

# Proceedings of the Sixteenth Meeting of the Austronesian Formal Linguistics Association (AFLA)

University of California, Santa Cruz

May 1 - 3, 2009

edited by Sandy Chung, Daniel Finer, Ileana Paul, and Eric Potsdam

## Table of Contents

	<i><u>Introduction</u></i>	i
Douglas Ball	<i><u>Hili-Clauses: Insights into Tongan Nominalizations</u></i>	1-15
William D. Davies and Eri Kurniawan	<i><u>Movement and Locality in Sundanese Wh-Questions</u></i>	17-28
Catherine Fortin	<i><u>On the Left Periphery in Indonesian</u></i>	29-43
Vincent Homer	<i><u>Backward Control in Samoan</u></i>	45-59
Laura Kalin	<i><u>News about the No-Phrase: Specificational Pseudoclefts in Malagasy</u></i>	61-75
Edward L. Keenan	<i><u>Voice Determines Co-Argument Anaphora in W. Austronesian</u></i>	77-91
Seongyeon Ko	<i><u>Two Types of Anticausatives in Acehnese</u></i>	93-107
Chao-Lin Li	<i><u>Voice, Agree and Locality in Paiwan</u></i>	109-123
Diane Massam	<i><u>The Structure of (Un)ergatives</u></i>	125-135
Justin Nuger	<i><u>On Downward-Entailing Existentials and Differential Object Marking in Palauan</u></i>	137-151
Ileana Paul	<i><u>Bare Nouns, Incorporation, and Scope</u></i>	153-164
Matt Pearson	<i><u>Another Look at No: Pseudo-Clefts and Temporal Clauses in Malagasy</u></i>	165-179
Norvin Richards	<i><u>The Tagalog Copula</u></i>	181-195
Yosuke Sato	<i><u>Radical Underspecification, General Number and Nominal Mapping in Indonesian</u></i>	197-209
Jozina Vander Klok	<i><u>Direct Adjectival Modification in Javanese</u></i>	211-225

## PREFACE

Although the Austronesian Formal Linguistics Association (AFLA) has been holding annual meetings since 1994, until now it has had no consistent approach to the publication of its *Proceedings*. Papers from AFLA 2 and AFLA 14 were published as edited volumes; in other years the local organizers published the *Proceedings* in their Department's Working Papers series; in still other years no *Proceedings* was published. The 16th annual meeting of AFLA was held May 1-3, 2009, at the University of California, Santa Cruz. During the business meeting, the idea was floated that the *Proceedings* henceforth be published electronically, in a consistent format, at the AFLA website (<http://ling.uwo.ca/afla/>), which is generously hosted by the University of Western Ontario. The initial result is this volume, which has emerged very quickly indeed—less than six months after AFLA 16 was held. Our hope is that on-line publication of this and future volumes of the *Proceedings of AFLA* will enable research on the formal linguistics of Austronesian languages to reach as wide a readership as possible.

We want to thank UCSC's Linguistics Department and its Linguistics Research Center for hosting AFLA 16, the authors for submitting their papers so efficiently, and the University of Western Ontario for hosting the website at which this volume is posted. We also wish to acknowledge the precedent set by the *Proceedings of AFLA 12*, which was published on-line as *UCLA Working Papers in Linguistics* No. 12, and whose stylesheet heavily influenced the stylesheet we constructed for the *Proceedings of AFLA*.

Sandra Chung  
Daniel Finer  
Ileana Paul  
Eric Potsdam

## VOICE DETERMINES CO-ARGUMENT ANAPHORA IN W. AUSTRONESIAN

Edward L. Keenan  
University of California at Los Angeles  
edward.keenan1@gmail.com

We review data from Toba Batak and Malagasy in which anaphors in nuclear sentences asymmetrically c-command their antecedents. We provide a language independent, semantic, definition of *anaphor* and show how to directly interpret expressions with c-commanding anaphors. Crucial here is the semantic interpretation of voice affixes. So for purposes of semantic interpretation it is not necessary for anaphors to be c-commanded by their antecedents. Abandoning Principle A we offer a mathematically precise definition of *structural invariant* of a grammar (based on Keenan & Stabler 2003) and claim as universals that both anaphors and the anaphor-antecedent relation are structural invariants of human language. We suggest, but lack the space to pursue, that the interpretative role of voice morphology in W. Austronesian is analogous to that of case morphology in languages like Korean and Greek. We conclude with a summary of the advantages of the approach we take.

### 1. Core Anaphora Patterns in Toba Batak and Malagasy

#### 1.1. Toba Batak (N. Sumatra)

The data below come from the papers in Schachter 1984a, Cole & Hermon 2008, and some modest consultant work. We see in (1) and (2) that for verbs built by prefixing *mang-* to a root, anaphors are c-commanded by the following Agent antecedent. But for verbs built from *di-*prefixed roots it is the Agent which forms a constituent with the verb, its anaphor c-commanding it, (2b).

- |  |  |
|--|--|
| (1)a. [Mang+ida si Ria] si Torus<br>mang-see art Ria art Torus<br>'Torus sees Ria' | b. [Mang+ida dirina] si Torus<br>mang-see self art Torus<br>'Torus sees himself' |
| (2)a. [Di-ida si Torus] si Ria<br>di-see art Torus art Ria<br>'Torus saw Ria'      | b. [Di-ida si Torus] dirina<br>di-see art Torus self<br>'Torus saw himself'      |

We ignore the internal structure of proper nouns (DPs) like [*si Ria*] where *si* functions as a proper noun article. In all four Ss above the immediate postverbal DP is tightly bound to the verb. Schachter (1984b) shows that adverbs, like 'yesterday' cannot separate them, but they do occur naturally between the two DPs. And more importantly, only the clause final DPs above can be relativized or wh-questioned by movement.

