Malagasy Clause Structure and Language Acquisition

2004

Edward L. Keenan,   Cecile Manorohanta

UCLA   Université Nord, Madagascar

abstract We argue that Malagasy (and related W. Austronesian languages[1]) has a positive setting for a macro-parameter RICH VOICE MORPHOLOGY which builds complex predicates that code the theta role of their argument: $S = [[_{Pred}V(\theta) + (X)] + DP]$. Manifestations of this parameter are: (1) Case and theta role are assigned in situ in nuclear clauses with no movement or co-indexing to a topic position. (2) Relative Clauses (and other "extraction" structures) satisfy the "Subjects Only" constraint, again with no movement or indexing. (3) UTAH is freely violated, as theta role assignment derives from compositional semantic interpretation. Predicates resemble lexical Ns in assigning case directly to arguments without using Prepositions and in combining directly with Dets to form DPs that include tense and negation (Keenan 1995, 2000). The major Predicate-Argument type is modeled on the Noun+Possessor one, not the Verb+Object one.

1. background Lexical verbs denote relations requiring that their arguments bear thematic (theta) relations such as *Agent, Theme, Goal, Experiencer* and *Means* to them. For example in *John chased Bill, Bill* bears the Theme relation to *chase*, meaning that the object it denotes[2] is affected by the chasing action. UTAH (Baker 1988:46) syntactically constrains the expression of theta roles:

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

We assume UTAH is updated, replacing "at the level of D-structure" by "in the derivationally prior structures at which theta roles have been assigned".

Formally UTAH seems arbitrary (if historically comprehensible). Would deriving constituents bearing the same theta role from non-isomorphic structures lead to a contradiction? Would we wrongly predict some theta equivalences? Deriving theta equivalence compositionally yields a negative answer to both questions. As a conceptual illustration first consider (1a,b) from Propositional Logic.

(1)   a.   (P or Q)   b.   not (not P and not Q)

(1a,b) are not syntactically isomorphic but are, like infinitely many other pairs, semantically interpreted the same in all models. This logical equivalence just depends on their compositional interpretation plus the same choice of lexical items, 'P' and 'Q'.

Similarly in Malagasy nuclear Ss built from verbs with different voice morphology may be theta equivalent even though constituents with the same theta role may never have been structurally identical. So UTAH fails here, which is unsurprising: theta role assignment can be a function of structure: "same structure,

same theta role", without being a one to one function: "different structures, different theta roles".

2. Theta Relations in Malagasy Here is a simple pair of theta equivalent Ss built from the same root, *enjika*. Verbs are given in morphemic decomposition followed by the derived form in standard orthography[3]. Square brackets mark the major constituent break (here, and later).

(2) a. [n+aN+enjika (nanenjika) ny jiolahy] Rabe
     past+aN+chase   the thief   Rabe
     Rabe chased the thief

    b. [n+enjika+ina+Rabe (nenjehin-dRabe)] ny jiolahy
     past+chase+Vna+Rabe.gen   the thief
     Rabe chased the thief (*The thief was chased by Rabe*)

Voice affixes are not glossed (except *-ina* as *-Vna* since the vowel varies). There is massive evidence that the bracketed string in each of (2a,b) is a constituent (Keenan 1976, 1995; Guilfoyle, Hung and Travis (GHT 1992); Paul (1999), Pearson (2001, 2003). Theory neutrally we call it a Predicate Phrase (PredP). The final DP in these Ss will be called the External Argument (EA), following Pearson (2001). Here are two types of evidence that support this constituency judgment.

First, various particles, such as the yes-no question particle *ve*, are placed at the right edge of the PredP, separating it from the EA.

(3) a.   nanenjika ny jiolahy ve Rabe?   b.   nenjehin-dRabe ve ny jiolahy?
     Did Rabe chase the thief?   *Did Rabe chase the thief?*

No other locus of *ve* is grammatical. Similarly the 'even...though' construction places *na* (*dia*) at the left edge of the PredP and *aza* at its right. Also negation, *tsy*, occurs at the left edge and npi's like *velively* 'at all' and *intsony* 'longer' at the right edge.

Second, Keenan (1995) shows that PredPs occur in diverse syntactic contexts with the same form and meaning: relative clauses (RCs), embedded questions, Raising to Object contexts, and tensed PredP nominalizations. We illustrate with RCs:

(4) **Relative Clause Formation: N = N + PredP**

    a.   zaza [nanenjika ny jiolahy]   b.   zaza [nenjehin-dRabe]
       child (who) [chased the thief]   child [chased-by-Rabe]

PredPs denote properties, functions from individuals to truth values. The property denoted by *nanenjika ny jiolahy* maps an individual x to True if x chased the thief, and to False otherwise. That denoted by *nenjehin-dRabe* maps x to True if x was chased by Rabe, and to False otherwise. In RCs PredPs are property denoting

expressions. In (4a) the children referred to are those that *nanenjika ny jiolahy* 'chased the thief' maps to True. In (4b) it is those *nenjehin-dRabe* 'was chased by Rabe' maps to True. RCs are interpreted as in (5a). (5b) is an instance of (5a).

(5) a. ⟦Noun + PredP⟧ = ⟦Noun⟧ ∩ {x| ⟦PredP⟧(x) = True}

    b. ⟦zaza nenjehin-dRabe⟧ = ⟦zaza⟧ ∩ {x|⟦nenjehin-dRabe⟧(x) = True}

                        = CHILD ∩ {x|(CHASED BY RABE)(x) = True}

Thus we compositionally interpret (4a,b)[4]. We note that demonstrative Dets and the definite article *ny* combine directly with PredPs yielding DPs: *ny nenjehin-dRabe* = "the (ones) chased by Rabe".

    Beginning with Keenan (1972) linguists have ethnocentrically modeled Malagasy RC formation and interpretation on the English pattern, yielding structures like (6) in which the post-N expression contains a full S and its PredP has an external argument (EA) that moves preclausally and is co-indexed with a trace in its original position:

(6) zaza [wh$_i$ [$_S$[$_{PredP}$ nanenjika ny jiolahy] [t$_i$]]].
     child who   chased   the thief

(6) is also compositionally interpretable but carries useless structure: an embedded S and an invisible EA. This is objectionable in Malagasy. To build and interpret RCs as per (4) and (5) the Malagasy child just uses Ns and PredPs already learned in interpreting simple Ss. What motivates the child to detect and interpret unpronounced categories? (We treat below a question of *expressive completeness* raised by this way of forming RCs).

    The English bias leads us to ask why *wh* does not occur as an internal argument, moving to Pred initial position yielding the ungrammatical (7) (extraction sites marked by case marked traces)? In informal classical terms, Why do only subjects relativize? The answer we give is that the post-N phrases do not occur as independent PredPs.

(7) a. *zaza nenjehina+t$_{gen}$ ny jiolahy   b. *zaza nanenjika t$_{acc}$ Rabe
       child (that) the thief was chased by   child (that) Rabe chased

    A more insightful perspective results from reversing the ethnocentric bias: Why does English not form RCs in the learning theoretically pleasing way that Malagasy does? A partial answer is given by comparing the subject and object RCs:

(8) a. (the) child who chased the thief   b. (the) child who Rabe chased

In (8a) the post-*who* phrase is a PredP interpretable as in Malagasy. But in (8b) *Rabe chased* is not independently a PredP. In *Rabe chased the child* it is not a constituent

and not assigned a meaning. The fact that RCs (wh-questions, clefts, etc.) in English force us to interpret strings which are not independently meaningful provides an explanation for why English speakers have recourse to variable binding operators and empty categories: they represent minimal logical extensions of what actually occurs that yields a correct compositional semantic interpretation.

But one can object here that the meaning of (8b) is expressible in a Malagasy way just by using passive PredPs, as in (*the*) *child who was chased by Rabe*. And this is correct. But as the range of syntactic contexts is increased we see that Malagasy presents a much more extensive repertoire of passive-like PredPs supporting the expressive completeness of this strategy in Malagasy. Here are six types of PredPs fully natural in Malagasy but marginal or ungrammatical as passives in English.

[1] PredPs with semantically oblique EAs. The PredP in (9a) admits of modification with many obliques, such as Benefactives and Instrumentals. The verb in (9b,c), formed by suffixing the verb in (9a) with *-ana* (see footnote 11), takes EAs corresponding to these obliques; the PredPs in (9b,c) are identical.

