The Typology of WH-words

—An Austronesian perspective—

Yuko Otsuka

1 Introduction

Languages of the world employ various strategies to form wh-questions. English, for example, demonstrates three different strategies, wh-movement, cleft, and pseudo-cleft (PC), as illustrated in (1).

(1a) What did John buy __? WH-MOVEMENT
(1b) What is it that John bought __?[1] CLEFT
(1c) What is the thing that John bought __? PSEUDO-CLEFT

The wh-movement strategy is illustrated in (1a), in which the wh-phrase occurs sentence-initially instead of the corresponding argument position (indicated by the underscore). In the tradition of transformational grammar, it is assumed that the wh-phrase that does not occur in the expected argument (or adjunct) position has moved out of that relevant position via the syntactic operation called wh-movement. Clefts are essentially a nominal construction containing an expletive in the subject position and a nominal predicate modified by a *that*-clause. In the cleft wh-questions, the wh-phrase occurs as a nominal predicate. It should be noted that the cleft strategy may, but does not necessarily involve wh-movement. In (1b), the wh-phrase occurs sentence-initially, because wh-movement is obligatory in English. In a language that does not require (or prohibits) wh-movement, the wh-phrase remains in situ, literally translated in English as '*It is <u>what</u> that John bought?*'. PC constructions have a similar structure to that of cleft constructions, but contains as the subject a headless relative clause (or one modifying a dummy head such as *the thing*) instead of an expletive *it*.

In Japanese, we find a different set of strategies. First, assuming wh-movement is not optional, (pure) wh-movement is not available in Japanese;

[1] While the underscore in (1b) and (1c) indicates the position in which the wh-phrase is interpreted, it is not the position it originates. Strictly speaking, cleft wh-questions like (1b) contain two gaps: What is it <what> [$_{CP}$ OP that John bought <OP>]? Similarly in PC wh-questions: What is the thing [$_{CP}$ OP that John bought <OP>? The first gap results from the wh-movement of *what*; the second gap is a result of a null operator movement within the *that*-clause (Chomsky 1977).

wh-phrases remain in situ, as shown in (2a), in an unmarked context. Second, the PC strategy is available in Japanese, as shown in (2b). Third, wh-phrases may occur in the sentence-initial position by means of focus fronting (2c).[2]

(2a) John-ga nani-o katta-no? WH-IN SITU
John-NOM what-ACC bought-Q
'What did John buy?'

(2b) John-ga ___ katta-no wa nani? PSEUDO-CLEFT
John-NOM bought-NO TOP what
'What is (the thing that) John bought?'

(2c) Nani-o John-ga ____ katta-no? FOCUS FRONTING
what-ACC John-NOM bought-Q
'What is it that John bought?' (lit. What, John bought?)

This paper examines the strategies for wh-question formation available in two Austronesian languages, Tagalog (Philippine) and Tongan (Polynesian). Wh-questions in Austronesian languages are typically formed using the PC strategy (Aldridge 2002, 2004 for Seediq; Chang 2000 for Tsou; Paul 2000, 2001 and Potsdam 2006a, 2006b for Malagasy; Cole et al. 2005 for Indonesian; Richards 1998 and Aldridge 2002, 2004 for Tagalog; Georgopoulos 1991 for Palauan; Bauer 1991, 1993 for Maori; Seiter 1980 for Niuean; Custis 2004 for Tongan; and Besnier 2000 for Tuvaluan; also see Potsdam and Polinsky 2011 for an overview). Tagalog and Tongan are similar in that respect. When examined carefully, however, the two languages exhibit intriguing differences as to what other strategies are available and for what kind of constituent questions (e.g., subject wh-questions, adjunct wh-questions, etc.).

2 Wh-strategies in Tongan and Tagalog: an overview

Tables 1 and 2 provide the summary of wh-question strategies in Tongan and Tagalog, respectively. Specific data are considered in Sections 4 (Tongan) and 5 (Tagalog). In both languages, the PC strategy is used for argument wh-questions. However, the PC strategy is not available for all kinds of wh-questions. In Tongan,

[2] Abbreviations used in this paper are as follows: ABS = absolutive, ACC = accusative, ANA = anaphor, DET = determiner, ERG = ergative, FUT = future, OBL = oblique, PRED = predicate marker, PRS = present tense, PST = past tense, Q = question marker, S = singular, SBJ = subject, TOP = topic.

predicative wh-questions cannot be formed using this strategy. In Tagalog, the PC strategy is limited to core argument wh-questions only. The two languages also differ quite drastically in other aspects of wh-question formation. First, while Tongan permits wh-phrases to remain in situ, this strategy is unavailale in Tagalog regardless of the type of wh-words involved. Second, while Tongan prohibits wh-movement, Tagalog requires anything other than core arguments to undergo wh-movement.

	PC	In situ	Wh-movement
Nominal wh	Yes	Yes	No
Adverbial wh	Yes	Yes	No
Predicative wh	No	Yes	No

Table 1. TONGAN WH-STRATEGIES

	PC	In situ	Wh-movement
Core argument wh	Yes	No	No
Oblique wh	No	No	Yes
Adverbial wh	No	No	Yes

Table 2. TAGALOG WH-STRATEGIES

It has been observed in the literature that languages generally divide into two classes, one that requires wh-movement and the other that disallows it. It is noteworthy that this generalization holds true for both Tongan and Tagalog despite the aforementioned differences. However, there are some issues that cannot be readily explained. The first set of questions concern the constraints on the use of PC strategy. Both languages show some constraints, but they are not the same kind of constraints. Second is a Tagalog-specific question and concerns the constraint on wh-movement.