- (3)a. Manjaha buku guru i  
 read book teacher Det  
 ‘The teacher reads the book’
- b. Dijaha guru buku i  
 read teacher book Det  
 ‘The teacher read the book’
- c. guru na manjaha buku i  
 teacher lnk read book Det  
 ‘the teacher who is reading a book’
- d. buku na dijaha guru i  
 book lnk read teacher Det  
 ‘the book the teacher read’
- e. \*buku na manjaha guru i  
 book lnk read teacher Det
- f. \*guru na dijaha buku i  
 teacher na read book Det

The well known “Subjects Only” phenomenon in (3e-f) is characteristic of W. Austronesian languages (Tagalog, Malagasy). Judgments on (3e-f) are strong and immediate. They reflect the basic interpretative mechanisms of nuclear clauses, not just an arbitrary constraint on what can be extracted. (3f) for example could only mean *the teacher that a book read*, which is nonsense.

Emmorey (1984) provides evidence of a different sort for the constituency in (1) and (2). She shows spectographically that the verb + its following DP form an intonation group in which the nuclear pitch accent falls on the stressed syllable of the last lexical item in the Predicate Phrase (PredP). Pitch accent placement identifies the immediate post-verbal DP in *mang-* and *di-* verbs as the last lexical item in their PredP’s. In conjoined PredP’s both conjuncts receive this accent (Emmorey op.cit.):

- (4)a. [[Manuhor baoang] jala [mangolompa mangga]] halak an  
 buy onions and cook mangos man Det  
 ‘The man buys onions and cooks mangos’
- b. [[Dituhor si Ore] jala [dilompa si Ruli]] mangga  
 di-buy art Ore and di-cook art Ruli mangos  
 ‘Ore buys and Ruli cooks mangos’

More could be said here, but all generative grammarians, despite theoretical differences, agree on the major constituency break in Ss like (1a-b) and (2a-b).

## 1.2. Malagasy (Madagascar)

The data here draw on extensive personal fieldwork, Rajemisa–Raolison 1971, Keenan 1995 and 2008, and Travis 2005 and 2006. Simple transitive Ss are given in (5).

- (5)a. [Namono azy] Rabe  
 past+maN+kill him.acc Rabe.nom  
 ‘Rabe was killing him’
- b. [Novonoin-dRabe] izy  
 past+kill+ina+Rabe.gen he.nom  
 ‘Rabe was killing him’

We choose the cumbersome progressive aspect in translating (5a-b) to acknowledge that the simple past tense in English is telic in a way in which the simple past in Malagasy is not (Travis 2005, Rasolofy 2006). The transitive verbs in (5a-b) are both built by affixing the root *vono* ‘kill’. That in (5a) is prefixed with *-aN-*, *N* a homorganic nasal which may trigger (as here) the loss of the initial consonant in a root, or its mutation to an affricate. This prefix is usually noted *man* in traditional Malagasy grammars (Malzac 1926; Rahajarizafy 1960). All verbs mark past tense with *n* (or *no-* = /nu/ if consonant initial) and future with *h* (or *ho*). Verbs built with *aN* take an *m* prefix in present tense and imperatives as do many other prefixes: the semantically neutral *i-*, reciprocal *if-* and causative *amp-*. Such verbs will be called *m-class* verbs.

In contrast the verb in (5b) is built by suffixing *ina* to the root *vono* yielding *vonoina* (with stress shift from the first syllable of *vono* = /VU.nu/ to the second: /vu.NU.i/na/. Then the agent phrase *Rabe* in (5b) combines with *vono+ina* in the morphophonemically complex way that possessors bind to their possessum (see Keenan & Polinsky 1998). We note that the *n-dR* in (5b) is standard orthography and represents a single phoneme—a prenasalized post-alveolar affricate. In distinction to Toba Batak, Malagasy distinguishes three cases in the pronoun system – nominative, accusative and genitive. Toba has distinct genitive forms, but does not distinguish nominative and accusative pronouns. Without exception, verbs built by suffixing a root are unmarked in present tense, and so are not *m-class* verbs.

Despite these morphological differences the alternation in (6a-b) behaves like that between *mang* and *di* verbs in Toba, and essentially the same interpretative mechanisms can be used to interpret the structurally variable position of anaphors:

- |  |   |
|--|---|
| (6)a. [Namono tena] Rabe<br>past+[aN+kill]self Rabe<br>‘Rabe killed himself’ | b. [[Novonoin-dRabe] ny tenany]<br>past+[[kill+ina]+Rabe.gen] the self.his<br>‘Rabe killed himself’ |
|--|---|

Of interest however is a morphosyntactically more complicated verb form which also allows object anaphors to c-command their antecedents. *M-class* verbs can be suffixed with *-ana* adding an argument to the theta grid of the predicate. The new argument is presented *PredP* externally and bears an oblique role to the predicate, such as *Instrument* (7b) or *Benefactive* (8b).

- |   |   |
|---|---|
| (7)a. Manao (m+aN+tao(v)) trano amin’ny birikinay Rabe<br>Pres+aN+make house with’the bricks+our Rabe<br>‘Rabe is building a house with our bricks’ | b. Anaovan-dRabe (aN+tao(v)+ana+Rabe) trano ny birikinay<br>[[aN+make]+ana]+Rabe.gen house the bricks.our<br>‘Our bricks are being used by Rabe to build a house’ |
| (8)a. Manao (m+aN+tao(v)) trano ho an-dRaso a Rabe<br>Pres+aN+make house for-Raso a Rabe<br>‘Rabe is building a house for Raso a’                   |   |

- b. Anaovan-dRabe (aN+tao(v)+ana+Rabe) trano Rasoa  
 [[aN+make]+ana]+Rabe.gen house Rasoa  
 ‘Rabe is building a house for Rasoa’

Semantically there are two important points to notice about *ana* verbs. Most obvious is that in (7a) and (8a), built with m-class verbs, the theta role of the oblique constituent is marked with a Preposition—*amina* in (7a) and *ho an-* in (7b). (*Ho an-* is treated as a single Preposition despite being written as two morphemes). These PPs are not at all obligatory and are naturally regarded as adjuncts, not arguments of the verb. But in the corresponding sentences (7b) and (8b) the PredP-external DP is nominative and is required, so it is an argument, though its theta role is not specifically specified. We just know that it is an oblique. In typical cases it is inferrable from context, as the kinds of objects that can be instruments are not those that can be benefactees, manners, etc. Sometimes though ambiguity or vagueness does arise. In (8b) *Rasoa* might be a proximate cause, having nagged Rabe so much that he built the house for Rasoa’s parents.

