VARIATION IN WESTERN AUSTRONESIAN CLAUSE STRUCTURE

MICHAEL DUKES

University of California, Los Angeles

1. Introduction

Recent analyses of W[estern] A[ustronesian] clause structure within G[overnment]-B[inding] theory have attempted to characterize the syntax of WA languages, such as Malagasy and Tagalog, in terms of underlying structures that are essentially identical to those proposed for more widely studied European languages such as English and French. In this paper, I want to discuss some inadequacies of this kind of approach to the study of WA syntax, focusing on some problems that arise within the recently proposed framework of Guilfoyle, Hung & Travis 1992 ('GHT'). I will conclude that the notion of a fully-specified 'underlying structure' is unhelpful and creates more problems than it solves. Furthermore, the concomitant application of movement rules to derive the observable word order and constituency data fails to account for robustly observable structural contrasts between the WA languages considered here, namely, Malagasy, Tagalog and Toba Batak. Finally, I sketch an account of the clause structure of these three languages within the framework of H[ead-driven] P[hrase] S[tructure] G[rammar] (Pollard & Sag 1987, 1993) which does not appeal to any notion of 'underlying structure' and thereby avoids the unwanted side-effects that arise under the GB approach. I conclude that a theory which seeks to describe the universal properties of human language within the same general frames should be based on an enriched notion of lexical structure rather than on the postulation of universal 'underlying structures'.

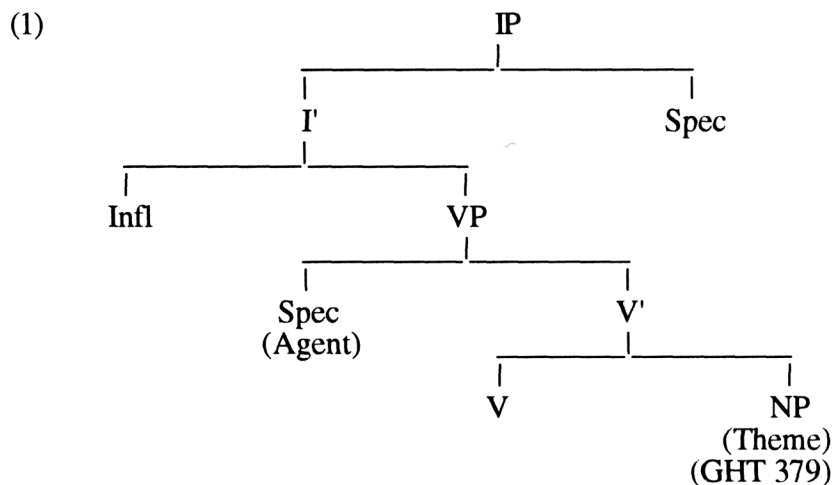## 2. Typical Properties of WA Clauses

Clause structure in WA languages is characterized by a number of distinctive properties. The large majority of WA languages (of those that have been well-studied, at least) are predicate-initial with one argument picked out via verbal morphology and/or nominal morphology and/or linear ordering as having a special grammatical prominence. This prominent argument has been variously identified as 'subject', 'topic', 'focus', or 'absolutive' amongst other labels, based on its properties in various WA languages. The correct choice of label for this argument is not an issue I will dwell on at length

43

here, except insofar as to mention that, as in Kroeger's LFG account of Tagalog (Kroeger 1991) and Davies' work on Javanese within Relational Grammar (Davies 1993), the HPSG analysis presented below defines this prominent argument as subject, so I will refer to it as 'subject' from now on.

The subject in WA languages is typically 'definite' or 'presupposed' and is usually the only argument available for grammatical processes, such as wh-movement, relativization and topicalization. Each language provides some system of voice morphology by the use of which a given argument can be promoted to subject and thus become available for the grammatical processes mentioned above. Finally, it is also typical of WA languages that so-called 'thematic roles' (or perhaps 'logical' grammatical relations), often more than surface grammatical relations, play an important part in the determination of conditions on semantically related phenomena, such as reflexive-binding and control. Any analysis of WA clause structure must provide some account of how these different properties are represented and how they interact in a particular language.

3. The GHT Account of WA Clause Structure

GHT posit a very orthodox D-structure for WA languages:

(1)

```
                          IP
             _____|_____
            |                             |
            I'                            Spec
     _____|_____
    |               |
   Infl             VP
            _____|_____
           |                   |
          Spec                 V'
         (Agent)         _____|_____
                        |             |
                        V             NP
                                   (Theme)
                                  (GHT 379)
```

They assume that voice marking already appears on the verb at D-structure, presumably via morphological rule. Depending on which voice marker appears on the verb, one argument is forced to move to [Spec, IP] to receive structural case from

Infl by S-structure. Oversimplifying somewhat, the idea is that 'active' voice case-marks the theme argument, while 'passive' voice case-marks the agent. The argument which does not receive case from V moves to [Spec, IP] to receive case and becomes the 'topic' (in GHT's terms; the 'subject' in the HPSG analysis presented below).

One of the principal goals of the GHT analysis is to account for the distribution of subject properties observed in WA languages. GHT adjudge that a 'split' in subject properties exists between two NPs, the 'topic' and the 'agent', on the basis of four properties; extractability, quantifier float, control of reflexives, and deletability in 'equi-NP' (or 'Control') constructions. As mentioned in the introduction, the subject (GHT's 'topic') is the only NP available for extraction and, in Philippines languages at least, it is the only NP available for quantifier float (Schachter 1976). By contrast, the agent, whether it is subject or not, appears to be the typical controller of reflexive pronouns. The status of NPs deleted under equi-NP is somewhat more unclear, but deleted NPs can apparently be either nonsubject agents or subjects.

GHT claim that their analysis provides a straightforward structural account of this 'split'. One position, [Spec, IP], is associated with extractability and quantifier float. The other position, [Spec, VP], in which a nonsubject agent remains in situ, is associated with control of reflexivization. However, it will be demonstrated below that the GHT analysis fails to account for the WA facts in at least two crucial ways. Firstly, there is clearly no absolute 'split' between [Spec, VP] and [Spec, IP] with respect to the properties that GHT discuss. This suggests that the positional [Spec, VP] / [Spec, IP] dichotomy is not the key to understanding the WA data. Secondly, even if there were such a split, the GHT analysis does not explain why or how some of these properties should be associated with one position rather than the other.[1]

4. The GHT Account Applied to Malagasy

The GHT analysis sketched above leads to the following general picture of clause structure in the VOS language, Malagasy. The three related sentences given in (2) provide illustration.

