

Voice and Relativization without Movement in Malagasy

abstract Like many W. Austronesian languages Malagasy (Madagascar) is verb initial with a rich voice system on which most major syntactic processes in the language depend¹. Here we derive and compositionally interpret nuclear Ss in different voices. Their “theta equivalence”, similar to English actives and passives, is a corollary. No A or A’ movement is used. We illustrate one major syntactic process, relativization, that builds on the voice system, and we derive the “subjects only” constraint in a new way. Again, no (operator) movement is used. We treat voice morphologies as structure building operations as in K&S (Keenan & Stabler 2003). They are special cases of Merge. We introduce one novel notion, **language invariant**, to illustrate structural constraints over non-isomorphic expressions.

This paper is organized as follows: **1.** presents a sample of the voice phenomena we analyze. **2.** reviews the two mainstream approaches to these phenomena. **3.** presents our own analysis. **4.** derives and interprets relative clauses and **5.** concludes.

1. Voice and Nuclear Clause Structure in Malagasy

Malagasy verbs are built by successive affixation of roots (which may or may not be words). For example from the root² *tolotra* ‘offer’ we construct among others the four verbs in (1). The translation of (1a) suffices for each. Exceptionally we mark stress, to show that it is preserved under prefixing but shifts right under suffixing.

- (1) a. [_{P1} ∅+man+tólotra (manólotra) vary ny vahiny amin'ny lovia vaovao] aho
pres+**man**+offer rice the guests on'the dishes new 1s.δ
I offer rice to the guests on the new dishes
- b. [_{P1} ∅+a+tólotra+ko (atólotro)ny vahiny amin'ny lovia vaovao] ny vary
pres+**a**+offer+1s.poss the guests on'the dishes new the rice
- c. [_{P1} ∅+ tólotra+na+ko (tolóranko) vary amin'ny lovia vaovao] ny vahiny
pres + offer+**ina**+1s.poss rice on'the dishes new the guests
- d. [_{P1} ∅+man+tólotra+ana+ko (anolóranko) vary ny vahiny] ny lovia vaovao
pres+[[**man**+offer]+**ana**]+1s.poss rice the guests the dishes new

The primary constituent division, bracketed, receives extensive support and is agreed upon by all generative grammarians (Keenan 1976, 1995; GHT 1992; Paul 1999; R&T 2000; Sabel 2003; Pearson 2005). We differ on the category and internal structure of the bracketed constituent, which theory neutrally we call a P1, *one place predicate phrase*. Tests for this constituency are: (1) the yes-no question particle *ve* is placed at its right

Abbreviations: 1 = 1st person, 3 = 3rd person, s = singular; pl = plural, AG = Agent, TH = Theme, GL = Goal, dem = demonstrative, poss = possessor; acc = accusative, Rec = reciprocal, Cause = Causative, pres = present; pst = past, Foc = focus, excl = exclusive.

- d. anolóry ([man+tolotra]+y) vary azy ny lovia vaovao (the **-ana** form)
 offer **man+offer+imp** rice them the dishes new
Use the new dishes to offer them rice

[4] *Present tense marking* is uniformly zero in (1). Future and past replace \emptyset by *h-/n-* (*ho/no* before consonants) but we also lose the initial *m* in (1a): *hanolotra* ‘will offer’, *nanolotra* ‘offered’. Yet treating *m* as a present tense marker is inaccurate (Builles 1984; Pearson 2005) as it also occurs in the, tenseless, imperative. An additional half dozen verbal prefixes use this *m*, and some 20 verbs (3.4.1) are formed by directly prefixing *m* to roots. Traditional grammars – [RR 1971], Rahajarizafy 1960 and Malzac 1926, treat *m* as an integral part of *m* initial prefixes.

[5] *Agent marking* groups (1b,c,d) together excluding (1a). In (1a) the semantic Agent of the verb is *aho*, the DP sister to the P1. Linguists often accord a special, and pretheoretically unclear, status to the DP sister of the P1, using terms such as “subject”, “topic”, “pivot”, or “trigger”. We just call it **the DP sister of the P1**, limiting ourselves to what we can structurally identify. In (1b,c,d) the Agent is expressed by a “possessive” suffix *-ko* on the verb. Malagasy has three pairwise disjoint case series of pronouns:

(3) Singular Pronouns	δ	acc	poss
I	<i>aho</i>	<i>ahy</i>	<i>-ko</i>
You	<i>ianao</i>	<i>anao</i>	<i>-nao</i>
He, she, it	<i>izy</i>	<i>azy</i>	<i>-ny</i> .

The *poss pronouns*, by definition, express possession in nominals, like *my mother*. Like their full DP counterparts, they are morphologically bound, partially illustrated below. In (4a’,b’,c’) *atao* is an *a-* prefix verb, like *atolotra* in (1b), built from the root *tao(v)* ‘do’.

(4) a.	trano	b.	ny tranoko	c.	ny tranon-dRabe ⁴
	house		the house+my		the house+of+Rabe
a’.	atao	b’.	ataoko	c’.	ataon-dRabe
	a+do		a+do+1s.poss		a+do+Rabe.poss

The possessive construction is morphophonemically complex (See Paul 1996, K&P and Ntelitheos 2005). The forms in (4c,c’) involve a nasal “linker” and if the host is weak ([3] above) its ending drops, but the non-continuant feature of its consonant is inherited by the possessor, triggering the mutation of an initial continuant to the closest non-continuant. There are 7 such mutations: *h* → *k*, *l* → *d*, *r* → *dʳ* (a post alveolar affricate), etc. (5c,c’) lack the nasal linker in (4c,c’).

(5) a.	tongotra	b.	ny tongotro	c.	ny tongo-dRabe
	foot		the foot+my ‘my foot’		the foot-Rabe ‘Rabe’s foot’
a’.	atolotra	b’.	atolotro	c’.	atolo-dRabe
	‘offers’		offer+my ‘I offer’		offer+Rabe ‘Rabe offers’

In addition to nominal possessors *poss* pronouns in Malagasy express Agent phrases of many verbs and most objects of prepositions.

The δ pronouns are used for DP sisters of P1s. They are also default forms in two cases (Pearson 2005). The first is with coordinate possessors:

- (6) a. ny tranon-dRabe sy Raso
the house-Rabe.poss and Raso
Rabe and Raso's house
- b. ny tranon-dRabe sy izy (*-ny)
the house-Rabe.poss and 3. δ (*-3.poss)
“*Rabe's and her house*”

Secondly, Malagasy augments 3rd person pronouns in the δ or acc series with plural demonstratives, like *ireo* ‘those’, or kin terms, like *mivady* ‘spouses’ or *mirahalahy* ‘brothers’. Bare third person pronouns (all three series) can be singular or plural but the poss form *-ny*, always unstressed, cannot host these augments. So we commonly find:

- (7) a. ny tranon'izy mivady
the house'3. δ spouses
The spouses house
- b. ny tanan'izy mirahalahy
the village'3. δ brothers
The brothers' village

Two prepositions, *noho* ‘against, than’ and *afa-tsy* ‘except’ take their complements in the δ series. (*Rabe* in (8), as DP sister to the P1, can itself be replaced with *izy*).

- (8) [Lehibe noho izy] Rabe *Rabe is bigger than him*
big against 3. δ Rabe

We do not read δ as “nominative” as that suggests that the DP sister to P1 has the properties of subjects in better known languages. But the split in subject properties between two structurally distinct DPs that Schachter 1976 demonstrated for Tagalog mostly carries over to Malagasy. So we will not bias our discussion with a terminology that favors one of the possible analyses.

Lastly, the *acc* pronouns are, by definition here, used for non-poss sisters of transitive verbs, usually interpreted as Theme or Goal. A few adjectives, such as *antonina* ‘suitable’ and *feno* ‘full’ and a few prepositions, such as *lavitra* ‘far (from)’ and *tahaka* ‘like’ take acc complements. acc forms also express predicate possessives, as in *Ahy io* ‘lit: me that’ = *That is mine*.

2. Mainstream approaches to nuclear clause structure in Malagasy

There are two mainstream approaches to nuclear clause structure in Malagasy, both clearly summarized in HNM. The first, GHT 1992, developed in Paul 1999, integrates Schachter’s “two subjects” into mainstream theory. The structure they propose for a transitive S is an IP with an empty Specifier. The complement of the head of IP is a VP whose Spec is assigned the Actor theta role. The complement of the head V is assigned Theme. Verbal morphology is basic: **man** assigns case to the Theme, whence the otherwise caseless Actor raises to Spec,IP to check case. **ina** assigns case to the Actor,

so the Theme raises to Spec,IP. The two subjects are the DPs in Spec,IP and Spec,VP, which share subject properties. Paul 1999 treats the *a* prefix forms, as in (1b).

The determining difference between their approach and ours in **3**. is that on theirs, theta roles are assigned in pre-movement structure to unique positions (a nearly constant property of passives from Chomsky 1957 through Mahajan 1994 to Collins 2005). So a DP is assigned the Agent theta role in Spec,VP and only there, regardless of whether it later raises to Spec,IP. Similarly DPs are only assigned Theme as complements of V. So DPs with the same theta role in Ss like (1a - d) receive it in the same pre-movement position. In contrast, on our approach Agents (and Themes) originate in the positions in which they audibly occur. No movement is needed, or used.

The second mainstream approach was pioneered in Richards 2000 for Tagalog and developed for Malagasy in Pearson 2005. They propose that the DP sister to P1 behaves like a topic in V2 languages such as Icelandic. Pearson treats a clause as a TopP (Topic Phrase) whose Spec is initially filled with a DP. The complement of the head of TopP is whP whose Spec is initially empty. The complement of its head is TP (Tense Phrase) and the complement of its head is the VP which includes all the arguments of the verb, one of which must be an empty operator. Theta roles *and* case are assigned in situ. The empty operator moves to Spec,whP where it is coindexed with the DP in Spec,TopP which inherits its case and theta role. Voice morphology, *man* and *ina*, is realized on the verb according to the position from which the empty operator was moved. So movement determines morphology, rather than morphology determining what can move. As in GHT/Paul, theta roles are assigned in fixed positions in the pre-movement structure. A DP interpreted as Agent is either in Spec,VP or is coindexed with an empty operator that originated there. Analogously for DP Themes.

