

SPEC OF IP AND SPEC OF VP:  
TWO SUBJECTS IN AUSTRONESIAN LANGUAGES\*

In this paper we claim that certain facts from four Austronesian languages provide striking support for the claim that there is a VP-internal subject position in addition to the traditional subject position of SPEC of IP. We show (i) that subject-sensitive properties may be split between SPEC of VP and SPEC of IP, theta-sensitive properties (such as binding and control) associated with the former and structure-sensitive properties (such as extraction and quantifier float) with the latter; (ii) that word order facts about these four Austronesian languages as well as variations among the languages follow from the assumption that the two subject positions may be licensed simultaneously at S-structure. Lastly, we show that the proposed analysis of Austronesian phrase structure allows the placement of Austronesian languages within a typology of passive.

In the past few years, a number of studies have taken advantage of the uniformity of the X' system of phrase structure to argue for the existence of two subject positions. More precisely, they have provided different types of evidence for establishing a VP-internal specifier position in addition to the traditional SPEC of IP position. These include Kuroda (1988), Kitagawa (1986), Fukui and Speas (1986), and Koopman and Sportiche (1988).

In this paper we present further evidence for postulating the existence of two subject positions using data from four Austronesian languages: Malagasy, Tagalog, Cebuano, and Bahasa (Malaysia and Indonesia).<sup>1</sup> The main attractions of assuming such an analysis are (i) that it suggests a *structural* explanation for the split in subject-like properties which has been a traditional problem for the analysis of subject in these languages;

---

\* The research presented in this paper is the result of a truly collaborative effort, and the authors owe a debt of gratitude to many people during the writing of this paper. The authors wish to thank the audiences at McGill University, Brandeis University, WCCFL VIII and GLOW 1989, where this material has been presented. We have received helpful comments from the other members of the Malayo-Polynesian group at McGill, Mark Campana and Anna Maclachlan; and from Sandra Chung, Hans den Besten, Carol Georgopoulos, Eric Hoekstra, Randy Kamp, Paul Kroeger, Barry Miller, George Rebuschi, Paul Schachter and three anonymous NLLT reviewers. We would also like to thank Zofia Laubitz and Ben Shaer for editorial help. All errors and omissions are the responsibility of the authors. The research for this paper was supported by FCAR grant #88EQ3630, and SSHRCC grant #410-87-1071.

<sup>1</sup> 'Bahasa' literally means 'the language (of)' so the terms Bahasa Malaysia and Bahasa Indonesia refer to the standard dialects of these two countries. Although there are some differences between the two dialects, the properties we discuss here are common to both.

(ii) that it accounts for the facts of word order (and word order variation) among these four languages; and (iii) that it provides a place for these languages within a typology of passive.

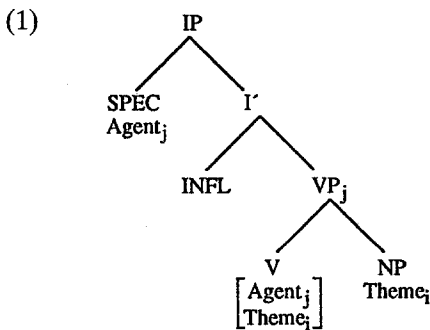
At the heart of this analysis are the assumptions that both IP and VP have specifier positions, that all arguments are base-generated within VP, and that the SPEC of IP is a non-theta position. If an argument is Case-marked in its theta position, it may remain there; if not, it must raise to the SPEC of IP position where it is Case-marked via SPEC-Head Agreement with INFL.

In addition to providing an account for word order facts, this analysis provides strong support for our original assumption that all theta assignment is carried out within VP at D-structure. Finally we extend this discussion by examining how Austronesian languages fit into the phrase structure typology of natural languages.

## 1. THEORETICAL BACKGROUND

### 1.1. *The Canonical Position of the Subject NP*

Within Government-Binding (GB) theory, one way of accounting for the behavior of subjects as opposed to objects has been to represent the subject as being 'external' to the verb phrase. The subject receives its theta role from the VP (rather than the verb), and is immediately dominated by IP (or S in earlier versions of the theory). This kind of analysis, which is assumed in Chomsky (1981) and most recent work in GB theory, is illustrated in (1) below:



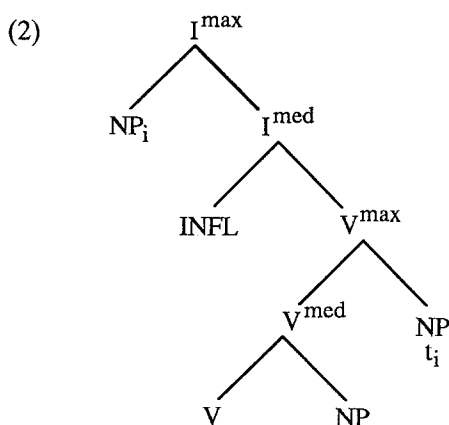
The subject position receives its theta role from VP via predication (see, e.g., Williams (1981)), while the object is directly theta marked by the verb. As the subject position is not necessarily theta marked by the VP,

it does not always receive a theta role. Thus, unlike any other position within IP, the subject position may be the target of NP movement, or may be occupied by an expletive NP.

### 1.2. *The VP-Internal Subject Hypothesis*

While the phrase structure shown in (1) above is still widely accepted, a number of researchers have proposed that subjects are actually generated within the projection of the verb phrase, and raise to Spec of IP at S-structure. This proposal is often termed the VP-internal subject hypothesis.

Kitagawa (1986), extending an idea of Kuroda's (published in Kuroda (1988)), proposes this kind of analysis for English. Using evidence from extraposition, quantification facts, and constituency tests, Kitagawa proposes that English sentences have the structure shown in (2) below:



In (2) the verb phrase is strictly head-initial and the subject is generated as the sister to  $V^{med}$ . The subject raises to the Spec of IP at S-structure in order to receive Case from INFL. A 'subject-raising' approach to English clause structure has also been adopted by Koopman and Sportiche (1988) and Fukui and Speas (1986). Koopman and Sportiche show that various types of extraction and quantification phenomena can be accounted for under the VP-internal subject analysis. Although the various analyses proposed by the authors mentioned above differ in many details, they all assume that English subject NPs raise from a position inside VP to the Spec of IP position, and that this movement is motivated by the Case requirements of the subject NP.

SPEC of IP receives Case at S-structure, and therefore the subject must

occupy this position at S-structure in order to receive Nominative Case and so avoid violating the Case Filter. However, Kuroda, Kitagawa, and Koopman and Sportiche all propose that subject raising is not obligatory in all languages. Kuroda and Kitagawa suggest that Japanese subjects may remain within the VP at S-structure. Koopman and Sportiche make a similar claim for Italian and VSO languages. The analyses proposed so far, then, assume that languages vary with respect to whether the VP or IP subject position is filled at S-structure.

In this paper we explore another logical possibility provided by the VP-internal subject hypothesis; namely, that languages may allow *both* subject positions to be filled at S-structure simultaneously. The assumption of this possibility allows us to explain many properties of subjects and word order in these languages which otherwise defy ready explanation, and to sharpen our notion of the differences between these two positions; thus it provides strong evidence for the VP-internal subject hypothesis.

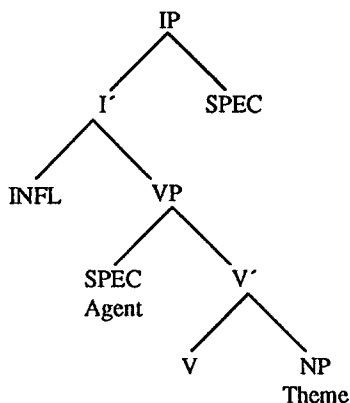
### 1.3. *SPEC of IP, SPEC of VP, and Theta Role Assignment*

Before considering the nature of these two subject positions, we outline our three major assumptions about theta assignment, which are the following: First, all theta roles of the verb are assigned at D-structure within the VP. This means that the SPEC of IP is a non-theta position. Second, theta roles are assigned in a uniform manner, as stated in the Uniformity of Theta Assignment Hypothesis of Baker (1988). Thus, morphologically related verb forms or, more specifically, the various complexes formed by affixing different 'voice' markers to a given stem, will assign their theta roles in an identical fashion. Third, hierarchical prominence relations within the theta grid (as discussed in Grimshaw and Mester (1988)) are mirrored in the structural relations within the VP. Thus, the most prominent theta role will be assigned to the SPEC of VP in such a way that it c-commands the rest of the arguments. (We might also point out here that while we will be using theta role labels such as Agent and Theme, these are probably a simplification of the actual mapping from theta roles to syntactic positions. For instance, although 'most prominent' theta role might be a more accurate term than Agent, we will continue to use these terms for the sake of convenience.)

Given these three assumptions, the picture which emerges is one where the SPEC of IP is not filled at D-structure. At S-structure, an NP which is not Case-marked in its base-generated position within VP may raise to SPEC of IP where it receives Case via SPEC-Head agreement with INFL. We propose, then, that the D-structure common to the Austronesian

languages we are discussing (Malagasy, Bahasa, Cebuano and Tagalog) has the following representation (modulo the orientation of the SPEC of IP).

(3) **The Proposed D-Structure<sup>2</sup>**



According to this tree there are two subject positions, the SPEC of VP, to which a theta role is assigned, and the SPEC of IP, to which an NP may move to receive Case. Which NP stays in the VP and which must raise are determined entirely by the Case-assigning properties of both V and INFL. In order to complete the picture, we assume that the verb moves to INFL in all four languages. This is evident in the three V-initial languages, Malagasy, Tagalog and Cebuano, and assumed in Bahasa although nothing hinges on this assumption.

#### 1.4. *Malagasy: NP Licensing and NP Movement*

In the previous section, we presented the assumptions that guide us in determining D-structure. In this and following sections we will see what mechanisms act on the D-structure to create the S-structure representations. We begin by looking at the active, passive, and circumstantial constructions in Malagasy, where it can be clearly seen how the presence of a particular morpheme in V or INFL can license material in a particular NP position via Case-assignment. A change in the morphological character of the V + INFL complex is thus accompanied by a change in its Case-

<sup>2</sup> Note that SPEC of IP is to the right and SPEC of VP to the left. We assume that this is a distinction between specifiers of functional categories and specifiers of lexical categories. Unfortunately, since there is no evidence for SPEC of CP (questions take the form of clefts or predicate structures (Seiter 1975)) or SPEC of DP (we assume that possessors remain in the SPEC of NP), we do not have strong support for this claim.

assigning properties, resulting in a change in the word order and structural analysis.

