analysis applies to Indonesian as well as to Malay.

Cole, Hermon and Norhaida contend that Malay WH questions employing *yang* have a structure quite different from those without *yang*. It is argued, following Mashudi (1976), that in

sentences like (31), the complementizer introduces a headless relative clause, and that *yang* is not a main clause complementizer.

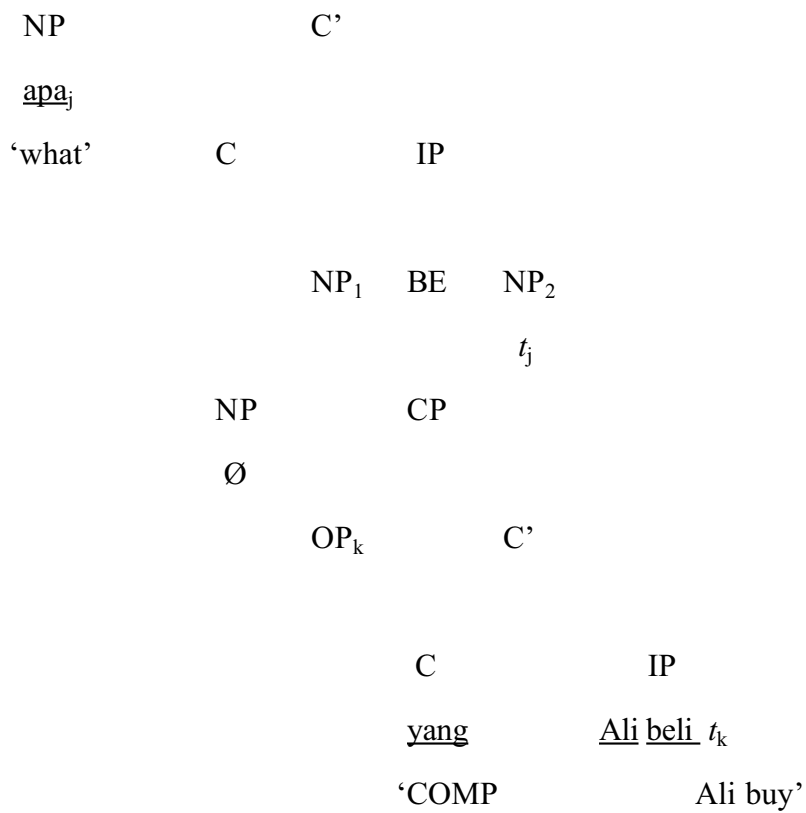
(31) Apa yang Ali beli?

what COMP Ali buy

'What did Ali buy?'

According to the Headless Relative Clause Hypothesis, questions like (31) are nominal sentences that are comprised of an IP (or, perhaps, a small clause) containing two NPs. In (31), the two NPs, a headless relative clause and the WH word *apa* 'what', are related in the structure as follows:¹

(32) CP



Returning to *yang* questions in SI, we argue that *yang* questions should be analyzed as having a structure similar to that of their Malay counterpart in (32). For example, sentence (33) has the structure in (34).

(33) Yang menghancurkan desa mereka apa?
COMP MEN-destroy-KAN village 3PL what
'What destroyed their village?'

(34)	IP				
	NP ₁	NP ₂			
		<u>apa</u>			
	NP	CP	‘what’		
	Ø				
	OP _k	C’			
		C	IP		
		<u>yang</u> t _k	menghancurkan	desa	mereka
		‘COMP	MEN-destroy-KAN	village	3PL’

Structure (34) conforms to the PH. The PH predicts that the topic NP will be the subject, namely NP₁ and the focus NP will be the predicate, namely NP₂. Thus, structure (34) does not violate the required parallelism of syntax and information structure imposed by the PH.

According to (34), *yang menghancurkan desa mereka* in (33) is simply a headless relative clause (hence an NP) and therefore should show the same distributional restrictions shown by nominal constituents in general. It is a distributional peculiarity of nominal relative clauses in SI that they can only relativize an argument of a verbal sentence (e.g. (36)), not of a nominal sentence (e.g. (35)).

- (35) *_{[IP [NP Bapakku] [NP e_i] [CP OP_i yang [IP [NP t_i] [NP guru]]]]].}
- | | | |
|------------|------|---------|
| father-1SG | COMP | teacher |
|------------|------|---------|
- 'My father is (the one) who is a teacher.'

- (36) [IP [NP Bapakku] [NP e_i [CP OP_i yang [IP [NP t_i] [VP menjadi guru]]]].
 father-1SG COMP MEN-become teacher
 'My father is (the one) who is a teacher.'

Let us return to the ungrammatical examples in (14a - b), repeated below as (37):

- (37) a. *Siapa yang Presiden RI keempat?
 who COMP president RI KE-four
 'Who is the fourth president of RI (Republic of Indonesia)?'
- b. *Apa yang jawabanmu atas pertanyaannya?
 what COMP reply-AN-2SG for PER-ask-AN-NYA
 'What is your reply to his question?'

We argue that the ill-formedness of the sentences of (37) is due to the distributional restriction on nominal relative clauses seen in examples (35) and (36). In other words, they are ill-formed because the headless relative clauses in those two sentences relativize the arguments of nominal sentences, which results in the ungrammaticality like that in (35).

3.5 The PH and Nominal Questions

We indicated earlier that, according to our analysis, *yang* questions and nominal questions have the same overall syntactic structure. Thus, a WH question like (38), has the structure shown in (39).

- (38) Pencurinya siapa?
 PEN-steal-NYA who
 'Who is the thief?'