I. Why is the PC strategy available for adverbial wh-questions in Tongan, but not in Tagalog?
II. Why is the PC strategy unavailable for predicative wh-questions in Tongan?
III. Why is the PC strategy available only for core argument wh-questions in Tagalog?
IV. Why is wh-movement prohibited for core arguments, but required for non-core arguments in Tagalog?

3 Theoretical background

This section provides some theoretical background that is assumed in the subsequent discussion.

3.1 Minimalist Program

The theoretical framework adopted in this paper is that of the Minimalist Program (Chomsky 2000 and subsequent work). In this framework, linguistic items, lexical as well as functional, are regarded as bundles of features. The kinds of features that are relevant to syntax are formal features, which fall into two classes: interpretable features (F), which have a specific value, and uninterpretable features (*u*F), which lack a specific value and are only specified for a feature type. An example of the former would be an agreement feature (called ϕ-feature) with specific values such as [ϕ: 1SG.F], whose corresponding uninterpretable feature is [*u*ϕ:], for which the specific value is left blank. Derivation of syntactic objects is motivated by the principle of Full Interpretation (FI), a condition that requires that syntactic objects consist only of interpretable features at the LF interface. Thus, the goal of syntactic operations is to eliminate *u*Fs within a given structure in the course of derivation. This is achieved by an operation called Agree, through which an *u*F (probe) receives a specific value from a matching interpretable feature F (goal). Movement is contingent on Agree and licensed by an EPP-feature on the relevant head. It is also assumed that movement leaves a copy of the item moved.

3.2 Three components of wh-questions

Before illustrating how wh-question formation is analyzed in this framework, it is necessary to understand the semantics of wh-questions and their syntactic realization. Wh-questions instruct to identify an individual out of a certain set (of people, objects, places, etc.). A wh-question thus consists of an instruction "Select X" and a proposition that is true of the individual to be identified. For example, the semantic interpretation of a wh-question *What did John buy?* (3a) would be the one given in (3b). Given that FI requires all elements that are necessary for semantic interpretation to be present in the relevant syntactic structure, it is assumed that a wh-question must have an operator/variable structure, in which a question operator binds a variable: [OPx [… x …]].[3] It is typically assumed that

[3] This is formulated by Cole & Hermon (1998) as "Variable Binding Condition".

wh-expression is the operator and that the remainder of the sentence contains a variable. The syntactic structure of (3a) would be the one provided in (3c).

(3a) *What did John buy?*

(3b) Select an item x from a set of objects such that John bought x.

(3c) [$_{CP}$ what [$_{C'}$ did [$_{TP}$ John buy <what>]]?

OPERATOR VARIABLE

Thus, wh-questions must have three components: instruction ("Select one"), an operator, and a variable. The principle of FI requires that each of these semantic components have a corresponding syntactic realization. It is generally assumed that the information about the sentence type is located in the C(omplementizer) head as a formal feature. For wh-questions, I assume there are two relevant features: [Q] to indicate that it is a question and [*u*WH] to indicate that it is a wh-question. I propose that the combination of these two features represent the instruction "Select one". Wh-questions must also contain an operator and a variable. In order to separate the operator function and the variable function, I assume they are linked to two separate formal features, [OP] and [WH], respectively. C's [*u*WH] requires a matching feature [WH], thereby ensuring the presence of a variable in wh-questions. Similarly, I assume C bears [*u*OP] to ensure the presence of an operator in wh-questions.[4]

SEMANTICS	SYNTAX
Select one	clause type [Q; *u*WH; *u*OP] on C
A particular individual x	operator [OP]
Such that … x …	variable [WH]

TABLE 3. THREE COMPONENTS OF WH-QUESTIONS

3.3 Analysis of wh-movement

Let us now turn to the syntactic derivation of wh-questions. It is typically assumed that in wh-movement languages such as English, the variable bound by the operator is its copy created as a result of wh-movement. This movement is licensed by an EPP-feature on C. To be specific, I assume C has an uninterpretable operator feature [*u*OP] and that wh-phrases (in wh-movement languages) are operators

[4] Treating a wh-feature as a variable feature is not a standard view. Wh-feature is often seen as a feature of an operator such as wh-phrases.

bearing a feature [OP]. C agrees with the wh-phrase and the latter moves to [Spec, CP] due to C's EPP-feature. In the resulting structure, the wh-phrase in [Spec, C] is an operator and its lower copy serves as a variable, as shown in (4).[5] In this analysis, we must assume that a wh-phrase such as *what* bears both the operator feature and the variable feature and that its interpretation is dependent on the structural position. The higher copy is interpreted as an operator and the lower copy, a variable.