Suppose we take the active and the passive of a regular transitive verb such as 'wash', which has two arguments, an Agent and a Theme. The relevant sentences are given in (4) below. Traditionally, the NP in sentence-final position has been called either the Topic or the Focus NP, and the particular morphology on the verb has been called either Topic or Focus morphology. As will become clear below, we assume that the Topic NP is in the SPEC of IP.<sup>3</sup>

- (4)a. M-an-sasa (manasa) ny lamba amin'ny savony **ny zazavavy**.  
*AT-wash the clothes with the soap the girl*  
 The girl washes the clothes with the soap.
- b. Sasa-na (sasan') **ny zazavavy** amin'ny savony **ny lamba**.  
*wash-TT the girl with the soap the clothes*  
 The clothes are washed with the soap by the girl.

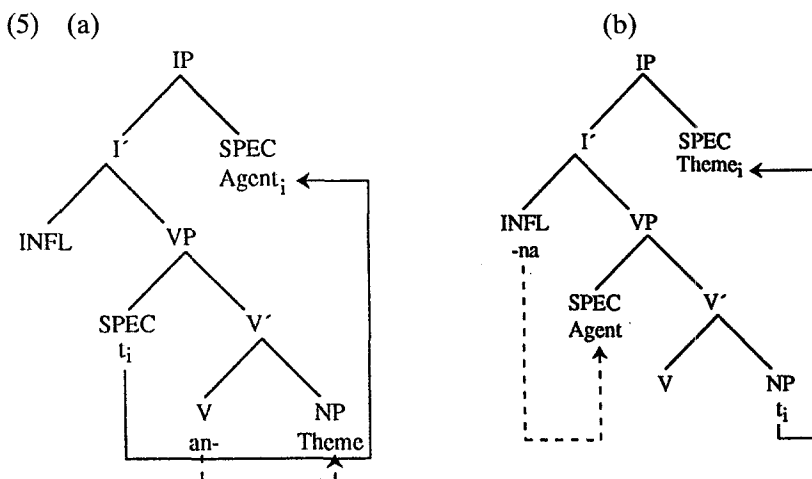
In (4a) the Agent is Topic, as indicated by the sentence-final position of the Agent, as well as by the particular morphology on the verb.<sup>4</sup> In (4b) the Theme is Topic. Assuming that these two verb forms have a common D-structure, we can derive the word order differences in the following way. Hung (1988) claims that the prefix *an-* of *manasa* in (4a) assigns Case to the Theme and hence licenses the Theme NP in its theta position. Since the Agent is not Case-marked in the SPEC of VP position, it is forced to move to the SPEC of IP where it is assigned Case via SPEC-Head Agreement with INFL. Hung claims further that the suffix *-na* on the verb *sasana* (in 4b) is in INFL and assigns Case to the Agent in the SPEC of

<sup>3</sup> In these examples, the Topic is underlined in the data, and the Agent is in boldface. The labels that we use may differ from the labels used by Schachter and others. We will, however, indicate verbal morphology as being AT (Agent-Topic, since Agent is in SPEC of IP) and TT (Theme-Topic, since Theme is in SPEC of IP). The abbreviations used in the paper are given below:

ACC – accusative	AT – Agent-Topic
BT – Benefactive-Topic	DT – Directional-Topic
GEN – genitive	IT – Instrumental-Topic
LK – linker	LT – Location-Topic
OBL – oblique	PL – plural
RP – recent past	TOP – topic
TT – Theme-Topic	XT – non-Agent/non-Theme Topic (Malagasy)

<sup>4</sup> Further tests for Topics are discussed in 2.1.1 and 2.1.2.

VP position, hence licensing the Agent NP in its theta position. The Theme, however, is not assigned Case and is forced to move to the SPEC of IP. The S-structure trees corresponding to (4a) and (4b) are shown in (5a) and (5b).<sup>5</sup>



The strongest evidence in favor of the claim that the presence of a particular morpheme licenses material in a particular NP position comes from the existence of a verb form that combines the elements of (4a) and (4b). When the verb bears the two affixes *an-* and *-na*, as in *anasana* (*an + sasa + na*), both Agent and Theme are licensed in their respective theta positions, and a third NP appears in the specifier of IP position.<sup>6</sup> This has been traditionally called the circumstantial form of the verb (see, e.g., Keenan (1976)). An example of this is given in (6) (we will call the circumstantial topic morphology XT since it may range over several theta roles).

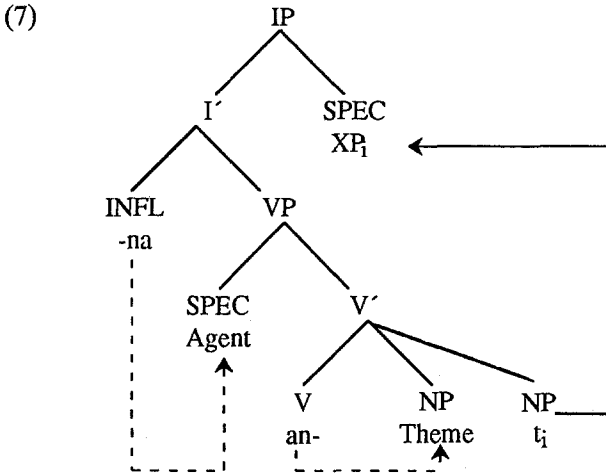
- (6) *An-sasa-na* (*anasan'*) **ny** *zazavavy* ny *lamba* ny savony.  
*XT-wash*                      *the girl*              *the clothes* *the soap*  
 The soap was washed (with) the clothes by the girl.

Hung's analysis of (6) is given in (7). Since both Agent and Theme are

<sup>5</sup> In these and subsequent trees, the solid arrows indicate movement and the dotted arrows indicate case assignment.

<sup>6</sup> For each verb one particular theta role is usually the unmarked choice for this third argument; this would be encoded in the lexical conceptual structure of the verb (see Hung 1988).

Case-marked within VP, a third NP, in this case an instrumental NP, must appear in the SPEC of IP position.<sup>7</sup>



Under this account, then, Topic morphology consists simply of Case-assigning morphemes.<sup>8</sup> AT *an-* assigns Case to the Theme, forcing the Agent to move to the SPEC of IP, TT *-na* assigns Case to the Agent, forcing the Theme to move. Circumstantial Topic, *an- . . . -na* is a combination of these two, meaning that both Theme and Agent receive Case within the VP, and a third NP moves to the SPEC of IP.

In all three constructions, verb movement to INFL also occurs. This verb movement, together with our analysis of Agent as being generated in the SPEC of VP, also explains why only the Agent is allowed to appear in a position immediately to the right of the verb, as in (4b) and (6), violating the strict adjacency condition which normally holds between the

<sup>7</sup> At this point we assume that the disappearance of the preposition in (6) is due to Preposition Incorporation (as in Baker's (1988) analysis of applicatives) in addition to the appearance of the two Case-assigning morphemes, *an-* and *-na*. By assuming that preposition incorporation is obligatory with the circumstantial form of the verb, we can explain why there must be an NP that moves to the SPEC of IP to get Case. We have represented the structure with ternary branching; however, we do this only for the sake of convenience since the actual structure within the V' is not relevant to our discussion.

<sup>8</sup> Note that this analysis of Topic morphology has an advantage over other analyses such as those that analyse Topic markers as agreement markers or incorporated pronouns. These analyses require that the Topic NP be indexed with the morphology on the verb and cannot neatly account for the combined effect of the morphology observed in the circumstantial construction.

verb and the Theme. As we can see in example (8) below, no other maximal projection may intervene between a verb and its object.<sup>9</sup>

(8)a. \*Nanasa omaly ny lamba ny zazavavy.

*past-AT-wash yesterday the clothes the girl*

The girl washed the clothes yesterday.

b. \*Nanasa ho an'ny ankizy ny lamba ny zazavavy.

*past-AT-wash for the children the clothes the girl*

The girl washed the clothes for the children.

In this analysis of Malagasy, even though the Agent appears to interrupt the adjacency condition on structural Case assignment, the trace of the verb is still adjacent to its object. This provides additional support for the assumption that verb movement to INFL takes place.<sup>10</sup>

### 1.5. Consequences of the Analysis

If we compare the NP positions that are, respectively, string adjacent to the verb and sentence-final, we see that the former can be occupied only by the NP bearing the most prominent theta role – usually the Agent – as long as it is licensed there. On the other hand, the NP occupying the sentence-final position may bear one of a number of theta roles. The theta consistency of the first position is explained by the fact that this position is filled at D-structure, in contrast to the theta variable Topic position, which is filled at S-structure. As we have said, the position adjacent to the verb at S-structure is the SPEC of VP, which is assigned the most prominent theta role by the verb at D-structure. The sentence-final position is the SPEC of IP, which is filled by NP movement for reasons related to Case Theory. In a way, both NPs are structural subjects. The SPEC of VP is the external argument of the V' and receives its theta role from the constituent which contains the verb and the other arguments of the verb. The SPEC of IP is also a subject, being in an argument position external

<sup>9</sup> Adverbs, however, may appear in this position as in *Manasa mafy ny lamba ny zazavavy* 'The girl washed the clothes hard'. This can be accounted for by assuming that adverbs adjoin to VP (see, e.g., Pollock (1989)), or that adverbs are heads adjoined to INFL (see Travis (1988)). Thanks to George Rebuschi for pointing out this possibility to us.

<sup>10</sup> We need the additional assumption here that internal arguments and adjuncts are generated to the right of the V, as in English. This again shows that 'demoted' Agents in a language like Malagasy are quite different from their counterparts in English, since they appear in a position in which no other argument or adjunct may appear.

to the VP where it is assigned case by INFL. As we see below, these two subject positions correspond to different syntactic phenomena.