(9) a. [m+aN+tao (manao)   trano ho an-dRasoa amin'ny birikinay] Rabe
        pres+aN+do   house for-Rasoa with'the bricks+our Rabe
        Rabe is making a house for Rasoa with our bricks

    b. [⌀+anaovan-dRabe trano] Rasoa
        pres+aN+do+ana+Rabe.gen house Rasoa
        Rabe is making a house for/because of Rasoa

    c. [⌀+anaovan-dRabe trano] ny birikinay
        pres+aN+do+ana+Rabe.gen house the bricks+our
        Rabe is making a house with our bricks

    d. ny vehivavy [anaovan-dRabe trano]
        the woman makes-Rabe house
        the woman Rabe is making a house for / because of
        *?the woman who is being made by Rabe a house for/because of

In (9d) the expression following *vehivavy* 'woman' independently denotes a property, as in (9b,c), but not so for what follows 'woman' in the English translation of (9d).

[2] Raising to Object, hosted by over 50 verbs (Paul and Ravaovololona 1998).

(10) a. manantena aho fa nanaovan-dRabe trano Rasoa
         pres+aN+hope 1.sg.nom that past+aN+do+ana+Rabe.gen house Rasoa
         I hope that Rabe made a house for Rasoa[5]

b. [manantena an-dRasoa ho nanaovan-dRabe trano] aho   (RTO)
pres+aN+hope acc-Rasoa as past+do+Cana-Rabe.gen house I
I hope Rabe made a house for Rasoa

c. [ø+antena+ina+ko (antenai-ko) ho nanaovan-dRabe trano] Rasoa
pres+hope+Vna+1.sg.gen as made+for-by Rabe.gen house Rasoa
I hope Rabe made a house for Rasoa

d. ny vehivavy [antenaiko ho nanaovan-dRabe trano]
the woman hoped+1sg.gen as built+for-Rabe.gen house
the woman that I hoped that Rabe built a house for
??the woman who is hoped by me to have been built a house for by Rabe

Complex predicates like those in (10c,d) are the major mechanism relativizing (questioning, etc.) arguments of complement clause verbs.

[3] Control context predicates whose EAs bear theta roles to the denotation of the controlled predicate, not the controlling one:

(11) a. [n+i+kasa h+aN+vaky (nikasa hamaky) io boky io] Rabe
past+i+intent fut+aN+read that book that Rabe
Rabe intended to read that book

b. [no+kasa+ina-dRabe (nokasain-dRabe) ho vaky+ina (vakina)] io boky io
past+intend+Vna+Rabe.gen fut read+Vna that book that
Rabe intended to read that book

c. ny boky nokasain-dRabe ho vakina
the book that Rabe intended to read
??the book that was intended by Rabe to be read (by Rabe)

[4] Object control contexts yield patterns similar to the RTO cases:

(12) a. [m+aN+ampy an-dRabe h+aN+petraka (hametraka) ny kodiarana] Rasoa
pres+aN+help acc-Rabe fut+aN+place the wheel Rasoa
Rasoa is helping Rabe change the tire

b. [ampian-dRasoa h+a+petraka+Rabe (hapetra-dRabe)] ny kodiarana
help+Vna-Rasoa.gen fut+a+place+Rabe.gen the wheel
Rasoa is helping Rabe change the tire

c. ny kodiarana ampian-dRasoa hapetra-dRabe
the wheel which Rasoa is helping Rabe change
* *the wheel which is being helped by Rasoa to be changed by Rabe*

[5] Verbs of motion induce PredPs similar to those in control contexts:

(13) a.   [niakatra hianatra teny vahiny tany Antananarivo] i Soa
         past+i+go+up fut+i+study lg foreign past+loc Antananarivo art Soa
         Soa went up to study foreign languages in Antananarivo

     b.   ny teny vahiny [niakaran'i Soa hianarana tany Antananarivo]
         the lg foreign past+i+go-up+ana+art Soa.gen fut+i+study+ana there Ant.
         the foreign languages Soa went up to Antananarivo to study
         **the foreign languages that were gone up by Soa to Antananarivo to study*

[6] PredPs with internal Agents binding anaphors:

(14) a.   [m+aN+vono (mamono) tena ho an'ny ankizy] ny ray aman-dreny
         pres+aN+kill self for'the children the father and-mother
         Parents kill themselves for their children

     b.   [⌀+amono+an'ny ray aman-dreny tena] ny ankizy
         pres+aN+kill+ana'the parents.gen self the children
         Parents kill themselves for their children

     c.   ny ankizy izay [amonoan'ny ray aman-dreny tena]
         the children that aN+kill+ana'the parents self
         the children that parents kill themselves for
         **the children that parents are killed by themselves for*

(15) a.   [n+aN+ampy (manampy) azy h+i+tsara ny fanadinana] isika
         past+aN+help 3acc fut+i+judge the exams we.incl
         We helped them grade the exams

     b.   [n+if+anampy hitsara ny fanadinana] isika
         past+recip+help judge the exams we.incl
         We helped each other grade the exams

     c.   [nifanampian-tsika hitsarana] ny fanadinana
         past+recip+help+ana-1.gen.pl.incl fut+i+judge+ana the exams
         We helped each other grade the exams

     d.   ny fanadinana [nifanampian-tsika hitsarana]
         the exams past+recip+help+1.gen.pl.incl fut+i+judge+ana
         the exams we helped each other grade
         **the exams that were reciprocally helped by us to be graded*

Pairs like (2a,b) do resemble English active-passive pairs in one respect: they are theta equivalent, with EA Agent in one case and Theme in the other. So we

sometimes translate non-AF forms as passives. But the analogy is limited. As we will see, Ss with Theme EAs are more like actives than passives.

Returning now to (2), the different verb forms take their internal argument DPs in different cases. Malagasy has a three case system labeled traditionally below. Here are the pronominal distinctions. See K&P (Keenan & Polinsky 1998) for details.

Singular:	1	2	3	Plural:	1excl	1incl	2	3
nom	aho	ianao	izy		izahay	isika	ianareo	izy[6]
acc	ahy	anao	azy		anay	antsika	anareo	azy
gen	-ko	-nao	-ny		-nay	-tsika	-nareo	-ny

nominative is the case of the EA and functions as a default (Pearson 2003). It is only distinctively marked in the pronouns. It is never selected by Vs, Ns or As in forming PredPs, NPs or APs. Only two Preps select **nom** (*noho* 'than, in comparatives)', and *afa-tsy* 'except, lit: "free-not").

accusative has distinctive pronouns and marks proper nouns, some kin terms and, optionally, demonstratives with *an-*. **acc** is selected by many Vs, some Ns (below), a few Ps (*tahaka azy* 'like him') and a few As (*feno azy* 'full (of) it'). DP$_{acc}$s also serve as predicate possessives: *Ahy io* "That is mine".

genitive is the case of nominal Possessors, non-external Agents of Vs and As (Phillips 2000) and most objects of Prepositions. DP$_{gen}$s, both pronominal and full, are morphologically bound to their heads (the latter by a process we call *n-bonding*). Except monosyllabic *-ko* and *-ny* pronouns may carry stress (phonemic). They may replace full DPs, so pronominal replacement is a test for the case of a DP.

In a text count based on two newspaper articles and selections from three novels, in Malagasy and in English, Keenan (1995) found case distributed as in (16)[7].

(16)	number of DPs	**nom**	**acc**	**gen**
Malagasy	1,237	33.6%	23.0%	43.4%
English	1,345	38.9%	47.1%	14.1%

Genitive is the most widely occurring case in Malagasy, compared with accusative in English, nominative falling in second place in the two languages.

Gen The dominant expression of the Predicate-Argument relation in Malagasy is the Head+Possessor one. In English it is the Verb+Object one.

Verb voices and Nuclear Clauses in Malagasy Primary voice affixes combine with roots to build verbs with different subcategorizations which build Ss with different phrase structures. We present a compositional semantic interpretation which yields as a corollary the correct judgments of theta equivalence. We repeat (2a,b) below.

(17) a. n+aN+enjika (nanenjika) ny jiolahy Rabe
         past+aN+chase   the thief   Rabe
         Rabe chased the thief

     b. n+enjika+ina+Rabe (nenjehin-dRabe) ny jiolahy
         past+chase+Vna+Rabe.gen   the thief
         Rabe chased the thief

In (17a,b) both verbs are marked past with *n-/no-* which is constant across voices. We often ignore tense marking (obligatory) as irrelevant to our concerns.