(4) [CP What [C' did [TP John buy <what>]]]?
[OP; WH] [*u*OP; *u*WH] [OP; WH]

3.4 Analysis of wh-in situ

In contrast, in wh-in situ languages like Japanese, wh-phrases do not move, but occur in the base position in an unmarked context. In the minimalist framework, the simplest explanation would be to say that C lacks an EPP-feature. Wh-words in situ therefore function only as a variable (cf. Nishigauchi 1990; Cheng 1991; Cole & Hermon 1998; Reinhart 1998). In other words, wh-words are not operators in wh-in situ languages. In the present analysis, this amounts to saying that wh-phrases in wh-in situ languages bear only the variable feature, [WH], but lacks the operator feature. Assuming that C nevertheless bears [*u*OP], this feature must be checked in some other way, namely, by generating a null operator directly in [Spec, C]. In the resulting structure (5), the operator-variable structure is obtained between the null operator and the *wh*-word in situ. The key claim of this analysis is that unlike in wh-movement languages, in which both [OP] and [WH] are located on a single head, the two are separated in wh-in situ languages. Thus, it is not simply the lack of EPP-feature that distinguishes the two types of languages. The crucial difference lies in the nature of wh-expressions, namely, their feature specification.

(5) [CP OP [C' [TP John-NOM what-ACC bought] C]]?
[OP] [WH] [Q; *u*WH; *u*OP]

4 Wh-questions in Tongan

Tongan is a predicate-initial language with a relatively free VSO-VOS alternation. Case marking shows an ergative-absolutive alignment with *'e* marking ergative

[5] See Tonoike (2015) for an alternative approach, in which the operator-variable relation is argued to hold within a DP and not as a result of movement.

and *ʻa* marking absolutive. Verbal constructions contain a tense-aspect-mood marker (TAM) in the clause-initial position (6a). Nominal predicate constructions lack a TAM, but instead have a predicate marker *ko* in the sentence-initial position (6b).

(6a) Naʻe kai ʻe Sione ʻa e ika.
PST eat ERG John ABS DET fish
'John ate a fish.'

(6b) Ko e faiako ʻa Sione.
PRED DET teacher ABS John
'John is a teacher.'

Tongan wh-words fall into three classes: nominal, adverbial, and predicative. Nominal wh-words co-occur with a case marker, preposition, or predicate marker. Adverbial wh-words are placed sentence-finally. Predicate wh-words occur in the predicate position, immediately after TAM.

Nominal		Adverbial		Predicative	
hai	'who'	*ʻafē*	'when.FUT'	*fēfē*	'how'
hā	'what'	*ʻanefē*	'when.PST'	*fiha*	'how many'
fē	'where'	(*hā*	'what')		

TABLE 4. WH-WORDS IN TONGAN

4.1 Wh-questions strategies in Tongan

First, Tongan permits wh-in situ for all kinds of wh-questions, as illustrated in (7).

(7a) Naʻe kai ʻa e **hā** ʻe Sione?
PST eat ABS DET what ERG John
'What did John eat?' (lit. 'John ate what?')

(7b) ʻoku ke saiʻia ʻia **hai**?
PRS 2S like in who
'Who do you like?' (lit. 'You like who?')

(7c) Te ke ʻalu ki **fē**?
FUT 2S go to where
'Where are you going?' (lit. 'You are going where?')

(7d) Te ke ‘alu **‘afē**?
FUT 2S go when.FUT
‘When are you going?’ (lit. ‘You are going when?’)

(7e) Na‘e **fēfē** ‘a e sivi?
PST how ABS DET exam
‘How was the exam?’ (lit. ‘The exam was how?’)

Second, as expected, wh-movement is prohibited for all types of wh-words.

(8a) *(‘a e) **hā** na‘e kai ‘e Sione?
ABS DET what PST eat ERG John
Intended: ‘What did John eat?’

(8b) *(‘ia) **hai** ‘oku ke sai‘ia (ai)?
in who PRS 2S like ANA
Intended: ‘Who do you like?’

(8c) *(ki) **fē** te ke ‘alu (ki ai)?
to where FUT 2S go to ANA
Intended: ‘Where are you going?

Third, as in many other Austronesian languages, the PC strategy is the most commonly used strategy for nominal wh-questions (9).

(9a) Ko e **hā** na’e kai ‘e Sione?
PRED DET what PST eat ERG John
‘What did John eat?’ (lit. ‘(The thing that) John ate is what?’

(9b) Ko **hai** ‘oku ke sai’ia ai[6]?
PRED who PRS 2S like ANA
‘Who do you like?’ (lit. ‘(The one) you like (him) is who?’)

The PC strategy can also be used to form adverbial wh-questions, although it is rarely used and only in a marked context (10).[7] Predicative wh-questions, however, cannot be formed using this strategy (11).

[6] In Tongan, relativization of oblique phrases requires resumptive pronoun, *ai*.

[7] According to my consultant, PC adverbial wh-questions are used to request the information that has already been mentioned in the preceding conversation, e.g., ‘What was the place you went to, again?’ or ‘You said you went there when?’, but not in an out-of-the-blue context.

(10a) ʔKo **fē** te ke ʻalu ki ai?
PRED where FUT 2S go to ANA
ʻWhere are you going?' (lit. ʻ(the place where) you are going is where?')

(10b) ʔKo **ʻafē** te ke ʻalu ai?
PRED when.FUT FUT 2S go ANA
ʻWhen are you going?' (lit. ʻ(the time when) you are going is when?')

(11a) * Ko **fēfē** na'e ʻa e sivi?
PRED how PST ABS DET exam
Intended: ʻHow was the exam?' (lit. ʻ(the way) the exam was is how?')

(11b) * Ko **fiha** ʻoku ʻa e tohi ni?
PRED how.much PRS ABS DET book this
Intended: ʻHow much is this book?