## 2. SUBJECTS IN AUSTRONESIAN LANGUAGES

In some Austronesian languages, no single NP consistently has all of the properties often associated with subjects. Some subject properties are characteristic of the NP which bears the theta role Agent, while others are characteristic of the NP which is identified by verbal morphology as the Topic of the sentence. This division of properties has led some researchers to claim that the Agent NP is the 'true' subject (Byma 1987, Carrier-Duncan 1985, Gerdtz 1988), while others argue that the Topic is the subject (Travis and Williams 1983). Schachter (1976) concluded his discussion of subject properties by saying that the notion of subject is irrelevant to any analysis of Philippine languages because subjects simply do not exist in these languages; thus the existence of subjects cannot be considered a universal in any sense. Within GB theory, given issues such as the Extended Projection Principle, the derivative notion of subject, the problem of *pro*-drop languages, etc., it is not clear what it means to say that a language is without a subject. In this section we consider how our analysis of word order in Austronesian allows us to explain the 'unusual' properties of Austronesian subjects without being forced to claim that these languages 'lack a subject'.

### 2.1. *Some Properties of Subjects*

In his discussion of the subject in Philippine languages, Schachter (1976) outlines a number of properties associated with subjects in the literature. These properties are, in some cases, tied to a certain syntactic theory but might be seen as guidelines for determining the subject in any given language. Four of these properties and their relevance for the languages investigated here are discussed below. Two syntactic phenomena, Wh-extraction and quantifier float, are sensitive to the Topic of the sentence in all four languages under discussion. In Tagalog and Cebuano, however, reflexivization is sensitive to the Agent NP, and Equi-NP deletion to either the Topic or the Agent.

#### 2.1.1. *Property I: Extraction facts*

In (9) to (15) below, we see that the Wh-questioning of an NP requires the verb to bear Topic morphology 'agreeing' with the theta role class of

the extracted NP. In (9) there are two examples from Bahasa where Agent-Topics have been extracted, but only one is grammatical. In the grammatical (9a), the Agent-Topic form of the verb *membaca* appears, whereas the ungrammatical (9b) has the Theme-Topic verb form *dibaca*.

(9) **Bahasa**

- a. Siapa yang membaca buku itu?  
*who COMP AT-read book the*
- b. \*Siapa yang buku itu dibaca?  
*who COMP book the TT-read*  
 Who read the book?

Similarly, in the Tagalog<sup>11</sup> and Malagasy examples shown in (10) and (11), only the Agent-Topic forms of the verbs are grammatical, because the Agent has been extracted in each case.<sup>12</sup>

(10) **Tagalog**

- a. Sino ang bumili ng damit para sa bata'?  
*who COMP AT-bought ACC-dress for OBL-child*
- b. \*Sino ang binili para sa bata' ang damit?  
*who COMP TT-bought for OBL-child TOP-dress*
- c. \*Sino ang ibinili ng damit ang bata?  
*who COMP BT-bought ACC-dress TOP-child*  
 Who bought the dress for the child?

(11) **Malagasy**

- a. Iza no mividy ny vary ho an'ny ankizy?  
*who COMP AT-buy the rice for the children*
- b. \*Iza no vidina ho an'ny ankizy ny vary?  
*who COMP TT-buy for the children the rice*

<sup>11</sup> Information questions in the Philippine languages may be analyzed as clefts or as nominalizations, as suggested by Seiter (1975). This is immaterial to the discussion at hand since, no matter what the analysis, it is always the Topic NP which is targeted.

<sup>12</sup> In the Tagalog examples, *ng* is glossed as ACC and as GEN. We assume that it is the same Case-marker, but that in the former case it is assigned by the V to its object, and in the latter by morphology in INFL to the SPEC of VP. Genitive Case (assigned by D to the SPEC of NP) is also *ng*. It is important to note, however, that we consider it a structural Case-marker and not a preposition.

- c. \***Iza** no ividiana ny vary ny ankizy?  
*who COMP XT-buy the rice the children*  
 Who bought the rice for the children?

In the examples shown in (12) to (14), a Theme-Topic has been extracted. Here the verb must bear Theme-Topic morphology for the sentence to be grammatical. Thus, only the (b) examples below are grammatical, since only they contain verbs bearing Theme-Topic morphology.

(12) **Bahasa**

- a. \***Apa** yang Ali membaca?  
*what COMP Ali AT-read*
- b. Apa yang dibaca **Ali**?  
*what COMP TT-read Ali*  
 What was read by Ali?

(13) **Tagalog**

- a. \***Ano** ang bumili para sa bata' ang tao?  
*what COMP AT-bought for OBL-child TOP-man*
- b. Ano ang binili **ng tao** para sa bata'?  
*what COMP TT-bought GEN-man for OBL-child*
- c. \***Ano** ang ibinili **ng tao** ang bata?  
*what COMP BT-bought GEN-man TOP-child*  
 What was bought for the child by the man?

(14) **Malagasy**

- a. \***Inona** no mividy ho an'ny ankizy ny lehilahy?  
*what COMP AT-buy for the children the man*
- b. Inona no vidin' **ny lehilahy** ho an'ny ankizy?  
*what COMP TT-buy the man for the children*
- c. \***Inona** no ividianan' **ny lehilahy** ny ankizy?  
*what COMP XT-buy the man the children*  
 What was bought for the children by the man?

In Malagasy and Tagalog, the verb may have Topic morphology that

designates a Topic other than the Agent or the Theme. In Tagalog, the verb has a number of different affixes which can identify Locations, Goals, Benefactives, and other relations as the Topic. Malagasy has circumstantial morphology which appears on the verb whenever any element other than Agent or Theme appears in Topic position. In (15) and (16), a Benefactive has been extracted; in each case only the (c) examples have verbs bearing the appropriate morphology, and thus only these are grammatical.

(15) **Tagalog**

- a. \*Sino ang bumili ng damit ang tao?  
*who COMP AT-buy ACC-dress TOP-man*
- b. \*Sino ang binili ng tao ang damit?  
*who COMP TT-buy GEN-man TOP-dress*
- c. Sino ang ibinili ng tao ng damit?  
*who COMP BT-buy GEN-man ACC-dress*  
 Who was bought the dress (for) by the man?

(16) **Malagasy**

- a. \*Iza no mividy vary ny lehilahy?  
*who COMP AT-bought rice the man*
- b. \*Iza no vidin' ny lehilahy ny vary?  
*who COMP TT-bought the man the rice*
- c. Iza no ividianan' ny lehilahy ny vary?  
*who COMP XT-bought the man the rice*  
 Who was bought rice (for) by the man?

The examples of Wh-movement given in (9)–(15) thus demonstrate that in these Austronesian languages, only the Topic can be the target of such movement.<sup>13</sup>

2.1.2. *Property II: Floating Quantifiers*

Floating quantifiers also single out the Topic NP since they must be associated with the Topic regardless of its theta role. In the Cebuano

<sup>13</sup> Relativization and clefting work in a similar manner.

examples in (17), we see two possible non-floated forms of the phrase meaning ‘all the children’, in which the quantifier *tanán* appears, respectively, before (as in (17a)) and after (as in (17b)) the head noun *bata* ‘children’. In (17c), however, the quantifier has ‘floated’ to a position outside the NP.<sup>14</sup>

(17) **Cebuano**

- a. Ningdagan ang tanán nga bata'. (Bell 1976, (38a))  
*AT-run TOP all LK child*  
 All the children ran off.
- b. Ningdagan ang batang tanán. (Bell 1976, (38b))  
*AT-run TOP child-LK all*  
 All the children ran off.
- c. Floated Form<sup>15</sup>  
 Ningdagan tanán ang bata'. (Bell 1976, (38c))  
*AT-run all Top child*  
 The children all ran off.

In (18) below, the quantifier has also floated out of its noun phrase – which in this example is the Theme-Topic. So, no matter whether the topic is the Agent, as in (17c), or Theme, as in (18), the quantifier is associated with the Topic.

(18) **Cebuano**

- Gibasa tanán sa mga istudiante ang mga libro ni Rizal.  
*TT-read all pl student TOP PL book Rizal*  
 Rizal's books were all being read by the students.  
 (Bell 1976, (39b))

Schachter presents the case for identifying Topic NPs as subjects, based on the evidence from extraction and quantifier float as presented here, and the relation of these phenomena to certain claims about subjects. These are that subjects will be affected by relativization if any grammatical relation is (Keenan and Comrie 1977); and that, in the framework of Relational Grammar, only ‘terms’ may launch floating quantifiers (Bell

<sup>14</sup> See Bell (1976) for a more detailed discussion of the data presented here.

<sup>15</sup> This quantifier is assumed to be positioned outside the NP since it precedes the Topic Marker *ang*. This is clearest in examples such as (18) where there is intervening material.

1976). These observations, Schachter suggests, may be handled “by identifying the topic as the subject” (Schachter 1976, p. 501). Schachter observes, however, that the picture becomes more complicated when reflexivization and Equi-NP deletion are included. As we discuss below, in a subset of the languages, reflexivization is sensitive to the Agent, and Equi-NP deletion to both the Topic and the Agent.

### 2.1.3. *Property III: Reflexivization*

In Tagalog and Cebuano, only the Agent can always be the antecedent of a reflexive (whether or not it is the Topic) and the Agent itself never bears reflexive morphology. In the examples in (19) below we see that the reflexive form *kanilang sarili* takes the Agent NP as its antecedent whether or not this NP is the Topic.

- (19) **Tagalog** (Schachter 1976, (22) and (23))
- a. Nagiisip            **sila**            sa kanilang sarili.  
*AT-think-about TOP-they OBL-their self*  
 They think about themselves.
  - b. Iniisip            **nila**            ang kanilang sarili.  
*TT-think-about GEN-they TOP-themselves*  
 They think about themselves.
  - c. \*Iniisip            **ng kanilang sarili** sila.  
*TT-think-about GEN-themselves TOP-they*  
 \*Themselves think about them.

Since reflexivization is often sensitive to subjecthood (e.g., *zibun* in Japanese), one might conclude that the Agent in Tagalog is always the subject. While this already complicates the picture that was beginning to emerge in which the Topic NP appeared to have subject-like properties, more complications arise when Equi-NP deletion is investigated.

### 2.1.4. *Property IV: Equi-NP Deletion*

In Tagalog and Cebuano, Equi-NP deletion may act on either the Topic or the Agent of the embedded clause. When the Topic of the embedded clause is dropped, it may bear a number of theta roles. This is shown in (20) below.