In (17a) *aN-* combines with the transitive root *enjika* to form a transitive verb[8], one which combines first with an accusative DP and then a nominative DP. We derive *aN-* verbs by a function AN mapping roots to verbs, illustrated in (18)[9]. We list leftmost in the subcategorization of the verb the category it combines with first (contra the convention in Keenan 1995). The resulting expression combines with a DP_{nom} forming an S. So the rightmost category is that of the EA of the verb.
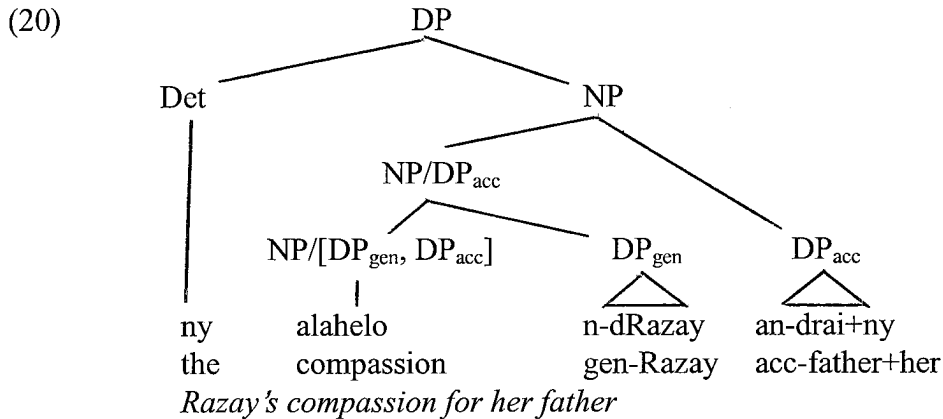
$$\text{AN}$$
(18)   enjika: RT_2 ===>   -anenjika: $S/[DP_{acc}, DP_{nom}]$

Syntax: *enjika* has category RT_2, a root with two theta roles. AN is a function mapping such expressions to predicates of category $S/[DP_{acc}, DP_{nom}]$, which Merge combines with the appropriate case marked DPs. The derivation of (17a) is (19).

(19)
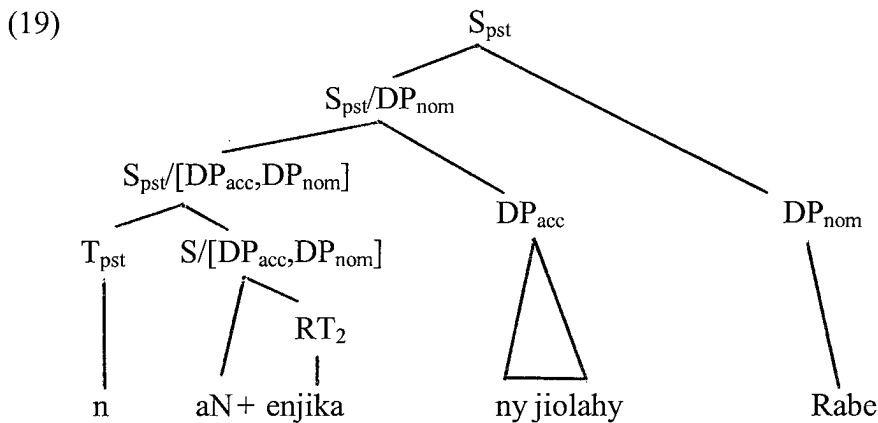
The use of 'S/...' in our category of verbs is not redundant. Nouns like 'fear' and 'compassion' subcategorize two case marked DP arguments to form a DP.

(20)

```
                              DP
                _____/‾‾_____
              Det                          NP
               |                    _____/‾‾_____
               |                NP/DP_acc             
               |              __/‾‾_____           
               |       NP/[DP_gen, DP_acc]   DP_gen      DP_acc
               |              |             /‾\         /‾\
              ny           alahelo       n-dRazay   an-drai+ny
              the          compassion    gen-Razay  acc-father+her
```

Razay's compassion for her father

Semantics We interpret the root as a binary relation, ENJIKA, with theta roles Theme and Agent. The interpretation of the verb AN(enjika) is given by (21), which says that the pair (x,y) is in the ENJIKA (CHASE) relation, y is its THEME and x its AGENT.

(21) ⟦AN(enjika)⟧(y)(x) = True iff

$$(x,y) \in ENJIKA \land THEME(y,ENJIKA) \land AGENT(x,ENJIKA)$$

Of course we want to give the syntax and semantics of *aN-* more generally. What roots does it combine with? Does it always build verbs whose EA is Agent? In fact the verbs *aN-* builds are usually (di)transitive. Examples are: *didy* ==> *mandidy* 'cuts', *vaky* ==> *mamaky* 'reads', *vono* ==> *mamono* 'hits, kills'. Examples of ditransitives are: *seho* ==> *maneho* 'shows', *tolotra* ==> *manolotra* 'offers', *ome* ==> *manome* 'gives'. AN however builds two types of intransitives. First a very few activity ones: *leha* ==> *mandeha* 'goes', *dihy* ==> *mandihy* 'dances', and *lohalika* ==> *mandohalika* 'kneels'. Finally, there is a non-trivial class of verbs it builds which do not present an Agent: *mangetaheta* 'is thirsty' (< *hetaheta*), *mangatsiaka* 'is cold' (< *hatsiaka*) and *mangovitra* 'shivers' (< *hovitra*) and *firatra* ==> *mamirapiratra* scintillates'. Relative to the Theta Hierarchy below, (22) covers all the cases.

Theta Hierarchy Agent > Experiencer > Goal /Theme/Patient > Other

Def   A verb whose EA is assigned the highest theta role in the verb's subcategorization is called *Actor Focus* (AF). Affixes which build AF verbs and nuclear Ss built from AF verbs are also called Actor Focus.

(22)   *aN-* is an Actor Focus prefix.

Language internal generalizations like (22) aid in acquisition by providing default interpretations for novel verbs.

Another AF voice prefix is *i-*. It builds many transitive verbs: *kapoka* ==> *mikapoka* 'beats', *vidy* ==> *mividy* 'buys', *orina* ==> *miorina* 'builds', *varotra* ==>

mivarotra 'sells', *laza* ==> *milaza* 'says', *tady* ==> *mitady* 'seeks', but most often *i*-builds intransitive verbs, and that in two distinct ways. First, it combines with roots taking just one theta role: *asa* ==> *miasa* 'works', *tsangana* ==> *mitsangana* 'stands up', *tomany* ==> *mitomany* 'cries'. Second, it combines with transitive roots that also accept the *aN*- prefix, usually omitting the Theme theta role: *enjika* ==> *manenjika* 'chases', *mienjika* 'flees'; *sasa* ==> *manasa* 'washes (tr)', *misasa* 'washes (intr), *hidy* ==> *manidy* 'closes (tr)', *mihidy* 'is locked'. So the following generalization holds:

(23)   *i*- is an Actor Focus affix; when *i*- and *aN*- combine with the same root the *i*-verb usually has lesser valence, lacking a Theme present in the *aN*- verb.

Consider now the verb *enjehina* in (17b), formed by suffixing the root *enjika* with *-ina*. It combines first with a DP_{gen}, interpreted as Agent, and then with a DP_{nom}, interpreted as Theme. (24) shows the syntactic action of this suffix, (25) gives its semantic interpretation, and (26) derives a nuclear S built from the INA-verb.

$$\text{INA}$$

(24)   enjika; RT_2 ==>   enjehina: $S/[DP_{gen}, DP_{nom}]$

(25) $[\![\text{INA(enjika)}]\!](u)(v) = \text{True iff}$

$$(u,v) \in \text{ENJIKA} \wedge \text{THEME(v,ENJIKA)} \wedge \text{AGENT(u,ENJIKA)}$$

(26)
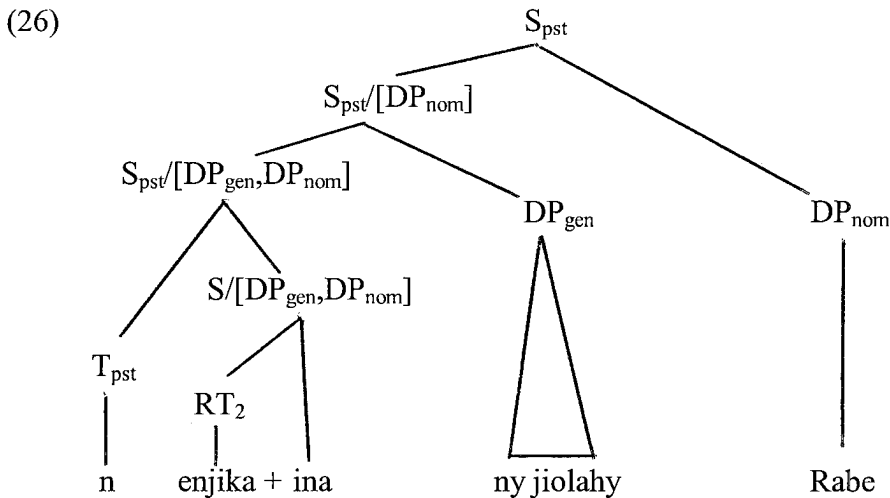
Let us call a verb τ Focus if its external argument is assigned the role τ[10]. So *enjehina* is a Theme Focus (TF) verb. Typically *-Vna* builds TF verbs. Notice now that (27a) follows immediately from (21) and (25), and in general (27b) holds when ROOT is in the domain of both AN and INA.