(39) IP

NP₁ NP₂
pencurinya siapa
'PEN-steal-NYA' 'who'

Similar to the account given for *yang* questions, the PH also holds of nominal questions like (38). According to the PH, the topic NP must appear in the subject position, and the focus in the predicate position. As shown in (39), NP₁ is the subject and NP₂ the predicate. Therefore, (39) is in conformity with the PH, thereby predicting the grammaticality of (38).

3.6 Predicate Movement

We have shown in Sections 3.4 and 3.5 that *yang* questions and nominal questions are both in fact nominal sentences, and that questions formed from nominal sentences presented in these sections do not violate the requirement that the subject be a topic. There are, however, also nominal questions which are grammatical despite the fact that the WH word gives the appearance of being in subject position, as shown in (8a), repeated as (40), and (13b), repeated as (41).

(40) Siapa yang akan menjadi wasit pertandingan itu?
who COMP FUT MEN-become referee PER-match-AN that
'Who will be the referee of that match?'

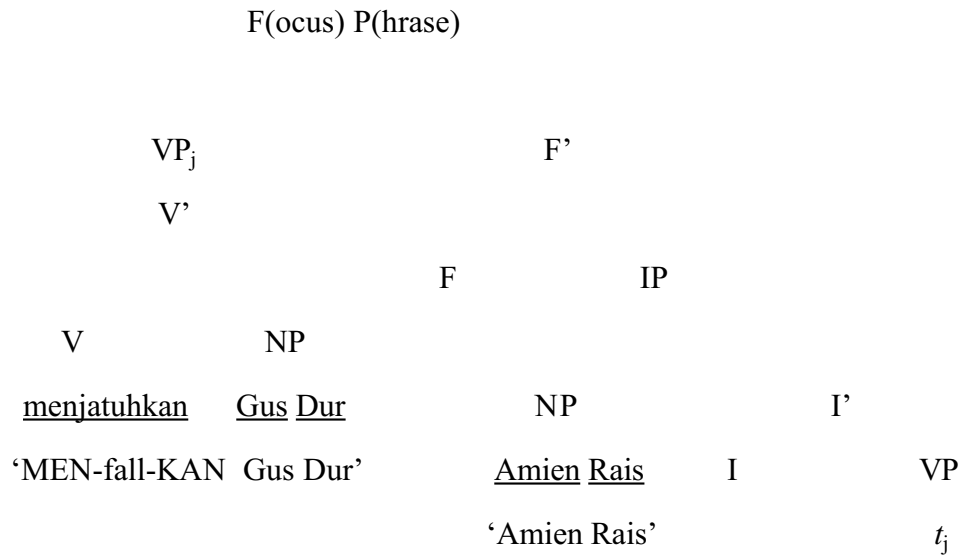
(41) Apa jawabanmu atas pertanyaannya?
what reply-AN-2SG for PER-ask-AN-NYA
'What is your reply to his question?'

It has already been established that the PH ((28)) requires that question words be located in the predicate, not the subject in SI. Thus, it follows that in (40) and (41), despite appearances to the contrary, the question words **cannot** be in subject position. We shall now argue that the order of constituents in these sentences is due to an independently motivated process of predicate fronting, which moves a focus-marked constituent from the predicate to a sentence initial position². Predicate fronting holds in general in SI; it applies not only to *yang* questions

and nominal questions, but also to verbal questions and declarative sentences. Compare (42a) - (43a) with (42b) - (43b).

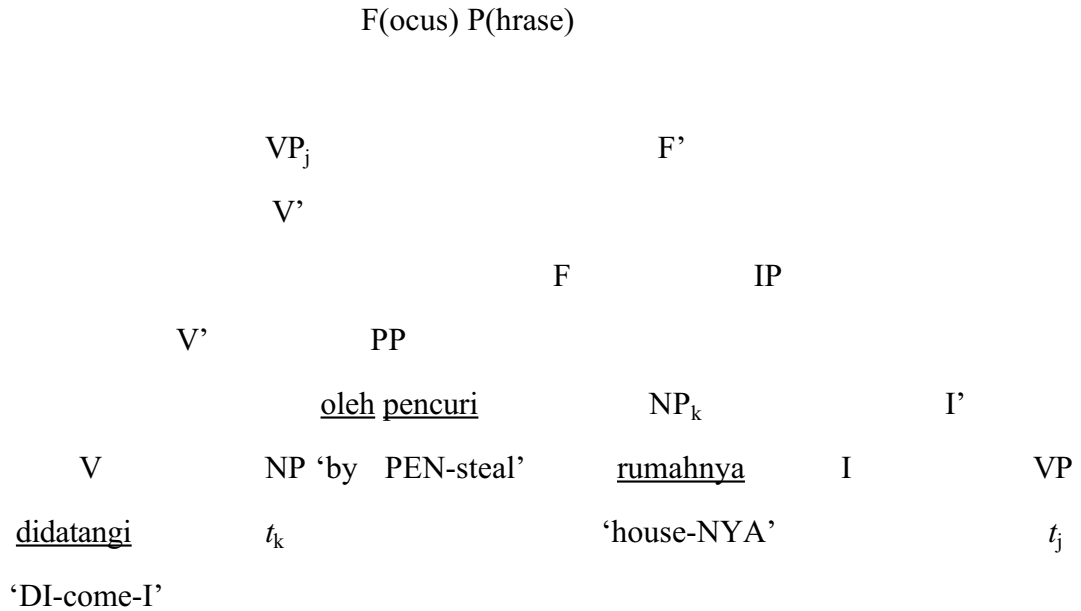
- (42) a. Amien Rais menjatuhkan Gus Dur.
Amien Rais MEN-fall-KAN Gus Dur
'Amien Rais overthrew Gus Dur.'
- b. [Menjatuhkan Gus Dur]_i Amien Rais *t_i*.
MEN-fall-KAN Gus Dur Amien Rais
'Amien Rais overthrew Gus Dur?'