(20) **Tagalog**<sup>16</sup>

- a. Natatakot ang batang maipakilala mo sa akin.  
*afraid TOP child-LK TT-introduce GEN-you OBL-me*  
 The child is afraid of being introduced by you to me.
- b. Natatakot kang ipakilala ang bata' sa akin.  
*afraid TOP-you AT-introduce TOP child OBL-me*  
 You are afraid of introducing the child to me.

When the Agent of the lower clause is dropped, it makes no difference which NP is Topic. In the sentences in (21) below we find that the Agent is always null regardless of whether the embedded verb bears Agent-Topic (21a), Theme-Topic (21b) or Direction-Topic (21c) morphology:

(21) **Tagalog** (Schachter 1976, (24a-c))

- a. Nagatubili siyang humiram ng pera sa bangko.  
*AT-hesitated he AT-borrow ACC-money OBL-bank*  
 He hesitated to borrow the money from the bank.
- b. Nagatubili siyang hiramín ang pera sa bangko.  
*AT-hesitated he TT-borrow TOP-money OBL-bank*  
 He hesitated to borrow the money from the bank.  
 (Literally: He hesitated the money to be borrowed from the bank by \_\_\_\_.)
- c. Nagatubili siyang hirañan ng pera ang bangko  
*AT-hesitated he DT-borrow ACC-money TOP-bank*  
 He hesitated to borrow the money from the bank.  
 (Literally: He hesitated the bank to be borrowed money from by \_\_\_\_.)

Within GB theory, the phenomenon of Equi-NP deletion – which is described in terms of the presence of PRO in control structures – follows from government theory. Since PRO may appear only as the subject of an uninflected INFL, its presence is, in effect, restricted to the SPEC of

<sup>16</sup> Thanks to a reviewer for supplying these data from Dell (1981), which were also accepted by our informants. See Kroeger (1991) for an interesting alternative analysis for the control facts in Tagalog.

IP position. In Tagalog, however, PRO appears in two different positions which, in our account, are SPEC of VP and SPEC of IP.

## 2.2. *Towards an Account of the Split in Subject Properties*

If we attempt to use the four subject properties discussed above as our only criteria for determining subjecthood, we are faced with a dilemma. According to properties I (extraction) and II (quantifier float), it is the Topic which is the subject; according to property III (reflexivization) it is the Agent, and according to property IV (Equi-NP deletion) it is both. This dilemma might be solved if we appeal to the analysis of Austronesian phrase structure discussed in Sections 1 and 2.

The split in Austronesian subject properties reflects the difference in properties associated with the SPEC of VP and SPEC of IP positions. Since Agents are always base-generated in the SPEC of VP, they will always exhibit the properties associated with that position. Likewise, Topics will exhibit the properties associated with the SPEC of IP position which they occupy.

It is interesting to observe, moreover, that such a split is not unique to the Austronesian language family. Noting again that reflexivization and control in certain Austronesian languages may refer to the theta sensitive position SPEC of VP, and extraction and quantifier float to the theta variable position SPEC of IP, we can see that a similar split has been observed in more well-studied languages such as English. Williams (1987, p. 151), in a work which “explore[s] the possibility that the binding theory (plus control) applies not to NPs, but to theta roles that are assigned to NPs”, comes to the conclusion that “the picture of grammar that emerges is one in which there are two components, one concerned exclusively with relations between theta roles (theta role assignment, the theta criterion, and binding theory) and one concerned with positions (quantification and Wh-movement).”

While the match between Williams' proposal for control of theta roles and our proposal for control of SPEC of VP is not perfect, it is significant that in both cases binding and control fall on one side of the dividing line, and Wh-movement and quantification on the other side. Unlike Williams, however, we wish to claim that both sets of relations are subject to structural positions, one to SPEC of VP and the other to SPEC of IP.

### 2.2.1. *Properties Associated with SPEC of VP*

The properties that we can associate with SPEC of VP are properties III and IV, i.e., reflexivization and control. Within GB theory these are properties that are subject to binding theory, and binding phenomena, according to the suggestion of Williams (1987), as just described, sometimes appear to be theta sensitive. This analysis is compatible with the data in (19) to (21). Thus we note that Agents are always possible antecedents of reflexives, and that they may appear as PRO.<sup>17</sup> However, as Agents are always base-generated in SPEC of VP, it is possible to maintain a purely structural account of these binding facts, with PROs and the antecedents of reflexives being associated with the *structural* position SPEC of VP.<sup>18</sup>

### 2.2.2. *Properties Associated with SPEC of IP*

The properties that we associate with SPEC of IP are properties I and II, i.e., extraction and quantifier float. In order to account for the interpretation of floating quantifiers, we assume that they are licensed in the same way as adverbs – in particular, subject-oriented adverbs (see, e.g., McConnell-Ginet (1982)).<sup>19</sup> Travis (1988) proposes that adverbs like *reluctantly* in example (22) below are adjoined to INFL and licensed by the AGR feature in INFL. Since this feature is coindexed with the NP (which is the subject) in the SPEC of IP position, it has the effect of relating the adverb to this NP, and thereby giving subject-oriented readings for the adverbs in (22). Thus, it is Mary who is reluctant in (22a), and the students in (22b).

<sup>17</sup> The examples in (21) suggest that the position SPEC of VP is not always governed or Case-marked by the verb in Tagalog, since PRO may appear in this position. We are, therefore, forced to say that Case-assignment to SPEC of VP is optional. A similar claim has also been made, however, for subjects in non-finite clauses in Irish by McCloskey (1984) and Chung and McCloskey (1987).

<sup>18</sup> We do not suggest, however, that our analysis will also account for the facts that Williams (1987) presents, where it appears that the Theme theta role may be controlled within nominals.

<sup>19</sup> Note that Sportiche's (1988) account of Quantifier Float does not have an obvious use in these languages. Floated quantifiers always appear immediately post-verbally, independent of what the theta role (and presumably D-structure position) of the related NP is.

- (22)a. Mary reluctantly was instructing the students.
- b. The students reluctantly were being instructed by Mary.

Given this analysis, it is not surprising that the quantifier in Austronesian languages should be sensitive to the NP in the SPEC of IP position, since it is this NP with which AGR will be coindexed. In the Tagalog example below, *lahat* 'all' is adjoined to INFL and licensed by the AGR feature in INFL. Since INFL is coindexed with the NP in the SPEC of IP, the quantifier will be interpreted with the Topic NP *mga libro* and not *mga bata*'.

- (23) **Tagalog** (Schachter and Otnes 1972, p. 148)

binasang lahat ng mga bata' ang mga libro.  
*TT-read-LK all ACC-pl child TOP-pl book*

- a. The children read all of the books.
- b. \*All the children read the books.

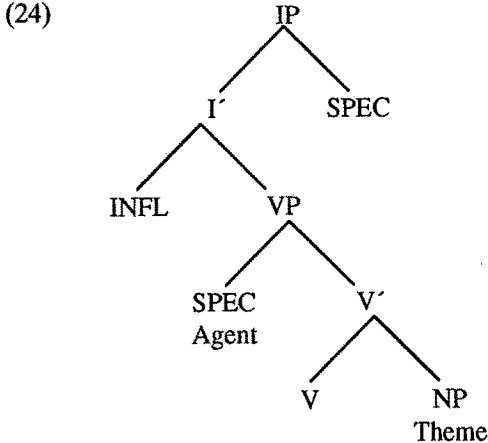
As for extraction, we want to explain why extraction only occurs from the SPEC of IP. Although an account of this is beyond the concerns of this paper, the phrase structure that we have proposed for these languages offers a clear direction for a solution. Extraction obviously requires some type of locality to prevent movement out of the VP across a filled SPEC of IP. This locality might be determined within the framework of relativized minimality (Rizzi 1990) or barriers (Chomsky 1986a); we leave this as an open question for future research.

### 2.2.3. *Summary*

What emerges from our analysis of extraction, quantifier float, reflexivization and control is that these phenomena, though sensitive to subjects in some languages, appear to be split between two NPs in some Austronesian languages. Extraction and quantifier float target Topic NPs, reflexivization targets Agent NPs, and Equi-NP deletion targets either. Such a split of properties is not surprising given the phrase structure we have proposed. Nor is our claim that both sets of phenomena have structural accounts, despite the apparent sensitivity of some of the properties to theta roles.

### 3. WORD ORDER VARIATION AMONG AUSTRONESIAN LANGUAGES

In this section we show that the presence of non-topic Agents in SPEC of VP can explain certain word order facts. Consider again the D-structure tree in (3), repeated below, which we claim to be basic to Malagasy, Bahasa, Tagalog and Cebuano (abstracting away from minor differences).



The discussion of Malagasy in Section 1.4 above suggests an idea of how the verbal morphology might play a role in the licensing of NPs. Now we will turn to a discussion of some of the properties of the other Austronesian languages which distinguish them from Malagasy.

#### 3.1. *Tagalog*

Tagalog has fairly free word order, and has been claimed to be a nonconfigurational language (Carrier-Duncan 1985). Nevertheless, it seems that it must be configurational at some level if we are to explain the constraints on Wh-questions, relative clauses, and cleft structures described in Section 2. We will assume that at D-structure the order of constituents is Agent-V-Theme-X, as shown in (24). Although factors such as the relative length of each word and the pronominal or full NP status of various arguments may influence surface word order, preferred orders (as judged by our informants) do emerge when these factors are controlled for. These orders are shown in (25) (based on Schachter 1976).<sup>20</sup>

<sup>20</sup> Different informants allow different amounts of variation. We assume that there must be some sort of scrambling rule and, in particular, PP extraposition, to account for the various permutations of the orders of post-verbal elements. Across informants, however, non-topic

- (25)a.(i) Mag-aalis ng bigas sa sako para sa bata'  
*AT-take-out ACC-rice OBL-sack for OBL-child*

ang babae.

*TOP-woman*

The woman will take rice out of the sack for the child.

- (ii) Mag-aalis ang babae ng bigas sa sako para  
*AT-take-out TOP-woman ACC-rice OBL-sack for*

sa bata'.