(27) a. $[\![\text{AN(enjika)}]\!](y)(x) = [\![\text{INA(enjika)}]\!](x)(y)$ and

   b. $[\![\text{AN(ROOT)}]\!](y)(x) = [\![\text{INA(ROOT)}]\!](x)(y)$ and

(28) **Theta Equivalence** Minimal Ss built from AN(ROOT) and INA(ROOT) are theta equivalent: the EA of each bears the same theta role as the internal argument of the other. But the Agent DPs in the two Ss were never in structurally comparable positions; nor were the Themes. Thus UTAH does not hold on this analysis.

Another primary TF affix is the prefix *a-*.

(29) a.   n+aN+tosika (nanosika) ny fiara Rabe   (AF)
          past+aN+push the car Rabe
          Rabe pushed the car

    b.   n+a+tosika+Rabe (natosi-dRabe) ny fiara   (TF)
          past+a+push+Rabe.gen the car
          Rabe pushed the car

tosika does not accept a *-Vna* suffix. However ditransitive roots with both Goal and Theme roles, (30), often have both *a-* and *-Vna* forms, in which case the *a-* verb assigns Theme to the EA and *-Vna* assigns it Goal:

(30) a.   n+aN+roso (nandroso)   vary ny vahiny Rasoa
          past+aN+serve   rice the guest Rasoa
          Rasoa served rice to the guests

    b.   no+roso+ana+Rasoa (norosoan-dRasoa) vary ny vahiny
          past+serve+Vna+Rasoa.gen   rice the guests
          The guests were served rice by Rasoa

    c.   n+a+roso+Rasoa (naroson-dRasoa) ny vahiny ny vary
          past+a+serve+Rasoa.gen the guests the rice
          The rice was served to the guests by Rasoa

(31) With ditransitive roots *a-* externalizes the "Intermediary": the Theme with roots of transmission, the Instrument otherwise. In the first case *-Vna* externalizes the Goal, and in the second the Theme. See Paul (1999).

Secondary verbal affixes build (tenseless) verbs from (tenseless) verbs, combining thus with items which already have (at least) one verbal affix. Of the four major such affixes, *if-* 'Reciprocal' and *amp-* 'Causative' are semantically specific and not our primary concern here (see K&R 2004).

(32)   m+if+aN+enjika (mifanenjika) Rabe sy Rakoto
        pres+Rec+aN+chase Rabe and Rakoto
        Rabe and Rakoto are chasing each other

(33) a. m+aN+sasa (manasa) lamba Rabe
pres+aN+wash clothes Rabe
Rabe is washing clothes

b. [m+amp+aN+sasa (mampanasa) lamba azy] Rasoa
pres+Cause+aN+wash clothes him Rasoa
Rasoa is making him wash clothes

c. [ø+amp+aN+sasa+ina+Rasoa (ampanasain-dRasoa) lamba] izy
pres+Cause+aN+wash+ina+Rasoa.gen clothes he
He is being made to wash clothes by Rasoa

(33c) shows that the primary suffix *-ina* also combines with causative verbs to form TF verbs. No other shape of *-Vna* will work here, nor do causatives accept an *a-* TF prefix. But There remains a central voice affix which is crucial to the expressive power constraints alluded to earlier. It is the suffix *-ana* in (9b,c,d). Compare:

(34) a. [n+aN+enjika (nanenjika) ny jiolahy t+amin'ny fiara] Rabe
past+aN+chase   the thief past+with'the car Rabe
Rabe chased the thief by means of the car

b. [n+enjika+ina+Rabe (nenjehin-dRabe) t+amin'ny fiara] ny jiolahy
past+chase+Vna+Rabe.gen   past+with'the car the thief
The thief was chased by Rabe by means of the car
(Rabe chased the thief by means of the car)

c. [n+aN+enjika+ana+Rabe (nanenjehan-dRabe) ny jiolahy] ny fiara
past+aN+chase+ana+Rabe.gen   the thief   the car
The car was used by Rabe to chase the thief

The EA in (34c) is the Means constituent is nominative, with no preposition. The verb in (34c) will be called CF, *circumstantial focus*, since its EA expresses some circumstance of the action or state denoted by the verb.

Syntactically a CF verb is built from an AF verb by suffixing *-ana*, which triggers stress shift to the right[11]. The resulting verb is not AF, and its first argument is bound to the verb as a genitive, just as with TF verbs, as in (17b). We note that all tense taking[12] AF verbs accept *-ana* suffixation yielding non-AF verbs. For example, from *enjika* 'chase' we may form (present tense, primary stress marked):

(35) miénjika   runs, flees   ienjéhana   circumstance of fleeing
manénjika   chases   anenjéhana   circ. of chasing
mifanénjika   chase each other ifanenjéhana   circ. of chasing e.o.
mampanénjika makes chase   ampanenjéhana circ. of causing to chase

-ana generally adds an argument for which its AF verb is not subcategorized (Paul

1999, Keenan 2000). All we know about it is that it will be interpreted as one expressible by an oblique constituent of the AF verb. *-ana* verbs are built by a function ANA, illustrated applying to a transitive verb in (36)[13].

ANA
(36) (verb$_{AF}$: S/[DP$_{acc}$, DP$_{nom}$]) ==> (verb+ana; S/[DP$_{gen}$, DP$_{acc}$, DP$_{nom}$])

Semantically oblique DPs and adverbials are predicate modifiers – functions F mapping predicate denotations P to predicate denotations F(P). Normally they are *restricting*: F(P) ≤ P, meaning that for all entities x, F(P)(x) = True then P(x) = True[14]. *for Max* is restricting in (37a), whence (37a) entails (37b). Replacing *for Max* by *with a screwdriver* preserves the entailment since it is also restricting.

(37) a.   Sue opened the box for Max   **(for Max)(open (the box))(Sue)**

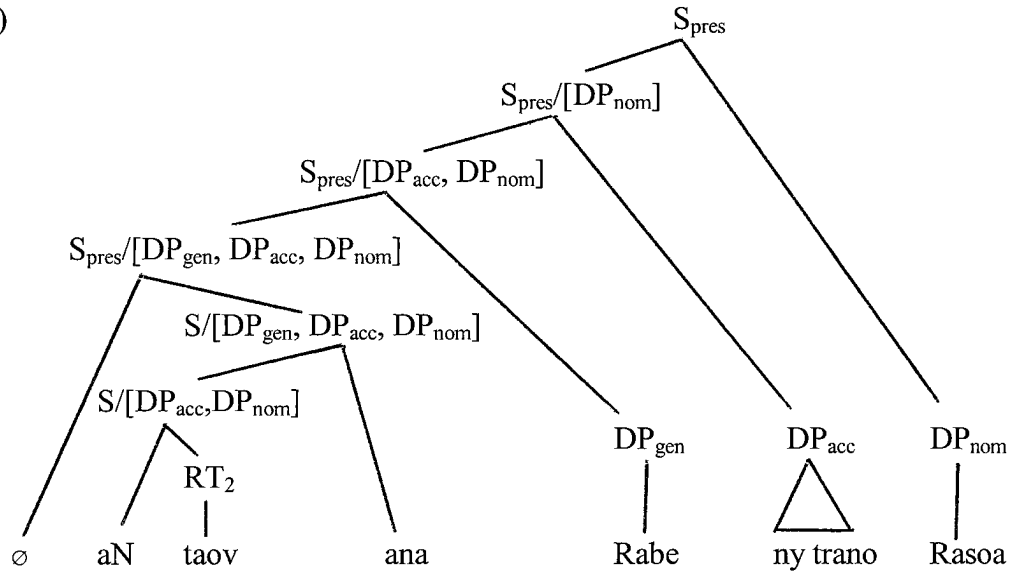   b.   Sue opened the box   **(open(the box))(Sue)**

So Prepositions are semantically functions **p** mapping entities to restricting modifiers. And the semantics for CF verbs quantifies over such functions, illustrated in (38) with the circumstantial of a transitive verb.