Second, *ana* verbs contain the m-class verb as a constituent. By Compositionality then we expect the interpretation of an *ana* verb to be dependent on the interpretation of the m-class verb it contains. Contrast now:

- (9)a. Mamono tena ho an’ny taranatsika isika  
 Pres+[aN+kill] self for’the descendents.our we.incl  
 ‘We kill ourselves (work hard) for our descendents’
- b. [[Amonoantsika] ([aN+vono+ana]+tsika) tena] ny taranatsika  
 [[pres+[aN+kill+ana]+1.incl.gen] self] the offspring+1.incl.gen  
 ‘Our offspring is the reason we kill ourselves (work hard)’

In (9b), from Rajemis-Raolison 1971:113, the relevant c-command relations are those between the agent phrase *-tsika*, the first plural inclusive genitive pronoun, bound to the verb *amonoana* and the c-commanding direct object *tena* ‘self’. So the reflexive *tena* ‘self’ asymmetrically c-commands its antecedent. Keenan 2008 provides compositional interpretations for the types of verbs in (9b) in which *-ana* suffixes to an already prefixed form *-aN+vono* = *-amono*.

Evidence for the gross constituent boundary is readily available in the literature. For example Subjects Only holds: only the PredP-external position can be relativized, and it is replaceable by nominative case pronouns. Also various particles, such as the polar question particle *ve?* (Pearson 2005), are placed at the right PredP boundary in all cases above. So in (9a) it would occur between *taranatsika* and *isika*; in (9b) between *tena* and *ny taranatsika*.

We turn now to a language-independent definition of *anaphor* and then to our direct compositional interpretation of Ss with anaphors that c-command their antecedent.

## 2. Defining Co-argument Anaphors

Given a domain E of (possibly abstract) objects, we treat n-ary predicates (Pn’s) as denoting n-ary relations over E (subsets of E<sup>n</sup>). DPs of any sort denote partial functions of *type-1*, the set of

functions that map  $n+1$ -ary relations to  $n$ -ary relations. So they are the functions that reduce arity by one. For simplicity we limit ourselves here to  $P_n$ 's,  $0 \leq n \leq 2$ .  $P_2$ s, such as transitive verb phrases, determine the smallest environment in which non-eliminable binding occurs.

Among DPs, R-expressions (in the Binding Theory sense) denote type-1 functions  $F$  whose values at binary relations  $R$  satisfy (10):

$$(10) \quad F(R) = \{a \in E \mid F(aR) = 1\}, \text{ where } aR =_{\text{def}} \{b \in E \mid (a,b) \in R\}.$$
<sup>1</sup>

So the values  $F$  takes at binary relations are determined by its values at unary relations (subsets of  $E$ ). So *admires all poets* denotes  $(\text{ALL POET})(\text{ADMIRE}) = \{x \in E \mid (\text{ALL POET})(x\text{ADMIRE}) = 1\}$ , the set of objects  $x$  such that (ALL POET) is true of the set of things  $x$  admires. The  $F$  which satisfy (10) are provably just those satisfying the AEC below (Keenan 1987):

### Accusative Extensions Condition (AEC)

For all  $x, y \in E$ , all binary relations  $R, S$  over  $E$ , if  $xR = yS$  then  $x \in F(R)$  iff  $y \in F(S)$

For example, Keenan 1996 identifies many classes of DPs which satisfy the AEC and are thus R-expressions. They include proper names, quantified DPs of the form Det+NP (*some/no student, most students John knows*), partitives (*most of John's friends, less than five of the students*), and many others. To check that *most of John's friends* is an R-expression, imagine a model in which Mary praised just the people Sue distrusts. Then *Mary praised most of John's friends* and *Sue distrusts most of John's friends* have the same truth value. So *most of John's friends* is an R-expression. In general if a DP  $d$  can take a  $P_1$  as argument, so it is interpreted as a function  $F$  mapping subsets of  $E$  into  $\{\text{True}, \text{False}\} = \{1, 0\}$  then it extends automatically to a function  $F_{\text{acc}}$  which maps each binary relation  $R$  to  $\{a \in E \mid (a,b) \in R\}$ . Usually we just write  $F$  in this case rather than the more tedious  $F_{\text{acc}}$ .

Observe now that *himself* and complex anaphors (*everyone but himself, only himself and the teacher, etc*) fail the AEC: If Mary praised just Bob, Pat, Ed, Sue, Dana and Luke, and those are just the people Sue distrusts, then *Mary praised herself* is False, but *Sue distrusts herself* is True. Accusative Anaphors do however satisfy the following weaker invariance condition:

### Accusative Anaphor Condition (AAC)

If  $xR = xS$  then  $x \in F(R)$  iff  $x \in F(S)$

For example, if Mary praised just the people she admires then *Mary praised herself (only herself, etc)* and *Mary admires herself (only herself, etc)* have the same truth value.

---

<sup>1</sup> We assume for simplicity that  $R$  is binary. When  $R$  is  $n+1$ -ary take  $x$  to be an  $n$ -tuple. Then  $xR = \{y \in E \mid \langle x, y \rangle \in R\}$  and  $F(R) = \{x \in E^n \mid F(xR) = 1\}$ .