*OBL-child*

The woman will take rice out of the sack for the child.

- b. Aalisin ng babae sa sako para sa bata'  
*TT-take-out GEN-woman OBL-sack for OBL-child*

ang bigas.

*TOP-rice*

The rice will be taken out of the sack for the child by the woman.

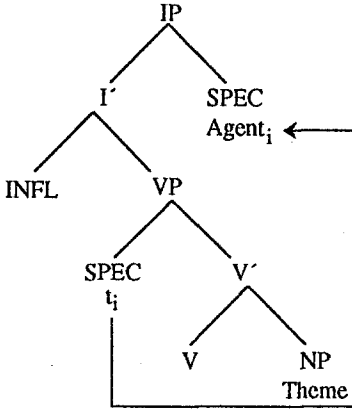
As these examples show, Tagalog is very much like Malagasy except for the particle *ang* which marks the Topic of the clause, and an interesting word order variation which occurs in Tagalog, but not Malagasy. This variation is shown above in (25ai,ii) where the verb has the same Agent-Topic form in both instances.<sup>21</sup> According to our analysis, while V must raise to INFL, the Agent may move to SPEC of IP, producing the order shown in (25ai), or remain in the SPEC of VP, producing the order shown in (25aai). The corresponding trees are given in (26).

---

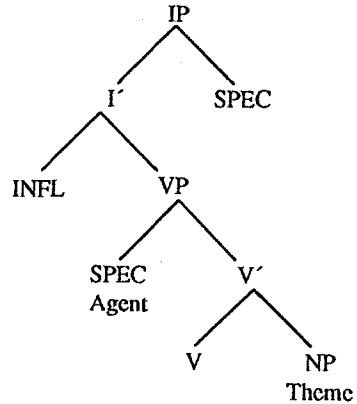
Agent is singled out as the preferred element to follow the verb. (Thanks to an anonymous reviewer for input on this question.) This is easily explained if these are always found within the SPEC of VP. Further, in cases where all the arguments of the verb are lexically realized, the only Topic that can appear adjacent to the verb is the Agent Topic. Again this follows from the fact that Agent is generated in SPEC of VP and is allowed to remain in this position even if it is the designated Topic.

<sup>21</sup> Cebuano appears to allow only the word order shown in (25aai). Our thanks to an anonymous reviewer for pointing this out to us.

(26)a. Agent moved to SPEC  
of IP



b. Agent remaining in  
SPEC of VP



The Theme, however, does not have a choice of positions when it is Topic. The unmarked word order is one in which the Topic NP has moved, as shown above in (25b). Our explanation of this asymmetry has to do with the relation of the Agent to INFL. In (26a) INFL Case-marks its specifier via SPEC-Head Agreement, as in Malagasy. In (26b), however, INFL governs down into the SPEC position of its VP complement and licenses the Agent NP in this position.<sup>22</sup>

Note that, after V has moved into INFL and before a Topic has moved to the SPEC of IP position, constituents exhibit the following order: V-Agent-Theme-X. This suggests that a sentence containing a verb that does not designate a Topic should also have this word order, which is what we find. Below is an example of just such a sentence, where the verb is in the 'recent past,' a form without any special topic-marking morphology.

- (27) Kakakain **ng leon** ng tigre.  
*RP-ate GEN-lion ACC-tiger*  
 The lion ate the tiger.  
 \*The tiger ate the lion.

If the word order were completely free, it ought to be possible for either *ng leon* or *ng tigre* to be the Agent.<sup>23</sup> In fact, *ng leon* must be the Agent, and the sentence is ungrammatical on the reading 'The tiger ate the lion'.

<sup>22</sup> The ability of INFL to assign Case to the SPEC of VP has already been proposed by Guilfoyle (1988) for VSO languages such as Irish.

<sup>23</sup> Genitive Case assigned by the V to its SPEC is homophonous with ACC assigned by the V to its complement.

These data suggest that the Agent precedes the Theme at an underlying level.<sup>24</sup>

### 3.2. *Bahasa Malaysia and Bahasa Indonesia*

Bahasa Malaysia and Bahasa Indonesia are examples of the SVO members of the Austronesian language family; they nevertheless resemble Malagasy and Tagalog in also permitting a non-demoted Agent alongside a non-Agent Topic. In this section we will discuss properties which are found in both dialects, and refer to these dialects collectively as Bahasa. Among its various sentence constructions, Bahasa has the active, the subjective passive, the canonical passive with the Agent in a *by*-type phrase headed by the preposition *oleh*, and the canonical passive with the Agent but without *oleh*. In this section we will attempt to show how a unified analysis of all four of these constructions follows from a common D-structure with two specifier positions: a theta position in VP and a non-theta position in IP.

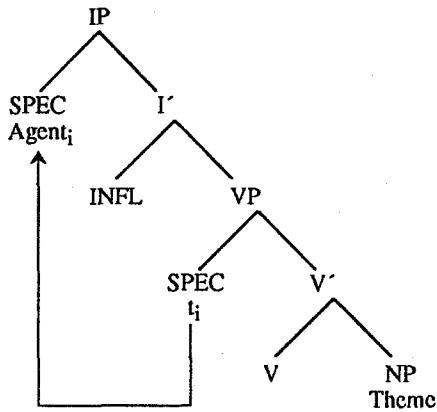
Unlike Malagasy and Tagalog, Bahasa is a subject-initial language. Except for this minor difference, we assume the same D-structure proposed earlier. The active construction is fairly straightforward since it coincides with the Actor-Topic construction, and the analysis should be familiar except for the SVO surface word order. (Data are from Hung (1987) unless otherwise indicated.)

- (28) Ali/ Saya/ Lelaki itu membaca buku itu dengan teliti.  
*Ali/ I/ boy the AT-read book the with care*  
 Ali/I/the boy read the book carefully.

In accordance with our analysis, the presence of the verbal prefix *meN-* is associated with the appearance of the Theme in VP and the Agent in the SPEC of IP behaving much like *an-* in Malagasy. While the existence of verb movement to INFL is not evident in the active construction, it will be necessary to assume that it takes place in the formation of the passive, as we will see. The S-structure representation of the active construction is shown below in (29). The Theme receives Case within the VP, while the Agent must move to the SPEC of IP to receive Case from INFL.

<sup>24</sup> While it might be argued that, in free word order languages, Agents precede Themes for pragmatic reasons, we are assuming here that there are also structural reasons for this word order preference.

(29)



For a detailed description of the passive construction in Bahasa, we refer the reader to Chung (1976) and Sie (1989). They report three types of passive: the subjective passive, the canonical passive without the agentive preposition *oleh*, and the canonical passive with the agentive preposition *oleh*. On this third type of passive, speakers are divided between the conservative and the liberal camps. The conservative speakers allow only a third person NP to appear as the object of *oleh*. The liberal speakers impose no such constraints. Let us now turn to each type of passive, noting the particular restrictions which apply to each.

We begin with the subjective passive, which has the following characteristics: the surface word order is Theme-Agent-Verb, the verb bears no morphology, and the Agent is restricted to being a pronoun, clitic, or proper name.

- (30)a. Anjing itu Ali/ saya/ ku- pukul.  
*dog the Ali/ 1 sg. pronoun/ 1 sg. clitic hit*

The dog was hit by Ali/me.

- b. \*Anjing itu lelaki itu pukul.  
*dog the boy the hit*

The dog was hit by the boy.

The canonical passive without *oleh* has the following characteristics: the surface word order is Theme-*di*-Verb-Agent, the verb bears the prefix *di-*, and there is a strict adjacency requirement which holds between the verb and the Agent.

- (31)a. Budak-budak itu telah ditangkap **polis**. (Othman 1981, p. 103)  
*child-PL the ASP TT-catch police*  
 The children have been caught by the police.
- b. Buku itu dibaca **lelaki itu**.  
*book the TT-read boy the*  
 The book was read by the boy.
- c. Ahmad dipukul **Ali** dengan kayu. (Lewis 1969, p. 49)  
*Ahmad TT-hit Ali with stick*  
 Ahmad was hit by Ali with a stick.
- (32)a. \*Uang itu dikirim kepada Tomo **Ali**. (Sie 1989, p. 62)  
*money the TT-send to Tomo Ali*
- b. Uang itu dikirim **Ali** kepada Tomo. (Sie 1989, p. 62)  
*money the TT-send Ali to Tomo*  
 The money was sent to Tomo by Ali.

The canonical passive with *oleh* has the following characteristics: the surface word order is Theme-*di*-Verb-*oleh*-Agent, the verb bears the prefix *di-*, and the object of the preposition *oleh* is interpreted as the Agent. Note that the entire PP may be absent, and that intervening material may appear between the verb and the PP, if it is present. Conservative speakers require that the object of *oleh* be in the third person; liberal speakers do not.<sup>25</sup>

- (33)a. Buku itu dibaca.  
*book the TT-read*  
 The book was read.
- b. Buku itu dibaca oleh **Ali/ lelaki itu**.  
*book the TT-read by Ali/ boy the*  
 The book was read by Ali/the boy.

<sup>25</sup> An anonymous reviewer has pointed out to us that the constraints observed by the set of conservative speakers holds true of Javanese.

- c. Sepuluh dolar sudah di-bayar kepada tukang rumput  
*ten dollars ASP TT-pay to worker grass*  
 oleh saya.<sup>26</sup>  
*by 1sg*

Ten dollars has been paid to the gardener by me.

It has been suggested that the presence versus absence of the agentive preposition might be due to a rule of *oleh* deletion or the generation of an empty preposition.<sup>27</sup> We reject these proposals on the grounds that neither can explain why intervening material can appear between the verb and the PP as in (33c) but not between the verb and the Agent as in (32a).

The analysis we propose to account for these different types of passive assumes that they are all derived from a common D-structure in which the verb assigns an Agent theta role to the NP in the SPEC of VP, and a Theme theta role to the NP in the complement position. Under our account, the canonical passive with or without *oleh* (excluding the dialect spoken by the more liberal speakers) is analyzed in the same way as the subjective passive. This is accomplished first by postulating that the prefix *di-* is the realization of the set of features representing the third person. Support for this comes from the fact that the third person singular subject pronoun is *dia*. The second assumption is that NPs are, in fact, dominated by DP, the maximal projection of the functional category DET, and that the Case-marking relation between a DP and a V<sup>0</sup> may be created by the movement of the head of the DP to a V in INFL.<sup>28</sup> Of course, due to locality constraints, not all DPs in the VP are eligible to enter into this relationship with INFL. The DP in the SPEC of VP bears a special relation to INFL, as we have seen in Malagasy and Tagalog.