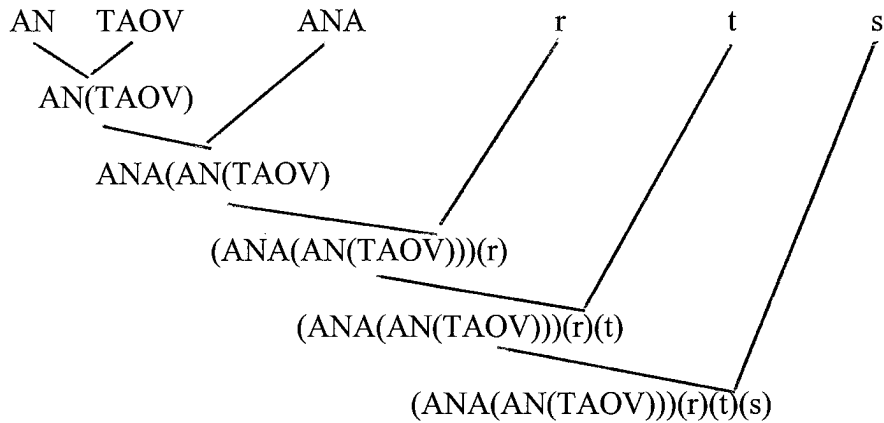(38) ⟦ANA(verb$_{AF}$)⟧(x)(y)(z) = True iff ∃**p** **p**(z)(⟦verb$_{AF}$⟧(y))(x) = True

The crucial point is that **p**(x) is restricting, so a nuclear circumstantial S like (39a) entails the corresponding AF one (39b) which lacks an oblique DP corresponding to the EA of the circumstantial. (39a) is derived in (40) and interpreted in (41).

(39) a.   ø+aN+taov+ana+Rabe (anaovan-dRabe) ny trano Rasoa
          pres+aN+make+ana+Rabe.gen the house Rasoa
          Rabe is making the house for / because of Rasoa

      b.   m+aN+taov (manao) ny trano Rabe
          pres+aN+make the house Rabe
          Rabe is making a house

(40)

S_{pres}

$S_{pres}/[DP_{nom}]$

$S_{pres}/[DP_{acc}, DP_{nom}]$

$S_{pres}/[DP_{gen}, DP_{acc}, DP_{nom}]$

$S/[DP_{gen}, DP_{acc}, DP_{nom}]$

$S/[DP_{acc}, DP_{nom}]$

RT_2

⊘	aN	taov	ana	DP_{gen}	DP_{acc}	DP_{nom}
				Rabe	ny trano	Rasoa

(41)

AN TAOV ANA   r   t   s

AN(TAOV)

ANA(AN(TAOV))

(ANA(AN(TAOV)))(r)

(ANA(AN(TAOV)))(r)(t)

(ANA(AN(TAOV)))(r)(t)(s)

The last line is True iff for some **p**, **p**(s)AN(TAOV)(t))(r) = True. And since **p**(s) is restricting the last line entails AN(TAOV)(t)(r), the interpretation of *Rabe is making the house*. Thus our semantics for the circumstantial S entails that Rabe is Agent of *make* and the house is Theme. Consider now the predictions of theta equivalence between a circumstantial S and a corresponding AF one with an overt oblique.
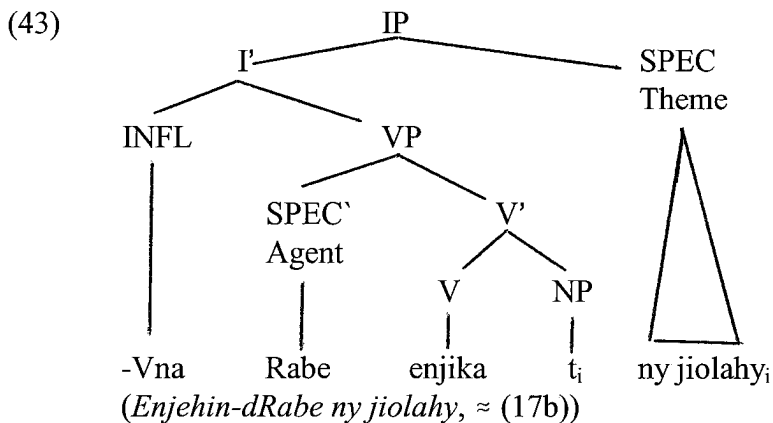
(42) a. ⊘+aN+taov+ana+Rabe (anaovan-dRabe) ny trano Rasoa   (CF)
         pres+aN+make+ana+Rabe.gen the house Rasoa
         Rabe is making the house for / because of Rasoa

     b. m+aN+taov (manao) ny trano ho an-dRasoa Rabe   (AF)
         pres+aN+make the house for Rasoa Rabe
         Rabe is make the house for Rasoa

Our semantic analysis shows that the genitive complement, Rabe, in (42a), has the same theta role, Agent, as the EA in (42b). Also the accusative *ny trano* 'the house' has the same theta role, Theme, in both Ss. But while *Rasoa* is a Benefactive in

(42b), forced by the choice of Prep, *ho an-* 'for', *Rasoa* is not assigned a theta role on our analysis in (42a). Rather our semantics merely quantifies over the set in which Preps denote, different theta roles for Rasoa correspond to different choices of possible Prep denotations. Most choices – Instrument, Location, Manner – are pragmatically implausible in (42a). But *Rasoa* could still be a Benefactee or a Cause.

Thus on our analysis the EA of CF Ss is vague, not ambiguous, with respect to theta role. We just know that it stands in some relation to the predicate which is of the sort expressible by a Preposition. On standard views in contrast the EA (or its co-indexed operator) is derived by movement from structures in which it has different theta roles, whence the CF S is semantically ambiguous. The point is subtle, but indicative of a deeper difference in approach. Namely, the only regular semantic relations between Ss differing just by verbs in different voices are those predictable from the verb. This includes theta equivalence and logical equivalence for Ss with individual denoting DPs, as the semantics of verbs is given in terms of individuals. But the behavior of adverbials and quantifier scope is not predictable.

3. Support for the K&M analysis We review first, briefly, some "standard" approaches to Malagasy clause structure – GHT (1992), Paul (1999) and Pearson (2001, 2003). (43) is the derivation of a nuclear TF S in GHT. Here *-Vna* occurs in INFL and by stipulation assigns case to the Agent in SPEC VP. The Theme complement to V lacks case and so moves to SPEC IP, where case is assigned (by stipulation). V moves to INFL to support the *-Vna* morphology. Then the Agent attaches to it in the same way that possessors n-bond to their heads.

(43)

```
                              IP
                I'_____
              ___|___                      SPEC
          INFL       VP                    Theme
           |       __|__                     /\
           |    SPEC`     V'                 /  \
           |    Agent    /\                 /    \
           |           V    NP             /      \
           |           |    |             /_____\
         -Vna  Rabe  enjika  t_i       ny jiolahy_i
       (Enjehin-dRabe ny jiolahy, ≈ (17b))
```

The derivation of a nuclear AF S differs from (43) in that *aN-* occurs in V and assigns case to its Theme. Nothing assigns case to the Agent in SPEC VP, so it moves to SPEC IP where it gets case. Later V moves to INFL string vacuously. (Both imperative and present tense indicative AF forms present an AF-specific *m-* in INFL, as Pearson notes. Hosting this *m-* would motivate the verb movement, but why does *m-* not assign case to the Agent in SPEC VP as *-Vna* in (43) does?)

Paul (1999)'s underlying structures are comparable to GHT's but richer. The *a*-prefix forms are explicitly treated, and *a-* is assigned a structural position different from *aN-* and *-Vna*. Movement of DPs to the EA position is motivated by case

assignment as in GHT. Pearson's approach shares with Paul and GHT the assumption that case is determined in different structural positions associated with the voice morphology. But for Pearson the EA is base generated in SpecTP (Topic Phrase) and by stipulation is co-indexed with an empty wh-operator in the next lower projection, SpecWhP. That position, by stipulation, licenses wh-operators, so the empty operator is moved there from a case position. So Pearson's movement is A' movement, to a non-case position, not A movement as in Paul and GHT. We turn now to three learning theoretic merits of our, K&M, analysis. For more narrowly grammar internal advantages see Keenan (1995).