Def 1. An expression is an *accusative anaphor* iff in all models all its denotations satisfy the AAC but in some model it fails the AEC

Def 2. Here we define some properly anaphoric functions which map binary to unary relations. Expressions which denote these functions are provably accusative anaphors as defined above.

$$\begin{aligned} \text{SELF}(R) &= \{a \in E \mid (a,a) \in R\}; & (\text{ONLY SELF})(R) &= \{a \mid aR = \{a\}\} \\ (\text{EVERY BOY BUT HIMSELF})(R) &= \{a \in E \mid aR \cap \text{BOY} = \text{BOY} - \{a\}\} \end{aligned}$$

### 3. Compositional Interpretation

Classically we think of proper names, like *Bob*, as denoting an object  $b$  in the domain  $E$  of objects under consideration. Pls, like *laugh*, *cry*, etc. are interpreted as subsets of  $E$ , and a sentence like *Bob laughed* is interpreted as True iff  $b \in \text{LAUGH}$ , that is, if and only if the object *Bob* denotes is an element of the set of things that *laugh* denotes. Here we treat proper names equivalently, as those type-1 functions we call *individuals*. Namely, for each  $b \in E$ ,  $I_b$ , the *individual generated by*  $b$ , is that type-1 function which maps each set  $A$  to True iff  $b \in A$ . And by the accusative extension mentioned earlier,  $I_b$  maps a binary relation  $R$  to  $\{a \in E \mid (a,b) \in R\}$ . So for example *knows Mary* denotes the set of people who know Mary.

#### 3.1. Interpreting Toba Batak

3.1.1. *Dirina* denotes SELF.

3.1.2. *Mang-* and *di-* denote the type lifting functions MANG and DI taking binary relations as arguments yielding as values functions taking type-1 functions as arguments, as follows:

$$\begin{aligned} (11) \text{ a. } \text{MANG}(R)(H)(F) &= F(H(R)) & R \text{ any binary relation, } F, H \text{ partial type-1 functions} \\ \text{ b. } \text{DI}(R)(F)(H) &= F(H(R)) & R \text{ any binary relation, } F, H \text{ partial type-1 functions} \end{aligned}$$

*Mang* and *di* predicates are independently generated and interpreted. Neither contains the other as a constituent. So the roots that *mang* and *di* can prefix might not be exactly the same. Consider now the interpretation of Ss like (1b) and (2b), the latter with a c-commanding reflexive *dirina*. We interpret the root *ida* of the transitive verbs *mangida* and *diida* as a binary relation and we interpret just the constituents of (12a-b).

$$\begin{aligned} (12) \text{ a. } & [[[\text{Mang - ida}] \text{ dirina}][\text{si Torus}]] \\ & \text{MANG IDA SELF } I_b \\ & = ((\text{MANG}(\text{IDA}))(\text{SELF}))(I_b) && \text{Compositional interpretation} \\ & = I_b(\text{SELF}(\text{IDA})) && \text{Def MANG} \\ & = \text{True iff } b \in \text{SELF}(\text{IDA}) && \text{Def individual} \\ & = \text{True iff } (b,b) \in \text{IDA} \quad (\text{i.e. iff } b \text{ sees } b) && \text{Def SELF} \end{aligned}$$



b. [[Di - ida][si Torus]] dirina]	
DI IDA I <sub>b</sub> SELF	
= ((DI(IDA))(I <sub>b</sub> ))(SELF)	Compositional interpretation
= I <sub>b</sub> (SELF(IDA))	Def DI
= True iff b ∈ SELF(IDA)	Def individual
= True iff (b,b) ∈ IDA (i.e. iff b sees b)	Def SELF

Note that *ida*, *si Torus*, and *dirina* have the same denotation in both Ss. Their interpretative difference lies in the differing interpretations of *mang* and *di*. So on the view of grammar presented here grammatical morphology is semantically interpreted (not “uninterpretable”). (It is also typically *structurally invariant*, defined in section 4.)

### 3.1.3. Theta Roles

The analyses of Schachter 1984b and Sugamoto 1984 are built on theta role assignment, which we have so far ignored. So let us add it in, since it is clearly part of the semantic relation between *mang* and *di* predicates we want to capture. We treat theta roles like AGENT and THEME as relations between entities and the n-ary relations denoted by predicates in the language (Keenan 2008). We require for example, that for all binary relations R, all a,b ∈ E,

$$(13)a. \text{MANG}(R)(I_b)(I_a) = \text{True} \Rightarrow \text{THEME}(b,R) \text{ and } \text{AGENT}(a,R)$$

$$b. \text{DI}(R)(I_b)(I_a) = \text{True} \Rightarrow \text{AGENT}(b,R) \text{ and } \text{THEME}(a,R)$$

So we interpret *mang* as the function in (12a) which meets the condition in (13a). Similarly *di* denotes the function in (12b) satisfying (13b). Of course a thorough study would discriminate among different roots according to the theta roles they assign to their arguments. And it would treat roots with different numbers of theta roles. Here we simply illustrate how theta role interpretation can be integrated into semantic interpretation.

### 3.1.4. Interim Conclusion

It is straightforward to base generate and interpret nuclear Ss in Toba Batak with anaphors c-commanding their antecedents (or not). This is made possible by the type lifting interpretation of the verbal affixes *mang* and *di*. Further our analysis is linguistically highly valued since we only interpret overtly present constituents, yielding a better account of why speakers speak the way they do. On the Minimalist analysis in Cole & Hermon 2008 the important c-command condition relating anaphors and their antecedents can be overridden in surface. But how can c-command, part of UG and so genetically determined, be both necessary and ignored? Why don't speakers speak in underlying forms (or LFs) which do satisfy the c-command condition?

### 3.2. Interpreting Malagasy

The Malagasy verbs in (6) behave like *mang/di* ones in Toba, but the binding in (9b) is new.