The subjective passive construction is derived from a similar structure. In a construction like *Anjing itu ku-pukul* 'I beat the dog', what is present at D-structure in the SPEC of VP is a set of features in the head of DP representing the first person singular Agent, *ku-*. These features, if realized in the SPEC of VP, will not be case-marked, since there are no potential Case-markers nearby. Instead, the head of DP, or DET, may raise to INFL and be realized there as *ku-* on the verb which has also been

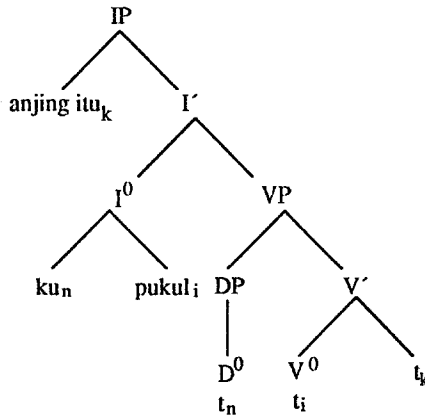
<sup>26</sup> This example was provided to us by an anonymous reviewer.

<sup>27</sup> Thanks to Georges Rebuschi for pointing this out to us.

<sup>28</sup> We will henceforth refer to NPs as DPs unless referring to traditional concepts such as NP-trace or the function of an NP.

raised. In this way, a Case-marking relation has been created between the Agent and Verb,<sup>29</sup> and the Theme moves to the SPEC of IP.

(34) Subjective Passive (see 30a)



Two comments need to be made. First of all, notice that in this analysis the verb is never truly 'bare'. It is always associated with some head, lexical (*meN-*) or functional (D). Perhaps this explains why we never find sentences like the ones below:

- (35)a. \*Buku itu **PRO** baca.  
*book the PRO read*
- b. \*Buku itu **PRO** baca oleh Ali.  
*book the PRO read by Ali*

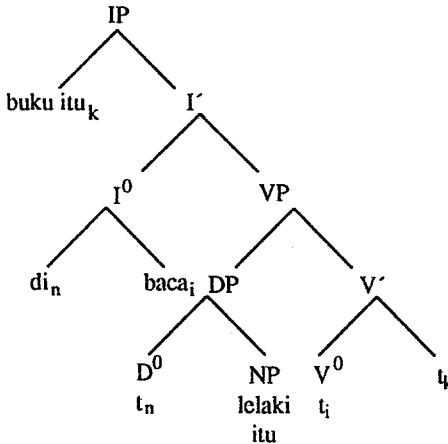
Second, we stated that only pronouns, clitics, and proper names like *Ali* could appear as the Agent in the subjective passive construction. This can be accounted for if we assume the claim made in Postal (1969) that pronouns and proper names are located in the position of the determiner.

The same thing happens in a canonical passive construction like *Buku itu di-baca lelaki itu* 'The boy read the book'. *Lelaki itu* 'the boy' is assigned the role of Agent in the SPEC of VP. The contents of the head of the DP are the features for the third person singular. (There is independent evidence that the determiner *itu* is only a modifier, not a functional head.) The head of the functional category DP, or DET, then raises to INFL, leaving behind any material that might be in the lexical

<sup>29</sup> This is similar to Baker's (1988) views on Case-assignment and incorporation. If a noun has been incorporated into a verb, it does not need to be Case-marked in the usual manner.

projection NP. In INFL, DET is spelled out as *di-* and combines with the verb which also raises to INFL, giving *di-baca*. Again, a Case-marking relation has been established between the Agent and the verb.<sup>30</sup> The Theme, *buku itu*, moves to the SPEC of IP in the usual manner.

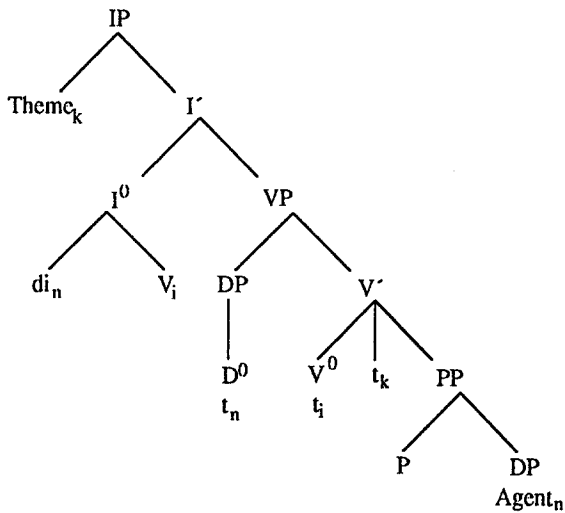
(36) Canonical passive without *oleh* (see (31b))



The adjacency requirement that holds between the verb and the Agent is by now a familiar one; it is a requirement that holds between INFL and the SPEC of VP.

We turn next to the canonical passive with the agentive preposition *oleh* of the conservative speakers. This construction offers an alternative way in which to add to the reference of an unspecified third person Agent, namely with the addition of an *oleh*-phrase. As with the cases described above, the set of third person features in DET raises to INFL where it is realized as *di-* on the verb which has also raised. The *oleh*-phrase may or may not be present. *Oleh* itself assigns an Agent theta role to its object, and Case-marks it as well. This specified Agent serves to double the unspecified Agent of the verb, so the two must be coindexed, and thus must agree in person and number.

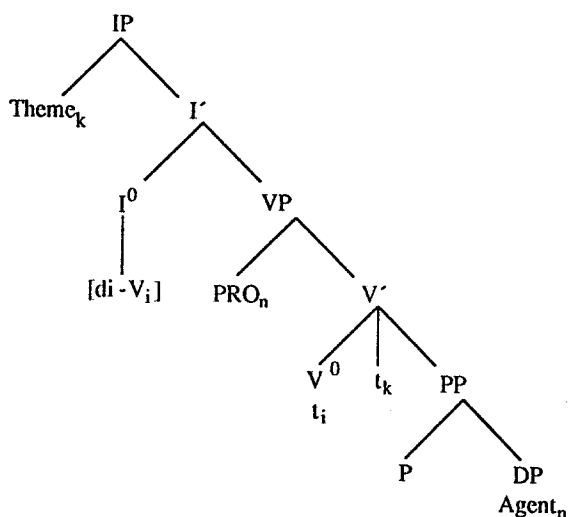
<sup>30</sup> Note that a proper name can also be 'stranded' as an NP, indicating that a proper name may or may not be analyzed as being located in the head of DP, e.g. *Buku itu di-baca Ali*, 'The book was read by Ali'.

(37) Canonical passive with *oleh* (conservative) (see (33b))

In this case, intervening material may occur between the verb and the Agent. This is because the Agent is in a prepositional phrase under VP or V' and bears no Case-marking relation to the verb which is in INFL.

Given this account, it is reasonable to say that the liberal speakers seem to have lost the original characterization of the morpheme *di-* as an expression of the third person.<sup>31</sup> For them, the morpheme *di-* is like the English passive morpheme *-en*, and we propose that the analysis is therefore much the same. Following Fukui and Speas (1986) and Guilfoyle (1990), we assume that the passive morpheme is in V, that it is not a Case-marker, and that the only thing that can appear in the SPEC of VP is therefore PRO (which crucially must not be Case-marked).

<sup>31</sup> If this is true, these same speakers would be expected to have also lost the canonical passive without *oleh*.

(38) Canonical passive with *oleh* (liberal) (see (33c))

That the various types of passive in Bahasa can be derived from a common D-structure provides further support for our initial hypotheses concerning phrase structure and theta assignment. We have shown that Bahasa differs from the other languages in that it does not have direct Case-marking of the SPEC of VP. Instead, it makes use of DET to INFL movement, an  $X^0$  movement rule which has also been proposed to account for cliticization facts in Romance (Kayne, 1989). This rule interacts with the V-to-INFL rule to establish a Case-marking relation between the Agent and the verb, as shown in (34), (36), and (37). In these three cases, the morpheme associated with the verb is the expression of the person and number features of the Agent. In the liberal dialect the origins of the morpheme *di-* have been lost, and instead *di-* is analyzed on a par with the English passive morpheme.

3.3. *Summary*

By assuming two subject positions, the SPEC of IP and the SPEC of VP, and by further assuming that they may be filled simultaneously, we have accounted for the following facts:

- (i) The Agent is the only maximal projection which may appear between a verb and its object at S-structure.
- (ii) No preposition is needed to license the 'demoted' subject in Malagasy and Tagalog.

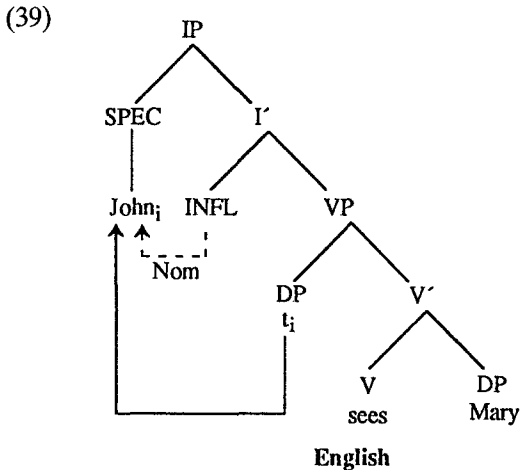
- (iii) There are two possible positions for an Agent-Topic in Tagalog.
- (iv) There are three forms of the passive in Bahasa.

#### 4. THE PHRASE-STRUCTURE OF AUSTRONESIAN LANGUAGES AND A TYPOLOGY OF PASSIVE

We have shown that our analysis can explain the split in subject properties which has long challenged researchers of Austronesian languages, as well as the word order facts in a number of these languages. We will now consider the implications of our analysis for the typology of natural languages. In particular, we will consider the non-Agent-Topic (i.e., non-active) forms found in Austronesian languages and compare them with passive structures in languages like English, showing why a language like English expresses the Agent argument of a passive as a *by*-phrase, while the Austronesian languages examined here – with the sole exception of Bahasa, as just described – lack *by*-phrases altogether.