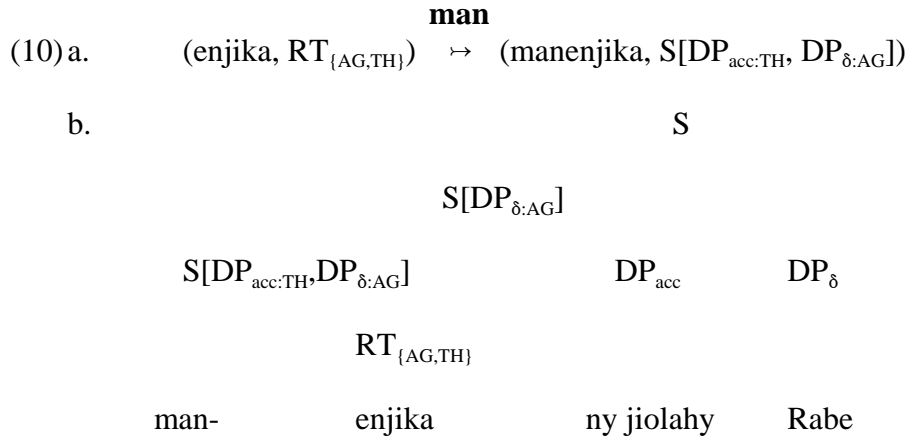
3. A proposal for the core voice system in Malagasy

We treat a language as a set of categorized expressions (s,C) – s a phonological string, C a category name – derivable from lexical (underived) expressions by structure building functions. The lexical expressions of Malagasy are listed in Abinal & Malzac 1963 and Rabenilaina 1993. The primary structure building function is Merge, particular cases of which is the merging of about 20 voice affixes, which, to be succinct, we treat as **predicate building functions** (PBFs)³. They form tenseless Pn's, *n-place predicate phrases*, by affixation. The core grammar of Malagasy is given by these affixes plus the language general Predicate-Argument and Modifier-Head cases of Merge. **3.1** presents enough Malagasy PBFs to show that they are not just an active/passive alternation. We also show how theta roles and theta equivalence derive from semantic interpretation. We start with P2s in the present tense. **3.2** presents general constraints on PBFs.

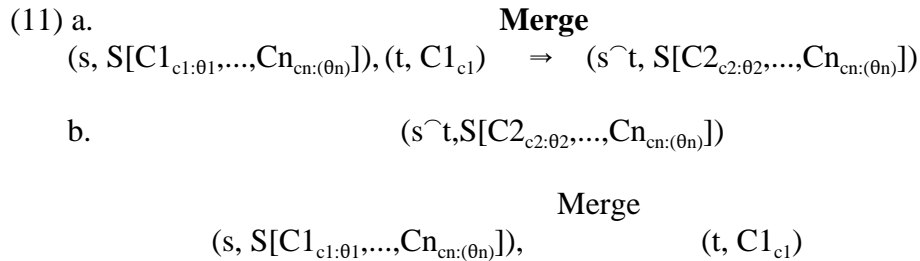
3.1.1 **man** is a PBF illustrated in (9)⁴. Its domain is a listed set of roots.

- (9) man+enjika (manenjika) ny jiolahy Rabe
man+chase the thief Rabe
Rabe chases/is chasing the thief

enjika is a root of category $RT_{\{AG,TH\}}$. **man** maps $(enjika, RT_{\{AG,TH\}})$ to the P2 *manenjika* of category $S[DP_{acc:TH}, DP_{\delta:AG}]$, which selects a DP_{acc} interpreted as Theme to form a P1, *manenjika ny jiolahy*, of category $S[DP_{\delta:AG}]$, which merges with a DP_{δ} interpreted as Agent to form a P0 (0 = no features left to check). The action of **man** is illustrated in (10a) and the derivation of (9) in (10b).



3.1.1.1 Syntax Merge concatenates a string s of category $S[C1_{c1:\theta1}, \dots, Cn_{cn:(\theta n)}]$ with a t of category $C1_{c1}$, satisfying a category (C1) and a case (c1) requirement, to form $s \hat{\ } t$ of category $S[C2_{c2:\theta2}, \dots, Cn_{cn:(\theta n)}]$. Pictorially in (11b):



So $\text{Dom}(\text{Merge})$, the domain of Merge, contains pairs $\langle u, v \rangle$ of expressions in which the category and case of v satisfy (are checked against) the category and case required by u . Theta role θ_i will be assigned to the i^{th} argument, see **3.1.1.2**. θ_n , but no other θ_i , is absent in one case (see **3.4.5**). A P0 derived by successive applications of PBFs has all arguments assigned case and all but possibly the last interpreted with a theta role.

3.1.1.2 Semantics For each domain E , $(enjika, RT_{\{AG,TH\}})$ denotes a set of pairs of elements of E . A P1 denotes a *property*, a function from E into $\{\text{True}, \text{False}\}^5$. And in general a P_{n+1} maps entities to P_n denotations. To simplify we use DPs interpretable as elements of E . Extending them to generalized quantifiers is standard (Keenan 1996).

Theta roles are relations between entities in E and n -ary predicate denotations, $n > 0$. We write $\llbracket d \rrbracket$ for the semantic interpretation of an expression d , and we simplify notation by omitting the category coordinate of an expression when clear from context. Here and later \in is the set membership relation, \wedge is *and*, and *iff* is *if and only if*. Then,

(12) $\llbracket \mathbf{man}(enjika, RT_{\{AG,TH\}}) \rrbracket(y)(x) = \text{True}$ iff

$$(x,y) \in \llbracket enjika \rrbracket \wedge \text{Theme}(y \llbracket enjika \rrbracket) \wedge \text{Agent}(x, \llbracket enjika \rrbracket).$$

We mostly wrote *enjika* instead of $\langle enjika, RT_{\{AG,TH\}} \rangle$ above. Some may want to read the right hand side of (12) as “x and y are participants in a chasing event $\llbracket enjika \rrbracket$, y is Theme of $\llbracket enjika \rrbracket$ and x is Agent”. All we require is compositionality: that the derived expression $\mathbf{man}(enjika)$ be interpreted as a function of what it is derived from, *enjika*.

From (12) it is the P2 $\mathbf{man}(enjika)$, not just the root *enjika* or the prefix *man-*, which matches arguments and theta roles. (9) is True (in E) iff the pair \langle the thief, Rabe \rangle is in $\llbracket enjika \rrbracket$, the thief bears the Theme relation to $\llbracket enjika \rrbracket$ and Rabe the Agent relation. Thus theta role assignment derives from semantic interpretation.

3.1.2 mi is a PBF whose domain is a listed set of roots. It prefixes them with *mi-* fully preserving the root, unlike **man**. It is the most frequent voice affix (K&M 2001) and the most productive. Hilda Koopman (p.c.) notes nonce uses like *miparticiper* ‘participates’ on the web. **mi** builds many P1s from roots of category $RT_{\{AG\}}$: *miasa* ‘works’, *miteny* ‘speaks’, *mihomehy* ‘laughs’, *milomano* ‘swims’, *mitomany* ‘cries’, *miezaka* ‘strive’. It also builds some P1s with Theme arguments: *miely* ‘be dispersed’, *mianjera* ‘collapse’, *mifaly* ‘to be happy, rejoice’, *migadra* ‘to be in chains, prison’.

mi is sometimes called an intransitive verb former, but that is a distortion. It applies to very many roots of category $RT_{\{AG,TH\}}$ building P2s of category $S[DP_{acc:TH}, DP_{\delta:AG}]$, like **man**. Here are 15 examples, another 15 are easily found: *mikapoka* ‘beats’, *mividy* ‘buys’, *mivarotra* ‘sells’, *mitady* ‘seeks’, *mifidy* ‘chooses’, *mikarakara* ‘takes care of’, *mitaona* ‘transports’, *mijery* ‘watches’, *miantso* ‘calls’, *mibaby* ‘carries’, *mitana* ‘holds’, *mitazana* ‘espies’, *mitily* ‘spies on’, *miteraka* ‘give birth to’, *mitelina* ‘swallow’, *misambotra* ‘capture’. None of these roots are also in the domain of **man**. Thus if we know of a Pn that it interprets its DP_{δ} as Agent we cannot predict the verbal prefix: it is *man-* with *enjika* ‘chase’ and many others, but *mi-* with *kapoka* ‘beat’ and many others. **mi** also prefixes many roots that are in the domain of **man**. Rahajarizafy 1960:50 notes:

(13) When **mi** and **man** apply to the same root r the **mi** verb usually has lesser valence.

Rahajarizafy 1960 notes a few cases where **mi**(r) and **man**(r) have the same valency, and two where **mi**(r) has greater valency. But (13) is far and away the norm, so it can help language learners. The precise relations between **mi**(r) and **man**(r) require a large scale study. Most commonly **man**(r) is a rough semantic causative of **mi**(r): *mihidy* ‘is locked’ / *manidy* ‘locks (tr)’, *milatsaka* ‘falls’ / *mandatsaka* ‘drops’, *milahatra* ‘line up (intr)’ / *mandahatra* ‘arrange in lines’; *miverina* ‘come back’ / *mamerina* ‘give back’; *mivory* ‘gather (intr)’ / *mamory* ‘gather (tr)’. But in several cases the two verb forms have sufficiently diverged in sense that we lose any compositional relation between them: *mianatra* ‘studies’ / *mananatra* ‘admonishes’; *mipetraka* ‘sits, resides’ / *mametraka* ‘puts’; *milaza* ‘says’ / *mandaza* ‘praises, celebrates’.

3.1.3 ina is a PBF whose domain includes both roots and *mamp*-prefixed verbs (such verbs are causatives)⁶. **ina** suffixes *-ina* (*-ena*, *-(a)na*) preserving root final consonants⁷. A *p* in $\text{Dom}(\mathbf{ina})$ has at least two theta roles to assign. (14) exhibits the action of **ina** on the root *enjika*. We mark stress to remind the reader that suffixing triggers stress shift. (15b) summarizes our derivation of (15a), and (16) gives its semantic interpretation.

ina

(14) $(\text{enjika}, \text{RT}_{\{\text{AG}, \text{TH}\}}) \mapsto (\text{enjehina}, \text{S}[\text{DP}_{\text{poss:AG}}, \text{DP}_{\delta:\text{TH}}])$

(15) a. *enjika* + **ina** + *Rabe* (*enjehin-dRabe*) *ny jiolahy*
 chase+**ina**+*Rabe.poss* the thief
Rabe chases the thief

b.

	S	
	S[DP _{δ:TH}]	DP _δ
	S[DP _{poss:AG} , DP _{δ:TH}]	DP _{poss}
	RT _{AG,TH}	
<i>enjika</i> + ina	<i>Rabe</i>	<i>ny jiolahy</i>
chase ina	<i>Rabe.poss</i>	the thief
<i>Rabe chases/is chasing the thief</i>		

We gloss *Rabe* as *Rabe.poss* to indicate that *Rabe* combines with *enjehina* in the complex way that possessors combine with their heads. And *enjehina* interprets its DP_δ argument as Theme (TH) and its DP_{poss} argument as Agent (AG). **Gen 1** is the general pattern.