c. Structure for (b)



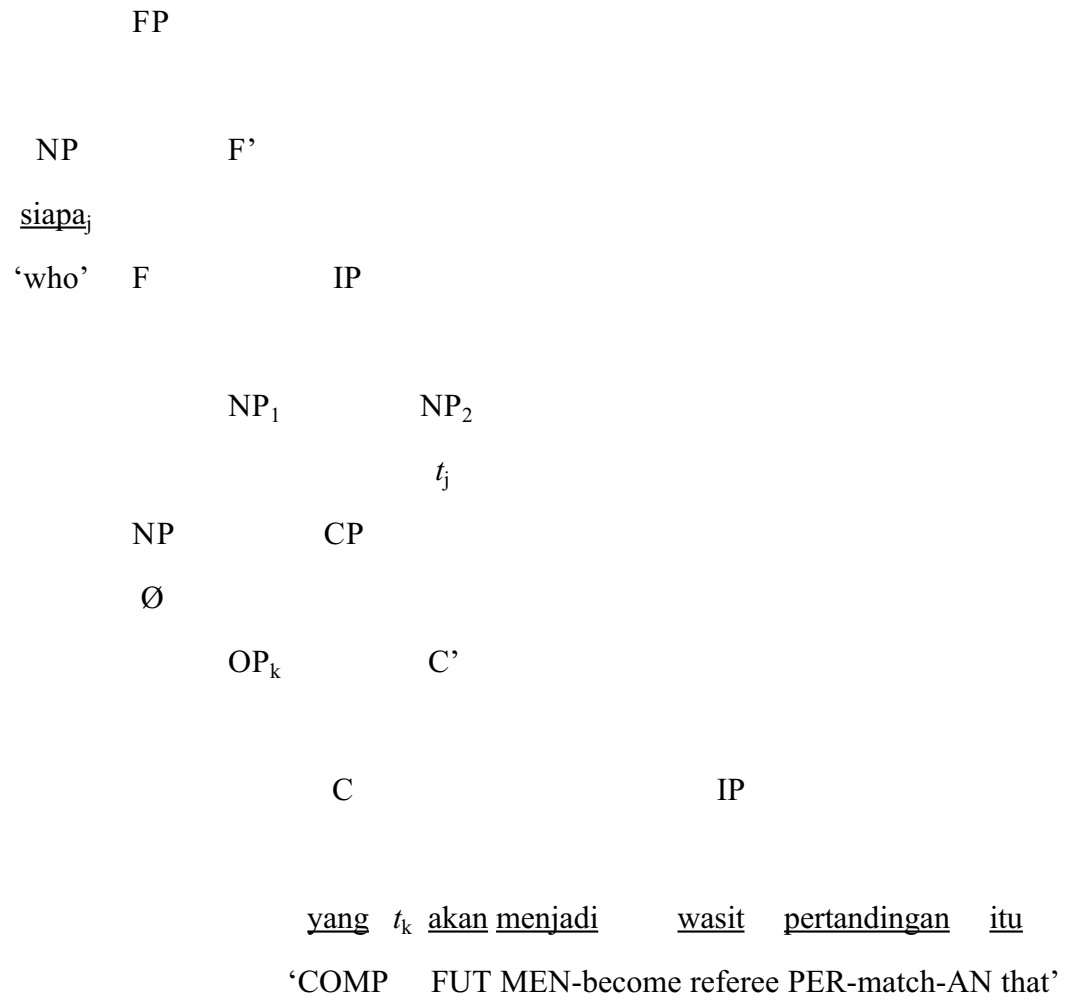
- (43) a. Rumahnya didatangi oleh pencuri.
house-NYA DI-come-I by PEN-steal
'His house was robbed.'
- b. [Didatangi oleh pencuri]_i rumahnya *t_i*.
DI-come-I by PEN-steal house-NYA
'His house was robbed.'

c. Structure for (b)

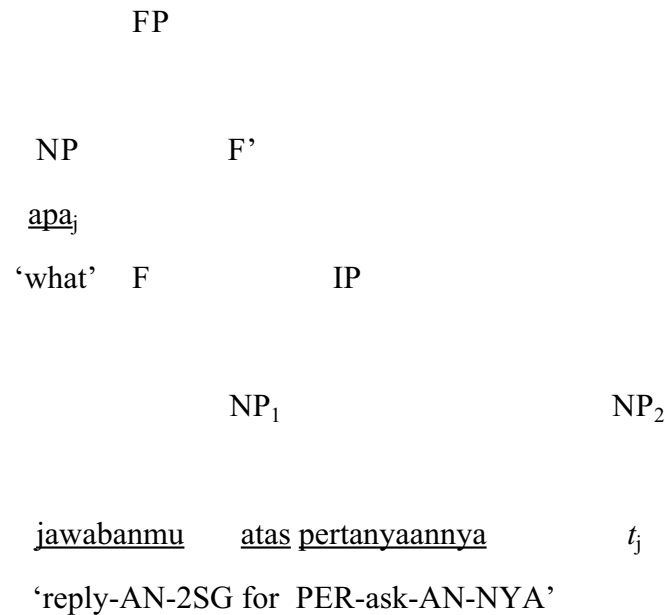


Returning to *yang* questions like those in (40) and to nominal questions like those of (41), we propose that these sentences, like (42b) - (43b), involve movement to the specifier of a focus projection. Thus, for example, the *yang* question in (40) and the nominal question in (41) will have the following derivations.