1. Our analysis correctly predicts several basic judgments of logical equivalence – (17a,b), (30a,b,c), and the judgment of entailment in (42a,b). Standard approaches lack an explicit semantic interpretation and thus cannot justify such predictions. Indeed the complexity of the structures assigned to simple Ss and the movement operations needed to derive the audible expressions present a non-trivial challenge to a compositional semantic analysis. Compositionality is basic from a learning theoretic perspective as it assures us that speakers can interpret complex expressions if they know what the component lexical items mean and how things built in that way take their meaning as a function of the meanings of their constituents. Compositionality is to date our only means of accounting for how speakers comprehend novel utterances.

2. On our view the voice affixes – *aN-*, *i-*, *ina*, *a-*, *-ana*, etc. are lexical entries with a semantic interpretation, different affixes applying to a given root yielding different meanings. So we are explicit regarding what children must learn to master their use and meaning. And given their interpretation, and hence that of the verbs they build, it is explicit what theta role each argument has. This is not a primary concern of standard approaches. It is also clear on our view why the generalizations in (22), (23), and (31) provide useful default reasoning guides to learning novel verbs built from the affixes considered. So our analysis better explains how Malagasy learn their language quickly with limited exposure to imperfect data....

In contrast on standard approaches the voice affixes are not meaningful lexical items, but rather reflexes of case assignment from different structural positions. This means that the morphemic identity of the case assigners is redundant. They could all be null and we could still claim that case X is assigned in configuration Y. On our approach the voice morphemes are not redundant – make them null and we can't distinguish and hence interpret differently verbs in different voices.

Note that we can not associate with each primary voice affix a specific case, given that they all form verbs of differing valencies. E.g. if *aN-* always assigns accusative, how do we generate the intransitive verbs built from it? *I-* is even more problematic. Nor can we bi-uniquely associate voice affixes with theta roles, a sufficient reason being that basic (di)transitive verbs have more theta roles than affixes. On our approach case and theta selection are determined by the affix plus the root. A given affix may map different roots to verbs of different valencies, hence

of different case and theta grids, and different affixes may map a given root to verbs with different case and theta grids.

3. Standard approaches and K&M differ regarding the syntactic complexity of primary AF and TF Ss. On K&M the *derivations* of (17a,b) are isomorphic, differing just by the choice of voice affix and the case of the internal argument[15]. But on standard approaches the AF S is simpler than the TF one, since the movement involved is shorter. In an AF S on the GHT/Paul view we only move the DP in SpecVP to SpecIP, "the next Spec up". But in TF Ss the complement of the V must move over SpecVP to get to SpecIP. A comparable distance gap arises with the movement of the empty operator on Pearson's analysis[16]. And in *a-* and *-Vna* forms of ditransitive Vs the EA (or empty operator) must move across two theta positions.

The symmetry between primary AF and TF verbs on the K&M approach extends beyond mere subcategorization and theta role assignment as Schachter (1976) first pointed out for Tagalog. Here are three important ones, all of which are slightly more natural using isomorphic K&M derivations, slightly less so on standard ones.

Imperatives The AF and non-AF verbs discussed all have imperative forms with the addressee phrase missed out. That will be the nominative EA for AF imperatives, (44a), but the genitive complement of non-AF forms in *a-*, *-Vna* and *–ana*, (44b,c):

(44) a. manolóra (m+aN+tólotra+a) vary azy!   (Stress marked)
offer   pres+aN+offer+imp rice them
Offer them rice

    b. tolóry (tólotra+y) vary izy   c. atolóry (a+tólotra+y) azy ny vary
offer   offer+imp rice they   offer a+offer+imp them the rice
Offer them rice   *Offer them the rice*

As always the EA, when present, must be definite. Note that imperative suffixes differ for AF verbs and non-AF ones, being the same for *a-*, *-Vna* and *-ana* verbs.

Selection Both AF and non-AF forms are selected by modal auxiliaries. *mahazo* 'permitted' and *mahay* 'able' select AF verbs, their TF roots *azo* and *hay* select non-AF verbs. Replacing AF by non-AF verbs in (45) and (46) yields ungrammaticality.

(45) a. mahazo manao (m+aN+tao(v)) izany aho
permitted   pres+aN+do that I
I can do that

    b. Tsy azoko (azo + ko) atao izany
Not   can + 1.sg.gen a+do that
I can't do that

(46) a. mahay manoratra (m+aN+soratra) amin'io penina io aho
         pres+can write   pres+aN+write with'that pen that I
         I am able to write with that pen

   b.   tsy haiko (hay + ko) anoratana (aN+soratra+ana) io penina io
         not able+1.sg.gen   aN+write+ana that pen that
         I can't write with that pen

Control With verbs of desire and intent control is vested in the Agent not the nominative (the two coinciding with AF verbs):

(47) a. n+i+kasa (ny) hamaky (h+aN+vaky) io boky io aho
         past+i+intend (the) fut+aN+read that book that I
         I intended to read that book

   b.   nokasaiko (no+kasa+ina+ko) (ny) hovakina (ho+vaky+ina) io boky io
         past+intend+Vna+1.sg.gen (the) fut+read+ina that book that
         I intended to read that book

Assuming that the child in general takes longer to learn complex things than simpler ones standard views should predict that the child learns AF transitive verbs before TF ones, *other things being equal*. This prediction is reinforced by the fact that the child's intransitive verbs will exhibit AF morphology (there being only one argument). This prediction is well supported for the active/passive distinction in English, where actives are plentiful in early child language and passives, especially agented ones, are scarce to non-existant (Borer & Wexler 1987; Grodzinsky and Fox 1998). Unfortunately there are many "other things" whose equality must be controlled to render this prediction informative. For example, the EA of verbs of any voice must be definite in Malagasy. To express "Rasoa is reading a paper" the simple TF form is not possible, since the Theme is indefinite[17]. And as we individuate inanimate objects less well than humans[18] and Agents are often human and Themes often inanimate, it is likely that Agents are more often definite in discourse than Themes, so these differences will influence the distribution of AF and TF forms in Malagasy texts. That said, predictions P1 – P4 below do seem to us well founded.

P1 On the standard views the proportion of AF forms to non-AF forms should be higher in child language than in adult language. On the K&M view the proportion should be roughly equal to the adult one.

P1 assumes that factors like definiteness which constrain the choice of voice when there is a choice are similar in child and adult contexts. Then the simplicity factor favors early AF learning on the standard treatments but not on the K&M analysis.

P2 a. CF verbs should be learned later than primary[19] AF ones
   b. CF verbs should be learned later than primary TF[20] ones

On the K&M view, P2a is a clear prediction since the AF verb is a proper constituent of the CF verb, so the latter cannot be formed or interpreted without the former. P2b follows from P2a plus P1: primary TF verbs are equal to AF ones in complexity and so should not be learned later. Standard views *should* also predict P2a on the grounds that CF forms require some movement in addition to whatever licenses the AF morphology. (But it is not clear what forces AF morphology and blocks TF morphology forms within CF verbs on Paul's and Pearson's approaches).

P3 From **Gen** the dominant expression of the Predicate-Argument relation will be the $Pred_{TF}+DP_{gen}$ one rather than the $Pred_{AF}+DP_{acc}$ one, so the former should outnumber the latter in the acquisition data.

P4 a. K&M expect that learning to form possessive nominals will be roughly simultaneous with the emergence of agented non-AF verbs since they are both instances of Possessor Formation. No such prediction follows on standard views as these two phenomena are independent.

b. Moreso than on standard views K&M's suggests that the genitive Agent of a non-AF verb will be present since it is a complement of the verb and forms a constituent with it. Neither is true on the standard view.

The child acquisition study We test P1 - P4 against the results of an acquisition project conducted by the second author. For a more systematic presentation of the analysis of that project see Hyams, Manorohanta, and Ntelitheos (2004).

Three children, TS, SO, and NI, were recorded regularly in their home settings for an 8 month period in each case. In all cases the children spoke the major dialect, Merina, that of the capital, Antananarivo, on which "Official Malagasy" is based. TS was recorded from age 24 months to 32 months, SO from18 months to 26 months, and NI from 22 months to 30 months.