(16) $\llbracket \mathbf{ina}(\text{enjika}) \rrbracket(u)(v) = \text{True iff } (u, v) \in \llbracket \text{enjika} \rrbracket \wedge \text{TH}(v, \llbracket \text{enjika} \rrbracket) \wedge \text{AG}(u, \llbracket \text{enjika} \rrbracket)$

Gen 1 For *p* a root with zero or more affixes, if $p \in \text{Dom}(\mathbf{ina})$ then

- (a) **ina**(*p*) is a P_n of category S[DP_{poss:θ1, ..., DP_{δ:θn}], for some $n > 1$, and}
- (b) if *p* selects an Agent then θ₁ = Agent

Many roots lie both in $\text{Dom}(\mathbf{man})$ and $\text{Dom}(\mathbf{ina})$: *haja* ⇒ *manaja/hajaina* ‘respects’, *vono* ⇒ *mamono/vonoina* ‘hits, kills’, *la* ⇒ *manda/lavina* ‘refuses’, *fonos-* ⇒ *mamono/fonosina* ‘wraps’, *vaky* ⇒ *mamaky/vakina* ‘cuts, reads’. Also many roots lie in both $\text{Dom}(\mathbf{mi})$ and $\text{Dom}(\mathbf{ina})$: *bata* ⇒ *mibata/bataina* ‘raises up’, *kapoka* ⇒ *mikapoka/kapohina* ‘beats’, *antso* ⇒ *miantso/antsoina* ‘calls’. We define:

Def 1 Nuclear Ss are *theta equivalent*, \equiv_{θ} , iff their arguments can be bijectively matched such that each argument in one and its match in the other have the same theta role. (‘nuclear S’ \equiv_{df} P₀ built just by Merging arguments with a P_n built by affixing roots).

Notice that (17a) follows immediately from (16) and (12), and in general (17b) holds

when *r* is in the domain of both **man** and **ina**. So the derived verbs are logical converses, from which the two propositions below follow.

- (17) a. $\llbracket \mathbf{man}(enjika) \rrbracket(y)(x) = \llbracket \mathbf{ina}(enjika) \rrbracket(x)(y)$
 b. $\llbracket \mathbf{man}(r) \rrbracket(y)(x) = \llbracket \mathbf{ina}(r) \rrbracket(x)(y)$

Proposition 1 For *r* of category $RT_{\{AG,X\}}$ in the domain of both **man** and **ina**,

- A. nuclear Ss built from **man**(*r*) and **ina**(*r*) are theta equivalent: the DP_{δ} argument of each has the same theta role as the non- DP_{δ} argument of the other.
- B. **mi**(*r*) and **ina**(*r*) build theta equivalent nuclear Ss when **mi**(*r*) is transitive.

Proposition 2 Nuclear Ss built from **man**(*r*) and **ina**(*r*), *r* of category $RT_{\{AG,X\}}$, are logically equivalent when their Agents are the same individual denoting DPs and their Themes are the same individual denoting DPs. Similarly for **mi**(*r*) and **ina**(*r*) in B.

3.1.4 The PBF **a**, as in (1b), prefixes *a-* to roots, and like **mi** and **man**, applies only to roots. **Gen 1** holds of **a**(*p*) just as it holds of **ina**(*p*). The DP_{δ} of a P1 built by **a** is an “Intermediary”: Theme with roots of movement or transmission, Instrument or Theme otherwise. See Paul (1999) for extensive discussion. (18b) illustrates its Theme usage.

- (18) a. [man+tosika (manosika) ny fiara] Rabe
man+push the car Rabe
Rabe pushes/is pushing the car
- b. [a+tosika+Rabe (atosi-dRabe)] ny fiara
a+push+Rabe.poss the car
Rabe pushes/is pushing the car

Thus we cannot predict the voice affix of a P1 which interprets its DP_{δ} as Theme and its DP_{poss} as Agent. If the root is *enjika* ‘chase’ we use *-ina*, if it is *tosika* ‘push’ we use *a-*. *tosika* is not in $\text{Dom}(\mathbf{ina})$: **tosihina*. Similarly *tao*(*v*) \Rightarrow *manao/atao* ‘do, make’ and *verina* \Rightarrow *mamerina averina* ‘return (sth)’ accept *man-* and *a-* but not *-ina*. *voaka* ‘exit’ derives *mivoaka* ‘go out’ and *mamoaka/avoaka* ‘expell’ but **voahina*; *lahatra* \Rightarrow *mandahatra* and *alahatra* ‘put in a line’ but **laharina*, and *tono* \Rightarrow *mitono* and *atono* ‘grills’ but **tonoina*. There are many more roots in $\text{Dom}(\mathbf{ina})$ not in $\text{Dom}(\mathbf{a})$: *lalao* ‘play’ \Rightarrow *lalaovina* but **alalao*; *la* ‘refuse’ \Rightarrow *lavina* but **ala*; *haja* ‘respect’ \Rightarrow *hajaina* but **ahaja*. But usually both **a** and **ina** apply to ditransitive roots of transmission, as in (1). **a** assigns Theme to its DP_{δ} and **ina** assigns Goal. Similar roots are *roso* \Rightarrow *aroso/rosoana* ‘serve’ and *solo* \Rightarrow *asolo/soloana* ‘replace, substitute’. This paradigm is not limited to verbs of transmission:

- (19) a. man+tondro (manondro) ny tranony amin’ny tehiny Rabe
man+point-out the house-his with’the cane-his Rabe
Rabe points out his house with his cane

- b. a+tondro+Rabe (atondron-dRabe) ny tranony ny tehiny
 a+point-out Rabe.poss the house-his the cane-his
- c. tondro+ina+Rabe (tondroin-dRabe) amin'ny tehiny ny tranony
 point-out+ina+Rabe.poss with'the cane-his the house-his

Here are the values of **man**, **a**, and **ina** at ditransitive roots r of category $RT_{\{AG, GL, TH\}}$. (20a) says that the P3 **man**(r) interprets its z argument as Theme, the resulting function interprets its y argument as Goal, and the final P1 function interprets its x argument as Agent. In contrast the **a** verb in (20b) interprets its first argument as Agent, its second as Goal and its last as Theme. And the **ina** verb in (20c) interprets its first argument as Agent, its second as Theme, and its last as Goal. R is the ternary relation denoted by r .

(20) a. $\llbracket \mathbf{man}(r, RT_{\{AG, TH, GL\}}) \rrbracket(z)(y)(x) = \text{True}$ iff

$$\langle z, y, x \rangle \in R \wedge \text{THEME}(z, R) \wedge \text{GOAL}(y, R) \wedge \text{AGENT}(x, R)$$

b. $\llbracket \mathbf{a}(r, RT_{\{AG, TH, GL\}}) \rrbracket(x)(y)(z) = \text{True}$ iff

$$\langle z, y, x \rangle \in R \wedge \text{AGENT}(x, R) \wedge \text{GOAL}(y, R) \wedge \text{THEME}(z, R)$$

c. $\llbracket \mathbf{ina}(r, RT_{\{AG, TH, GL\}}) \rrbracket(x)(z)(y) = \text{True}$ iff

$$\langle z, y, x \rangle \in R \wedge \text{AGENT}(x, R) \wedge \text{THEME}(z, R) \wedge \text{GOAL}(y, R)$$

(20b,c) are theta equivalent, as are (1b,c). So theta equivalence doesn't just apply to pairs we might analogize to active vs passive in English. And we now extend **Proposition 1**:

- C.** For a two theta role root r in the domain of both **man** and **a**, nuclear Ss built from **man**(r) and **a**(r) are theta equivalent, the DP_δ argument of one bearing the same theta role as that of the non- DP_δ argument of the other.
- D.** For r a ditransitive root of transmission in the domains of **man**, **a**, and **ina**, the three nuclear Ss they build from r are theta equivalent, as in (20) above.

3.1.5 Interim summary We have exhibited several structure building affixes which derive theta equivalent P0s (Ss). None uses **Move**. And we have seen how an explicit semantics *can* guarantee theta equivalence when DPs with the same theta role originate in structurally different positions. This is a usual semantical pattern: (21a,b) for example are syntactically non-isomorphic but logically equivalent, which follows from their

$$(21) \quad \text{a. } (P \ \& \ Q) \qquad \qquad \qquad \text{b. } \neg(\neg P \ \text{or} \ \neg Q)$$

compositional interpretation – specifically the meanings of *and*, *or*, and *not* (\neg). And what we have done above is show that structurally non-isomorphic nuclear Ss are semantically equivalent in a certain respect – namely, they are theta equivalent.

3.2 General Constraints on Case and Theta Role Assignment

We consider further PBFs below, but the four we have presented – **man**, **mi**, **ina**, and **a** already illustrate some general properties of case and theta role assignment that Malagasy satisfies. We call them *axioms* as they are true of Malagasy, don't follow from other assumptions, and they constrain our further analysis. As indicated, some apply to languages other than Malagasy. We use F to range over affixal PBFs, r over roots, p over roots with zero or more affixes; $\Theta(p)$ is the set of theta roles associated with p⁸.

Axiom 1 limits to $\Theta(r)$ the theta roles a predicate F(r) built from r can assign. And all arguments must be assigned a theta role, but not all elements of $\Theta(r)$ need be assigned, as in *milatsaka* 'fall' vs *mandatsaka* 'drop'. Axiom 1 applies to languages which affix roots to form Pn's (Semitic and Bantu come to mind).

Axiom 1 For all F(r) of category $S[C1_{c1:\theta1}, \dots, Cn_{cn:\thetan}]$, $\theta1, \dots, \theta n \in \Theta(r)$

Axiom 2 For all F(p) of category $S[C1_{c1:\theta1}, \dots, Cn_{cn:(\theta n)}]$,

- a. $\theta_i \neq \theta_j$, all $1 \leq i \neq j \leq n$ and
- b. θ_j is absent iff $j = n$ and $F = \mathbf{ana}$, all $1 \leq j \leq n$

So Pns derived by PBFs never *assign* the same theta role to distinct arguments. Only the last argument may lack a theta role, and that iff $F = \mathbf{ana}$ (below). If $F \neq \mathbf{ana}$ then axiom 2 + Merge guarantees a bijective relation between argument expressions of affixal Pns and their assigned theta roles, so in this weak sense the Theta Criterion holds for all but one affix. But note that nothing blocks a given real world object from being both Agent and Theme of a given predicate, as in *Namono tena Rabe* 'Rabe killed himself'.