(2)a. M- AN-asa   ny lamba   amin' ny savony   ny zazavavy
     pres- ACT-wash   the clothes   with the soap   the   girl
     'The girl is washing the clothes with the soap'
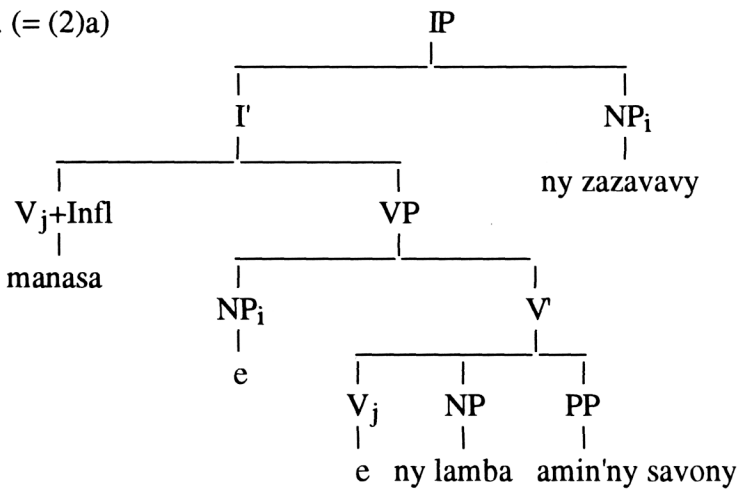   (Active)

b. Sasa-N'   ny zazavavy   amin' ny savony   ny lamba
   wash-PASS the girl   with the soap   the clothes
         'The clothes are being washed with the soap by the girl'
                                                        (Passive)

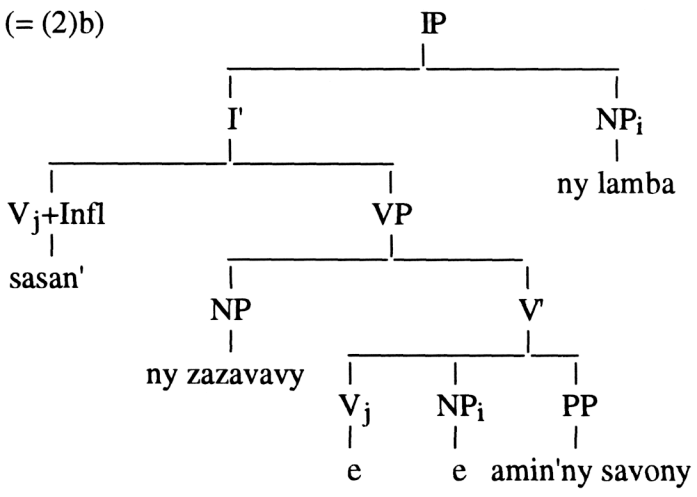c. AN-asa-N'   ny zazavavy   ny lamba   ny savony
   CIRC-wash-CIRC the girl   the clothes   the soap
          'The soap is washing with the clothes by the girl'
                                                  (Circumstantial)

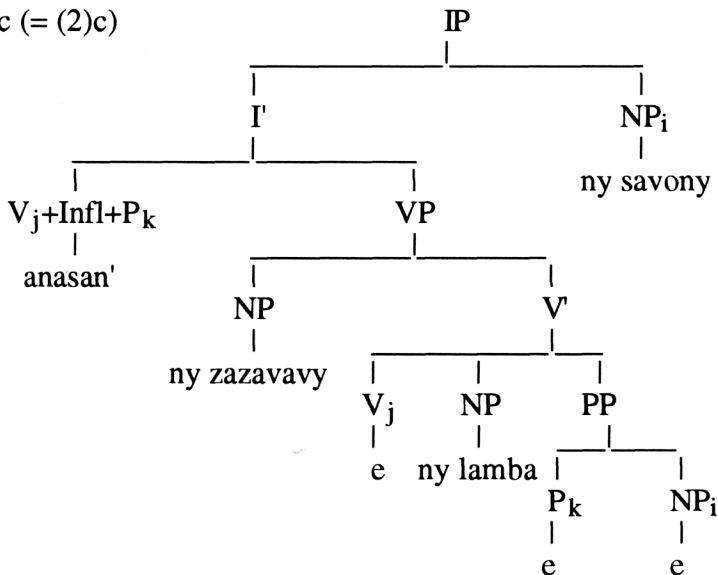The examples given above are each associated with an S-structure given in (3):

(3)a. (= (2)a)

(3)b. (= (2)b)

(3)c (= (2)c)

```
                              IP
                   _____|_____
                  |                        |
                  I'                       NPᵢ
          _____|_____                |
         |                 |            ny savony
      Vⱼ+Infl+Pₖ          VP
         |            _____|_____
      anasan'        |            |
                     NP           V'
                     |        _____|____
                ny zazavavy   |    |    |
                             Vⱼ   NP   PP
                              |    |   _|_
                              e ny lamba |    |
                                        Pₖ   NPᵢ
                                         |    |
                                         e    e
```

Example (2/3)a illustrates an active sentence in which the 'agent' NP moves to [Spec, IP] to receive nominative case as the 'topic'. The verb is assumed to raise to Infl to pick up tense inflection. The 'theme' object receives accusative case in situ from the 'active' voice morphology and the instrumental NP receives case as the object of the preposition. (2/3)b illustrates the passive situation in which the passive voice morphology case-marks the agent so that it remains in situ. The theme, on the other hand, does not receive case in situ, so it moves to [Spec, IP] where it receives nominative and becomes the 'topic'.[2] In the so-called 'circumstantial voice' (Keenan 1976), both the agent and theme are claimed to receive case in situ from the voice morphology on the verb. The preposition which marks an instrumental, benefactive, or some similar additional argument is assumed to incorporate into the verb (as in (3)c), supposedly depriving its object of case.[3] The object of the preposition thus raises to [Spec, IP] to receive case, becoming the 'topic'.

4.1 Problems with the GHT Analysis of Malagasy

The GHT analysis derives the correct word order for Malagasy, providing a straightforward account of the position of the subject (GHT's 'topic'). Beyond these generalities though, the analysis runs into a number of problems, both from

a descriptive point of view and from a theory-internal perspective. I will review two descriptive problems that impinge directly on the GHT analysis before turning to a review of some comparative data from Tagalog and Toba Batak.[4]

4.2 The Status of the Nonsubject Agent

One set of descriptive problems arising under the GHT analysis of Malagasy revolves around the treatment of non-subject agent phrases. The GHT proposal entails a major constituent break between the V+Infl complex and the [Spec, VP] position in which the agent NP appears in passive and circumstantial clauses. But Keenan 1991 has shown convincingly that verbs form extremely tight constituents with non-subject agents. Thus while adverbs of various sorts may intervene between the subject and the VP, or even between the object and the verb, ABSOLUTELY NOTHING may intervene between a nonactive verb and the agent NP. As Keenan points out, it is not plausible to appeal to a 'late' phrase phonological rule to create a phonological unit out of these separated elements (as a defender of the GHT analysis might propose). *He demonstrates that the relationship between a nonactive verb* and its agent is strongly parallel to that between a noun and its possessor. The same morphophonological process of 'n-bonding' that occurs between nouns and possessors is observed between nonactive verbs and their agents:

(4) 'n-bonding' between nouns and possessors:

> ny trano-N-dRabe 'Rabe's house'
> ny boki-N-ny mpianatra 'the student's book'

(5) 'n-bonding' between nonactive verbs and agent NPs:

> vita-N-dRakoto 'finished by Rakoto'
> azo-N-ny ankizy 'understood by the children'
>   (examples from Keenan 1992)

Furthermore, the rules of phonological liaison holding between a noun having one of the so-called 'weak endings', -*na*, -*ka*, and -*tra* and its pronominal possessor are identical to the rules holding between a nonactive verb with a weak ending and its pronominal agent phrase:

(6) Weak endings on possessed nouns:
  entana + ko > entako   'my luggage'
  ny mpampianatra + nao > ny mpampianatrao 'your teacher'

(7) Weak endings on nonactive verbs:
 soratana + ko > soratako   'written by me'
  fantatra + nao > fantatrao   'known by you'
                              (examples from Keenan 1992)

These parallels suggest that some significant structural property is shared by possessors and non-subject agent NPs, a property which the GHT analysis fails to capture in any obvious way, assuming (uncontroversially, I think) that Malagasy possessors form a constituent with their nominal heads.
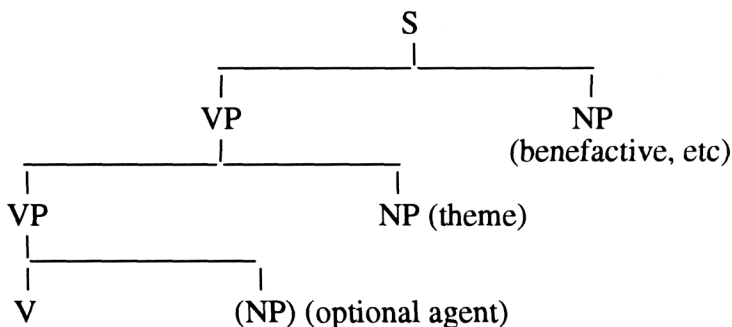
The structural similarity between possessors and nonsubject agents manifests itself in an additional characteristic; both possessors and nonsubject agents are optional. Just as *by*-phrases are almost always optional with passivized verbs in English, a passive or circumstantial verb in Malagasy need not appear with an adjacent agent NP. Thus examples (2)b and c above are still perfectly grammatical if the agentive NP *ny zazavavy* 'the girl' is deleted. As in English, the sentences would have a straightforward interpretation in which the agent would simply be understood existentially. This fact is difficult to reconcile with the GHT claim that agents appear in an A[rgument] position (i.e. [Spec, VP]) because the optionality of the agent suggests it is some kind of modifying adjunct rather than an (obligatory) argument.

To summarize, the data regarding non-subject agents presented above suggests a (surface) structure like (8) for Malagasy passive sentences:

(8)
                              S
          _____|_____
          | |
          VP   NP (theme)
    _____|_____
    | |
    V   (NP) (optional agent)
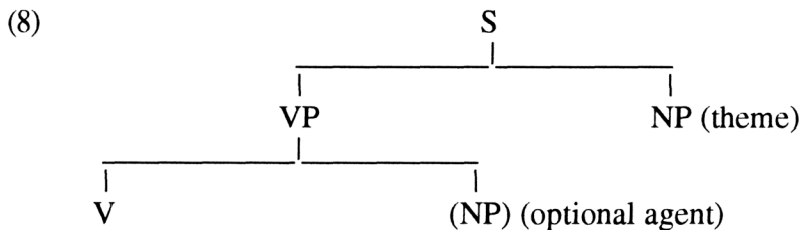
and a structure like (9) for circumstantial sentences:

(9)

```
                              S
         ┌────────────────────┴──────────────┐
         │                                    │
        VP                                   NP
   ┌─────┴──────────────┐              (benefactive, etc)
   │                    │
  VP                  NP (theme)
   │
   ├──────────────────┐
   │                  │
   V              (NP) (optional agent)
```

The special possessor-like status of the agentive NP is addressed in the analysis presented in section eight, where the nonsubject agent is treated as a kind of restricting specifier akin to a determiner rather than as an argument. The S-structure, *which GHT propose for active sentences, seems to be* essentially correct, modulo the status of the 'empty' [Spec, VP] and the choice of categorial labels designating certain nodes.

4.3 Reflexivization in Malagasy

One of the key properties that GHT take as diagnostic of the contrast between 'topic' [Spec, IP] and 'agent' [Spec, VP] is reflexivization. The ability to control a reflexive is considered a property of the agent. Under their account, a topic may be a reflexive pronoun as long as it is not an agent-topic. Furthermore, an agent may never be a reflexive no matter whether it is a topic or not. These facts are correct, as far as I am aware, for the Philippines languages and for Toba Batak. Although it is never actually spelled out by GHT, one assumes that this generalization is supposed to follow from the fact that [Spec, IP] is a 'non-argument' (A') position in WA languages. Therefore, under standard GB assumptions, it does not count as relevant for the checking of reflexive-binding constraints (which must be checked relative to A-positions). However, the reflexivization data from Malagasy (which are mysteriously absent from GHT's discussion of reflexivization (GHT 389)) do not completely follow the typology laid out by GHT. Keenan 1992 shows that the subject (= GHT's 'topic') cannot be a reflexive pronoun even if it is bound by a nonsubject agent, as the following examples illustrate:

(10) a. * Hajain-dRabe   tena
         respects(pass)-Rabe   self
         'Himself is respected by Rabe'

b. *Nividianan-dRabe   mofo   tena
   bought(circ)-Rabe   bread self
   'Himself was bought bread for by Rabe'
                      (examples from Keenan 1992:32)

According to Keenan's generalizations regarding these data, *tena* 'self' can be an (accusative) object, but it cannot be a (nominative) subject or a (genitive) nonsubject agent. Thus we are forced to the conclusion that constraints on reflexivization in Malagasy must somehow account for the binding-theoretic prominence of [Spec, IP] - something which the GHT analysis fails to do.