4.2 Analysis of wh-questions in Tongan

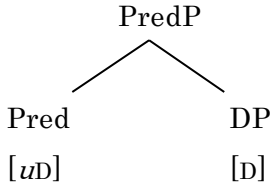
To recapitulate, Tongan behaves as expected of a wh-in situ language: wh-movement is banned; wh-in situ is permissible for all kinds of wh-questions. Based on this, I claim that wh-words in Tongan bear only the variable feature [WH] and that wh-questions contain a null operator base generated in [Spec, C]. There is, however, an interesting dichotomy between predicative wh-words and non-predicative ones, as summarized in Table 1. Notably, predicative wh-questions may not be formed using the PC strategy.

Two questions arise. First, why is the PC strategy unavailable for predicative wh-questions? Second, why is it available for adverbial wh-questions? That is, why is the contrast not one between nominal and non-nominal?

To answer the first question, let us recall that PC is a construction with a nominal predicate: DP_{PRED} DP_{SBJ}. I propose that the nominal predicate marker *ko* has a c-selectional feature [*u*D], thereby selecting only a DP as its complement.[8] Since predicative wh-words lack a categorial D-feature by definition, they cannot be merged with *ko*; hence the impossibility of forming predicative wh-questions using the PC strategy.

[8] Here I assume *ko* is the Pred head for the ease of exposition. There is a possibility that $Pred^0$ is phonetically null and *ko* is an equivalent of case markers (see Otsuka 2000).

(12)



The answer to the second question lies in the categorial status of what we have been calling "adverbial" wh-phrases. While their function is clearly adverbial (modifying the action/state), their morphosyntactic distribution likens that of a nominal wh-word *hai* 'who'. Note first that the two nominal wh-words, *hai* 'who' and *hā* 'what' behave differently with respect to the kind of morphemes they can and must co-occur. While *hā* behaves like other noun, requiring both a determiner and a case marker (or a preposition), *hai* cannot take a determiner. I take this to suggest that *hai* is a D-head, while *hā* is a N-head. The distributional behavior of *fē* is quite similar to that of *hai*: it cannot take a determiner, but must always be preceded by a preposition (though, unlike *hai*, it cannot co-occur with a case marker). Based on this, I propose that *fē* is actually a locational pronoun belonging in the paradigm provided in Table 5. Being a D-head, *fē* can freely merge with the predicate marker *ko* to occur as the predicate of a PC construction.

proximal	medial	distal	interrogative
heni	*hena*	*hē*	*fē*

TABLE 5. TONGAN LOCATIONAL PRONOUNS

Time adverbial wh-words, *'afē* 'when (future)'and *'anefē* 'when (past)' are slightly different from *fē* on the surface in that they never co-occur with a preposition. However, it should also be noted that the locative preposition *'i*, which is used for temporal nouns as well as locational nouns, can be and often is omitted when followed by a determiner, as illustrated in (13). I propose that *'afē* 'when (future)'and *'anefē* 'when (past)' are temporal pronouns and bear a D-feature and that the preposition *'i* is obligatorily omitted before them.

(13a)	Te	u	'alu	('i)	he	taimi-ni
	FUT	1S	go	in	DET	time-this
	'I'm going now.'					
(13b)	Na'a	ku	ha 'u	('i)	he	'aho-ni
	PST	1S	come	in	DET	day-this
	'I came today.'					

Table 6 below summarizes the inventory of wh-words in Tongan and their feature specification.

	hā[9] 'what'	*hai* 'who'	*fē* 'where'	*ʻafē/ ʻanefē* 'when'	*fēfē* 'how'	*fiha* 'how many'
[D]	–	+	+	+	–	–
[PRED]	–	–	–	–	+	+
[WH]	+	+	+	+	+	+
[OP]	–	–	–	–	–	–

TABLE 6. FEATURE SPECIFICATION OF TONGAN WH-WORDS

5 Wh-questions in Tagalog

Like Tongan, Tagalog is predicate-initial. In Tagalog, NPs are marked by one of the prenominal markers which inflect for case, specificity, and personhood (Table 7), except when occurring as predicates in nominal constructions. Descriptively, Tagalog has two-way case system, core, marking core arguments, and oblique (OBL). Verbal morphology correlates with the semantic role of *ang*-marked NP, e.g., ***bum****ili* (*ang* agent/actor), ***bin****ili* (*ang* patient/theme), and ***bin****ili****han*** (*ang* location/goal).

	CORE [+SPECIFIC]	CORE [–SPECIFIC]	OBL	GEN
Common nouns	*ang*	*ng*	*sa*	*ng*
Personal names	*si*	*ni*	*kay*	*ni*

TABLE 7. TAGALOG PRENOMINAL MARKERS.

Wh-words can be divided into three classes: core, oblique, and adverbial (Table 8).[10] Traditionally, *sino* 'who' and *ano* 'what' are regarded as equivalent to *ang*-forms and *nino* 'who', *ng*-forms. Oblique wh-forms are also used in combination with a preposition: e.g., *na sa ano* 'in/with what', *para sa ano* 'for what', *para kanino* 'for whom'.

[9] The distributional property of *hā* suggests that it is unspecified for the category; it must co-occur with a category-defining functional category, either a determiner, the predicate marker *ko*, or even a TAM.