(44) Structure for (40): *Siapa yang akan menjadi wasit pertandingan itu?*



(45) Structure for (41): *Apa jawabanmu atas pertanyaannya?*



Thus, we claim that sentences like those of (40) and (41) are due to predicate fronting, and are not instances of WH in situ or of WH movement. As a result, these examples do not constitute true counter examples to the PH.

It should be noted that the process of predicate fronting moves a predicate phrase, not an NP or adverbial, to initial position in SI. Thus, predicate fronting can be distinguished from WH movement, which moves a WH constituent, NP or adverbial, leaving the verb in situ. Notice that predicate fronting applies in both questions and non-questions. In contrast, WH movement is restricted to elements which are [+WH]. It is only in the case of nominal sentences, in which the WH NP is the predicate, that there is the appearance of either WH in situ or WH movement. But when sentences like those in (40) and (41) are placed in the context of examples like those of (412b - 43b), it can be seen that (40) and (41) are in fact instances of predicate fronting.

4 The Distribution of WH Argument Words in CJI

In this section we shall turn from SI to CJI. We shall show that the distributional constraints on WH in situ in subject position observed in SI do not hold in CJI.

4.1 Questioning in Verbal Sentences

In CJI, as in SI, WH in situ is employed to form WH questions in both root and embedded clauses.

(46) a. **WH in Direct Object Position**

Dia ini ngeliat siapa? (Jakarta Field Station database)

3SG this N-see who

'Who is he looking at?'

b. Gua nggak tau Ali bakalan ngawinin siapa.

1SG NEG know Ali FUT N-marry-IN who

'I don't know who Ali will marry.'

(47) a. **WH as Possessor**

Kita mo ke rumahnya siapa nih? (Jakarta Field Station database)

1PL want to house-NYA who this

'Whose house are we going to?'

b. **WH as Indirect Object**

Bokir udah nawarin siapa aja kerjaan itu?

Bokir PAST N-offer-IN who exactly work-AN that

'Who exactly has Bokir offered the job to?'

c. **WH as Object of Preposition**

Kemarin perginya sama siapa, Hizkia? (Jakarta Field Station database)

yesterday go-NYA with who, Hizkia

'With whom did you go yesterday, Hizkia?'

d. **WH as Agent in Passive**

Kalo naik sepeda, digonceng siapa? (Jakarta Field Station database)

if go.up bike DI-get.ride who

'If you go by bike, who gives you a ride?'

However, the constraint which bans a WH word from occurring in the spec of IP does not seem to hold in CJI³. Compare (5) - (7), which are ungrammatical in SI, with (48) - (50) below:

(48) a. **Siapa tahu?** (Jakarta Field Station database)

who know?

'Who knows?'

b. **Siapa, Dik, mandi, Dik?** (Jakarta Field Station database)

who younger.brother take.a.bath younger brother

'Younger brother, who is taking a bath?'

c. **Siapa belum makan?** (Jakarta Field Station database)

who not.yet eat

'Who hasn't eaten yet?'

d. **Boleh gua tanya [siapa bakalan nggantiin Gus Dur]?**

may 1SG ask who FUT N-replace-IN Gus Dur

'May I ask who will replace Gus Dur?'

- e. Apa sih bikin elu gembira banget hari ini?
 what SIH make 2SG happy very day this
 'What makes you so happy today?'
- f. Gua penasaran [apa sih bikin elu gembira banget hari ini].
 1SG anxious.to.know what SIH make 2SG happy very day this
 'I wonder what makes you so happy today.'
- (49) a. Apanya dibuka? (Jakarta Field Station database)
 what-NYA DI-open
 'Which one should be opened?'
- (50) a. Apa sih mesti gua beliin buat elu?
 what SIH must 1SG buy-IN for 2SG
 'What exactly must I buy for you?'
- b. Coba elu jelasin [apa sih mesti gua beliin buat elu]?
 Try 2SG explain-IN what SIH must 1SG buy-IN for 2SG
 'Please explain to me what exactly I must buy for you?'

4.2 Questioning in Nominal Sentences

As in SI, in CJI WH words appear in apparent subject position in nominal sentences:

- (51) a. [IP [NP2 warna apa] [NP1 ikannya]]? (Jakarta Field Station database)
 color what fish-NYA
 'What color is the fish?'

- b. [[NP2 **Apa**] [NP1 namanya]]? (Jakarta Field Station database)
 what name-NYA
 'What is this called?'

CJI, however, seems to differ from SI in allowing *yang* to follow the WH words in nominal sentences. That is, CJI appears to permit relative clauses based on nominal as well as verbal sentences. Consider the following examples.

- (52) a. Bunga apa yang warna merah? (Jakarta Field Station database)
 flower what COMP color red
 'What flower is red?'
 b. Tas apa yang kulit buaya?
 bag what COMP skin crocodile
 'What bag is the one made of crocodile skin?'

CJI, therefore, raises two questions: 1) Why is WH in situ possible in subject position in CJI, but not in SI?, and 2) Why are apparent instances of relativization of subjects of nominal sentences possible in CJI, but not in SI?

5 Why WH in situ in Subject Position is Grammatical in CJI

In this section we will give an account of why WH in situ in subject position ((48) - (50)) is grammatical in CJI despite the fact that it is ill-formed in SI. We shall consider two hypotheses: (a) the PH applies to both SI and CJI, but *yang*-less questions in (48) - (50) contain covert headless relative clauses [henceforth *the Covert HRC Analysis*] with a phrase structure similar to (44), and should be analyzed as nominal sentences, and (b) the PH does not apply in CJI [henceforth *the Non-Parallelism Analysis*]. We shall argue in favor of the second hypothesis, that the PH does not apply in CJI.