5. The GHT Analysis Applied to Tagalog

One of the most salient aspects of the GHT proposal is its highly articulated tree structure. NP positions can be uniquely identified by their location in the phrase structure representation. Furthermore, the GHT theory seeks to unify the similar behaviour of agents and subject/topics in the WA language family almost exclusively via reference to the similarity of hierarchical tree structures in that family. We have seen that the clause structure of Malagasy differs in important ways from the one GHT propose, with the nonsubject agent being located very low in the tree, although Malagasy clauses do seem to have a highly articulated structure (as proposed by GHT). When we turn to Tagalog however, the evidence for any hierarchical structure is extremely small. Kroeger (1991:187-96) notes (amongst a range of other problems which he discusses for the GHT analysis) that the freedom of word order in Tagalog argues against the kind of articulated hierarchical structure that GHT propose. He claims instead that Tagalog VP structure is 'flat' so that the relative grammatical prominence of NPs is not encoded directly in phrase structure representations.

In order to derive the various grammatical word order permutations of Tagalog sentences, GHT (pp. 394-5) are forced to assume the existence of 'scrambling' rules in Tagalog (although no attempt is made to spell out any actual analysis of how these rules would work or what constraints there might be on their operation). Thus the following permutations of the same sentence elements are all grammatical:

(11) Nagbigay   ng libro   sa babae   ang lalaki
      gave   gen-book   dat-woman   nom-man

      Nagbigay   ng libro   ang lalaki   sa babae
      Nagbigay   sa babae   ang lalaki   ng libro
      Nagbigay   sa babae   ng libro   ang lalaki
      Nagbigay   ang lalaki   ng libro   sa babae
      Nagbigay   ang lalaki   sa babae   ng libro

'The man gave the woman a book'
                    (example from Schachter & Otanes 1972:83)

But once we allow for the existence of 'scrambling' rules within the GHT analysis of Tagalog, a number of troubling questions arise, the most important of which is the following: why does Tagalog have scrambling rules when Malagasy does not? Under the 'scrambling' approach, the rather significant contrasts between the syntax of Malagasy and the syntax of Tagalog are reduced to a set of stipulations about which categories in Tagalog may 'scramble'. Such an approach is unlikely to lead to deep insights into the similarities and differences observed between related languages.

An alternative approach to the contrasts between Tagalog and Malagasy involves the acceptance of a more significant structural difference between them, namely, that Tagalog lacks the hierarchical structure found in Malagasy (pace Kroeger 1991). The flat structure found in Tagalog can be considered a prerequisite for freedom of order, a prerequisite which is clearly lacking in Malagasy. In this way, extreme freedom of order (for example, order determined by pragmatic constraints like 'newness' and 'oldness', as found in a number of Native American languages as well as Tagalog) can be related directly to a 'parameterized' property of grammars--whether or not the arguments of predicates are represented hierarchically in phrase structure. Such an approach actually leads to a prediction of sorts, rather than the kind of stipulation required under GHT. Apparent 'scrambling' should only be possible in languages that lack hierarchical VPs. An added consequence of such an approach is that 'underlying' phrase structure representations become somewhat otiose. Once we eliminate 'scrambling' via movement and allow for differences in the degree of configurationality of related languages, it becomes impossible to assume that they possess the same D-structure, the functions of which (to the extent that there are any) must be taken on by some other structural representation.

6. The GHT Analysis Extended to Toba Batak

The GHT analysis is intended as a very general account of WA clause structure. It therefore seems important to consider to what degree it extends to other WA languages beyond those considered by GHT. The grammar of Toba Batak provides an interesting test case because it seems to be intermediate between Malagasy and Tagalog in certain ways.
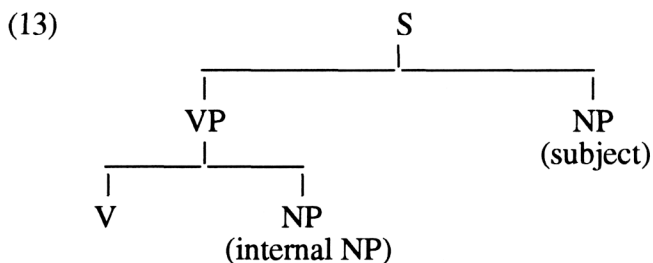
Word order is rather rigidly constrained in Toba Batak, as it is in Malagasy. There is a considerable body of evidence for a VP constituent in Toba Batak, a summary of which can be found in Schachter 1984. Oversimplifying a little, there are two basic voices found in Toba Batak which I will denote 'active' and 'passive'.[5] Examples of each are given in (12):

(12)a. Mang-ida   si Ria   si Torus
        Act-see   PM Ria   PM Torus
              'Torus sees Ria'

    b. Di-ida   si Torus   si Ria
        Pass-see   PM Torus   PM Ria
              'Torus saw Ria'

As can be seen in the above examples, the change in voice morphology on the verb corresponds to an obligatory change in the linear order of the arguments. As in Malagasy, the ordering of NPs in the sentence is rigidly fixed. The subject (i.e., the element which has all the properties discussed in the introduction) is linearly the second NP. Thus Toba Batak can reasonably be described as a VOS language like Malagasy.

Using a number of constituency tests, Schachter 1984 demonstrates that the 'internal NP' (the NP which appears adjacent to the verb) forms an extremely tight constituent with the verb, reminiscent of the relationship between the verb and the nonsubject agent in Malagasy. As in Malagasy, NOTHING may intervene between the verb and the internal NP. Malagasy and Toba Batak differ in one important respect however. While the close relationship between nonsubject agent and verb in Malagasy is restricted to sentences containing nonactive verbs, both active and passive verbs form constituents with the internal NP in Toba Batak. Schachter establishes that internal NPs in active and passive sentences have exactly the same syntactic properties and concludes that the structure of active and passive sentences is exactly the same. All that differs is the

choice of voice marking on the verb and the linking of semantic roles to the arguments. In an active sentence, the subject is the agent and the internal NP is the nonagent (typically a theme or patient); in a passive sentence, the subject is the nonagent and the internal NP is the agent:

(13)

```
                        S
       _____|_____
      |                 |                 |
      VP                               NP
  ____|____                          (subject)
 |         |
 V         NP
      (internal NP)
```

If Schachter is right in his claims about the syntax of Toba Batak, then an interesting problem is raised for the GHT proposal.[6] Since internal NPs have the same properties in active and passive clauses, they should bear the same structural relation to the verb in both types of clauses. But the GHT analysis crucially assumes that NPs bearing agentive semantic roles must start off in D-structure in a higher position than nonagents. Unless there is some A' position other than [Spec, IP] to which internal NPs can move, there is no way that agents and nonagents can have an identical (or even similar) structural relation to the verb. Furthermore, just as in Malagasy, there is no way that an agentive NP in [Spec, VP] can even form a constituent with the verb once the verb raises to Infl to collect its tense/aspect morphology. Interestingly, the structural parallels between active and passive clauses in Toba Batak do not correlate with the binding facts of the language. Binding in Toba Batak is totally insensitive to the effects of voice morphology. 'Agent' NPs may not be bound by 'non-agents' whether they are internal or external NPs. Thus the arboreal structure given in (13) is irrelevant for binding purposes, as the following reflexive instantiations of (13) show:

(13')a. Mang-ida   dirina   si Torus
        Act-see   himself   PM Torus
              'Torus sees himself'

b. Di-ida   si Torus   dirina
   Pass-see   PM Torus   himself
   'Torus saw himself'
                   (examples from Schachter 1984:130)

An external NP (subject) may be bound by an internal NP despite the lack of c-command. The same phenomenon is observed in Tagalog, where surface grammatical relations do not provide the correct domain for the statement of conditions on reflexive binding. As Schachter 1976 shows, reflexivization in Tagalog is conditioned by the same kind of thematic constraints that apply in Toba Batak. An agent can never be bound no matter what surface grammatical relation it bears.[7]

7. The Problem Summarized

The data reviewed above have highlighted a recurring inadequacy of the GHT analysis. The structure proposed by GHT as common to WA languages does not correlate with the very basic structural facts of the three languages examined. Consequently, significant contrasts in their configurational properties are not accounted for. This problem appears all the more serious when one takes into account the fact that the GHT analysis is a PURELY phrase-structural one which sinks or swims on the correctness of the proposed structures.

The inadequacies of the GHT analysis appear to derive from the overly restrictive notion of 'syntax' assumed within GB. The insensitivity of the analysis to DIFFERENCES between languages comes about because it is assumed that what languages have in common must be represented via a 'universal' D-structure. Once the D-structure is fixed, there is no possibility of deriving differences in structure except via adhoc 'scrambling' rules and language-particular exceptions to otherwise obligatory constraints (such as those made by GHT to account for unexpected properties of Tagalog and Malay (GHT 410-12, 400-4)).
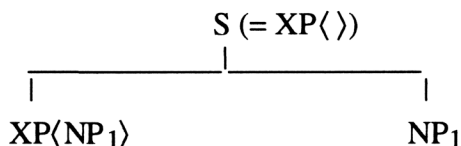
8. A Lexically Based Approach to WA Clause Structure

While GB seeks to account for universal syntactic tendencies via highly restrictive underlying phrase-markers, several alternative frameworks within generative grammar have begun to converge on an approach which stresses the key role of lexical information. HPSG is one such theory. Within HPSG, phrase structure must satisfy all relevant lexical requirements, but it does not itself provide the representations
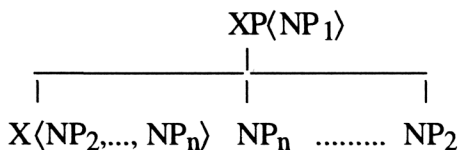
upon which phenomena such as grammatical relations and reflexive binding constraints are defined. All such phenomena are defined in terms of lexical entries. The influence of NONSYNTACTIC information on syntax can also be perspicuously represented via the cross-referencing of information in different parts of lexical entries; something which is impossible in a theory where all constraints on syntax must be stated via reference to tree structures. The following sketch of the three languages discussed above illustrates some of the advantages of such an approach.

The basic phrase structural properties of WA languages can be described by a set of putatively universal I[mmediate] D[ominance] Schemata (Pollard & Sag 1993):
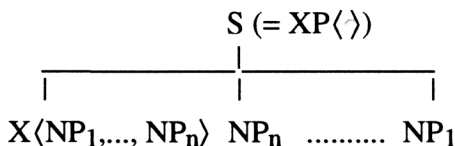
(14)a. ID Schema 1 ('Subject - Predicate Schema')

$$S \ (= XP\langle \ \rangle)$$

$$XP\langle NP_1 \rangle \qquad\qquad NP_1$$

  b. ID Schema 2 ('VP Schema')

$$XP\langle NP_1 \rangle$$

$$X\langle NP_2,..., NP_n \rangle \quad NP_n \ \dots\dots\ NP_2$$

  c. ID Schema 3 ('Flat Schema')

$$S \ (= XP\langle \ \rangle)$$

$$X\langle NP_1,..., NP_n \rangle \quad NP_n \ \dots\dots\ NP_1$$

Pollard & Sag use exactly these three schemata to account for the relationship between heads and arguments in their account of English phrase structure. I will show that the same schemata can be used to account for the basic properties of clause structure in WA languages.

Each predicate (denoted 'X(P)' above) is provided with an ordered subcategorization (SUBCAT) list that is given in angle brackets to the right of the predicate. Each element of the SUBCAT list is indexed to an argument that must appear in a structural position determined by the kind of schemata that

apply in a given language. Each argument that appears in some position that satisfies the requirements of its predicate is deleted from the SUBCAT list of the predicate until the predicate has an empty SUBCAT list (denoted 'XP⟨ ⟩' above). Arguments must be discharged from the SUBCAT list in ordered fashion from the bottom of the list. A 'subject' is DEFINED as the element at the top of the SUBCAT list. In (14) above and in the examples to follow, the subject will typically be the element indexed with the subscript '$_1$'. A 'sentence' is simply a predicate whose arguments have all been discharged. Thus a predicate which requires no arguments (such as weather and environmental predicates ('rain', 'earthquake', etc) in many Austronesian languages) is essentially a lexical sentence. A predicate requiring one argument is 'intransitive', and a predicate requiring two arguments is 'transitive'. Structural case-marking requirements are given in SUBCAT lists as additional constraints on the form of arguments. Thus in a language which marks arguments of transitive predicates by nominative and accusative case, the SUBCAT list for a verb like 'hit' might be V⟨NP[nom]$_1$, NP[acc]$_2$⟩.

ID schemata 1 and 2 together produce the hierarchical clause structure familiar from English. These schemata will also be applied in the treatment of Toba Batak and Malagasy, which both distinguish the subject via phrase-structural means. In Tagalog on the other hand, 'subject' is distinguished via case-marking but not hierarchically. The 'flat' ID schema 3 will therefore be employed in the treatment of Tagalog.

ID schemata do not say anything about word order within a constituent. Word order is imposed by L[inear] P[recedence] constraints which may make reference to any information carried by syntactic categories or their denotations. The operation of these constraints is relatively straightforward and will be illustrated below.

Voice morphemes can be viewed as markers of the ordering of elements in the SUBCAT list of a predicate. The reordering that takes place in passive sentences for example, can be viewed as the HPSG equivalent of promotion and demotion in Relational Grammar (Perlmutter & Postal 1983, Blake 1990). This reordering can be illustrated with basic lexical entries for an active/passive pair from Toba Batak:

(15)a. *mangida*   V   SUBCAT $\langle NP_1, NP_2 \rangle$

                   CONTENT   RELATION see
                                     SEER   [1]
                                     SEEN   [2]

    b. *diida*   V   SUBCAT $\langle NP_2, NP_1 \rangle$

                   CONTENT   RELATION see
                                     SEER   [1]
                                     SEEN   [2]

The active verb form differs from the passive one only in that the elements on the SUBCAT list have been switched in the passive case. This reordering is not accompanied by any changes in case marking or requirements for prepositions (unlike English for example, where the 'demoted' subject must be realized as the object of *by* if it appears at all). The CONTENT matrices given below each SUBCAT list represent the semantic content of each predicate. The numerical index to the right of each semantic role indicates which (coindexed) NP bears that role. It can be observed that changes in SUBCAT order need not correlate with any change in basic semantic content, since both verb forms have the same semantics matrix.