#### 3.2.1. Predicate Modifiers (Pmods)

Pmods like *with our bricks* or *for Rasoaa* combine with  $P_n$ 's ( $n > 0$ ) to form  $P_n$ 's. They denote restricting functions  $F$  mapping  $n$ -ary relations to  $n$ -ary relations, that is,  $F(R) \subseteq R$ . The set of objects denoted by *build a house with our bricks* is a subset of that denoted by *build a house*. If Rabe is building a house with our bricks we can infer that Rabe is building a house. Similarly the denotation of *build a house for Rasoaa* is a subset of that of *build a house*. Prepositions then denote functions from type-1 functions into the set of restricting modifiers. Compare:

- (14) a. [[ Man+tao(v) [ilay trano]] Rabe 'Rabe is building that house'  
 $\begin{matrix} \text{AN} & \text{R} & \text{F} & \text{G} \\ \text{AN}(\text{R})(\text{F})(\text{G}) \\ = \text{G}(\text{F}(\text{R})) \text{ from the denotation of } aN \end{matrix}$
- b. [[ Man+tao(v) [ilay trano]] [ho an-dRasoaa]] Rabe 'Rabe builds that house for Rasoaa'  
 $\begin{matrix} \text{AN} & \text{R} & \text{F} & \text{q} & \text{H} & \text{G} \\ \text{AN}(\text{R})(\text{F})(\text{q}(\text{H}))(\text{G}) \\ = \text{G}(\text{q}(\text{H})(\text{F}(\text{R}))) \end{matrix}$  AN(R)(F) is a unary relation, q(H) maps  $n$ -ary relations to  $n$ -ary relations
- c. Anaovan- dRabe ilay trano Rasoaa 'For some function p from individuals to restricting functions, Rabe is making that house p Rasoaa'  
 $\begin{matrix} \text{[[aN+tao]+ana]} & \text{G} & \text{F} & \text{H} \\ \text{ANA}(\text{AN}(\text{R})) \\ \text{ANA}(\text{AN}(\text{R}))(\text{G})(\text{F})(\text{H}) \\ = \exists p(\text{G}(\text{p}(\text{H})(\text{AN}(\text{R})(\text{F})))) \\ = \exists p(\text{G}(\text{p}(\text{H})(\text{F}(\text{R})))) \end{matrix}$
- d. [[amonoan+ tsika] tena] ny taranatsika  
 $\begin{matrix} \text{[[aN+vono]+ana]} & + & \text{G} & \text{SELF} & \text{H} \\ \text{ANA}(\text{AN}(\text{R})) \\ \text{ANA}(\text{AN}(\text{R}))(\text{G})(\text{SELF})(\text{H}) \\ = \exists p(\text{G}(\text{p}(\text{H})(\text{AN}(\text{R})(\text{SELF})))) \\ = \exists p(\text{G}(\text{p}(\text{H})(\text{SELF}(\text{R})))) \end{matrix}$

One checks that (14b) entails (14c), which entails (14a). (14b) and (14c) differ only in that (14b) specifies the theta relation of Rasoaa, Benefactee. (14c) just quantifies over such relations. And (14d) is the interpretation for (9b) showing that SELF is within the scope of the Agent phrase G. It instantiates the same interpretative pattern in (14c). Again it is the verbal morphology that type lifts the transitive predicate yielding the correct result. For deeper discussion of the semantics of the Malagasy voice system see Travis 2005, Rasolofo 2006, and Keenan 2008.

#### 4. How does English Differ from W. Austronesian?<sup>2</sup>

Cole & Hermon (2008:145) claim that Passive Agents in English can't antecede subject anaphors because they have not c-commanded them at any stage of the derivation, whereas Agents of *di*-verbs in Toba can antecede them on the grounds that they c-commanded them at an intermediate stage in the derivation. Their analysis posits many movement operations which change c-command relations and which they acknowledge lack independent motivation. Moreover, as we have shown, binding of reflexives simply does not require c-command.

For us the crucial difference is that Toba and Malagasy present verbal type lifting expressions (*mang*, *di*, *aN*, etc.) which determine the composition and theta roles of two or more argument functions. English lacks such type lifting morphology, punkt. Nuclear clauses in English differ in structure from those in Toba and Malagasy. And in English passives, the closest English gets to *di/ina/ana* verbs, Agent phrases are not arguments but optional adjuncts absent 80% to 96% of the time (Keenan & Manorohanta 2004). We have tried treating the passive *en* morphology in English as a type lifter, along the lines of (15), but without success.

(15)  $EN(R)(F) = 1$  iff for some individual  $y$ ,  $y \in F(R)$

For  $F = \text{SELF}$  this would say that (16) holds.

(16) *Himself was criticized* is True iff for some  $y$ ,  $y$  criticized himself.

We cannot assess (16) since *Himself was criticized* is not used meaningfully in standard English. But we get more clearly incorrect results when *himself* is replaced with a mere quantified DP:

(17)a. *No worker was criticized* is True iff for some  $x$ ,  $x$  criticized no worker.

b. *Just one worker was criticized* is True iff for some  $x$ ,  $x$  criticized just one worker

(17a-b) are clearly false, as they get the scope wrong. *No worker was criticized* means for no worker was there an  $x$  who criticized him (not *someone didn't criticize any worker*). But we can't put the existential quantification inside the scope of EN in (15) since that would eliminate an argument so neither passive nor reflexive could apply.

#### 5. Grammatical Invariants

Having lost the c-command constraint relating anaphors and antecedents we seek some new structural regularities. We propose two, slightly different in character than standard Binding Theory. We define the notion structural *invariant* and claim that both anaphors, as defined, and the anaphor-antecedent relation, are universal invariants of natural language. Invariants are defined in terms of structure preserving substitutions (Keenan & Stabler 2003).

---

<sup>2</sup> Thanks to Matt Pearson for pushing me on this point.