5.1 The Covert HRC Analysis

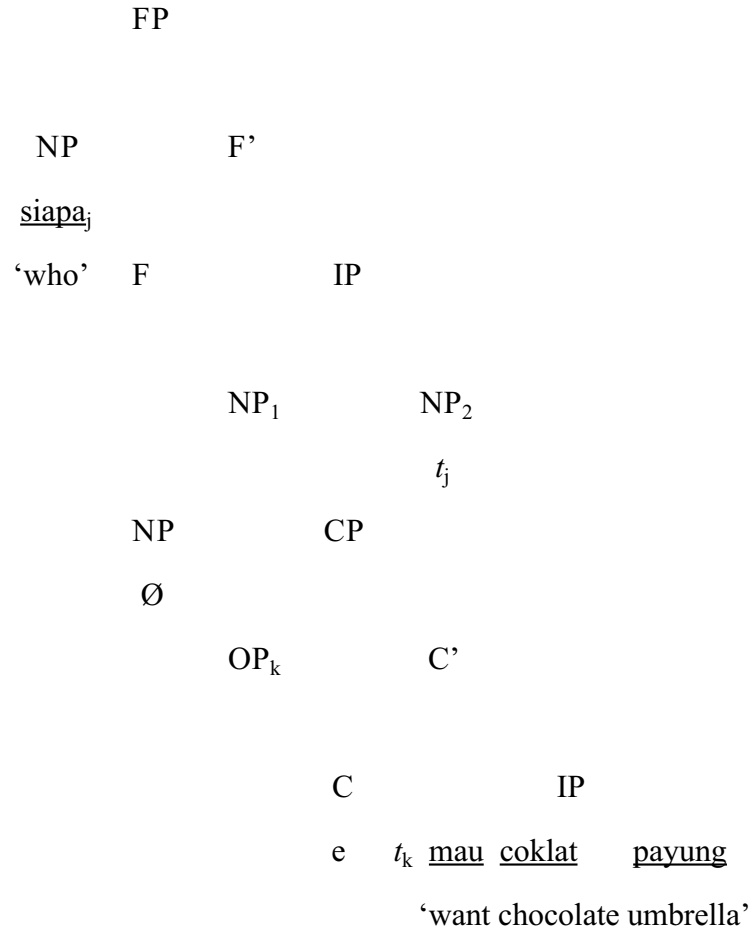
According to the Covert HRC analysis, *yang*-less questions such as (53) would have the following derivation.

- (53) Siapa mau coklat payung?

who want chocolate umbrella

'Who wants an umbrella chocolate?'

(54) Structure for (53)



Thus, according to the covert HRC hypothesis, (53) has essentially the same structure as (40) [tree (44)], except that in (53) there is no overt complementizer. Under this analysis, (53) is compatible with the PH. Thus, according to the covert HRC analysis, SI and CJI differ only with respect to whether *yang* is obligatory in HRCs.

5.2 The Non-Parallelism Analysis: Evidence Against the Covert HRC Analysis

In this section we will argue for hypothesis (b) and against (a). We will show that CJI is in fact similar to English, in being a language that does not require a strict parallelism between subject and topic. To support our claim, we shall present evidence that questions without *yang* like those in (48) - (50) and (53) are not instances of the same construction as those with *yang*, and

they should therefore be analyzed as WH questions based on verbal sentences, like their counterparts in English, rather than as covert nominal sentences.

There are a number of reasons why the covert HRC Analysis is not tenable. First of all, headless relative clauses in CJI (like SI) are ill-formed without the presence of the overt complementizer *yang*. Consider the sentences in (55), which contains a headless relative clause in subject position.

- (55) a. [NP1 Yang diri di belakang] [NP2 teman gua].
 COMP stand in back friend 1SG
 '(The one) who is standing in the back is my friend.'
- b. [NP1 Yang teriak] [NP2 mahasiswa dari Delaware].
 COMP shout university.student from Delaware
 '(The one) who is shouting is a student from Delaware.'

According to the covert HRC Analysis, *yang* should be optional in the examples of (55). Thus, the covert HRC Analysis predicts that the examples of (56) are grammatical⁴. The ill-formedness of those sentences, however, shows that this prediction is not borne out.

- (56) a. *[NP1 [CP \emptyset diri di belakang] [NP2 teman gua].
 COMP stand in back friend 1SG
 '(The one) who is standing in the back is my friend.'
- b. *[NP1 [CP \emptyset teriak] [NP2 mahasiswa dari Delaware].
 COMP shout student from Delaware
 '(The one) who is shouting is a student from Delaware.'

Examples like those of (56) provide prima facie evidence that *yang* cannot be omitted in HRCs. However, the examples of (57) appear, on the face of it, to provide evidence that *yang* **can** be omitted in HRCs.

- (57) a. Yang diri di belakang teriak.
 COMP stand in back shout
 '(The one) who is standing in the back is (the one who is) shouting.'

b. Yang teriak **diri** **di belakang**.

COMP shout stand in back

'(The one) who is shouting is (the one who is) standing in the back.'

The covert HRC analysis would claim that the bold face phrases in (57) can be parsed as HRCs and have the structure in (58).

- (58) a. [NP1 [CP yang diri di belakang] [NP2 [CP \emptyset **teriak**].
 COMP stand in back COMP shout
 '(The one) who is standing in the back is (the one) who is shouting.'
- b. [NP1 [CP yang **teriak**] [NP2 [CP \emptyset **diri** **di** **belakang**].
 COMP shout COMP stand in back
 '(The one) who is shouting is (the one) who is standing in the back.'