      Very recent work in HPSG has assumed that predicates may also bear information on a second list similar to the SUBCAT list (or alternatively, SUBCAT is abandoned in favour of multiple lists). A second list (which here I will denote 'LEX' to indicate its relatedness to the notion of invariant lexical content shared by all forms of a particular lexeme) can be used as the lexical equivalent of an 'underlying representation', but without the unwanted side-effects inherent in an underlying level of syntax. I assume here that the ordering defined on LEX derives largely from considerations of a semantic nature, mimicking the 'thematic hierarchy' of Jackendoff 1972 and others. Essentially, LEX provides a list of indexed NPs which is unaffected by voice morphology, and which is therefore common to all word forms sharing the same verbal base.[8] I will claim that constraints on reflexive-binding in some languages are stated in terms of the ordering defined by LEX, rather than (or in addition to) the ordering defined by SUBCAT (unlike English (Pollard & Sag 1992)).

8.1 Malagasy Clause Structure in HPSG

On the basis of the framework sketched above, an account of the data from the three WA languages considered can now be given. Active clauses in Malagasy can be straightforwardly described by implementing ID schemata 1 and 2, just as in English active clauses. The differences between active clauses in Malagasy and English are accounted for by reference to differences in LP constraints; subjects in English precede predicates, whereas all arguments follow predicates in Malagasy. The relevant LP constraint is given in (16), which will also serve to account for basic head/argument word order in Tagalog and Toba Batak:

(16)   **WA LP Constraint 1:**

$$X \ [\text{SUBCAT} \ \langle ...YP_i...\rangle] \quad < \quad YP_i$$

ID schemata 1 and 2 are also employed in the treatment of passive and circumstantial clauses. In fact, the treatment of demoted agents aside, all three clause types have the same tree structure. The differences between the sentence types arise largely from the reordering of elements on SUBCAT lists. The Malagasy examples given in (2)a.-c. are respectively associated with the verbal categories given in (17):

(17)a. *manasa*, V[+ active]   [SUBCAT $\langle NP_1, NP_2, PP_3\rangle$]

   [LEX $\langle NP_1, NP_2, NP_3\rangle$]

   b.   *sasan"*, V[- active]   [SUBCAT $\langle NP_2, PP_3\rangle$]

   [LEX $\langle NP_1, NP_2, NP_3\rangle$]

   c.   *anasan '*, V [- active]   [SUBCAT $\langle NP_3, NP_2\rangle$]

   [LEX $\langle NP_1, NP_2, NP_3\rangle$]

Because nonsubject agents do not behave much like arguments in Malagasy, I assume that nonactive verbs do not syntactically select an agentive argument; hence, NP_1 does not appear on the SUBCAT list of the nonactive verbs.

Nonsubject agents in Malagasy are treated here in a manner analogous to possessive determiners. Following Pollard and Sag's (1993) treatment of determiners, I analyze the nonsubject agent as a species of MARKER which marks a nonactive verb by fixing the index of the first element on the
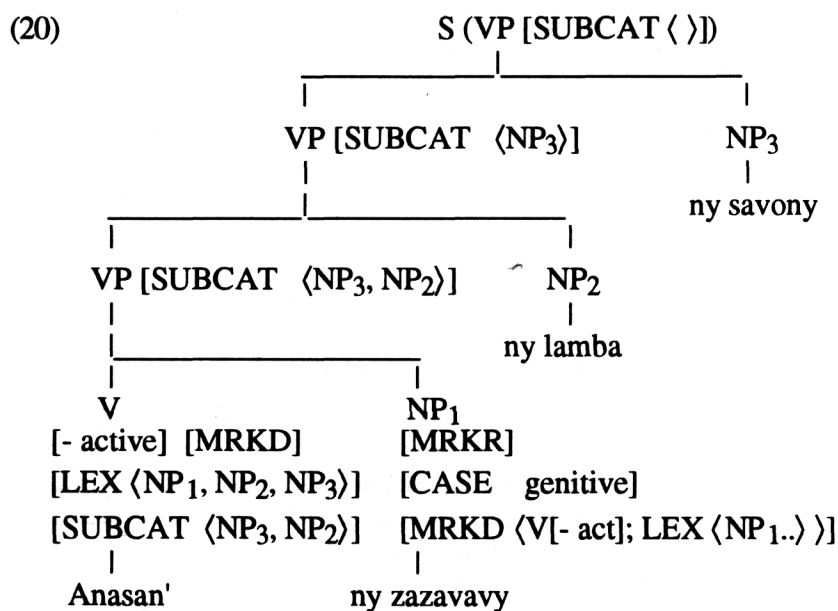
verb's LEX list to match its own. The following informally specified category for a genitive NP guarantees that it forms a constituent with the nonactive verb (putting aside the question of how the NP as a whole is constructed from its constituent parts, including case-marking):

(18) NP$_1$ [MRKR]
　　　　　　[CASE　　　　genitive]
　　　　　　[MRKD　　　　⟨V[- active]; LEX ⟨NP$_1$ ... ⟩ ⟩]

The genitive NP is specified as requiring a nonactive verb to mark whose most prominent LEX element it is coindexed with. MRKD is a list-valued feature which encodes this information. Adopting the LP statement in (19), which requires marked elements to precede markers,

(19) X [MRKD]　　　<　　　Y [MRKR]

we can generate an example like (2)c with the optimal structure suggested in (9).[9] The tree is given in (20):

(20)　　　　　　　　　　　　S (VP [SUBCAT ⟨ ⟩])
　　　　　　　　　　　　　|
　　　　　　|　　　　　　|　　　　　　　　　　|
　　　　　VP [SUBCAT　⟨NP$_3$⟩]　　　　　　NP$_3$
　　　　　　|　　　　　　　　　　　　　　　|
　　　　|　　　　　|　　　　　　　　　　ny savony
　　|　　　　　　　　　　　　　|
　VP [SUBCAT　⟨NP$_3$, NP$_2$⟩]　　NP$_2$
　　|　　　　　　　　　　　　|
　|　　　　　　　　　　　ny lamba
　|　　　　　|
　V　　　　　NP$_1$
[- active] [MRKD]　　[MRKR]
[LEX ⟨NP$_1$, NP$_2$, NP$_3$⟩]　[CASE　genitive]
[SUBCAT ⟨NP$_3$, NP$_2$⟩]　[MRKD ⟨V[- act]; LEX ⟨NP$_1$..⟩ ⟩]
　|　　　　　　　　　|
Anasan'　　　　ny zazavavy