Axiom 3 For all s of category $S[C1_{c1:\theta1}, \dots, Cn_{cn:(\theta n)}]$,

- a. $c_j = \text{poss} \Rightarrow j = 1$ all $1 \leq j \leq n$ ($\Rightarrow =$ "implies", here and later)
- b. $c_j = \delta$ iff $j = n$ all $1 \leq j \leq n$

So from (3a) if a Pn selects a DP_{poss} then that is its first argument. So a Pn can have at most one DP_{poss} argument. (3b) says that the last argument of a Pn is in the default case δ . No other argument is *assigned* δ case and Malagasy lacks quirky case. This is as close as we get to the EPP. Axiom 3 requires a Pn to have exactly one DP_{δ} and not more than one DP_{poss} . But they may have more than one DP_{acc} . **man**(*ome*) = *manome* 'gives' has two, as do causatives of P2s, not discussed here.

Malagasy verbs are divided into two major syntactic classes according as they select a DP_{poss} or not. Axiom 4 says that this is decided entirely by the choice of voice affix:

Axiom 4 For all F, either for all $p \in \text{Dom}(F)$, F(p) selects a DP_{poss}
or for all $p \in \text{Dom}(F)$, F(p) does not select a DP_{poss}

So since **man**(*enjika*) does not select a DP_{poss} no other verb built by **man** does either. And since **ina**(*enjika*) does select a DP_{poss} then all other verbs built from **ina** do too.

Def 2 F is (poss) **bound** iff for some p, F(p) selects a DP_{poss} . F is (poss) **free** iff for some p, F(p) does not select a DP_{poss} . The voice affix F introduces and the predicate it builds are also called *bound* (*free*).

By axiom 4 Def 2 is a well definition. Proposition 3 now follows from axiom 3 given that $DP_{\text{poss}} \neq DP_{\delta}$. Further Gen 2 holds and is massively helpful in acquisition.

Proposition 3 For all bound F, F(p) selects at least two arguments (DP_{poss} and DP_{δ}).

Gen 2 F is free iff the affix it introduces is *m* initial in the present tense.

Thus **man** and **mi** are free (as are **ma**, **maha**, **mana**, **m+ø**, **mamp**, **mif**, mentioned briefly later); **a** and **ina** are bound (as are **voa**, **tafa**, **in** mentioned briefly later).

Axiom 5 tells us how the semantic notion Agent is coded in Malagasy syntax. That is, we can identify the Agent of P_n as a function of the structure of the P_n.

Axiom 5 For F(p) of category $S[DP_{c1:\theta1}, \dots, DP_{cn:(\theta_n)}]$ and $\text{Agent} \in \Theta(p)$,

a. F is bound $\Rightarrow \theta_1 = \text{Agent}$ and **b.** F is free $\Rightarrow \theta_n = \text{Agent}$

Axiom 5 blocks deriving an F(p) of category $S[DP_{\text{acc:AG}}, DP_{\delta:\text{TH}}]$ since F would be free so θ_2 must be Agent, not Theme. Also blocked is an F(p) of category $S[DP_{\text{poss:TH}}, DP_{\delta:\text{AG}}]$ since this F is bound, the theta role associated with DP_{poss} must be Agent.

Axiom 5 is Malagasy specific, mentioning whether P_n's select a possessive DP or not. But we can generalize it in a way which helps us understand better the relation between language general properties and their realization in any given language. We first define *isomorphic* expressions. Let $G = (\text{Lex}, \mathbf{F})$ be a grammar, with Lex a lexicon, \mathbf{F} a set of structure building functions. $L(G) = L$ the set of expressions derivable by starting with those in Lex and adding in iteratively all those obtained by applying structure building functions to ones already in the set. An **automorphism** h of L is a bijection from L to L which doesn't change how expressions are built up: if $z = F(x,y)$ iff $h(z) = F(h(x),h(y))$, any $F \in \mathbf{F}$ (say $F = \text{Merge}$). So z and h(z) are built in exactly the same way. Expressions x,y are *isomorphic*, $x \approx y$, iff for some automorphism h, $h(x) = y$ (in which case the automorphism h-inverse maps y to x). We now generalize Axiom 5:

Axiom 5* Let P and P' have categories $S[C1_{c1:\theta1}, \dots, Cn_{cn:\theta_n}]$ and $S[C1_{c1:\theta1'}, \dots, Cn_{cn:\theta_n'}]$ respectively. Then $P \approx P' \Rightarrow$ for all $1 \leq j \leq n$, $\theta_j = \text{Agent}$ iff $\theta_j' = \text{Agent}$

Axiom 5* makes it evident that **man**(*enjika*) should not be isomorphic to **ina**(*enjika*), since the former assigns Agent to C2, and the latter to C1. And indeed while they are theta equivalent, they are not isomorphic⁹: If they were then the P1s they build would be,

whence automorphisms could map DP_{poss} to DP_{acc} and conversely, an impossibility since they have different distributions. Similarly **man**(r) is always in $\text{Dom}(\mathbf{ana})$, 3.4.5 below, and **ina**(r) never is, hence again neither can be mapped to the other as automorphisms preserve the domains of the structure building functions (K&S).

Axiom 5* applies to more languages than Axiom 5 but tells us less about them. It does not entail Axiom 5, it just guarantees that the semantic notion of Agent is syntactically coded, but does not specify how. Axiom 5* does not really depend on our notation for the argument/case selection matrix. We can paraphrase it by saying that if two n-place predicates are isomorphic then for any j, the jth argument of one is assigned Agent iff the jth argument of the other is. 5* holds (vacuously) if neither P_n assigns Agent, and Axiom 2a guarantees that at most one argument is assigned Agent.

The relation between Axioms 5 and 5* is of general interest. Axiom 5 does state just how Malagasy satisfies Axiom 5* for P_n's built by successive affixation of roots (as opposed to P_n's formed by coordination, or combination with adverbs for example). This pattern seems to me typical. Linguists often speak as though universal constraints appear as such as part of the grammar of each language. But more usually any given language satisfies language specific conditions which are stronger than universal ones, as is the case here. And we might generalize Axiom 5 further, to treat all theta roles alike:

Axiom 5** Let P and P' have categories $S[C1_{c1:\theta1}, \dots, Cn_{cn:\thetan}]$ and $S[C1_{c1:\theta1'}, \dots, Cn_{cn:\thetan'}]$ respectively. Then $P \approx P' \Rightarrow$ for all $1 \leq j \leq n$, $\theta_j = \theta_j'$.

Axiom 5** is consistent with but weaker than current assumptions in PP/MP theories. It does not require that the relevant structures in different languages be similar. Malagasy may code theta relations in terms of verbal affixes, Korean in terms of nominal case marking, Eskimo in terms of agreement, All we require is that for each grammar G, the automorphisms of G respect theta relations in the way these axioms say. We turn now to a consequence of our axioms. It uses the fundamental notion of **invariant**.

Def 3 A binary (n-ary) relation R on a language L is **invariant** iff [$\langle x, y \rangle \in R$ iff $\langle h(x), h(y) \rangle \in R$, all automorphisms h]. This says that $h(R) = R$, where $h(R)$ is_{df} the set of pairs $\langle h(x), h(y) \rangle$ for each pair (x,y) in R. So the *invariant (structural)* relations among expressions of L are just those which are **fixed** (mapped to themselves) by all the automorphisms. They are what you cannot change without changing structure. As a special case a **grammatical constant** ("function word") is an expression that is mapped to itself by all automorphisms¹⁰. For example:

Gen 3 The Malagasy voice affixes *man*, *mi*, *a*, *ina*, are automorphism invariant (grammatical constants) as are ones we list later: *ma*, *mana*, *mif*, *mamp*, *tafa*, *voa*, *in*, and \emptyset (the zero prefix). And we conjecture:

Gen 3* In any language derivational morphology is invariant.

Note that expressions, properties (sets) of expressions, relations among expressions, etc. are invariant (structural) in exactly the *same* sense: fixed by the automorphisms.

Theorem 1 Theta Equivalence is Invariant. That is, for all nuclear Ss Φ, Ψ

$$\Phi \equiv_{\Theta} \Psi \Rightarrow \text{for all automorphisms } h, h(\Phi) \equiv_{\Theta} h(\Psi)$$

Theorem 1 says that *syntactic equivalence* respects *theta equivalence*: if $\Phi \equiv_{\Theta} \Psi$ then anything that Φ is isomorphic to is theta equivalent to anything that Ψ is isomorphic to. To see what this excludes, we show that *logical equivalence* (having the same truth value in every model) is not invariant in English. We find Φ, Ψ with $\Phi \equiv \Psi$ but $h(\Phi) \neq h(\Psi)$ for some automorphism h . Now there is no reason to think that the numerals *seven* and *eight* are syntactically different, so let h be an automorphism that interchanges them and fixes all other lexical items. (22a) and (22b) are logically equivalent but their values under h (got by swapping *eight* and *seven*) are not, so logical equivalence is not preserved under replacement by isomorphic expressions. (NOT a surprise; whether two Ss are logically equivalent depends on their meaning, not their syntactic isomorphism class).

- (22) a. At least eight students came to the party
 b. More than seven students came to the party

But some semantic relations are coded in the syntax. K&S argue that the *possible antecedent of an anaphor* relation is of that sort: in any natural language, “ x is a possible antecedent of an anaphor y in z ” holds iff $h(x)$ is a possible antecedent of an anaphor $h(y)$ in $h(z)$, all automorphisms h . And Theorem 1 says that theta equivalence is that kind of relation. Isomorphic expressions assign theta roles comparably, as Axiom 5**¹¹ says.

Further it is still true on our approach that theta roles are assigned as a function of structure. To say that the *man-* verb in (1a) assigns the same theta role to its DP_{δ} argument as the *a-* verb in (1b) does to its DP_{poss} one is like saying that $8^2 = 4^3$. Different functions at different arguments may have the same value. So we don’t need to assume that DPs in different Ss with the same theta role originate in the same configuration.

Finally, Theorem 1 does not claim or assume that the relevant structures of one language are similar (much less isomorphic) to those of another, or even that they share grammatical categories. In contrast the claim that Agent is always and only assigned to Spec,vP does make such assumptions. In fact Theorem 1 is just a way of making explicit the idea that theta role assignment is *structure dependent* (Chomsky (1965:55-56; 1975:30 – 35, Radford 1997:11 – 15) – an early notion never formally defined.

3.2.1 Other Uniformity Parameters? The following candidate parameters for PBFs fail in Malagasy – though (23B) challenges Axiom 5** and does not fail by much.