For this analysis we draw on Guilfoyle's (1991) argument that the form of passivization in a given language is closely related to its phrase structure, and, more specifically, on her (1990) argument that languages permitting the SPEC of VP to be Case-marked, and hence filled by any lexical DP at S-structure, will lack Agent *by*-phrases, whereas languages which, in passives, can only Case-mark SPEC of IP will have Agent *by*-phrases. Whether or not SPEC of IP receives Case depends on the directionality of Case assignment in the language and on the value of the setting of the head-initial/head-final parameter.

To see how this analysis works for English, consider the tree below.



In English INFL assigns Case to the left only; therefore, the only DP types that can occupy the SPEC of VP at S-Structure are PRO and NP trace, which do not appear in Case-marked positions. A lexically realized subject DP will always raise from its D-structure position.

In Section 1.3 we argued that, in Austronesian languages, the non-Agent DP that has moved to the SPEC of IP and the Agent DP that has been base-generated in the SPEC of VP are both licensed at S-structure. If we consider Theme-Topic constructions in particular, we note that they are equivalent to English passives except for having their Agent in the SPEC of VP rather than having a *by*-phrase. Thus the Theme-Topic sentences in (40) have the S-structure represented in (41).

(40)a. **Tagalog**

Aalisin      **ng babae**      sa sako      para sa bata'

*TT-take-out GEN-woman OBL-sack for OBL-child*

ang bigas.

*TOP-rice*

The rice was taken out of the sack for the child by the woman.

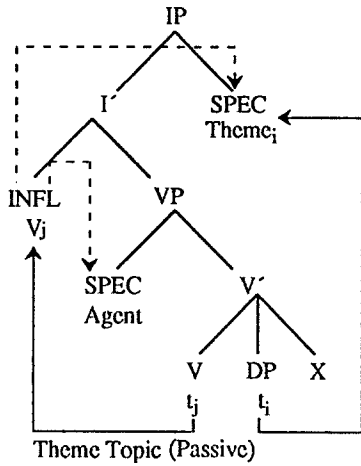
b. **Malagasy**

Vidin' **ny lehilahy** ho an'ny ankizy ny vary.

*TT-buy the man for the children the rice*

The rice was bought by the man for the children.

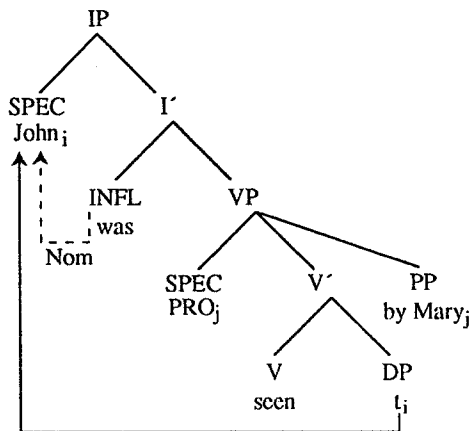
(41)



Malagasy and Tagalog

According to the analysis of Guilfoyle (1990, 1991),<sup>32</sup> the expression of Agent by means of a *by*-phrase in the English passive stems from the inability of the SPEC of VP to receive Case. The SPEC of VP receives an Agent theta role, and may only be occupied by a PRO coindexed with an adjunct *by*-phrase. The S-structure representation of an English passive is given in (42) below:

(42)



In contrast, Austronesian does permit the SPEC of VP to be Case-marked. Bahasa falls somewhere between English and the other Austrone-

<sup>32</sup> This view of English passive has also been proposed by Fukui and Speas (1986).

sian languages. The Bahasa canonical passive, as we have seen, consists of two subtypes: one in which the Agent in the SPEC of VP is licensed only because DET can move into INFL (and not because of direct Case-marking, as in Malagasy), and one in which there is no such option and the Agent can only be PRO. Our analysis of the latter subtype is similar to that of English passive structures; the Agent theta role is assigned to PRO, which may be coindexed with an adjunct PP.

The analysis of passive presented above has important theoretical implications, as it relates to two proposals in the literature: Burzio's (1986) generalization regarding Case and theta role assignment, and Baker's (1988) UTAH.

Baker (1988) proposes that the assignment of theta roles is mediated by UTAH, given in (43) below:

(43) The Uniformity of Theta Assignment Hypothesis (UTAH)

Identical thematic relationships between items are represented by identical structural relationships between those items at the level of D-structure.

(Baker 1988, p. 46)

This hypothesis predicts that particular phrase-structure positions are associated with particular theta roles. Under most theories of passive formation the assignment of the Agent theta role presents a difficulty for UTAH. While Themes are always associated with the complement of V, Agents are associated with the subject position only in active sentences. In passive structures the Agent is assumed to be suppressed and is not assigned to the subject position. This problem does not arise under the analysis presented here, because the Agent is assigned to SPEC of VP in both active and passive structures. In passive structures it may appear as PRO or as a lexical DP in a given language depending on whether or not SPEC of VP can receive Case in that language. Adopting the proposed analysis of passive formation in English and Austronesian allows us to maintain a strong version of UTAH, because the Agent theta role is always associated with the SPEC of VP.

Moreover, our analysis is incompatible with Burzio's generalization, which links the assignment of Case to the object DP with the assignment of a theta role to the subject DP. Burzio's generalization is given in (44) below:

(44) Burzio's Generalization (adapted from Burzio 1986, p. 178)

A verb may assign a theta role to its subject iff it assigns (accusative) Case to its object.

Burzio's generalization is undesirable on both theoretical and empirical grounds. On theoretical grounds it is objectionable because it is not clear how the mechanisms of Case and theta role assignment are related, seeing that they occur at different levels of the grammar, and, therefore, how they conspire to associate external theta roles with accusative Case. Burzio's generalization is inadequate on empirical grounds because it cannot account for the facts of passivization in Austronesian languages like Tagalog and Malagasy. In these languages both the Agent and the Theme are expressed in passive structures. Thus, while the verb does not assign accusative Case to its object, it does theta mark its subject.

Under our analysis Burzio's generalization is eliminated, as passivization has no effect on theta assignment whatsoever; rather, it affects only the ability of the verb to assign accusative Case to the Theme. The Agent theta role is assigned to the SPEC of VP in the usual way; however, some languages will not permit this DP to be lexically realized because they do not permit the SPEC of VP to receive Case.

Finally, strong predictions can be made about the presence or absence of adjunct *by*-phrases in various languages depending on the status of the SPEC of VP and the SPEC of IP in them. More specifically, one should never find a language where, in a given construction, the SPEC of VP is Case-marked, and yet an adjunct *by*-phrase may still appear. We relate the presence of these adjunct *by*-phrases directly to the inability of the SPEC of VP to be licensed in English. In languages like Malagasy and Tagalog, 'passive' morphology and the verb licenses a lexical DP in the SPEC of VP at S-structure. In languages like English, SPEC of VP is never licensed and thus can be occupied only by PRO. Again, Bahasa falls somewhere between the two language types. In Bahasa, the SPEC of VP is not licensed by direct Case-marking. In this sense it is like English, in that the *by*-phrase is available but only to specify the reference of a third person Agent. Unlike English, Bahasa can also express an Agent in a Theme-Topic construction without the use of the adjunct PP.

## 5. CONCLUSION

In this paper, we have proposed that the two NPs in Austronesian that behave like subjects – the Agent and the Topic – are associated with two different SPEC positions: the SPEC of VP and the SPEC of IP, respectively. This proposal, as we have shown, allows for a straightforward

solution, within the Government-Binding framework, to the problems of word order and the notion of subject that have perplexed Austronesianists for many years, and provides a place for Austronesian languages within a typology of passives.

#### APPENDIX

The foregoing analysis of Austronesian languages has attempted to account for specific word order patterns and passivization forms while achieving the greatest generality in its conclusions about the ‘shape’ of these languages. One specific pattern, however, that was not treated was the alternate order of Tagalog Agent-Topic structures, in which the Agent DP remains in the SPEC of VP (see (25a<sub>ii</sub>)). We will argue here that this order raises difficult but certainly not unanswerable questions for our analysis. One question is simply an empirical one. When the Topic has not moved to the SPEC of IP, do the properties claimed to be sensitive to the SPEC of IP still pick out the Topic? The answer appears to be yes, if we examine the one instance in which it occurs – namely, when the Topic is the Agent and remains in the SPEC of VP. What is required for well-formedness here is the appearance on the V of Agent-Topic morphology: with this condition met, only the Agent DP may be extracted (45a–c), or may launch a floated quantifier (45d).

#### (45) Tagalog

- a. Magsalis            ang babae    ng bigas sa sako para  
*AT-will-take-out TOP-woman ACC-rice OBL-sack for*  
 sa bata’.  
*OBL-child*

The woman will take rice out of the sack for the child.

- b. Sino ang        magsalis            ng bigas sa sako para  
*who COMP AT-will-take-out ACC-rice OBL-sack for*  
 sa bata’?  
*OBL-child*

Who will take rice out of the sack for the child?

- c. \*Ano ang        magsalis            ang babae    sa sako  
*what COMP AT-will-take-out TOP-woman OBL-sack*

para sa bata'?

*for OBL child*

What will the woman take out of the sack for the child?

- d. Magsalis            **lahat ang mga babae** ng mga libro  
*AT-will-take-out all TOP-pl woman ACC-pl book*

sa sako    para sa bata'.

*OBL-sack for OBL-child*

- i. All the women will take books out of the sack for the child.
- ii. \*The women will take all the books out of the sack for the child.

We might say that the Agent is simply behaving as if it had moved, a suggestion which can be made more precise in the following claim: even when there is no movement, the topic morphology of the verb still Case-marks the SPEC of IP, licensing a pleonastic *pro* in this position.<sup>33</sup> This pleonastic will be coindexed with the Topic, preventing movement of any other DP.