General properties of the data As expected, there are very few functional elements: one occurrence of the question particle *ve*, two occurrences of the definite article *ny*, no occurrences of the focus particle *no*, no relativizers or complementizers, and almost no prepositions and no modal auxiliaries which serve to build complex predicates (A few do occur as main verbs, without a verbal complement). There are no syntactically complex constructions, such as relative clauses, sentence complements, focus/cleft constructions including wh-questions, or nominalized predicates. Negation occurs, as do imperatives and a few prohibitives. Utterances are often phonologically collapsed. The complex morphophonemics associated with possessive formation as well as compounding and object incorporation is almost entirely absent. We turn now to P1 – P4.

First, P2 is verified. We find only four CF verbs with a total of 7 occurrences. This is most compatible with the K&M analysis, but not incompatible with the

standard analysis. And additional factors may contribute to the relative absence of CF forms: A CF verb must have at least two arguments, and will have three if its argument AF verb is transitive. This is pretty much a maximum in adult speech. Further several of the structures that force a CF form are syntactically complex (RCs, Clefts/Focus, nominalizations) and so independently absent. Finally, many oblique DPs whose externalization requires a CF form cannot occur sentence finally (Rajemisa-Raolison 1971) but must be focused.

The absence of CF forms has the effect that the total of AF and TF forms have greater frequencies in child speech than adult speech. K&M (2001) compute voice distribution in two romance novels. (48) compares the voice distribution in those adult sources with that in the children studied here:

(48)　　　　　　　Children　　　　　Adults (Novels)

AF	890	(66.2%)	5601 (62.3%)
TF	447	(33.3%)	1846 (20.5%)
CF	7	(0.5%)	1532 (17.1%)
Total	1344		8979

These data support the K&M view over the standard view since the percentage of AF forms in the children's speech, 66%, is about the same as in the adult speech, 62%. Thus the absence of CF forms in the children's speech is largely compensated by an increase in TF forms. This is the opposite of what we would expect on the standard view where AF forms are less complex than TF ones and so should dominate early learning. But on the K&M view we expect, P1, the ratio of AF to non-AF forms to be roughly equal in the two groups. And the comparison supports this expectation.

More revealing than the gross frequency comparison of AF and TF forms is the one obtained by comparing the number of *transitive* AF verbs in the child data with the number of TF ones. This comparison is closer to telling us what verb form children use when they have a grammatical choice:

(49) Child data:

	Number of X	Number of occurrences of X
X = TF verbs	107	448
X = AF transitive verbs	40	131

So children choose TF verbs more than two and half times as frequently as AF ones, and they use them more than three times as frequently. Clearly TF verbs in Malagasy are not distributed like passives in English, despite some thematic similarities.

These facts also support P3 (that $Pred_{TF}+DP_{gen}$ structures dominate $Pred_{AF}+DP_{acc}$ ones) since it shows that TF verbs, which can all host a DP_{gen} complement, strongly outnumber the AF ones that can host a DP_{acc} complement. Other factors, not well understood even for the adult language, determine whether DP_{gen} and DP_{acc} complements are overt. One factor that is understood is imperative deletion.

AF verbs built with *aN-* and *i-* (and causatives and reciprocals built from them) have imperative forms built by suffixing *–a*. But equally non-AF verbs, *a-* and *-Vna* and CF ones, have imperative forms, built by suffixing *-o* (or *-y* when an *o* occurs in the root). In fact imperatives are overwhelmingly in non-AF form in adult speech.

(50) a. novonjen-dRabe aho
       past+save+Vna-Rabe.gen 1.sg.nom
       Rabe saved me

     b. vonjeo aho!
        save(imp) 1.sg.nom
        Save me!

(51) a. Namonjy ahy Rabe
       past+aN+save 1.sg.acc Rabe.nom
       Rabe saved me

     b. ?Namonjia (namonje) ahy!
        Save(imp)   1.sg.acc
        Save me!

(52) In the child data there are 26 imperative verbs with 144 tokens. *All* are TF forms

The existence of TF imperatives is incomprehensible if TF verbs are regarded as Passives. It helps here to realize that what is deleted in an imperative is not the "subject" but the Addressee Phrase, which must then be the Agent and thus the EA with AF verbs, but the internal, genitive, argument with non-AF verbs. We turn now to the distribution of Agents with TF verbs in order to test P4 and further evaluate P3.

First, P4a is solidly established. Each child in our sample presents both significant numbers of TF predicates with Agent, and also possessive DPs consisting of a head N and a possessor, both pronominal and full DP. For example: TS at 25 months produces Agented TF verbs, as in (53), and DP possessives, as in (54).

(53)   a. titako for tsy hitako 'not seeTF+1.sg.gen' = not seen by me
       b. labiby for lanin'ny biby 'swallowed by the animal'
       c. tintsioso for tian'i Tsoso 'liked by Tsoso'

(54)   a. teleko for ny teleko 'the television+1.sg.gen' (my television set)
       b. bibian zaza for biberon'ny zaza 'bottle of the baby'
       c. basy dopi for basin'i Droopy 'Droopy's gun'

In all cases there is simplification of the juncture between the V or N and the DP_{gen}, except when mere concatenation is required (as with *hitako = hita+ko* 'seen by me').

Similarly SO first produces possessives at 21months: *kasetinini* for *kasetin'i Nini*, 'cassette of Nini', *zazatati* for *zazan'i Tati* 'child of Taty' and *mapa nini* for *kapan'i Nini* 'shoes of Nini'. The first verb + genitive agent occurs at the same time: *bubela nani tati* for '*du beurre lanin'i Taty* 'butter finished by Taty', and more productively by 23 months: *nalefana* for *nalain'i Fara* 'taken by Fara'. Lastly NI also presents both genitive DPs and $Pred+DP_{gen}$ by 23 months, (55).

(55) a. kakajoko for akanjoko 'clothes+1.sg.gen'
    b. silipiniseheno for silipon'i Sehena 'Seheno's underpants'

(56) a. tsy ti papakoa for tsy tian'ny papako 'not liked by my Papa'
    b. titko for tsy tiako 'not liked by me'

Thus P4a is supported. The early simultaneous appearance of nominal and verbal DP$_{gen}$s supports K&M's position that the two expression types are instances of the same structure. And in adult speech the two share more than complex morphology and pronoun forms. For example DP$_{gen}$s in both license empty categories:

(57) a. Adidin'i Doda ny manaiky ny ray aman-dreni+ny
       duty of art Doda.gen the obey the father and mother+his
       Doda's duty is to obey his father and mother

    b. Kasain'i Doda ny hamaky io boky io
       intend+ina'art Doda.gen the fut+aN+read that book that
       It was intended by Doda to read that book

Finally, concerning P4b, we compute:

(58) Of the 447 TF verb occurrences in the child language data, 224 or 50% have overt Agents. (Imperatives are counted as +Agent here).

This figure is high compared with English and Dutch passives, which present agent phrases from 14% to 20% of the time (K&M 2001) and so supports our treatment of the DP$_{gen}$ as a complement of the verb rather than as an adjunct. Note that in possessive DPs in adult speech the genitive behaves as a complement not a specifier:

(59) a. ity trano+n-dRabe ity   b. ny vadi+n-dRanaivo
       this house+of-Rabe this   the spouse+of-Ranaivo
       This house of Rabe's   *the spouse of Ranaivo*

    On the other hand in Malagasy novels agents are present in TF forms 67% of the time (K&M 2001)[21]. And in several cases counted as agentless the missing agent is controlled. So the children lag behind the adult pattern here. We have no explanation for this lag aside from the obvious: their utterances are all short with much omitted.
    There is also a last relevant respect in which the profile of the children's speech does not quite conform to that of the adult: the case distribution of DPs. Here we just counted pronouns, as usually their form is distinctive (though not always in compressed speech). Their case distribution is:

(60)   Genitive: 106   Nominative: 121   Accusative: 5

As in adult speech the accusatives are the least well represented. But we have seen, (16), that in adult speech Genitives dominate Nominatives 43% to 34%. There are at least two obvious factors that contribute to this disparity. First, short utterances dominate the child corpus, many Ss are just an intransitive verb + EA, perforce nominative. Second, Prepositions, which overwhelmingly take genitive complements, are almost entirely absent from the children's speech.

In sum, the language acquisition data in Malagasy support our analysis and interpretation of simple clauses.

References

Abinal, R.P. et V. Malzac (1963) Dictionnaire Malgache-Français. Mission Catholique de Tananarive. Paris: Editions Maritimes et d'Outre-Mer (reprint from 1888 edition).

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

Borer, Hagit and Kenneth Wexler. 1987. The maturation of syntax. In *Parameter Setting* Thomas Roeper and Edwin Williams (eds). Reidel. Dordrecht.