- (23) A. For $r, r' \in \text{Dom}(F)$, $F(r)$ and $F(r')$ have the same number of arguments
 B. For $r, r' \in \text{Dom}(F)$, $\text{Agent} \in \Theta(r)$ iff $\text{Agent} \in \Theta(r')$.

(23A) fails: **man** builds both P2s, like *manenjika* ‘chases’ in (10a), and also P3s, as in (1a) *manolotra* ‘offers’. *man-* is the second most common voice affix (K&M, 2001), and the majority of roots *r* in its domain have category $RT_{\{AG,TH\}}$. **man** maps them to **man**(*r*) of category $S[DP_{acc:TH}, DP_{\delta:AG}]$. Some present tense examples are: **man**(*didy*) = *mandidy* ‘cuts’, **man**(*vaky*) = *mamaky* ‘reads’, **man**(*vono*) = *mamono* ‘kills’, **man**(*fonos-*) = *mamono* ‘envelopes’. **man** also applies to roots *r* of category $RT_{\{AG,TH,GL\}}$, mapping them to **man**(*r*) of category $S[DP_{acc:TH}, DP_{acc:GL}, DP_{\delta:AG}]$, as in: **man**(*roso*) = *mandroso* ‘serves’, **man**(*tolotra*) = *manolotra* ‘offers’, **man**(*ome*) = *manome* ‘gives’.

More challenging, **man** builds two types of intransitive verb. It applies to a few roots of category $RT_{\{AG\}}$: **man**(*leha*) = *mandeha* ‘goes’, **man**(*dihy*) = *mandihy* ‘dances’, **man**(*lainga*) = *mandainga* ‘lies’, **man**(*lohalika*) = *mandohalika* ‘kneels’. (These four examples may exhaust the agentive (unergative) intransitives **man** builds).

man also has some roots of category $RT_{\{TH\}}$ in its domain: **man**(*hetaheta*) = *mangetaheta* ‘is thirsty’, **man**(*hatsiaka*) = *mangatsiaka* ‘is cold’, **man**(*firatra*+*firatra*) = *mimirapiratra* ‘scintillates’. These **man** verbs do not select an Agent, whence (23B) fails. We might analyze them as unaccusatives, but even so (23B) remains false. And the regularities that unaccusative analyses account for in Dutch, French, Italian, and Turkish do not seem to carry over to Malagasy. Malagasy has no auxiliaries of the *have/be* sort nor do verbs agree with arguments in ϕ features. There is no analogue of *ne* cliticization, and Malagasy seems to have no (overt or null) expletives. The closest is *ny andro* ‘the day’ in weather expressions, as in *mangatsiaka ny andro* ‘lit: ‘cold the day’. But one says in the same sense *mangatsiaka aho* ‘I am cold’. And lastly, unaccusative verbs in other languages do not receive a uniform voice coding in Malagasy. Nor do weather expressions¹².

Not adopting an unaccusative analysis allows the possibility that *mandihy izy* ‘He dances’ and *mangatsiaka izy* ‘He is cold’ with Agent and Theme arguments respectively are isomorphic but not theta equivalent, contra Axiom 5**. But the two verbs classes do differ: The activity ones have standardly formed imperatives – *mandéha* \Rightarrow *mandehána* ‘Go!’, *mandíhy* \Rightarrow *mandihíza* ‘Dance!’, *mandohálíka* \Rightarrow *mandohaléha* ‘Kneel!’, and *mandáinga* \Rightarrow *mandaingá* ‘Lie!’. But the Theme argument predicates lack imperatives. Thus the former lie in $\text{Dom}(\mathbf{imp})$ and the latter don’t, so they aren’t isomorphic.

Thus **man** intransitives illustrate an important property of automorphisms. Namely, K&S, the trees representing the derivation of two expressions may be isomorphic qua ordered labeled trees, but the derived expressions may fail to be isomorphic. Reason: the tree isomorphism might map some *X* to a *Y* acceptable in that tree. But a language automorphism must map *all* expressions to (well-formed) expressions, and replacing *Y* by *X* in another expression may fail to preserve grammaticality. Here is a hypothetical example. Assume that *Nigel shot Trevor* and *Nigel shot himself* have isomorphic derivations. That isomorphism maps *Trevor* to *himself*. But a language automorphism can’t do that since it would map *Trevor cried* to **Himself q’ed* (*q* a P1), contradicting that automorphisms map expressions in the language to expressions in the language.

3.3 Predicate Building Functions and the Active - Passive Distinction

We have illustrated two free PBFs, **man** and **mi**, and two bound ones, **ina** and **a**. They build a significant variety of theta equivalent Ss. Many linguists (Rahajarizafy 1960, Keenan 1976, Paul 1999) have analogized from **man/mi** to “active” and **ina/a** to “passive” in English or French. In a typical **man** S the DP sister to the P1 is Agent, but in its theta equivalent **ina** S it is Theme, with the Agent DP part of a constituent with the verb, as in English passives. So the DP sister of the P1 is analogized to subjects.

However as work on Malagasy has deepened with contributions from many scholars over the last 15 years, the disanalogy between European passives and **ina/a** verbs in Malagasy has become increasingly apparent. It is well acknowledged in Pearson 2005. We review some of these differences here, as our view is one in which primary (= just one PBF affix) free and bound verbs are comparably basic (though not isomorphic).

3.3.1 Imperatives (again) If the forms in (1b,c,d) were passives it would be surprising to find them in the imperative. But we find them in abundance when a verb has both bound and free primary forms. Strikingly for example (K&M 2004) of the 26 imperatives in the child language data of HNM, all were bound forms. Forcing a passive translation on **ina** verbs may lead to jarring interpretations not in fact present in the Malagasy. Compare:

- (24) a. novonjen-dRabe aho
Pst.save-**ina**-Rabe 1.δ
I was saved by Rabe
Rabe saved me
- b. vonjeo aho!
be+saved+**imp** 1.δ
“Act that I may be saved!”
Save me!

The form in (24b) is the normal way to say “Save me!”. The **man** imperative in (25b) exists but is at best a pragmatically marked form.

- (25) a. namonjy ahy Rabe
pst.**man**+save 1s.acc Rabe
Rabe saved me
- b. mamonjé ahy!
man+save+Imp 1s.acc
Save me!

3.3.2 Position If **ina/a** forms are passives it is surprising, as Pearson 2005 points out, that the position of the Agent phrase is rigorously positioned relative to the verb. Usually passive Agents have the positional freedom of optional obliques.

3.3.3 Frequency K&M’s counts show that the Agent phrase in **ina** forms is normally present, supporting that **ina** verbs are better thought of as transitive (as here).

3.3.4 Control with verbs of desire and intent is vested in the Agent for both **man/mi** verbs and **ina/a** verbs. Equally Agents of **ina** (and **man**) verbs can be controlled

- (26) a. n+i+kasa hamaky (h+aN+vaky) io boky io aho
past+**mi**+intend fut+**man**+read that book that 1s.δ
I intended to read that book

- b. nokasaiko (no+kasa+ina+ko) hovakina (ho+vaky+ina) io boky io
 past+intend+**ina**+1s.poss fut+read+**ina** that book that
I intended to read that book
 (??) *That book was intended by me to be read*

Even the forced passive translation of (26b) above is not accurate, since in English the Agent of *be read* is not bound. Note too that in both morphologies the verb *read* is selected in the future, governed by the verb ‘intend’ (in the past).

Coordinate structures, as in (27), from Malagasy novels (Keenan 1995) show possessor Agent control (with new verb types), unexpected if it is a non-active agent.

- (27) a. araka izao [[ahitana- \emptyset_i sy ahafantara-nareo \emptyset_i] ahy] izao
 according this [[[see+**ana** and know+**ana**]+2.poss.pl] 1s.acc] this
according to this which you see and understand (of) me
- b. ataoko fa [voa+jery- \emptyset_i sy voa+dinikao \emptyset_i] tsara ireo sary teo
 I+think that [[**voa**+see and **voa** examine+2s.poss] good those picture there]
I think that you have seen and examined well those pictures there

3.3.5 Selection Both free and bound verbs are selected by modal verbs. *mahazo* ‘can’ and *mahay* ‘able’ select free verbs, their roots *azo* and *hay* select bound ones. Replacing free by bound verbs in (28a) and (29a) is ungrammatical.

- (28) a. mahazo manao (man+tao(v)) izany aho
 permitted **man**+do that I
I can do that
- b. tsy azoko (azo + ko) atao izany
 not can + 1s.poss **a**+do that
I can't do that
- (29) a. mahay manoratra (man+soratra) amin'io penina io aho
 can write **man**+write with'that pen that 1s. δ
I am able to write with that pen
- b. tsy haiko (hay + ko) anoratana (man+soratra+ana) io penina io
 not able+1s.poss **man**+write+**ana** that pen that
I can't write with that pen

In sum, treating **ina/a** forms as European passives breeds confusion. Outside of Indo-European (Haspelmath 1990) aux+participle passives are little attested.

3.4 Further predicate building functions in Malagasy

- | | | |
|---|---|--|
| (33) a. Adala izy
crazy he
<i>He is/was crazy</i> | b. *Nadala / *No adala izy
Past+crazy / Past crazy he
<i>He was crazy</i> | c. Ho adala izy
Fut crazy he
<i>He will be crazy</i> |
|---|---|--|

In (32a) the use of past tense *no* would be ungrammatical, whereas it is natural with the **ina** form of the verb in (32b). And while there are semantic differences between the stative and dynamic bound forms they appear to form theta equivalent pairs. However, as a reader suggests, possibly a finer analysis of theta roles would distinguish Cause in the bare root case from Agent in the suffix one, losing theta equivalence. Similarly then the roots and their corresponding **man** or **mi** forms would not be theta equivalent. The free and bound forms in the imperative would remain theta equivalent:

- | | | | |
|--|----------|--|-----------|
| (34) a. Reseo
Defeat+ ina + imp 3.δ dem.pl
<i>Defeat them!</i> | izy ireo | b. Mandresea
man +defeat+ imp 3.acc dem.pl
<i>Defeat them!</i> | azy ireo! |
|--|----------|--|-----------|

3.4.3 Free Roots? There are only three free roots in common usage: *tonga* ‘arrive’, *avy* ‘come’, and *lasa* ‘gone (past tense)’. They do not take the initial *m* characteristic of free forms in the present and they do not directly form *-ana* forms (below).

- | | | |
|---|---|---|
| (35) a. Tonga izy
arrive he
<i>He is arriving/has arrived</i> | b. Avy ny orana
come the rain
<i>It is raining / rained</i> | c. Lasa izy
gone he
<i>He is has left</i> |
|---|---|---|

So free roots are few but bound ones many, at least 25 (possibly many more). Equating bound forms with passives, these data suggest that passive is the more basic form.