While this sort of pleonastic coindexation may seem very different from CHAIN formation between pleonastics and VP-internal DPs in languages such as English (see, e.g., Chomsky (1986b)), there are at least two arguments in favor of such an analysis. First, it explains a minimal difference between the alternate Agent-Topic and the Recent Past constructions in Tagalog, which consists in the absence in the latter of any Topic morphology on the verb. According to our analysis, the latter construction would not license the SPEC of IP through Case-marking, and thus would have no *pro* in the SPEC of IP and no coindexation between it and the SPEC of VP. This would, in turn, lead to the prediction that any VP-internal DP may extract through the truly empty SPEC of IP (which would behave very similarly to SPEC of CP). This prediction is in fact borne out in the examples below (based on McGinn 1988):

- (46) Kabibili lang **ni Pedro**    ng tela  
*RP-buy just GEN-Pedro ACC-cloth*  
 Pedro just bought some cloth.

<sup>33</sup> It has been suggested (Lasnik 1989) that in *there* constructions in English, both the pleonastic and the DP in the VP are assigned Case. This is also what we would be saying for these constructions in Tagalog.

- (47)a. **Sino** ang kabibili lang ng tela?  
Who has just bought some cloth?
- b. **Ano** ang kabibili lang **ni Pedro**?  
What has just been bought by Pedro?
- c. **Sino** ang kabibili lang **ni Pedro** ng tela?  
Who has just been bought some cloth by Pedro?

A second argument in favor of such an analysis is based on data from Kalagan, another Philippine language. In this language, Topic DPs, though marked in the same fashion as those in Tagalog and Cebuano, must remain in the VP whether or not they are Agents. Collins (1970, p. 4)<sup>34</sup> briefly describes the word order facts of Kalagan, claiming that the underlying order is: V-Agent-Object-Instrument-Benefactive-Locative-Time. She adds that “the one regular exception is that when the *ya*-phrase [Topic] is not the agent, it immediately follows the agent, all other phrases keeping their places.” In our account this implies that the Agent Topic remains in SPEC of VP, and all other Topics would appear in object position, i.e., adjacent to the trace of the verb. There are two questions that need to be answered here: (i) why are the circumstantial Topics (instrumental, benefactive, locative, time) found in object position, and (ii) why do all Topics remain within the VP?

As an answer to the first question, we would propose that the P-incorporation which is part of oblique topic constructions forces the Topic DP to appear adjacent to the trace of the verb, much like applicative constructions in Chichewa (Baker 1988). To answer the second question, we would claim that the coindexation that we have argued for in Tagalog between SPEC of IP and SPEC of VP is extended in Kalagan to include coindexation between SPEC of IP and object position. This suggests that languages differ with regard to such coindexation possibilities. In Tagalog, the Agent optionally remains within the VP. In Cebuano, the Agent always remains within the VP. In Kalagan, all Topic DPs remain within the VP. In every case, however, the Topic will act as if it were in the SPEC of IP position due to coindexation with the pleonastic *pro*.

#### REFERENCES

- Baker, Mark: 1988, *Incorporation: A Theory of Grammatical Function Changing*, University of Chicago Press, Chicago.

<sup>34</sup> We would like to thank an anonymous reviewer for directing us to this work.

- Bell, Sarah: 1976, *Cebuano Subjects in Two Frameworks*, unpublished Ph.D. dissertation, MIT.
- Burzio, Luigi: 1986, *Italian Syntax: a Government-Binding Approach*, Reidel, Dordrecht.
- Byma, Gary: 1987, *Government and Binding in Tagalog: an Ergative Analysis*, unpublished M.A. Thesis, University of Calgary.
- Carrier-Duncan, Jill: 1985, 'Linking of Thematic Roles in Derivational Word Formation', *Linguistic Inquiry* 16, 1-35.
- Chomsky, Noam: 1981, *Lectures in Government and Binding*, Foris, Dordrecht.
- Chomsky, Noam: 1986a, *Barriers*, The MIT Press, Cambridge, MA.
- Chomsky, Noam: 1986b, *Knowledge of Language*, Praeger, New York.
- Chung, Sandra: 1976, 'On the Subject of Two Passives in Indonesian', in C. N. Li (ed.), *Subject and Topic*, Academic Press, New York, pp. 57-99.
- Chung, Sandra and Jim McCloskey: 1987, 'Government Barriers and Small Clauses in Modern Irish', *Linguistic Inquiry* 18, 173-238.
- Collins, Grace C.: 1970, *Two Views of Kalagan Grammar*, unpublished Ph.D. dissertation, Indiana University.
- Dell, François: 1981, 'On Certain Sentential Complements in Tagalog', *Philippine Journal of Linguistics* 12, 11-28.
- Fukui, Naoki and Margaret Speas: 1986, 'Specifiers and Projections', *MIT Working Papers in Linguistics* 8, 128-172.
- Gerds, Donna: 1988, 'Antipassives and Causatives in Ilokano: Evidence for an Ergative Analysis', in R. McGinn (ed.), *Studies in Austronesian Linguistics*, Ohio University Press, Athens, OH, pp. 295-321.
- Grimshaw, Jane and Armin Mester: 1988, 'Light Verbs and Theta-Marking', *Linguistic Inquiry* 19, 205-232.
- Guilfoyle, Eithne: 1988, 'Parameters and Functional Projection', *Proceedings of NELS XVIII*.
- Guilfoyle, Eithne: 1990, *Functional Categories and Phrase Structure Parameters*, unpublished Ph.D. dissertation, McGill University.
- Guilfoyle, Eithne: 1991, 'Phrase-Structure and Passive', in D. Bouchard and K. Leffel (eds.), *Views on Phrase-Structure*, Kluwer Academic Publishers, Dordrecht, pp. 137-156.
- Hung, Henrietta: 1987, *Functional and Lexical Categories in Bahasa Malaysia*, unpublished M.A. thesis, University of Toronto.
- Hung, Henrietta: 1988, 'Derived Verbs and Nominals in Malagasy', unpublished ms., McGill University.
- Kayne, Richard: 1989, 'Null Subjects and Clitic Climbing', in O. Jaeggli and K. Safir (eds.), *The Null Subject Parameter*, Kluwer Academic Publishers, Dordrecht, pp. 239-261.
- Keenan, Edward: 1976, 'Remarkable Subjects in Malagasy', in C. N. Li (ed.), *Subject and Topic*, Academic Press, New York, pp. 247-301.
- Keenan, Edward and Bernard Comrie: 1977, 'Noun Phrase Accessibility and Universal Grammar', *Linguistic Inquiry* 8, 63-99.
- Kitagawa, Yoshihisa: 1986, *Subjects in Japanese and English*, unpublished Ph.D. dissertation, UMass, Amherst.
- Koopman, Hilda and Dominique Sportiche: 1988, 'Subjects', unpublished ms., UCLA.
- Kroeger, Paul: 1991, *Phrase Structure and Grammatical Relations in Tagalog*, unpublished Ph.D. dissertation, Stanford University.
- Kuroda, S.-Y.: 1988, 'Whether We Agree or Not: a Comparative Syntax of English and Japanese', in W. Poser (ed.), *Papers from the Second International Workshop on Japanese Syntax*, CSLI, Stanford, pp. 103-143.
- Lasnik, Howard: 1989, 'Case and Expletives: Notes Toward a Parametric Account', talk given at the Second Princeton Workshop on Comparative Grammar, Princeton University.
- Lewis, M. Blanche: 1969, *Sentence Analysis in Modern Malay*, Cambridge University Press, Cambridge.
- McCloskey, James: 1984, 'Raising, Subcategorization, and Selection in Modern Irish', *Natural Language and Linguistic Theory* 1, 441-485.

- McConnell-Ginet, Sally: 1982, 'Adverbs and Logical Form', *Language* 58, 144-184.
- McGinn, Richard: 1988, 'Government and Case in Tagalog', in R. McGinn (ed.), *Studies in Austronesian Linguistics*, Ohio University Press, Athens, OH, pp. 275-293.
- Othman, Arbak: 1981, *Tatabahasa Bahasa Malayasia*, Sarjana Enterprise, Kuala Lumpur.
- Pollock, Jean-Yves: 1989, 'Verb Movement, Universal Grammar and the Structure of IP,' *Linguistic Inquiry* 20, 365-424.
- Postal, Paul: 1969, 'On So-Called Pronouns in English', in D. Reibel and S. Schane (eds.), *Modern Studies in English*, Prentice-Hall, Englewood Cliffs, NJ, pp. 201-224.
- Rizzi, Luigi: 1990, *Relativized Minimality*, The MIT Press, Cambridge, MA.
- Schachter, Paul: 1976, 'The Subject in Philippine Languages: Topic, Actor, Actor-Topic or None of the Above', in C. Li (ed.), *Subject and Topic*, Academic Press, New York, pp. 491-518.
- Schachter, Paul and Fe Otnes: 1972, *Tagalog Reference Grammar*, University of California Press, Berkeley.
- Seiter, William: 1975, 'Information Questions in Philippine Languages', unpublished ms., University of California, San Diego.
- Sie, Ing Djiang: 1989, *The Syntactic Passive in Bahasa Indonesia*, unpublished Ph.D. dissertation, University of Amsterdam.
- Sportiche, Dominique: 1988, 'A Theory of Floating Quantifiers and its Corollaries for Constituent Structure', *Linguistic Inquiry* 19, 425-450.
- Travis, Lisa: 1988, 'The Syntax of Adverbs', *McGill Working Papers in Linguistics, Special Issue on Comparative Germanic Syntax*, 280-310.
- Travis, Lisa and Edwin Williams: 1983, 'Externalization of Arguments in Malayo-Polynesian Languages', *The Linguistic Review* 2, 57-78.
- Williams, Edwin: 1981, 'Argument Structure and Morphology', *The Linguistic Review* 1, 81-114.
- Williams, Edwin: 1987, 'Implicit Arguments, the Binding Theory and Control', *Natural Language and Linguistic Theory* 5, 151-180.

Received 21 November 1989

Revised 1 August 1991

Eithne Guilfoyle  
 Department of Linguistics  
 University of Calgary  
 2500 University Drive NW  
 Calgary, Alberta  
 Canada T2N 1N4

Henrietta Hung  
 Linguistics Program  
 Department of Psychology  
 Brandeis University  
 Waltham, MA 02254

Lisa Travis  
 Department of Linguistics  
 McGill University  
 1001 Sherbrooke West  
 Montreal, PQ  
 Canada H3A 1G5