The above claim is based on the idea that the examples of (58), which lack the complementizer *yang*, have structures like (59) except that the complementizer is null in (58).

- (59) a. [NP1 [CP yang diri di belakang] [NP2 [CP **yang** **teriak**].
 COMP stand in back COMP shout
 '(The one) who is standing in the back is (the one) who is shouting.'
- b. [NP1 [CP yang **teriak**] [NP2 [CP **yang** **diri** **di** **belakang**].
 COMP shout COMP stand in back
 '(The one) who is shouting is (the one) who is standing in the back.'

We shall argue, however, that the putative *yang*-less headless relative clauses in (58) (shown in bold face) cannot be assigned the structure of headless relative clauses which lack *yang*, but rather are only parsable as main clause verb phrases. Our argument that *teriak* and *diri di belakang* are not relative clauses is based on the distribution of *adalah* 'to be' and *bukan* 'NEG'. As is well known (Alwi et al 1998), in nominal sentences *adalah* can optionally be used as a copula, and *bukan* can be used instead of *nggak/ndak*⁵ as clausal negation. Thus if the predicates of the sentences in (58) are analyzed as headless relative clauses with *yang* omitted, then it should be possible for either *adalah* 'to be' or *bukan* 'NEG' to precede the putative *yang*-less headless relative clauses. Note that *adalah* 'to be' and *bukan* 'NEG' are well-formed in the sentences of (60), sentences, in which the predicate is clearly a HRC.

- (60) a. Yang diri di belakang **adalah** [NP [CP yang teriak].
 COMP stand in back be COMP shout
 '(The one) who is standing in the back is the one who is shouting.'

- b. Yang teriak **adalah** [_{NP} [_{CP} yang diri di belakang]].
 COMP shout be COMP stand in back
 '(The one) who is shouting is the one who is standing in the back.'
- c. Yang diri di belakang **bukan** [_{NP} [_{CP} yang teriak]].
 COMP stand in back NEG COMP shout
 '(The one) who is standing in the back is not the one who is shouting.'
- d. Yang teriak **bukan** [_{NP} [_{CP} yang diri di belakang]].
 COMP shout NEG COMP stand in back
 '(The one) who is shouting is not the one who is standing in the back.'

When we turn to (61), a different pattern is observed: The presence of *adalah* 'to be' and *bukan* 'NEG' is ill-formed, a pattern which is incompatible with the HRC hypothesis, which assigns a nominal status to the bracketed constituents.

- (61) a. *Yang diri di belakang **adalah** [_{NP} [_{CP} \emptyset teriak]].
 COMP stand in back be COMP shout
 '(The one) who is standing at the back is the one who is shouting.'

- b. *Yang teriak **adalah** [NP [CP \emptyset diri di belakang].
 COMP shout be COMP stand in back
 '(The one) who is shouting is the one who is standing at the back.'
- c. *Yang diri di belakang **bukan** [NP [CP \emptyset teriak]⁶.
 COMP stand in back NEG COMP shout
 '(The one) who is standing at the back is not the one who is shouting.'
- d. *Yang teriak **bukan** [NP [CP \emptyset diri di belakang].
 COMP shout NEG COMP stand in back
 '(The one) who is shouting is not the one who is standing at the back.'

Furthermore, sentences analogous to (61a) - (61b) are grammatical, if *adalah* 'to be' does not occur, as illustrated in (57), or if *nggak/ndak*, the negative words for VPs, are instead employed, as illustrated in (62):

- (62) a. Yang diri di belakang **nggak/ndak** teriak.
 COMP stand in back NEG shout
 '(The one) who is standing at the back is not shouting.'
- d. Yang teriak **nggak/ndak** diri di belakang.
 COMP shout NEG stand in back
 '(The one) who is shouting is not the one who is standing at the back.'

Thus, we conclude that the bold face phrases in (57) are VPs and not NPs which are instances of covert HRCs, and that in CJI, as well as in SI, HRCs must be introduced by the complementizer *yang*. We argue that the sentences of (57) have the structure in (63) rather than (58).

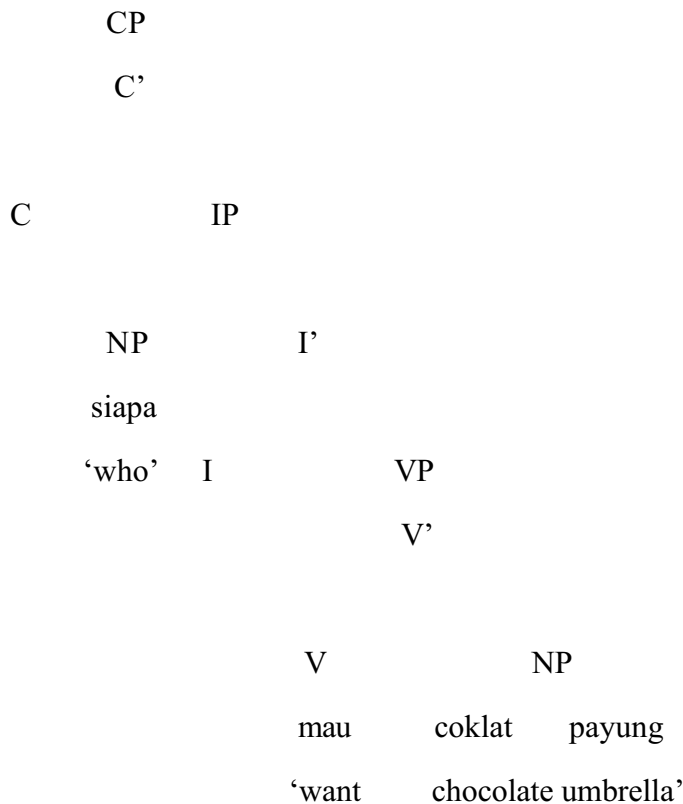
- (63) a. [_{NP} Yang diri di belakang] [_{VP} teriak].
 COMP stand in back shout
 '(The one) who is standing at the back is shouting.'
- b. [_{NP} Yang teriak] [_{VP} diri di belakang].
 COMP shout stand in back
 '(The one) who is shouting is standing at the back.'