5.1. A grammar  $G$  is given by a finite set of categorized expressions  $\text{Lex}_G$ , the *lexicon* of  $G$ , and a set of structure building partial functions called *rules*.  $L(G)$ , the set of expressions generated by  $G$ , is the closure of the lexicon under the rules; that is, all expressions derivable from the lexical items by finitely many applications of the rules. A *structure (preserving) map* for  $G$  (see K&S 2003:20) is a bijection from  $L(G)$  to  $L(G)$  which does not change how expressions are derived. If  $z$  is derived by applying a rule  $F$  to some  $x$  and  $y$ , that is,  $z = F(x,y)$ , then  $h(z)$ , the expression  $h$  maps  $z$  to, is  $F(h(x),h(y))$ . So  $z$  and  $h(z)$  are derived in exactly the same way. Structure maps are standardly called *automorphisms*. Expressions  $x$  and  $y$  are *isomorphic* iff some automorphism maps one to the other. And an *invariant* of  $G$  is a *linguistic object* that is mapped to itself (is invariant) by all the automorphisms (auts). A linguistic object is an expression, a property (set) of expressions, a relation between expressions, a (partial) function on expressions, etc. So to say that an expression  $d$  in  $L(G)$  is invariant just says that  $h(d) = d$ , all auts  $h$  of  $G$ . In general we expect that expressions we call *grammatical morphemes*, *heads of functional projections*, etc. are automorphism invariant. They are the expressions you cannot replace by any others without changing how expressions are derived.

A set  $P$  of expressions is invariant iff  $h(P) = P$ , all auts  $h$ , where  $h(P)$  is  $\{h(d)|d \in P\}$ . So for  $P$  invariant, an automorphism just permutes members of  $P$  with themselves, but does not map anything out of  $P$  or map anything into  $P$  that was not already there. Similarly for  $R$  a (binary) relation among expressions,  $h(R)$  is just  $\{<h(c),h(d)>| \text{all } <c,d> \in R\}$ . This just says that, for  $R$  invariant,  $c$  is related by  $R$  to  $d$  iff for all auts  $h$ ,  $h(c)$  is related by  $R$  to  $h(d)$ .

These definitions may be unfamiliar but the linguistic objects they identify are not. As linguists we largely agree on the expressions in a language we call “functional”. Automorphism invariance just gives a formal characterization of that common intuition. To see this explicitly here is a “mini-grammar” for Toba to illustrating these definitions.

## 5.2. TB, a Mini-Grammar of Toba Batak

We treat a possible expression as a pair  $(s,C)$ , for  $s$  is a string of vocabulary items, and  $C$  a category name. This is similar to the use of labeled bracketings [<sub>DP</sub> *John's doctor*]. It is helpful when defining the rules of  $G$  and it eliminates some ambiguities. Thus  $(\text{love},N)$  and  $(\text{love},V)$  are distinct expressions since their second coordinates differ. The vocabulary of TB is the set of strings that appear as left hand coordinates in lexical items or are introduced in the values of any of the structure building functions below.

$\text{Cat}_{\text{TB}}$       P2, P2a, P2n, P1a, P1n, P0, Vaf, Vpf, NP, NPrefl, CJ

$\text{Lex}_{\text{TB}}$

P2:	saw, praised, flattered	(that is, $(\text{saw},P2) \in \text{Lex}_{\text{TB}}$ , etc.)
P1n:	laughed, cried, sneezed	CJ: and, or
Vaf:	mang-	Vpf: di-
NP	John, Bill, Sam	NPrefl self

Rule<sub>TB</sub> There are three structure building functions: VM (Voice marking), PA (Predicate Argument), and COORD (Coordination)

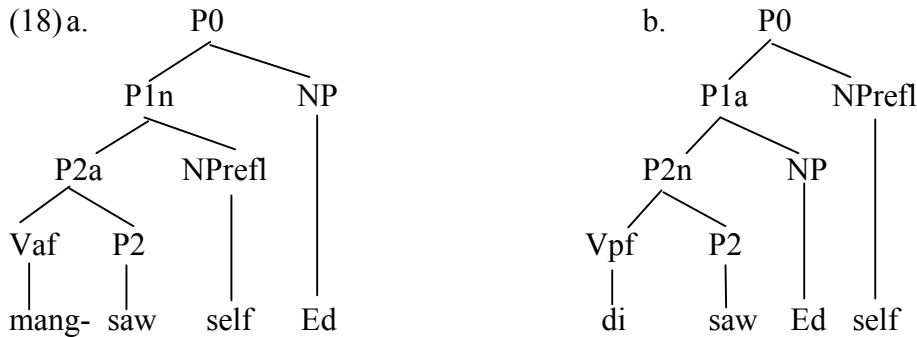
	VM	
(mang, Vaf), (t, P2)	→	(mang+t, P2a)
(di, Vpf), (t, P2)	→	(di+t, P2n)

So VM maps the string *mang* of category Vaf and any string *t* of category P2 to form *mang+t* of category P2a. Similarly VM builds strings of category P2n from *di* of category Vaf and *t* of category P2. The domains and actions of PA and COORD are understood similarly:

	PA	
(s, P2x) (t, NP)	→	(s+t, P1y)     all $x \neq y \in \{n, a\}$
(s, P2a) (t, NPrefl)	→	(s+t, P1n)     all $x \in \{n, a\}$
(s, P1x) (t, NP)	→	(s+t, P0)     all $x \in \{n, a\}$
(s, P1a) (t, NPrefl)	→	(s+t, P0) <sup>3</sup>

	COORD	
(u, CONJ) (s, C) (t, C)	→	(v+++s+u+t, C)     all $C = P2, P2a, P2n, P1a, P1n, P0, NP$ v = both if u = and; otherwise v = either
(u, CONJ) (s, C) (t, C')	→	(v+s+u+t, NPrefl) all $C \neq C' \in \{NP, NPrefl\}$ , v, u as above

The trees below summarize proofs that their terminal strings of category P0 are in L(TB).



(18a-b) are isomorphic as ordered labeled trees, but not isomorphic in TB: no aut can map one to the other since such an aut would have to map (Ed, NP) to (self, NPrefl), so that aut or a variant would map (Ed laughed, P0) to (self laughed, P0), which is not in L(TB), a contradiction.

## 6. Some Empirical Invariants of Toba Batak

6.1. (mang-, Vaf) is invariant; as is (di-, Vpf) and (dirina, NPrefl).