[10] See Schachter and Otanes 1972 for a comprehensive list of Tagalog wh-words. One of the important forms omitted in this table is genitive, *nino* [+person] and *ng ano* [-person], to be discussed in Section 5.2 below.

Core		Oblique		Adverbial	
sino	'who', 'whom'	*kanino*	'to whom'	*kailan*	'when'
(*nino*	'who')	*sa ano*	'to what'	*saan*	'where'
ano	'what'			*paano*	'how'

TABLE 8. TAGALOG WH-WORDS

5.1 Wh-question strategies in Tagalog

Wh-words cannot occur in situ in verbal constructions, as illustrated in (14).

(14a) Declarative

Binili ng babae ang bigas sa tindahan.
bought DET woman DET rice OBL store
'A/the woman bought the rice at the store.'

(14b) *Patient wh-in situ

*Binili ng babae **ano**?
bought DET woman what.ANG
Intended: 'What did a/the woman buy?'

(14c) *Adverbial wh-in situ

*Binili ng babae ang bigas **saan**?
bought DET woman DET rice where
Intended: 'Where did a/the woman buy the rice?'

However, wh-in situ is permitted if the wh-word is the predicate of a nominal construction, as illustrated in (15).

(15a) **Ano** iyon
what.ANG that
'What is that?' (lit. 'That is what?')

(15b) **Sino** si Pedro?
who.ANG DET Pedro
'Who is Pedro?' (lit. 'Pedro is who?')

Since the in-situ strategy is not available in verbal construction, we may expect wh-movement in Tagalog, as least in verbal constructions. This prediction is only partially borne out. While wh-movement is required of oblique and adverbial wh-forms (Richards 1998, Aldridge 2002, 2004) as shown in (16), core argument

wh-words may not undergo wh-movement (17).[11] These examples are formed from the corresponding declarative sentence by simply moving the wh-phrase to the sentence-initial position.

(16a) ADVERBIAL WH-FRONTING

Saan	binili	ng	babae	ang	bigas ___?
where	bought	DET	woman	DET	rice

'Where did a/the woman buy the rice?'

(16b) OBLIQUE WH-FRONTING

Kanino	mo	ibinigay	ang	pera	____?
who.OBL	2S	gave	DET	money	

'Who did you give the money to?'

(16c) OBLIQUE WH-FRONTING

Sa ano	mo	ibabalot	ang regalo	____?
OBL what	2S	wrap.FUT	DET present	

'What will you wrap the present in?'

(17a) *CORE ARGUMENT WH-FRONTING

***Ano**	binili	ng	babae	____?
what.ANG	bought	DET	woman	

Intended: 'What did a/the woman buy?'

(17b)

***Nino**	binili	____	ang	bigas?
who.NG	bought		DET	rice

Intended: 'Who bought the rice?'

(17c)

***Sino**	bumili	____	ng	bigas?
who.ANG	bought		DET	rice

Intended: 'Who bought (the) rice?'

Thus, the PC strategy is the only means to form core argument wh-questions in Tagalog. Furthermore, the PC strategy is constraint in the reverse fashion: the PC strategy must be used with core argument wh-words, but may not be used with other kinds of wh-forms (18).[12]

---

[11] The use of *nino* as an actor wh-word is possible, but very rare (Schachter & Otanes 1972: 512; Kroeger 1993: 212; Richards 2010: 181-182). Actor-wh questions are usually formed as a PC construction, using *sino* as in (18a) below.

[12] On the surface, wh-words occur sentence-initially in both examples of wh-movement and those of PC. The only apparent difference is that in PC examples, wh-words are followed by *ang*. There is

(18a) Core-wh (*ang*) PC

[PRED Sino] [SBJ ang bumili ____ ng bigas]?
who.ANG DET bought DET rice
'Who bought (the) rice?' (lit. '(the one who) bought (the) rice is who?')

(18b) *Adverb-wh PC

*[PRED Saan] [SBJ ang binilihan ng babae ang bigas ____]?
where DET bought DET woman DET rice
Intended: 'Where did a/the woman bought the rice?'

(18c) *Oblique-wh PC

*[PRED kanino] [SBJ ang ibinigay mo ang pera ____]?
who.OBL DET gave.PF 2S DET money
Intended: 'To whom did you give the money?'

Note also that unlike the *ang*-form *sino*, the *ng*-form *nino* cannot occur in a PC wh-question, regardless of the verbal morphology, as shown in (19). We will return to this point shortly.

(19a) * Core (*ng*)-wh PC with actor voice (AV) verb

*[PRED **nino**] [SBJ ang bumili ____ ng bigas]?
who.NG DET bought.AV DET rice
'Who bought (the) rice?' (lit. '(the one who) bought (the) rice is who?')

(19b) *Core (*ng*)-wh PC with patient voice (PV) verb

*[PRED **nino**] [SBJ ang binili ____ ang bigas]?
who.NG DET bought.PV DET rice
'Who bought (the) rice?' (lit. '(the one who) bought the rice is who?')

5.2 Analysis of Tagalog wh-questions

Table 9 summarizes the available strategies for and constraints on wh-questions in Tagalog. Notably, Tagalog is not well behaved either as a wh-movement language or a wh-in situ language. Movement is required of certain wh-questions, while it is prohibited for certain others. Like in Tongan, there is a dichotomy. However, in Tagalog, the contrast is between core arguments (specifically, *sino* and *ano*) and others.