3.4.4 voa is a bound PBF which prefixes *voa* to roots. Its domain overlaps significantly with that of **ina**: *voatolotra/tolorana* ‘offered’, *voafitaka, fitahina* ‘deceived’, *voavonjy, vonjena* ‘saved’. But **voa** verbs focus on the final state of an activity, **ina** on the process. **Voa** verbs lack imperatives, and like bare roots are not marked for past tense. They select a DP_{poss} Agent naturally, but it is often absent in discourse.

- | | | |
|---|--|---|
| (36) a. Voa+kapoka (voakapoka) ilay alika
voa + beat that (aforementioned) dog
<i>That dog was beaten</i> | b. no+laza+ina+ko (nolazaiko) izany
past+say+ ina +1s.poss that
<i>I said that</i> | c. voa+laza+ko (voalazako) izany
Voa +say+1s.poss that
<i>I said that</i> |
|---|--|---|

There are yet further affixes which apply primarily to roots, but in the interests of space we just list them: the prefix *tafa-* and the no longer productive infix *-in-/-on-*, as in *tafatolotra* and *tonolotra*, create bound verbs, whereas *(m)a-*, *(m)ana-* and the very productive *(m)aha-* are free. *(m)aha-* has abilitative and indirect causative meanings (Phillips 2000). We turn now to the highly productive suffix *-ana* in (1d).

3.4.5 ana suffixes *-ana* (rarely *-anana*) to all affixed free verbs¹³, but not to bound ones or to roots. It creates bound verbs. **ana** applies to free verbs built from semantically rich secondary affixes, such as *(m)if-* ‘Reciprocal’ (Keenan & Razafimamonjy 2004) and *(m)amp-* ‘Causative’ (Randriamasimanana 1986; Andrianierenana 1996), which build free verbs from already affixed free verbs less initial *m*.

(37) a. [n+man+enjika (nanenjika) ny jiolahy t+amin’ny fiara] Rabe **man**
 past+**man**+chase the thief past+with’the car Rabe
Rabe chased the thief by means of the car

b. [n+[man+enjika]+ana+Rabe (nanenjehan-dRabe) ny jiolahy] ny fiara **ana**
 past+[[**man**+chase]+**ana**]+Rabe.poss the thief the car
Rabe chased the thief in/with/...the car

(38) a. man+taov (manao) ny farafara ho an-dRasoana Rabe
man+do the bed for-Rasoana Rabe
Rabe is building the house for Rasoana

b. \emptyset +[[man+taov]+ana]+Rabe (anaovan-dRabe) ny farafara Rasoana
 pres+**man**+make+**ana**+Rabe.poss the bed Rasoana
Rabe is making the bed for / because of Rasoana

ana verbs form present, past, future and imperatives as do bound verbs in general. The DP_{δ} of an **ana** verb usually bears an oblique theta role – Benefactive, Instrument, Cause, Location, Time, Manner, Price,..., not selected by the free verb **ana** applies to (see K&P, and Paul 1999). The Theta role of the DP_{δ} of an **ana** verb is not uniquely determined. In (38a) the Prep *ho an’* forces a Benefactive role for *Rasoana* but in (38b) the DP_{δ} could be any appropriate oblique. Usually few are appropriate, (Instruments are not Benefactees) but in this case Cause is possible (Maybe Rasoana nagged Rabe so much he gave in). (39a) illustrates the application of **ana** to an Agent/Theme P2. (39b) is the general statement.

ana

(39) a. (v, S[$DP_{acc:TH}$, $DP_{\delta:AG}$]) \mapsto (v+ana, S[$DP_{poss:AG}$, $DP_{acc:TH}$, DP_{δ}])

ana

b. (v, S[$DP_{c1:\theta1}, \dots, DP_{\delta:\theta n}$]) \mapsto (v+ana, S[$DP_{poss:\theta n}$, $DP_{c1:\theta1}, \dots, DP_{cn-1:\theta n-1}$, DP_{δ}])

Note that $c1 \neq poss$, otherwise the derived **ana** form has two DP_{poss} , violating Axiom 3a. Also DP_{δ} there is assigned no theta role, which is inferred from context, usually different from the selected theta roles. So an **ana** sentence is not quite theta equivalent to a free form with an overt PP. (40) illustrates some possible interpretations of **ana** forms.

(40) miénjika	runs away	ienjéhana	circumstance of fleeing
manénjika	chases	anenjéhana	circ. of chasing
mifanénjika	chase each other	ifanenjéhana	circ. of chasing each other
mampifanénjika	make chase e.o.	ampifanenjéhana	circ. of causing to chase e.o.

(41) Inona no nifanenjehan-dRabe sy Ravao omaly?
 what Foc past+Rec+**man**+chase+**ana** -Rabe.poss and Ravao yesterday
Why were Rabe and Rasoa chasing each other yesterday?

In isolation *inona* is just *What?*, but given the **ana** verb it must be questioning some circumstance of the action, and Cause/Reason is plausible here.

We derive **ana** verbs by extending the domain of Merge to concatenate PPs and adverbials with Pn's to form Pn's (independently needed). PPs and adverbials denote predicate **modifiers** – functions H mapping Pn denotations P to Pn denotations H(P). Usually H is *restricting*: $H(P) \leq P$, meaning that for all n-tuples of entities x, if $H(P)(x) = \text{True}$ then $P(x) = \text{True}$. *For Max* is restricting in (42a), so (42a) entails (43b).

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

So Prepositions denote functions **p** mapping entities to restricting functions H, and **ana** verbs quantify over such **p** as in (43), the **ana** form of a transitive verb. (39b) is derived as in (44) and compositionally interpreted as in (45), tense omitted.

(43) $\llbracket \text{ANA}(v) \rrbracket (x)(y)(z) = \text{True}$ iff $\exists \mathbf{p}(\mathbf{p}(z)(\llbracket v \rrbracket)(y)(x) = \text{True} \wedge \mathbf{p} \neq \theta_1, \theta_2)^{14}$

(44)

				S		
				S[DP _δ]		
			S[DP _{acc} , DP _δ]			
		S[DP _{poss} , DP _{acc} , DP _δ]				
	S[DP _{acc} , DP _δ]					
				DP _{poss}	DP _{acc}	DP _δ
		RT _{AG,TH}				
	aN	taov	ana	Rabe	ny farafara	Rasoa
(45)	AN	TAOV	ANA	r	f	s
		AN(TAOV)				
		ANA(AN(TAOV))				
		(ANA(AN(TAOV)))(r)				
		(ANA(AN(TAOV)))(r)(f)				
		(ANA(AN(TAOV)))(r)(f)(s)				

Crucially the semantic interpretation of the **ana** verb uses that of the **man** verb it takes as argument. And since **p(x)** is restricting, a nuclear **ana S** like (38a) entails the **man** (46) which lacks a PP corresponding to the DP_{δ} argument of the **ana** form.

(46) **man**+taov (manao) ny farafara Rabe
man+make the bed Rabe
Rabe is building the bed

The last line of (45) 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)(b)(r)**, the interpretation of *Rabe is making the bed*. Thus our semantics for **ana Ss** entails that Rabe is Agent of *make* and the bed is Theme. Consider now the predictions of theta equivalence between an **ana S** and a corresponding **man** one with an overt oblique, as in (38a,b).

On our semantics (38a) entails (38b), but the converse fails. (38b) is true in a model in which Rasoa's nagging prompted Rabe to build a bed for sale not for her. Our analysis also shows that the possessor complement *Rabe* in (38b), has the same theta role, Agent, as the P1 argument in (38a), and *ny farafara* 'the bed' is Theme in both Ss. But while *Rasoa* is a Benefactive in (38a), forced by the choice of Prep *ho an-* 'for', *Rasoa* is not assigned a theta role in (38b). So (38a,b) are not fully theta (or logically) equivalent. Our semantics just quantifies over functions from entities to restricting functions, different theta roles for Rasoa correspond to different choices of function. Thus on our analysis an **ana S** is *vague*, not ambiguous, regarding the theta role of its DP_{δ} . We just know that it bears some non-Agent, non-Theme relation to the predicate denotation. The case is similar to *the bridge where the spies meet*, which is unspecific but not ambiguous according as the spies meet under, on, or next to the bridge.

The syntax and semantics of **ana** verbs has proven the most challenging for all approaches to voice affixes in Malagasy (though causative *mamp-*, also valency increasing, is not without problems). We will see them in best relief in relative clauses, 4 below, so here let us just summarize what we are committed to so far:

- (47) a. Semantically an **ana-P1** maps DP_{δ} denotations to truth values, assigning them a novel oblique relation: Place, Time, Means/Instrument, Cause/Reason, Manner, Benefactee, Location, Purpose, Price, or Point of view [RR:112-3], and
 b. Syntactically, oblique relations are denoted by Prepositions or locative deictics, but like all P1s, **ana-P1s** select a DP_{δ} not a PP or a locative as last argument.

(47) prompts, I claim, a surface pragmatic principle, which helps avoid excessive vagueness when forced to assign a theta role to an overt DP.

Oblique Choice Constraint (OCC)

Given a P0 consisting of an **ana-P1** + an overt DP_{δ} d
 interpret d as *Means*, *Benefactee*, or *Location*

The OCC is not a semantic principle. **ana-P1s** quantify over relations other than Means,

Benefactee, and Location. They are used in (48c,d) whose **ana**-P1s lack overt DP_δs. Only (48b) with an overt DP_δ is uninterpretable, as happiness isn't a location, benefactee, or instrument/means. We return to these cases after discussing relative clauses.

- (48) a. miarahaba anao amin-kafaliana lehibe izahay
mi+greet 2s.acc with-happiness big 1.excl
We greet you with great happiness
- b. #iarahaba+ana+nay (iarahabanay) anao ny hafaliana lehibe
greet+**ana**+1.excl.poss 2s.acc the happiness big
- c. amin-kafaliana lehibe no iarahaba+ana+nay (iarahabanay) anao
with-happiness big Foc greet+**ana**+1.excl.poss 2s.acc
It is with great happiness that we greet you
- d. ny hafaliana lehibe (izay) iarahaba+ana+nay anao
the happiness big (which) greet+**ana**+1.excl.poss 2s.acc
the great happiness with which we greet you

4. Relative Clauses (RCs) and “Subjects only”

The derivation and interpretation of RCs in Malagasy has been assumed modeled on the *wh*-movement approach used for English and other languages.