The truth of 6.1 is a matter of proof (omitted here for reasons of space) not of intuition. We have

an explicit definition of invariant, so given a grammar  $G$  and  $d \in L(G)$ ,  $d$  is either fixed (mapped to itself) by all automorphisms of  $G$  or not. Intuition plays no role here.

6.2.  $\text{Lex}_{\text{TB}}$  is invariant.

So lexical items are mapped to lexical items (not necessarily themselves) by all automorphisms. We can make up  $G$  in which  $\text{Lex}_G$  is not invariant, so 6.2 is empirical not definitional.

6.2.1. For all  $C \in \text{CAT}_{\text{TB}}$ , the property of having category  $C$  is invariant.

That is, all auts map each  $\text{PH}(C)$  to  $\text{PH}(C)$ , where  $\text{PH}(C) =_{\text{def}} \{d \in L(\text{TB}) \mid (\text{Cat}(d) = C)\}$

6.2.2. The set of anaphors in  $L(\text{TB})$  is invariant.

For  $d \in L(\text{TB})$ ,  $d$  is provably an anaphor iff  $\text{Cat}(d) = \text{NPrefl}$ . (*either both himself and John or both himself and Bill*,  $\text{NPrefl}$ ) is an anaphor.

6.2.3. The AA (anaphor-antecedent) relation in TB is invariant, AA defined ad hocly by:

(19) For  $x, y, z \in L(\text{TB})$ ,  $x$  is a possible antecedent of an anaphor  $y$  in  $z$  iff  $y$  is an anaphor and there is a constituent  $z'$  of  $z$  such that for some  $u$  of category P2,

Either  $z' = \text{PA}(\text{PA}(\text{VM}(\text{mang-}, u), y), x)$  or  $z' = \text{PA}(\text{PA}(\text{VM}(\text{di-}, u), x), y)$ <sup>3</sup>

The category coordinates of *mang* and *di* are omitted as predictable from the string coordinate. In fact for any  $(s, C)$  and  $(s, C') \in L(\text{TB})$ ,  $C = C'$ . This isn't true of English, with lexical items such as (*respect*, N) and (*respect*, V) or complex expressions such as *Ma's home cooking*, both a DP and an S. That the AA relation in TB as defined above is invariant is actually trivial as we defined it in purely syntactic terms. That this definition does capture our pretheoretical intuitions of binding in TB is not trivial, but an empirically significant fact about TB. We need (but do not quite have) a language general definition of the AA relation analogous to that of *anaphor* so we know that different syntactic characterizations of that relation in different grammars are actually characterizing the same relation.

In proving claims about invariants we can rely on universal invariants, linguistic objects that are uniformly definable for all  $G$  and whose invariance follows just from the definition of invariant. Here are some useful examples, noting first two lemmas (K&S 2003:21).

---

<sup>3</sup> We could replace the last disjunct by:  $z' = \text{PA}(\text{PA}(\text{VM}(\text{di-}, u), v), w)$ ,  $v \neq w \in \{x, y\}$ . This would take into account Cole & Hermon's observation that *dirina sandiri* may occur with reflexive meaning as the first argument of a *di* verb (with reduced acceptability). So TB allows *dirina* to be either argument of a P2a (but not both at once). We haven't added *sandiri* itself to our grammar as we don't know much about its syntax – but we could add it easily in a new category, say  $\text{NPrefl}/\text{NPrefl}$ , something which combines with  $\text{NPrefl}$ s to form  $\text{NPrefl}$ s.

Lemma 1: If  $A \subseteq B \subseteq L(G)$  then for all automorphisms  $h$ ,  $h(A) \subseteq h(B)$

Lemma 2: For  $A \subseteq L(G)$ ,  $A$  is invariant iff for all automorphisms  $h$ ,  $h(A) \subseteq A$ .

### 6.3. Some Invariants for Arbitrary Grammars $G$

1. Of course  $L(G)$  itself is invariant: each aut  $h$  maps  $L(G)$  to  $L(G)$  since  $h$  is onto. This says that the property of being in  $L(G)$ , that is, of being grammatical in  $G$ , is invariant. (Were this not the case our definition of *invariant* would be flawed).

2. Again trivially each structure building function (thought of as a set of sequences of expressions) is invariant. By definition  $h$  maps each such function to itself.

3. For each structure building function  $F$ , the set of tuples of expressions in  $L(G)$  that constitute the domain of  $F$  is invariant. This is an important source of invariants since the rules are virtually always properly partial functions, so they can't "see" everything, and what they can see, those aspects of structure they are sensitive to, is structurally significant. Similarly the range of  $F$ , the set of its values at tuples of expressions in  $L(G)$ , is invariant.

4. The relations *is an (immediate) (proper) constituent of* are invariant. We define these relations more generally than usual as follows: The *immediate constituents of* a derived expression  $F(d_1, \dots, d_n)$  are just the  $d_i$ . *s is a proper constituent of t* iff there is a finite sequence of two or more expressions with  $s$  the first,  $t$  the last, and each non-final expression an immediate constituent of the next one. And finally, *s is a constituent of t* iff  $s = t$  or  $s$  is a proper constituent of  $t$ . *s is a sister of t* iff  $s \neq t$  and for some  $u$ ,  $s$  and  $t$  are immediate constituents of  $u$ . Finally *s c-commands t* iff  $t$  is a constituent of a sister of  $s$ .

5. The set of invariant subsets of  $L(G)$ , or for that matter of  $L(G)^n$ , is closed under set theoretical operations:  $\cap$ ,  $\cup$ ,  $-$ ,  $\times$ , and power setting.