　　　Malagasy reflexive-binding facts can be stated via reference to the SUBCAT and LEX lists. Recall that according to Keenan's (1992) generalizations, neither a subject nor a

nonsubject agent can be a reflexive. The problem for the GHT analysis was that there was no way to state the simultaneous prominence requirements of [Spec, IP] and [Spec, VP] in Malagasy at one level of representation. Furthermore, the binding-theoretic prominence of [Spec, IP] ran counter to the 'split' in subject properties assumed by GHT, since prominence for binding purposes was claimed to be a property of [Spec, VP]. The Malagasy binding constraints can be easily stated in the present analysis via reference to relative obliqueness on the SUBCAT AND the LEX lists. The notion of RELATIVE OBLIQUENESS defined in HPSG derives primarily from well known work in Relational Grammar (Perlmutter & Postal 1977, Keenan & Comrie 1977) and Categorial Grammar (Dowty 1982). As we saw earlier, the subject of a predicate is the first or LEAST OBLIQUE element on the SUBCAT list of that predicate. The object is the second least oblique, etc. The HPSG binding theory for English given in Pollard & Sag (1992) was stated in terms of relative obliqueness solely on the SUBCAT list (i.e. in terms of grammatical relations as they are defined within the theory). An informal version of the condition governing English reflexives is given in (21):

(21)   An anaphor must be coindexed with a less oblique
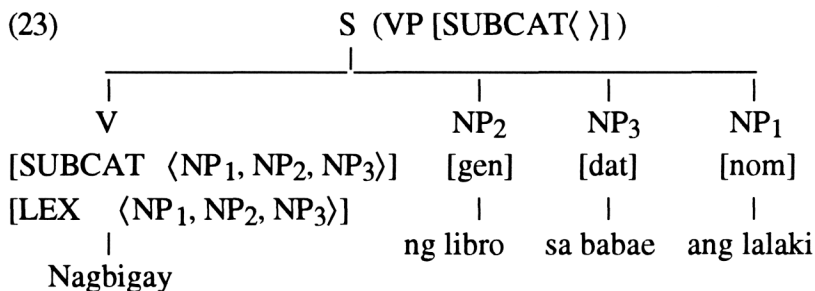   coargument, if there is one.   (Pollard & Sag 1992:266)

To account for the binding conditions in Malagasy, we need to ensure that the obliqueness requirement applies to both the SUBCAT and LEX lists. This requires a conjunctive formulation of the reflexivization constraint for Malagasy:

(22)   **Malagasy Reflexivization Constraint:**
         A reflexive in Malagasy must not be coindexed with a
         more oblique coargument (on the LEX or SUBCAT
         lists) but must be coindexed with at least one
         coargument.

Because subjects are, by definition, the least oblique elements on the SUBCAT list, they can never satisfy (22). Because agents are, by hypothesis, the least oblique elements on the LEX list, they can never satisfy (22) either. Only an element which satisfies both obliqueness requirements simultaneously, such as an accusative object, can be a reflexive. Thus the conjunctive binding requirements observed in Malagasy can be stated in a way that naturally extends the intent of the HPSG binding theory.

8.2 Tagalog Clause Structure in HPSG

The flat structure observed in basic Tagalog clauses can be obtained by implementing ID schema 3. The permutations of NP order observed in (11) can be generated by underspecifying the LP constraints on relative ordering of sister NPs. I assume there are none. The only LP constraint we need is the one given in (16) which ensures that predicates precede their arguments. The first sentence in (11) thus has the structure in (23):

(23)

$$S\ (VP\ [SUBCAT\langle\ \rangle])$$

V	NP$_2$	NP$_3$	NP$_1$
[SUBCAT $\langle NP_1, NP_2, NP_3\rangle$]	[gen]	[dat]	[nom]
[LEX $\langle NP_1, NP_2, NP_3\rangle$]			
	ng libro	sa babae	ang lalaki
Nagbigay			

Any of the other examples can be straightforwardly derived by reordering the NPs in the sentence since no grammatical constraint is violated. [10]

The binding facts of Tagalog can be stated via reference to obliqueness on the LEX list since SUBCAT appears to be irrelevant to binding in Tagalog (as in Toba Batak). Thus as in Malagasy, the fact that an agent cannot be a reflexive follows from its position at the top of the LEX list. The reflexivization constraint for Tagalog is stated in (24):

(24)   **Tagalog Reflexivization Constraint:**
A reflexive in Tagalog must be coindexed with a less oblique coargument on the LEX list.
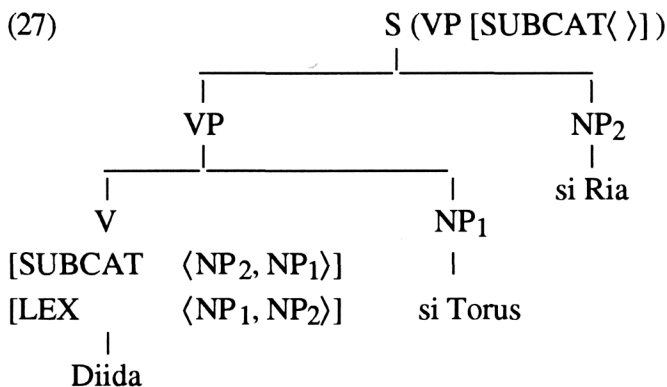
Thus examples like (25), in which a nominative nonagent is reflexive, are ruled in, whereas examples like (26), in which an agent is reflexive, are ruled out:

(25)   Iniisip   nila   ang kanilang sarili
       think about (- active) they (gen)   their self (nom)
                'Themselves are thought about by them'

(26) *Iniisip　　　　　　sila　　　　　ng kanilang sarili
　　　think about (- active)　they (nom)　their self (gen)
　　　'They are thought about by themselves'
　　　(examples adapted from Schachter 1976:503-5)

8.3 Toba Batak Clause Structure in HPSG

　　　Toba Batak clauses can be described by implementing ID schemata 1 and 2 (as in Malagasy and English). Thus active and nonactive clauses alike are assigned the optimal structure suggested in (13). The structure of (12)b, for example, is given in (27):[11]

(27)　　　　　　　　　　　　S (VP [SUBCAT⟨ ⟩])
　　　　　　　　　　　_____|_____
　　　　　　　　　　　|　　　　　　　　　　　　|
　　　　　　　　　　VP　　　　　　　　　　　NP$_2$
　　　　　　_____|_____　　　|
　　　　　|　　　　　　　　　　　|　　　　si Ria
　　　　　V　　　　　　　　　　NP$_1$
[SUBCAT　　⟨NP$_2$, NP$_1$⟩]　　|
[LEX　　　　⟨NP$_1$, NP$_2$⟩]　　si Torus
　　|
　　Diida

Despite the obvious differences in arboreal structure between Tagalog and Toba Batak, constraints on reflexive-binding in the two languages are very similar (as was noted earlier). Thus the Tagalog binding constraint stated in (24) will also serve to give the correct results for Toba Batak. In both languages, SUBCAT is irrelevant for binding purposes.