---

independent morphosyntactic evidence to suggest that oblique/adverbial wh-questions are monoclausal (hence involving wh-movement to [Spec, C]) and that nominal wh-questions are bi-clausal (hence wh-words are not in [Spec, C]). See Aldridge 2004 for more discussion.

	CORE		OBL	Adverb
	ang	*ng*		
In situ	No	No	No	No
Movement	No	No	Yes	Yes
Pseudo-cleft	Yes	No	No	No

TABLE 9. SUMMARY OF WH-QUESTION STRATEGIES IN TAGALOG

This distribution of wh-question strategies raises three questions. First, what makes it impossible to use the PC strategy to form oblique and adverbial wh-questions? Second, why are *ano* and *sino* disallowed to undergo wh-movement while oblique and adverbial wh-words are required to do so? And third, why is *nino* unable to undergo wh-movement or to occur in a PC construction?

Let us consider the first question. There is an independent reason why oblique and adverbial wh-words cannot occur in PC wh-questions. Recall that the subject NP of a PC contains a relative clause. The predicate NP corresponds to the gap in the relative clause modifying the subject NP. In Tagalog, relativization is constrained in such a way that only *ang*-marked NPs can be relativized. This immediately explains why oblique wh-phrase cannot occur in PC wh-questions: the relevant construction contains an illicit relative clause, as oblique-phrases cannot be relativized. As briefly mentioned above, in Tagalog, various verbal inflections designate a particular NP as the *ang*-marked NP. Therefore, locative/goal NPs can be *ang*-marked with appropriate verbal inflections (locative and banefactive voice, respectively). This predicts that locative/goal wh-questions may be formed using the PC strategy with appropriate verbal morphology. This turns out to be true. When this happens, however, an *ang*-form (*ano/sino*) must be used instead of the oblique forms (*saan*/*kanino*), as shown in (20).[13]

[13] According to Nozomi Tanaka (pers.comm. September 2015), the use of *saan* instead of *ano* in (20a) is acceptable to some speakers, but with a different interpretation. With *ano*, the question is about the kind of place and a felicitous answer would be something like 'at the shop/market'. On the other hand, with *saan*, the question is about the general geographical location: 'Where is the place at which she bought rice located?' The felicitous answer in that case would be 'in Manila/town' and cannot be a specific place such as 'store'. Prescriptively, in locational wh-questions 'where is X?', another wh-form *nasaan* is used: *nasaan si Pedro*? 'Where is Pedro?' It should be noted that not all speakers accept the use of *saan* in (20a) (Ivan Bondoc, pers.comm. November 2015). It is likely that for those speakers who permit *saan* to occur in (20a), *saan* and *nasaan* are interchangeable; and that when *saan* is used in (20a) it is intended as a locational construction.

(20a) Locative voice (LV) PC wh-question

[PRED Ano] [SBJ ang binilihan ng babae ng bigas ___]?

what.ANG DET bought.LV DET woman DET rice

'What is (the place where) a/the woman bought (the) rice?'

(20b) Goal voice (GV) PC wh-question

[PRED Sino] [SBJ ang binigyan mo ng pera ____]?

who.ANG DET gave.GV 2S DET money

'Who is (the person to whom) you gave money?

Now, let us turn to the real mystery: the distribution of wh-movement and wh-in situ. First, the movement strategy must be used to form oblique and adverbial wh-questions. I take this to mean that these wh-forms bear both the operator feature [OP] and the variable feature [WH], and that C has an EPP-feature to license the movement. The hypothesis that C has an EPP-feature runs into a problem when we consider the fact that *ano/sino* cannot undergo wh-movement. To circumvent this problem, I propose that *ano/sino* lack the operator feature, hence cannot agree with C in that respect. This in turn predicts that *ano/sino* should be allowed to occur in situ (with a null operator generated in [Spec, C]). However, this again is only partially true. While they can occur in situ as a nominal predicate of a PC construction, *ano/sino* cannot remain in situ in verbal constructions. This suggests that something else prohibits *ano/sino* from occurring in an argument position.

As mentioned above, traditionally, *ano* and *sino* are regarded as the *ang*-form of 'what' and 'who'. They are the only forms that can be used in PC wh-questions, where the gap in the relative clause can only correspond to the *ang*-marked argument. However, this view fails to provide an elegant solution to the aforementioned mystery concerning wh-questions. Thus, I propose instead that *ano* and *sino* are not nominal at all; rather, they are predicative forms and therefore lack a categorial D-feature.[14] Assuming that arguments are DPs and that their merge with a verbal head (V or v) is licensed by a c-selectional feature [*u*D], this explains why *ano* and *sino* cannot occur in an argument position. It also explains why they can occur as a predicate in PC constructions.

The distribution of *sino* clearly supports this analysis. *Sino* can only occur in

[14] Thanks to Shigeo Tonoike's (pers. comm. September, 2015) insightful observation that *ano/sino* can appear only in the predicate position.

predicate positions. In other positions, the other form, *nino* is used as in *kanino* (oblique) and *nino* (genitive). *Ano* is not that straightforward, however. In fact, Schachter and Otanes (1972: 507-509) note that *ano* has multiple functions in addition to the predicative use. For one thing, it can be combined with a preposition, e.g., *sa ano* 'to what' and *para sa ano* 'for what'. Intriguingly, when used as a nominal, *ano* apparently must always co-occur with the oblique marker *sa*. And when combined with *sa*, it must undergo wh-movement, suggesting that in this case, *ano* bears an operator feature. A more accurate description of the distribution of *ano* is, then, that unlike other nouns, it cannot co-occur with *ang* or *ng*, i.e., as a core argument of a verb. This is obviously a stipulation, but not an outrageous one.