Returning to (53), we argue that this question cannot contain a headless relative clause since it lacks *yang*. In fact, *mau coklat payung* is a VP. This is confirmed by the fact that it can only be made negative with *nggak/ndak* 'not', the negative words for VPs, and not *bukan* 'not', the negative word for NPs.

- (64) a. *Siapa **bukan** mau coklat payung?
 who NEG want chocolate umbrella
 'Who does not want an umbrella chocolate?'
- b. Siapa **nggak/ndak** mau coklat payung?
 who NEG want chocolate umbrella
 'Who does not want an umbrella chocolate?'

Thus, in our analysis (53) has the derivation in (65), rather than the derivation in (54).

(65) Structure for *siapa mau coklat payung?* in CJI:



We conclude that unlike SI, in CJI the question word *siapa* in (53) can occur in the spec of IP.

The fact that question words in *yang*-less questions in CJI can be located in the spec of IP leads us to the conclusion that the non-parallelism analysis is the correct account for WH in situ in subject position in CJI.

To conclude this section, let us summarize the difference between SI and CJI regarding the subject questions: The questioning of subjects in SI is **not** possible, and apparent instances of subject questions with *yang* are in fact instances of fronted WH predicates rather than WH subjects. This state of affairs is due to the fact that SI requires a strict parallelism between topic and subject, which blocks WH in subject position. The in situ questioning of subjects in CJI, however, is possible because CJI does not require a parallelism between topic and subject (the Non-parallelism Analysis). Therefore, the question words can reside in the spec of IP.

7 Conclusions

We have shown in this paper that seemingly contradictory restrictions on the distribution of WH question words in SI are due in fact to a requirement of parallelism between syntactic and information structure. Because WH words have inherent focus, they cannot fill the subject position, which is reserved for topics. We have also argued that *yang* questions and nominal questions in SI are both instances of nominal sentences, with a structure [IP [NP1] [NP2]]. The first NP is the argument and the second one is the predicate. To account for the word-order variability of nominal sentences in SI, we claim that WH-initial nominal questions are instances of a general process of predicate fronting, and do not instantiate either WH in situ or WH movement.

Turning to CJI, we have argued that WH subject questions without *yang* should be analyzed as instances of WH in situ in subject position, and that the strict parallelism between topic and subject that is required in SI does not apply to CJI. *Yang*-less questions cannot be analyzed as containing headless relative clauses because headless relative clauses in Indonesian are only well-formed when the complementizer *yang* is overtly present.

We conclude that the differences in restrictions on WH in situ in SI and CJI are not localized in the grammar in the strict sense, but are due to the interplay between grammar and pragmatics, specifically information structure. It is of some interest that the parallelism requirement (28) states the relationship between two components, one of which is purely linguistic in nature (the formal syntax) and the other of which is non-linguistic (knowledge of how information is structured). What the parallelism requirement states is the required relationship between the domain of formal linguistics and a domain exterior to language per se. While the parallelism requirement itself is non-linguistic in the technical sense used in Generative Grammar, it is a central part of the systematic knowledge that a Standard Indonesian speaker has about his language.

Put differently, to know Standard Indonesian you must have an unconscious knowledge of this meta-requirement relating grammar and pragmatics. That is, the knowledge of parallelism is part of the competence of the native speaker. (Here competence is used to mean 'unconscious,

systematic knowledge of the language' rather than 'knowledge of the principles and parameters of universal grammar as reflected in the language'. A comparison of SI and CJI is reminiscent of Comrie's (1989, pp. 74-85) comparison of Russian and English. Comrie notes that in many ways the grammars of Russian and English are similar. For instance, both languages encode grammatical relations and use them in a variety of grammatical processes. Russian, however, differs from English in that much of what is systematic and, hence, linguistically interesting is not in the grammar per se, but rather in the semantics and pragmatics, and in the relationship between grammar and semantics/pragmatics. Thus, while it arguably is desirable to view grammar (in the technical sense) as a formal system divorced from pragmatics etc., this does not imply that all the interestingly systematic knowledge we have about our language is part of the grammar (so-defined). Rather, there is strong evidence that much of what is linguistically interesting is external to the grammar per se.

Notes

* This paper is based on data from two varieties of Indonesian, Standard Indonesian (as spoken in Jakarta) and Colloquial Jakarta Indonesian. The data from Standard Indonesian (SI) is based on the judgements of one of the co-authors (Tjung) and was confirmed by consulting other speakers of this register from Jakarta. The data on Colloquial Jakarta Indonesian (CJI) is primarily drawn from naturalistic conversations collected as part of the Max Planck Institute of Evolutionary Anthropology project on the acquisition of Jakarta Indonesian and carried out at the Jakarta Field Station of the Institute (<http://monolith.eva.mpg.de/~gil/jakarta/project-child.html>). The naturalistic data are supplemented by the judgements of Jakarta speakers, including one of the co-authors. The facts in SI are largely parallel to those in Javanese. The analysis of Javanese is discussed in Cole et al (to appear). We would like to thank our Javanese consultants, Alexander (Eric) Rabda and Lily Widjaja for the Javanese data, and Satoshi Tomioka, Li-May Sung, Chonghyuck Kim, Chang-Yong Sim, Yaping Tsai, and Kozue Inoha for their many helpful suggestions.