9. Conclusions

　　　The approach outlined above provides enough range of variation in structure to account for the arboreal contrasts between WA languages, while at the same time providing the kind of theoretical machinery needed to account for their commonalities. The overriding conclusion to be drawn from the data reviewed above is one which runs counter to the expectations of a theory like that outlined by GHT. Phrase structural relations do not provide an adequate basis for explaining the grammatical properties of WA languages. Implicating 'underlying' tree structures, where surface structures fail to account for the facts, simply leads to

unmotivated and frequently incorrect representations for the languages in question. Put simply, languages DO differ in syntactic structure, even when they share many inherited properties, as in the case of Malagasy, Tagalog, and Toba Batak.

Notes

1. A particularly glaring problem in this respect (which I do not have time to discuss in depth here) is the exclusive extractability of the subject [Spec, IP] in WA languages. Nothing in GB theory leads us to expect that [Spec, IP] should be the sole locus of extraction. Indeed, the Empty Category Principle (Chomsky 1981) leads to the expectation that [Spec, IP] CANNOT be extracted from. GHT offer only a vague suggestion as to how this problem might be solved. Until such a solution actually comes along, the special status of [Spec, IP] with respect to extraction stands as an obvious hole in the GHT analysis.

2. It is far from clear in what sense an element in [Spec, IP] is a 'topic'. Why is [Spec, IP] a 'topic' position in WA, but a 'subject' position in English? Why should a 'topic' have to receive nominative case?

3. As Keenan 1992 points out, it is far from obvious why preposition incorporation should deprive the stranded NP of case. Keenan shows that incorporated prepositions still assign case to their objects in Greek, Latin, and Russian.

4. Lack of space precludes a comprehensive summary of all the problems I am aware of. The reader is referred to Keenan 1991 and 1992 for additional discussion of a number of problems with the GHT analysis. See also Kroeger 1991.

5. For objections to the appropriateness of the labels 'active' and 'passive' in Toba Batak, see Schachter 1984. The behavior of voice morphology is actually considerably more complicated than I am assuming here for presentational purposes. See Nababan (1981: chapter 4) for a fuller presentation of the voice morphology.

6. I am aware of only one potential contrast in the properties of active internal NPs versus passive internal NPS. Nababan (1981:69) notes that agentive NPs can be freely deleted when they are internal NPs in passive clauses, but no mention is made of the possibility of deleting active internal NPs. I assume that this deletion occurs under discourse conditions where the agent is unknown, or at least not previously mentioned in the discourse, since Schachter (1984:128)

explicitly notes that discourse-controlled internal NPs are not deletable (in contrast to 'external' NPs ('subjects' in my terms)).
7. For a reconstruction of the notion of 'agent' within HPSG, see section 8, where 'agent' (or 'actor') is construed as the highest element on the LEX list.
8. The difference between LEX and SUBCAT can be conveniently thought of as being determined by both grammatical and nongrammatical requirements. SUBCAT can be viewed as an 'online' version of LEX which is selected in response to pressures imposed by changing discourse and processing situations. Voice morphology overtly registers the way in which the SUBCAT list differs from the LEX list (if at all).
9. Note that the analysis of nonsubject agents proposed here suggests a natural extension to the treatment of nominal possessors using essentially the same machinery given in (18) and (19). All that is required is an extension of the formalism to treat nominal and verbal possessors as a natural syntactic class.
10. The LP constraints could be augmented by the use of pragmatically based features if that turned out to be desirable. Thus, a language like Tagalog could have ordering of argument NPs partially determined by features such as [+/- topic], etc. I have not investigated the desirability of this approach in any depth.
11. It should be noted, however, that some additional extensions to the theory appear necessary to deal with dative verbs in Toba Batak, an issue I do not investigate here. See Schachter 1984 for details.

References

Blake, Barry. 1990. Relational grammar. Routledge: London.
Chomsky, Noam. 1981. Lectures on government and binding. Foris: Dordrecht.
Davies, William D. 1993. Javanese subjects and topics and psych verbs. Linguistics 3: 239-277.
Dowty, David. 1982. Grammatical relations and Montague grammar. P. Jacobson & G. Pullum (eds.), The Nature of syntactic representation. Dordrecht: Reidel.
Guilfoyle, Eithne, Henrietta Hung & Lisa Travis. 1992. Spec of IP and Spec of VP: two subjects in Austronesian languages. Natural language and linguistic theory 10.3:375-414.
Jackendoff, Ray. 1972. Semantic interpretation in generative grammar. MIT Press: Cambridge.

Keenan, Edward L. 1976. Remarkable subjects in Malagasy. C. Li (ed.), Subject and Topic. Academic Press: New York.

Keenan, Edward L. 1991. Malagasy X' structure. UCLA ms.

Keenan, Edward L. 1992. VP nominative structures & western Malayo-Polynesian languages. The Malagasy case. UCLA ms.

Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. Linguistic Inquiry 8:63-99.

Kroeger, Paul. 1991. Phrase structure and grammatical relations in tagalog. Stanford University dissertation.

Nababn, P. W. J. 1981. A grammar of Toba Batak. Pacific Linguistics, Series D no. 37. ANU: Canberra.

Perlmutter, David & Paul Postal. 1983 (1977). Toward a universal characterization ofpassive. David Perlmutter (ed.). Studies in relational grammar 1. Chicago University Press: Chicago.

Pollard Carl & Ivan Sag. 1987. Information-based syntax and semantics, volume 1. CSLI: Stanford.

Pollard, Carl & Ivan Sag. 1992. Anaphors in English and the scope of binding theory. Linguistic Inquiry 23.2.

Pollard, Carl & Ivan Sag. 1993. Information-based syntax and semantics, volume 2. CSLI: Stanford.

Schachter, Paul. 1976. The Subject in Philippine languages: topic, actor, actor-topic, or none of the above. C. Li (ed.), Subject and Topic. Academic Press: New York.

Schachter, Paul. 1984. Semantic-role-based syntax in Toba Batak. Paul Schachter (ed.), Studies in the structure of Toba Batak. UCLA Occasional papers in linguistics no. 5.

Schachter, Paul & Fe Otanes. 1972. A Tagalog reference grammar. UC Press: Los Angeles, Berkeley.