Erwin, Sean. 1996. Quantity and moras: An amicable separation. In *The Structure of Malagasy* vol I. M. Pearson and I. Paul (eds). Dept. of Linguistics, UCLA. pp. 2 – 31.

Fox, Danny and Yosef Grodzinsky. 1998. Children's passive: a view from the *by-*phrase. *Linguistic Inquiry* vol 29.2:311 – 332.

[GHT] Guilfoyle, Eithne, Henrietta Hung and Lisa Travis. 1992. Spec of IP and Spec of VP: two subjects in Austronesian languages. *Natural Language and Linguistic Theory* 10:375 – 414.

Hyams, Nina, Cecile Manorohanta and Dimitris Ntelitheos. 2004. The acquisition of the Malagasy voicing system. Dept. of Linguistics, UCLA.

Keenan, Edward L. 1972. Relative Clause Formation in Malagasy (and some related some not so related languages. *The Chicago Which Hunt* University of Chicago.

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

Keenan, Edward L. 1995. Predicate-argument structures in Malagasy. In *Grammatical Relations* C. Burgess, K. Dziwirek, and D. Gerdts (eds). CSLI. Stanford. pp.171 – 217

Keenan, Edward L. 2000. Morphology is structure: A Malagasy test case. In *Formal Issues in Austronesian Linguistics* I. Paul, V. Phillips, and L. Travis (eds). Kluwer. pp.27 – 49.

[K&M] Keenan, Edward L. and Cecile Manorohanta. 2001. A quantitative study of voice in Malagasy. in *Oceanic Linguistics* vol 40:67 – 85.

[K&P] Keenan, Edward L. and Miriam Polinsky. 1998. Malagasy Morphology. In *The Handbook of Morphology*. A. Spencer and A. Zwicky (eds) Blackwell. pp. 563 – 624.

[K&R] Keenan, Edward L. and Jean Paulin Razafimamonjy. 2004. Reciprocals in Malagasy. in *Oceanic Linguistics* vol. 43.1:177 – 207.

[K&S] Keenan, Edward L. and Edward P. Stabler. 2003. *Bare Grammar: Lectures on Linguistic Invariants*. CSLI, Stanford.

Kroeger, Paul. 1988. Verbal focus in Kimaragang. In *Papers in Western Austronesian Linguistics* No. 3:217 –140. Pacific Linguistics, A-78.

Paul, Ileana. 1999. *Malagasy Clause Structure*. PhD Diss. Dept. of Linguistics, McGill University.

Paul, Ileana and Lucie Rabaovololona. 1998. Raising to object in Malagasy. In *The Structure of Malagasy* vol II I. Paul (ed). Dept. of Linguistics, UCLA. pp.50 –65.

Pearson, Matthew. 2001. *The Clause Structure of Malagasy: A Minimalist Approach* UCLA Dissertations in Linguistics. No. 21.

Pearson, Matthew. 2003. The Malagasy subject/topic as an A' element. To appear in *Natural Language and Linguistic Theory*

Phillips, Vivianne. 2000. The interaction between prefix and root: the case of *Maha-* in Malagasy. In *Formal Issues in Austronesian Linguistics* I. Paul, V. Phillips, L. Travis (eds). Kluwer, Dordrecht. pp. 85 – 105.

Prentice, D.J. 1971. *The Murut Languages of Sabah* Pacific Linguistics Series C – No.18. The Australian National University. Canberra, Australia.

Rabenilaina, Roger-Bruno. 1985. *Lexique-grammaire du malgache: Constructions transitives et intransitives*. Thèse de doctorat d'état. Paris Université Paris VII.

Rahajarizafy, Antoine. 1960. *Essai sur la grammaire malgache* Imprimerie Catholique, Antanimena, Tananarive

Rajemisa-Raolison, Regis. 1971. *Grammaire malgache*. Fianarantsoa.

Sabel, Joachim. 2003. Malagasy as an optional multiple wh-fronting language. In *Multiple Wh-Fronting* Cedric Boeckx and Kleanthes Grohmann (eds). John Benjamins, Amsterdam. pp. 229 – 254.

Schachter, Paul. 1976. The subject in Philippine languages: topic, actor, actor-topic, or none of the above. in Charles Li *Subject and Topic* (1976). Academic Press. pp. 491–518.

Endnotes

[1]E.g. Tagalog, Timugon (Prentice 1971) and Kimaragang Dusun (Kroeger 1988).

[2]Or quantified over if the object DP is quantificational.

[3]The major orthography-phonology correspondences are: o = /u/; word internal -i- = word final -y = /i/; *tr* is a voiceless prepalatal affricate, *dr* (or *dR*) its voiced counterpart; *j* = /dz/ with *ts* its voiceless counterpart. nC = ^nC, mC = ^mC

[4]*izay* sometimes separates the N and PredP in a RC. We interpret it as follows: IZAY(F)(A) = A $\cap$ {x|F(x) = True}. A is a N denotation, F a PredP one.

[5]Note the VSO order here. We assume a rightward CP extraposition rule.

[6]*Izy* can be augmented to force a plural reading: *izy* ireo '3 dem+pl' = they; *izy mivady* '3 spouse' = they spouses; *izy roa lahy* '3 two man' = they two men.

[7]We used pronominal replacement as a test for case. In *He left on Monday*, *Monday* was not counted accusative (in English) or genitive (in Malagasy translation) as pronominal replacement is unnatural *He left on it. Counting such cases would have increased the proportion of accusatives in English and genitives in Malagasy.

[8]Standard grammars of Malagasy (e.g. Rahajarizafy 1960) give the *aN*- and *i*-prefixes as *man* and *mi*. The initial *m* is specific to present tense AF verbs.

[9]Minimally an expression is a pair (s,C), s a string, C a category name. A language is the set of expressions derivable from a Lexicon by iterated application of some structure building functions. Each voice affix, like *aN*-, is associated with such a function. We also assume a function Merge, which maps a predicate of category S/[C1,...,Cn] and an expression of category C1 to one of category S/[C2,...,Cn].

[10]A verb with no Agent subcategorized may still be Actor Focus, as *matahotra* 'fears', where the EA is Experiencer. By the Theta Hierarchy an Agent Focus verb is always Actor Focus. An intransitive verb like *mangatsiaka* 'to be cold' is both Actor Focus and Theme Focus but not Agent Focus.

[11]In suffixing, an extra consonant often appears, as the *v* in (9). Following Erwin (1996) we treat it as part of the root. It remains in derived forms until the CV template to which Malagasy conforms is applied. At that point it drops if no suffix provides a vowel to follow it. So we have indicative *manao* (AF) and *atao* (TF), but imperative *manaova* and *ataovy*. We continue to use standard orthography.

[12]There are four degenerate roots which are AF verbs but which are not "tense taking", meaning they do not mark present or past tense or use an AF prefix in nuclear clauses. There are about 20 AF verbs that do fully mark tense but do not take an AF affix. They all have normal circumstantial forms. E.g. *isy* $\Longrightarrow$ *m+isy* 'exists'; *isiana* 'circumstance of existing'.

[13]Of course ANA takes AF verbs of all valencies as arguments.

[14]For P an n > 0 place predicate denotation and F restricting, F(P) $\leq$ P means that if F(P)(x) = True then P(x) = True. In the text we treat x as a "1-tuple".

[15]But, as defined explicitly in K&S, AF Ss and their TF counterparts are not isomorphic since an isomorphism would map DP_{gen}s to DP_{acc}s or vice versa, which cannot happen as they have quite different distributions.

[16]The problem is more serious on the GHT/Paul view, since movement of the Theme is an instance of A movement but must cross a theta position, which is in general not possible. For Pearson the movement is an instance of A' movement,

which applies unproblematically across a co-argument (*the man he saw*) [Dimitris Ntelitheos pc]..

[17]To present an indefinite Agent with an AF verb an Existential S would be used: *Misy zaza mitomany* 'Exist child cry', for "There are children crying"

[18] "Which two?" is a reasonable response to "I just interviewed two of your students" but not to "I just smoked two of your cigarettes".

[19]We use 'primary' here to put aside complex AF and TF verbs, such as (passives of) causatives. They virtually don't occur in our child data in any event.

[20]We use TF to cover Goal Focus forms as well. In fact there is only one GF verb in our data, *ome* 'give' and it has the same TF and GF forms: *omena*.

[21] This figure was given as 65% in K&M (2001). Re-examination showed that a few imperatives were omitted from the +Agent category there.