## 7. Conclusions

We propose (20) as universal for natural language grammars  $G$ :

(20) Binding Theoretic Invariants

1. The property of being an anaphor is invariant in  $G$
2. The Anaphor-Antecedent relation in  $G$  is invariant

Invariant 1 says that a natural language grammar  $G$  morphosyntactically distinguishes anaphors in such a way that they cannot be replaced by non-anaphors preserving structure, that is, preserving how expressions are built up. Invariant 2 says that we can always identify the possible antecedents of an anaphor in terms of morphosyntactic structure. In the languages treated here identifying antecedents crucially uses the voice morphology on the verb (not something we would have been led to notice on the basis of English). In case marking languages such as Korean, Greek and Hindi we make crucial reference to case suffixes. (21a-b) from Korean are illustrative and suggestive.

- (21) a. (Motun) haksayng-tul-i caki-casin-ul piphanhayssta  
all student-pl-case<sub>1</sub> self-emph-case<sub>2</sub> criticized  
'All the students criticized themselves'
- b. Caki-casin-ul (motun) haksayng-tul-i piphanhayssta  
self-emph-case<sub>2</sub> all student-pl-case<sub>1</sub> criticized  
'All the students criticized themselves'

The case<sub>1</sub> suffix *-i* takes the shape *-ka* when suffixing a vowel final nominal. Writing *nom-* to generalize over these phonologically conditioned shapes we see that this morphology can serve the same type lifting function that *di-/ina* affixes served in Toba and Malagasy:

- (22) a.  $\text{nom}(F)(R)(G) = F(G(R))$ , R a binary relation, G a type-1 function, F a map from P1 denotations to P0 ones  
b.  $\text{nom}(\text{ALL STUDENT})(\text{CRITICIZE})(\text{SELF}) = (\text{ALL STUDENT})(\text{SELF}(\text{CRITICIZE}))$

This is essentially the same interpretation as for Ss built from *di* verbs with c-commanding reflexives in Toba. So we have a generalization: in head final Korean the lifting morphology is carried by Nouns, in head initial Toba and Malagasy it is carried by Verbs. How general is this pattern? It is certainly common that head final languages have extensive nominal case marking. But many also have voice marking on the verb, though we know of no pattern like that found in Toba or Malagasy. Similarly a few verb initial languages (not that common) have extensive nominal marking (Tagalog) though we do not know if they ever carry the type lifting load we have associated (implicitly) with *nom* marking in Korean. Lastly, we summarize what we see as merits of the analysis we have proposed:

(23) Merits of our direct interpretation analysis

1. Semantic interpretation is explicit and compositional; not mere syntactic regimentation.
2. Audible morphology is meaningful and a non-trivial part of structure.
3. Our analysis is simpler than the Minimalist alternative in Cole & Hermon 2008.
4. Our analysis resolves the c-command inconsistency: why is it needed if it can be easily overridden? Answer: it is not needed.

## References

- Cole, Peter and Gabriella Hermon. 2008. VP raising in a VOS language. *Syntax* 11.2:144–197.
- Emmorey, Karen. 1984. The intonation system of Toba Batak. In *Studies in the structure of Toba Batak*, ed. Paul Schachter, 37–59. UCLA Occasional Papers in Linguistics 5.
- Keenan, Edward L. 1987. Semantic case theory. Reprinted in *Semantics and contextual expression*, ed. Renate Bartsch, Johan van Benthem, and Peter van Emde Boas, 33–57. Dordrecht: Foris. 1989.



- Keenan, Edward L. 1995. Predicate-argument structure in Malagasy. In *Grammatical relations: Theoretical approaches to empirical questions*, ed. C. Burgess, K. Dziwirek and D. Gerds, 171 – 217. Stanford: CSLI.
- Keenan, Edward L. 1997. The semantics of determiners. In *The Handbook of contemporary semantic theory*. ed. Shalom Lappin, 41–63. Oxford: Blackwell.
- Keenan, Edward L. 2008. Voice and relativization without movement in Malagasy. In *Natural Language & Linguistic Theory* 26.3:467– 497.
- Keenan, Edward L. and Cécile Manorohanta. 2001. A quantitative study of voice in Malagasy. *Oceanic Linguistics* 40:67–85.
- Keenan, Edward L. and Maria Polinsky. 1998. Malagasy morphology. In *The Handbook of morphology*, ed. Andrew Spencer and Arnold M. Zwicky, 563–624. Oxford: Blackwell.
- Keenan, Edward L. and Edward P. Stabler. 2003. *Bare grammar*. Stanford: CSLI.
- Malzac, R.P. 1926. *Grammaire malgache*. 2e édition. Paris: Société d'éditions géographiques, maritimes et coloniales.
- Pearson, Matthew. 2005. The Malagasy subject/topic as an A' element. *Natural Language and Linguistic Theory* 23:381–457.
- Rahajarizafy, Antoine. 1960. *Essai sur la grammaire malgache*. Imprimerie catholique Antanimena-Tananarive.
- Rajemisa-Raolison, Régis. 1971. *Grammaire malgache*. 7e édition. Fianarantsoa, Madagascar.
- Rasolofo, A. 2006. Malagasy relative clause types and their functions. Doctoral dissertation, University of Oregon, Eugene.
- Schachter, Paul, ed. 1984a. *Studies in the structure of Toba Batak*. UCLA Occasional Papers in Linguistics 5
- Schachter, Paul. 1984b. Semantic-role-based syntax in Toba Batak. In *Studies in the structure of Toba Batak*, ed. Paul Schachter, 122–150. UCLA Occasional Papers in Linguistics 5.
- Sugamoto, Nobuko. 1984. Reflexives in Toba Batak. In *Studies in the structure of Toba Batak*, ed. Paul Schachter, 150–172. UCLA Occasional Papers in Linguistics 5.
- Travis, Lisa. 2005. Passives, states, and roots and Malagasy. *UCLA Working Papers in Linguistics* 12, 395–409.
- Travis, Lisa. 2006. Voice morphology in Malagasy as clitic left dislocation or 'Through the looking glass: Malagasy in Wonderland.' In *Clause structure and adjuncts in Austronesian Languages*, ed. Hans-Martin Gärtner, Paul Law and Joachim Sabel, 281–318. Berlin: Mouton de Gruyter.