- (49) a. manasa (man+sasa) lamba amin'io savony io ny tovolohy
washes (**man**+wash) clothes with'that soap that the young.man
- b. (ny) tovolahy (izay) manasa lamba amin'io savony io
(the) young.man (that) washes clothes with'that soap that
- c. (ny) tovolahy [wh_i [_{P0} [_{P1} manasa lamba amin'io savony io][t_i]].
(the) young.man who washes clothes with'that soap that t
- d. $b \in \llbracket (wh_i)\phi \rrbracket(B)$ iff $b \in B$ & $\llbracket \phi \rrbracket[t_i/b] = \text{True}$

(49d) says that for ϕ a PO of any complexity, $\llbracket (wh_i)\phi \rrbracket$ maps a set B – in (49c) the set of young men – to the set of those b in B that ϕ is true of when t_i is set to denote b. The semantics in (49d) is compositional and extensionally correct. That ϕ be arbitrary is important. For then if the position of t_i in ϕ violates Subjacency, Coordinate Structure, etc. $(wh_i)\phi$ is still interpreted, and ungrammaticality is due to syntactic constraints.

- (50) a. *ny lamba (izay_i) manasa t_i amin'io savony io ny tovolahy
the clothes (that) **man**+wash t with'that soap that the young.man
- a'. ny lamba (izay_i) sasan'ny tovolahy amin'io savony io t_i
the clothes (that) wash+**ina**'the young.man.poss with that soap that

something, so again autonomous movement of houses is implied.

Second, a reader queries whether we can interpret reflexives correctly in POs built from bound verbs since it would appear that the anaphor asymmetrically c-commands its antecedent. And it does. One of the main empirical points of K&S is to show how binding among arguments works independently of how the AA (Anaphor-Antecedent) relation is coded. A full discussion of binding in Malagasy would be a paper in its own right. So here we just adapt K&S to show how the c-command issue disappears.

Define a function **self** which maps binary relations R to the set of objects b such that the pair (b,b) ∈ R. Now let r be a RT_{AG,TH} in Dom(**man**) ∩ Dom(**ina**). Then **man**(r) and **ina**(r) are both two place functions taking individuals as arguments. We extend their domains so that **man**(r) takes **self** as argument yielding the one place function of individuals given in (52a). **Ina**(r) still takes individuals b as arguments to yield functions taking **self** type things as arguments, as in (52b)¹⁷:

$$(52) \text{ a. } (\mathbf{man}(r)(\mathbf{self}))(b) = \mathbf{man}(r)(b)(b) \quad \text{and} \quad b. (\mathbf{ina}(r)(b))(\mathbf{self}) = \mathbf{ina}(r)(b)(b)$$

That **man**(r)(b)(b) and **ina**(r)(b)(b) are the same truth value follows from Proposition 2. One sees easily that (52a,b) can provide compositional interpretations. The tough case in terms of c-command is:

$$(53) \quad \begin{array}{ll} [\text{novonoin-dRabe (no+vono+ina+Rabe)}] & \text{ny tena+ny} \\ \text{past+kill+ina+Rabe.poss} & \text{the self+3.poss} \\ \text{Rabe killed himself} & \end{array}$$

Ignoring tense and treating *ny tenany* as denoting **self** the interpretation of (53) follows the formula in (52b): ((**ina**(kill))(Rabe))(**self**).

Third, there remains a learnability issue concerning the interpretation of **ana**-P1s in RCs. Semantically they select an oblique relation on their argument. But the OCC in practice rules out overt arguments interpreted as Manner, Cause/Reason, etc., only allowing Means, Benefactee, and Location. So what language exposure prompts the learner to assign the Manner meaning to (54a) and the Cause/Reason meaning to (54b) given that DPs with those meanings don't occur as overt arguments of the **ana**-P1s?

$$(54) \text{ a. } \begin{array}{ll} \text{ny hafaliana lehibe niarahabanay (n+mi+arahaba+ana+nay)} & \text{anao} \\ \text{the happiness big} & \text{past+mi+greet+ana+1.excl.poss 2s.acc} \\ \text{the great happiness with which we greeted you} & \end{array}$$

$$\text{ b. } \begin{array}{ll} \text{ny antony nanaovako (n+man+taov+ana+ko)} & \text{izany} \\ \text{the reason} & \text{past+man+do+ana+1s.poss that} \\ \text{the reason I did that} & \end{array}$$

And relatedly, what drives the OCC? Specifically, why select Means, Benefactee and Location? Selecting just Manner and Cause/Reason would also reduce vagueness.

Two questions, two responses: First, there are other contexts that select P1s and in which **ana**-P1s occur with an explicit indicator of their oblique relation. Second, there are independent constraints on P1 arguments that disfavor certain oblique interpretations.

One other P1 selecting context is given by the very widely used Focus construction, illustrated in (55) and studied in Paul 2001.

- (55) a. Rabe no [nanenjika ny jiolahy tamin'ny fiara]_{P1}
 Rabe Foc pst+**man**+chase the thief pst+Means'the car
It was Rabe who chased the thief using the car
- b. Ny jiolahy no [nenjehin-dRabe tamin'ny fiara]_{P1}
 the thief Foc pst+chase+**ina**-Rabe.poss pst+Means'the car
It was the thief that Rabe chased using the car
- c. (Tamin') ny fiara no [nanenjehan-dRabe ny jiolahy]_{P1}
 (pst+Means'the car Foc pst+**man**chase+**ana**-Rabe.poss the thief
It was by means of the car that Rabe chased the thief
- d. Tamin'ny fahazotoam-be no [nanenjehan-dRabe ny jiolahy]_{P1}
 Pst+with'the zeal- big Foc pst+**man**+chase+**ana**-Rabe.poss the thief
It was with great zeal that Rabe chased the thief

In all the examples the post *no* constituent is a P1 whose argument is semantically bound by the focus (pre-*no*) constituent which makes the oblique relation explicit. In the first three the pre-*no* constituent could occur as an overt post-P1 argument (less the Prep in (55c)). But this is not possible in (55d), where the Prep is obligatory. The post-*no* P1 in (55d) is identical to that in (55c). Here are some other examples with other Preps.

- (56) a. Momba ny fifidianana no niresahanay (n+mi+resaka+ana+nay)
 about the elections Foc pst+**mi**+conversation+**ana**+1.excl.poss
It was about the elections that we were speaking
- b. Noho ny andro alina no odiako (∅+ody+ana+ko) haingana
 cause the day night Foc 0+go.home+**ana**+1s.poss quickly
It is because it is dark that I'm going home quickly
- c. Mandritra taona iray no hitoeranay (h+mi+toetra+ana+nay) any Betafo
 during year one Foc fut+**mi**+reside+**ana**+1.excl.poss loc Betafo
We will reside in Betafo for one year

Also noteworthy here is that focusing does not require definiteness (Paul 2001), whence the naturalness of forming DP_δ questions from the cleft structure (Sabel 2003).

(57) a. vato no hanaovako (h+man+taov+ana+ko) ny trano+ko (Malzac:71)
 stones Foc fut+man+do+ana+1s.poss the house+1s.poss
It is with stones that I will build my house

a'. Inona no hanaova+nao ny trano+nao?
 what Foc fut+man+do+ana+2s.poss the house+2s.poss
What will you build your house from?

b. lobaka no sasan-dRasoana (sasa+ina+Rasoana) (fa tsy pataloaha)
 shirts Foc wash+ina+Rasoana.poss but not pants
It is shirts that Rasoana is washing (not pants)

In general then what follows the focus particle *no* in a cleft is a P1¹⁸. And this is a basic paradigm which leads language learners to interpret **ana**-P1s as P1s which impose an oblique interpretation, even when they don't take a corresponding overt DP sister. And once we can interpret **ana**-P1s as P1s we can use them in RCs:

(58) a. ny vehivavy anaovako (man+taov+ana+ko) trano
 the woman **man**+do+**ana** 1s.poss house
the woman for whom I am building a house

b. io alina nodiako (n+∅+ody+ana+ko) haingana io
 that night past+0+go.home+**ana**+1s.poss quickly that
that evening when I went home quickly

c. ny vato hanaovako (h+man+taov+ana+ko) trano
 the stone fut+**man**+do+**ana**+1s.poss house
the stone from which I will make my house

Secondly, why does the OCC select Means, Benefactive and Locative obliques as the unmarked interpretations for DP arguments of **ana**-P1s? This is, I feel, due to an independent constraint on P1 arguments. Alas I don't know precisely what the constraint is, but it is tied to the well recognized "definiteness" requirement on P1 arguments in Malagasy (Tagalog, ...). Contrast:

(59) a. manenjika (ny) voalavo (*ny) akoho
 chases (the) rat the chicken
The chicken is chasing rats, a rat, the rat(s)

b. misy akoho manenjika voalavo
 exist chicken chases rat
There is a chicken / (some) chickens which is / are chasing (some) rats

The Theme DP in (59a) may be a bare N in which case it has the expected indefinite interpretation. (Number is not marked on Ns). But the P1 argument *akoho* 'chicken' must have a definite Det. Simply omitting its *ny* 'the' is roundly rejected by everyone.

The intended sense would be expressed as an existential, as in (59b). This basic paradigm is well known. But I think it extends a little further and deeper.

One component of definiteness is *referent identifiability*. And instruments (shovels, knives, needles), Benefactees (human beings), and Locatives (houses, villages, rivers) are physical objects and thus visually identifiable. But the other oblique relations mentioned in (47) – Times, Causes, Manners, Purposes, Points of View, Prices are not visually identifiable. So the OCC restriction makes sense in terms of referent identifiability, rendering it less ad hoc. Perhaps this perspective can be extended to “classifier Ns” such as *antony* ‘reason’, etc. in (60):

(60) ny antony / fotoana / fomba nandehanany (n+man+leha+ana+ny)
 the reason / time/ manner pst+**man**+go+**ana**+3.poss
the reason / time he went; the manner in which he went

Such Ns denote oblique properties the argument of an **ana**-P1 might bear and the whole DP denotes whatever has that property. As soon as we hear (in Malagasy) a DP that begins ‘the reason that...’, ‘the time that...’ we know that there will be a subordinate clause verb in **ana** form. The cases are reminiscent of ones from Schachter 1973 discussed in Bhatt 2002 in which the RC must be present to identify the head N:

(61) a. the clever way (*in which he solved the problem) impressed me

b. Nahagaga ahy ny fomba (*nanaovany (n+man+taov+ana+ny) ny trano+ny)
 surprised 1s.acc the manner pst+**man**+do+**ana**+3.poss the house+his
*The manner [*in which he built his house] surprised me*

4.2 Two syntactic challenges to RC = P1

Second, there are at least two common RC types that appear not to fit our format. One is possessor RCs, as in *the student whose cat I fed*. In fact Malagasy can relativize a variety of possessors easily, due to its ability to form P1s whose DP sister binds a possessor position within the P1. K&R [Keenan and Ralalaoherivony 2000] provide extensive documentation. An illustrative example is (62), built from the P1 in (63b).