In fact, when comparing the non-adverbial wh-forms and prenominal markers (Table 7), it appears that core as well as oblique argument wh-forms are all derived from *ano*. This is obvious in the non-personal set, in which the oblique and genitive forms are clearly bimorphemic, e.g., *sa* (OBL) + *ano*. The personal set can also be shown to be fused forms of *ano* and a prenominal marker: *sino* → *si* [core;+specific; +person] + *ano*; *nino* → *ni* [GEN; +person] *ano*; *kanino* → *kay* [OBL; +person]+ *ano*. This suggests that *ano* is simply a morphological realization of a variable feature [WH] and must be merged with some other features to be turned into a lexical item. An interesting twist is that *ano* can actually be used with *ang/si* as in (21). In that case, however, *ano* can only be interpreted as indefinite (Schachter & Otanes 1972).[15] Based on this, I propose that *ano* is a root bearing only a variable feature [WH] and that the interrogative *ano* is derived by adding an operator feature [OP] to this root.[16]

---

[15] Neither *sino* nor *nino* cannot be used for this purpose. Nor can they be interpreted as a multiple wh-question, 'Who is where?; for that meaning, *sino* must occur as the predicate as in (iii). This further supports the hypothesis that *sino* is a predicate, not a nominal. I thank Nozomi Tanaka and Ivan Bondoc (pers.comm. September 2015) for sharing the relevant data.

(i) *na-saan sino?
where who.ANG
Intended: 'Where is whatshisname?

(ii) *na-saan si nino?
in-where DET who
Intended: 'Where is whatshisname?'

(iii) sino ang na-saan ___?
who.ANG DET in-where
'Who is where?' (lit. 'The one that is where is who?')

[16] Tagalog roots are argued to be category free: generally the same form can be used as N, V, or Adj (Himmelmann 2008, Kaufman 2009 among others).

(21a) na-saan ang ano?
where DET what
'Where is whatchamacallit?' but not 'where is what?'

(21b) na-saan si ano?
in-where DET what
'Where is whatshisname?' but not 'where is who?'

Table 10 summarizes the classification of wh-words in Tagalog. Tagalog wh-words are operators and therefore must undergo wh-movement. However, *sino* 'who' and *ano* 'what' can only occur in a predicate position due to their [+PRED] feature (and the lack of D-feature). This results in apparent prohibition on wh-movement from an argument position.[17] Adverbial wh-forms cannot occur in PC constructions due to an independent constraint that restricts relativization to *ang*-marked NPs.

	PRED	OBL	GEN
[–person]	*ano*	*sa ano*	*ng ano*
[+person]	*sino*	*kanino*	*nino*

TABLE 10. FEATURE SPECIFICATION OF TAGALOG WH-WORDS

6 Conclusion

In this paper, I have proposed that wh-questions consist of three semantico-syntactic components: an operator, a variable, and C with a set of features [Q], [*u*WH], and [*u*OP]. The combination of these features is interpreted as an instruction to select an individual from a set. The two uninterpretable features on C ensure that the structure contains a variable [WH] and an operator [OP] in order to establish an operator-variable structure. I have also argued that all wh-words are variables, bearing a variable feature [WH], but only some of them are operators, bearing a feature [OP]. Following the standard analysis of the typology of wh-strategies, I have proposed that wh-movement occurs when a wh-word bears a feature [OP] and C has an EPP-feature. On the other hand, wh-phrases remain in situ when they lack [OP], in which case, C's EPP-feature is checked by a null operator base-generated in [Spec, C].

With this background, I examined Tongan and Tagalog wh-questions. Tongan

[17] I remain agnostic about the position of these predicative wh-words.

turns out to be a well behaved wh-in situ language. Wh-movement is strictly prohibited and all kinds of wh-words remain in situ. With respect to the PC strategy, Tongan makes a strange distinction between nominal and adverbial wh-questions on the one hand and predicative wh-questions on the other. This turns out to be due to the fact that what appear to be "adverbial" such as 'where' and 'when' in Tongan are in fact nominal, specifically locational and temporal pronouns. Table 11 summarizes the wh-question strategies in Tongan.

	[OP]	in situ	movement	PC
Nominal wh	–	Yes	No	Yes
Predicative wh	–	Yes	No	No

TABLE 11. WH-QUESTION STRATEGIES IN TONGAN

The situation in Tagalog is a bit more complicated. First, wh-movement is obligatory for oblique and adverbial wh-questions, but prohibited for core argument wh-questions. Second, wh-phrases are not allowed to stay in situ. Third, the PC strategy is only available for core argument wh-questions. The main mystery is why core argument NPs cannot undergo wh-movement. I argued that this puzzling behavior of core argument wh-questions is due to the fact that what appear to be nominal wh-phrases are actually predicates; due to their categorial feature [PRED], they fail to merge with V or *v* as an argument. The unavailability of wh-movement of *ano*/*sino* is due to their failure to be base generated in an argument position to begin with.