This research is supported in part by funding from the National Science Foundation (SBR-9729519) and the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

¹ 'BE' occurs only for expository purposes. We take NP₂ to be the predicate. The question word can appear in situ (in the NP₂ position) or move to the spec of CP in Malay. That is, the in situ version of (30) is also grammatical:

- (i) Yang Ali beli apa?
COMP Ali buy what
'What did Ali buy?'

² A similar line of analysis is provided by Sie (1988). He argues that in a question like (i), it is not the subject NP which is questioned. Rather (i) is the result of moving the predicate from the base position seen in (ii) to initial position.

-
- (i) Siapakah yang makan di dapur? (ii) Yang makan di dapur siapakah?
 Who-KAH COMP eat in kitchen COMP eat in kitchen who-KAH
 'Who is eating in the kitchen?' 'Who is eating in the kitchen?'

³ Norhaida (1999) also noted that the constraint does not hold either in the colloquial variety of Singapore Malay:

- (i) Sapa/siapa makan agar-agar?
 who eat jello
 'Who ate the jello?'

⁴ The sentences of (56) are well-formed if they are interpreted as instances of predicate fronting. For example, sentence (ii), like those of (42b) - (43b), is derived by applying predicate fronting to

- (i).
 (i) [NP Teman gua [VP diri di belakang]. (ii) [VP Diri di belakang] [NP teman gua].
 friend 1SG stand in back stand in back friend 1SG
 'My friend is standing in the back.' 'Standing in the back is my friend.'

Sentence (ii), however, lacks the interpretation 'The one who is standing in the back is my friend'.

⁵ *Nggak* and *ndak* are the CJI equivalents of Standard Indonesian *tidak/tak* 'not'.

⁶ The sentences of (61c-d) are well-formed if they have the interpretation 'The one who is standing in the back is not shouting (but screaming)' and 'The one shouting is not standing in the back (but standing in the middle of the row)' respectively.

References

- Alwi, Hasan. Soenjono D., H. Lapoliwa and Anton M. Moeliono. 1998. *Tata Bahasa Baku Bahasa Indonesia*. Jakarta: Balai Pustaka.
- Chomsky, Noam. 1981. *Lectures on Government and Binding: The Pisa Lectures*. Dordrecht: Foris Publications.
- _____. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.
- Chomsky, Noam. 1995. *The Minimalist Programme*. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 1998. *Minimalist Inquiries: The Framework*. MIT Occasional papers in linguistics: 15.
- Chomsky, Noam. 1999. *Derivation by Phase*. MIT Working Papers in Linguistics: 18.

-
- Cole, Peter and Gabriella Hermon. 1995. Is Wh in Situ Really in Situ? Evidence from Malay and Chinese. In *The Proceedings of the Thirteenth West Coast Conference on Formal Linguistics*. Center For the Study of Language and Information.
- Cole, Peter and Gabriella Hermon. 1998. The Typology of *Wh*-movement: *Wh*-Questions in Malay. *Syntax* 1.33:221-258
- Cole, Peter and Gabriella Hermon. 2000. Partial *Wh*-Movement: Evidence from Malay. In Uli Lutz, Gereon Muller and Arni von Stechow (eds.), *Wh-Scope Marking*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Cole P., G. Hermon, and Norhaida Bt. Aman. To appear. Clefted Questions in Malay. In David Gil (ed.), *Malay/Indonesian Linguistics*. London: Curzon Press.
- Cole, Peter, G. Hermon, Kozue Inoha and Yassir Tjung. To appear. A Constraint on WH in situ in Javanese. In Ileana Paul, Andra Rackowski, and Norvin Richards (eds.), *Proceedings of Austronesian Formal Linguistics (AFLA) 8*, MIT WPELFL, Fall 2001.
- Comrie, Bernard. 1989. *Language Universals and Linguistic Typology*. 2nd Edition. Chicago: The University of Chicago Press.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
- Mashudi Kader. 1976. The Syntax of Malay Interrogatives. Doctoral Dissertation: Simon Fraser University.
- Norhaida Bt. Aman. 1999. How to Ask What in Malay: The Acquisition of WH-Questions in Singapore Malay. Ph.D. Dissertation: University of Delaware.
- Poedjosoedarmo, Gloria. 1977. Thematization and Information Structure in Javanese. In Amran Halim (ed), *NUSA Vol. 3, Miscellaneous Studies in Indonesian and Languages in Indonesia: Part II*, 34-43, Jakarta.
- Ross, J.R. 1967. Constraints on Variables in Syntax. Ph.D. Dissertation: MIT.

-
- Saddy, Douglas. 1991. Wh Scope Mechanism in Bahasa Indonesia. In Lisa Cheng and Hamida Demirdash (eds.), *More Papers on Wh-Movement*. *MITWPL* 15:183-218.
- _____. 1992. A versus A-bar Movement and *Wh*-Fronting in Bahasa Indonesia. Ms., University of Queensland, Australia.
- Sie, Ing Djiang. 1988. *The Syntactic Passive In Bahasa Indonesia*. Doctoral Dissertation: University of Amsterdam.
- Sneddon, James Neil. 1996. *Indonesian: A Comprehensive Grammar*. London: Routledge.
- Soemarmo, Marmo. 1970. *Subject-Predicate, Focus-Presupposition, and Topic-Comment in Bahasa Indonesia and Javanese*. Doctoral Dissertation: UCLA.
- Van Valin, Robert D. and Randy J. Lapolla. 1997. *Syntax: Structure, Meaning, and Function*. Cambridge: Cambridge University Press.