The second mystery concerning the inability of oblique and adverbial wh-phrases to occur in PC wh-questions can be readily explained in terms of an independent, language-specific constraint on relativization: only *ang*-marked core arguments can be relativized. Since oblique and adverbs cannot be relativized as such, it is impossible to form the headless relative that serves as the subject of the presumed PC construction. See Table 12 for summary.

	[OP]	in situ	movement	PC
PRED	+	No	No	Yes
OBL	+	No	Yes	No
Adverbial	+	No	Yes	No

TABLE 12. WH-QUESTION STRATEGIES IN TAGALOG

Altogether, Tongan and Tagalog data support the crosslinguistic

generalization that only those wh-forms that are operators can and must undergo wh-movement. Apparent prohibition on wh-movement in Tagalog core argument wh-questions has shown not to be an exception to the rule; rather, it is due to their categorial status of predicate and the absence of D-feature. It is also notable that while the use of PC strategy is limited in both languages, the limitation is due to independent factors that are relevant to the structure of PC (i.e., the subject containing a relative clause and the predicate being a nominal). This seems to suggest that the PC strategy is not an alternative to complement the unavailability of a particular strategy (be it wh-movement or wh-in situ), but is expected to be generally available in all language that allows PC in general.

7 References

Aldridge, Edith. 2002. Nominalization and wh-movement in Seediq and Tagalog. *Language and Linguistics* 3: 393–427.

Aldridge, Edith. 2004. Ergativity and word order in Austronesian languages. Doctoral dissertation, Cornell University.

Bauer, Winifred. 1991. Maori *ko* again. *Te Reo* 24: 31–36.

Bauer, Winifred. 1993. *Maori*. London: Routledge.

Besnier, Niko. 2000. *Tuvaluan*. London: Routledge

Chang, Melody Y. 2000. On Tsou wh-questions: movement or in situ? *Language and Linguistics* 1: 1–18.

Cheng, Lisa. 1991. On the typology of wh-questions. Doctoral dissertation, MIT.

Chomsky, Noam. 1977. On wh-movement. In *Formal Syntax*, ed. by Peter Culicover, Tom Wasow and Adrian Akmajian, 71-132. New York: Academic Press.

Chomsky, Noam. 2000. Minimalist Inquiries: The Framework. In *Step by step: Essays in Minimalist Syntax in honor of Howard Lasnik*, ed. by Robert Martin, David Michaels and Juan Uriagereka, 89-155. Cambridge, MA: The MIT Press.

Cole, Peter and Gabriella Hermon. 1998. The typology of *wh*-movement: *Wh*-questions in Malay. *Syntax* 1: 221-258.

Cole, Peter, Gabriella Hermon, and Yassir Tjung 2005. How irregular is WH in situ in Indonesian? *Studies in Language* 29: 553-581.

Custis, Tonya. 2004. Word order variation in Tongan: a syntactic analysis. Doctoral dissertation, University of Minnesota.

Georgopoulos, Carol. 1991. *Syntactic variables: resumptive pronouns and*

A'-binding in Palauan. Dordrecht: Kluwer.

Himmelmann, Nikolaus P. 2008. Lexical categories and voice in Tagalog. In *Voice and grammatical relations in Austronesian languages*, ed. by Simon Musgrave and Peter Austin, 247–293. Stanford, CA: CSLI Publications.

Kaufman, Daniel. 2009. Austronesian Nominalism and its consequences: A Tagalog case study. *Theoretical Linguistics* 35:1–49.

Kroeger, Paul R. 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford: CSLI Publications.

Nishigauchi, Taisuke. 1990.*Quantification in the theory of grammar*. Dordrecht: Kluwer Academic.

Otsuka, Yuko. 2000. Ergativity in Tongan. Doctoral thesis, University of Oxford.

Paul, Ileana. 2000. Malagasy clause structure. Doctoral dissertation, McGill University.

Paul, Ileana. 2001. Concealed pseudo-clefts. *Lingua* 111: 707–727.

Potsdam, Eric. 2006a. The cleft structure of Malagasy wh-questions. In Clause structure and adjuncts in Austronesian languages, ed. by Hans-Martin Gärtner, Paul Law, and Joachim Sabel, 195–232. Berlin: Mouton de Gruyter.

Potsdam, Eric. 2006b. More concealed pseudoclefts and the Clausal Typing Hypothesis. *Lingua* 116: 2154–2182.

Potsdam, Eric and Maria Polinsky 2011. Questions and Word Order in Polynesian. In *Topics in Oceanic Morphosyntax*, ed. by Claire Moyse-Faurie and Joachim Sabel, 83-109. Berlin: Mouton de Gruyter.

Reinhart, Tanya. 1998. Wh-in-situ in the framework of the minimalist program. *Natural Language Semantics* 6: 29–56.

Richards, Norvin. 1998. Syntax versus semantics in Tagalog wh-extraction. In UCLA occasional papers in linguistics 21: Recent papers in Austronesian linguistics, ed. by Matthew Pearson, 259–275. Los Angeles: University of California, Los Angeles Department of Linguistics.

Richards, Norvin. 2010. Uttering trees. Cambridge, MA: MIT Press.

Schachter, Paul and Fe T. Otanes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.

Seiter, William J. 1980. *Studies in Niuean syntax*. New York: Garland.

Tonoike, Shigeo. 2015. A general theory of wh-questions. Paper presented at the Linguistics Department Tuesday Seminar. March 2015. Honolulu,

University of Hawai‘i at Mānoa.