(62) ny mpianatra [hadinoko anarana]_{P1}
 the student forget+1s.poss name
the student whose name I forget

(63) a. [Hadinoko]_{P1} [ny anaran’ny mpianatra]
 forget+1s.poss the name’the student.poss
I forget the name of the student

b. [Hadinoko anarana]_{P1} [ny mpianatra]
 forget+1s.poss name the student
“The student is name forgotten by me”

- c. [manadino ny anaran'ny mpianatra]_{P1} aho
man+forget the name'the student.poss 1s.δ
I forget the name of the student

Hadino is a bound root; the free form is *manadino*, illustrated in (63c). It is tempting, with K&R, to think that (62b) is derived by “DP raising” (which includes possessor raising as a special case). But this approach does not clearly guarantee that what is left after DP Raising is a P1. But they are: interrogative *ve* placement is correctly predicted, the following DP is in δ case, and they coordinate with other clear P1s:

- (64) a. [Be sandry sady mihetraketraka] Rabe [K&R:5]
 big arms and **mi**+look+for+trouble Rabe
Rabe has big arms and looks for trouble
- b. [Very hevitra ka miherikerika fotsiny] Ranaivo
 lost thought and+so **mi**+look+around only Ranaivo
Ranaivo is confused and so just looks around
- c. [Miasa loha sady mifoka loatra] i Soa
 works head and **mi**+smoke too+much art Soa
Soa is preoccupied and smokes too much

If the left conjuncts are P1s we can coordinate them with the P1s on the right. But how can we derive (65c) for example if *miasa loha* is only created in contexts like (65b)?

- (65) a. ??Miasa ny lohan'i Soa “Soa’s head is working”
 work the head’art Soa.poss
- b. Miasa loha i Soa “Soa is preoccupied”
 work head art Soa
- c. Miasa loha ve i Soa? “Is Soa preoccupied?”

And if we derived (65c) from the polar question based on (65a) the question particle *ve* would immediately follow *Miasa*, ungrammatical in (65c). Without answering this query, it is constructive to consider how P1s are formed rather than how DPs are moved. For example, the P1s in (63b) and (64a,b,c) require an argument satisfying a requirement internal to an N they are built from. Even more surprising cases cited in K&R are:

- (66) a. Maro [ny raharaha sahaniko (sahana+ina+ko)]
 many the tasks I+face confront+**ina**+1s.poss
The tasks I am confronting are numerous
- b. [Maro raharaha sahanina]_{P1} aho
 many tasks confront+**ina** 1s.δ
I have many things to do

(67) a. Fohy [ny andro niveloman-dRabe (n+i+velom+ana+Rabe)]
 short the day past+mi+live+ana+Rabe.poss
The days during which Rabe lived were few

b. [Fohy andro nivelomana]_{P1} Rabe
 short day lived+ana Rabe
Rabe had a short life

Deriving (66b) and (67b) by DP movement violates standard constraints since it is movement from a relative clause. Important here is that the bracketed portions of (66b) and (67b) satisfy our P1 tests, including coordination above. An even more striking case of P1s formed with NP fragments is (68b). (68c) illustrates the use of the P1 as an RC.

(68) a. Be [ny asa manahirana ahy] “*The work which bothers me is extensive*”
 much the work man+bother 1.acc

b. [Be asa manahirana] aho “*I (have) much bothersome work*”
 much work bothers 1s.δ

c. ny [mpampianatra [be asa manahirana]]
 the teachers (who) (have) much bothersome work

So Malagasy has a variety of ways of forming P1s who DP sisters bind an argument properly within the P1. Still in simple cases stage-level adjectives do not form P1s with possessor heads freeing the possessor to be the argument of the derived predicate. Compare the natural (69b) with the unacceptable (70b).

(69) a. Fotsy [ny nifin-dRabe]
 white the teeth-Rabe.poss
Rabe’s teeth are white

b. [Fotsy nify] Rabe
 white teeth Rabe
Rabe is white toothed

(70) a. Tezitra [ny mpiasan-dRabe]
 angry the workers-Rabe.poss
Rabe’s workers are angry

b. *[Tezi-mpiasa] Rabe
 angry-worker Rabe
 *?*Rabe is angry workered*

This observation supports a P1 derivational view – it is a constraint on P1 interpretations, not on which DPs can raise. Thus (71a) below is natural, (71b) much less so.

(71) a. ny tovolahy fotsy nify
 the young.man white teeth
the white toothed young man

b. *ny tale tezi-mpiasa
 the boss angry worker
the boss with angry workers

To express modifiers as in (71b) Malagasy can always use an overt verb of possession. The simplest choice is *misy* ‘exist, have’, as in *kafe misy siramamy* ‘coffee with sugar’ (lit: ‘coffee (that) has sugar’). This option works for (71b) as well: *ny tale misy mpiasa tezitra* ‘the boss with angry workers’. With the possession of inanimates it would be

more natural to use *manana* ‘possess’. So ‘doctor whose house is large’ = *dokotera manana trano lehibe* ‘lit: doctor (who) has house large’.

The second potentially problematic case is relativization into complement clauses. How do we say *the thief who Rabe thinks that Rasoa captured*? Surprisingly Malagasy has two regular answers to this query, and both maintain **RC = P1**. The first is:

(72) ny jiolahy [heverin-dRabe ho nosamborin-dRasoa]
 the thief think+**ina**+Rabe.poss Comp past+capture+**ina**+Rasoa.poss
the thief Rabe thinks that Rasoa captured

The Ss below may help the reader to parse the bracketed P1 in (72). Note that when the matrix verb ‘think’ is free the complement clause, introduced in (73a) by *fa*, follows its DP_δ. All the Ss in (73) have the same cognitive meaning. The use of poss, δ, and acc indicates the class of pronoun that would replace the full DPs. We replaced ‘the thief’ with the appropriate pronoun in (73b,c,d) to help readability. The Raising to Object relation between (73b) and (73c) is very productive in Malagasy, Paul and Rabaovolona 1998 cite some 50 verbs which host it).

- (73) a. mihevitra Rabe fa nisambotra ny jiolahy Rasoa
mi+think Rabe.δ that past+**mi**+capture the thief.acc Rasoa.δ
Rabe thinks that Rasoa captured the thief
- b. mihevitra Rabe fa nosamborin-dRasoa izy
mi+think Rabe.δ that past+capture+**ina**+Rasoa.poss he.δ
- c. [mihevitra azy ho nosamborin-dRasoa]_{P1} Rabe
mi+think him.acc Comp past+capture+**ina**+Rasoa.poss Rabe.δ
- d. [heverin-dRabe ho nosamborin-dRasoa]_{P1} izy
 think+**ina**-Rabe.poss Comp past+capture+**ina**+Rasoa.poss he.δ

The bracketed constituents in (73c,d) unproblematically pass our P1 tests – interrogative *ve* occurs at the right edge, and the following DP is in δ case.

A second way of relativizing into a complement clause, adapted from [K&R] in (74), appears less convincingly a case of RC = P1. Many speakers hesitate slightly at this expression, but then accept it.

(74) ny tovolahy [nolazain-dRabe fa mpangalatra]
 the young.man past+say+INA+Rabe.poss Comp thief (er+steal)
the young man (that) Rabe said was a thief

If (74) is to fit our RC = P1 format the bracketed string in (74) must be a P1. It looks like the P1 in (d) save that it uses the complementizer *fa* instead of *ho*. But that difference signals a significant difference in structure. Consider the S in (75a). As both K&R and Pearson 2005 note, two placements of the question particle are acceptable:

- (75) a. Nolzain-dRabe fa mpangalatra ny tovolahy
 Past+say+**ina**+Rabe.poss that thief the young.man
Rabe said that the young man was a thief
- b. [Nolzain-dRabe]_{P1} ve fa mpangalatra ny tovolahy?
 Past+say+**ina**+Rabe Q that thief the young.man
Did Rabe say that the young man was a thief?
- c. [Nolzain-dRabe fa mpangalatra]_{P1} ve ny tovolahy?
 Past+say+**ina**+Rabe that thief Q the young.man
Was the young man said by Rabe to be a thief?

Ve placement marks the bracketed string as a P1 in (75c). And (75a) is ambiguous according as ‘the young man’ is the DP_θ argument of the lower P1 or the entire P1.

5. Conclusion This concludes our discussion of the role of P1s in RC formation. We note only that in addition to RCs there are several other major structure types that build off of P1s: We have already hinted at their role in Clefts, which includes *wh*-questions as a special case. Another very large class of DPs can be described by the equation: **DP = Det + P1 + (Det)** illustrated in (76). See Ntelitheos (2006) for extensive discussion.

- (76) Faly amin’ [izao [h+_o+ananan’i Soa tokantrano]_{P1} izao]_{DP} izahay
 happy with this fut+0+have+**ana**’art Soa.poss household this we.δ.excl.
We are happy about Soa’s future having a household (getting married)

W. Austronesian languages are good at building predicates. Moving DPs is derivative.

References

- Abinal, R.P. et R.P.. Malzac (1963) Dictionnaire Malgache-Français. Mission Catholique de Tananarive. Paris: Editions Maritimes et d’Outre-Mer (reprint from 1888 edition).
- Aldridge, Edith. 2002. Nominalization and WH-movement in Sediq and Tagalog. *Language and Linguistics* 3.2 Special Issue: Nominalization in Formosan Languages, Elizabeth Zeitoun (ed). Academia Sinica. Pp. 393 – 427.
- Andrianierenana, L. 1996. Morphological causatives in Malagasy. In *The Structure of Malagasy* M Pearson and I. Paul (eds). UCLA Occasional Papers in Linguistics 17:58 – 76.
- Bhatt, Rajesh. 2002. The raising analysis of relative clauses: evidence from adjectival modification. *Natural Language Semantics* 10:43 – 90.
- Builles, Jean-Michel. 1984. *Le Merina* Thèse pour le doctorat d’état. Paris V.
- Chomsky, Noam. 1957. *Syntactic Structures*. Mouton.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. MIT Press.
- Chomsky, Noam. 1975. *Reflections on Language*. Pantheon.
- Chomsky, Noam. 1995. *The Minimalist Program*. The MIT Press.
- Collins, Chris. 2005. A smuggling approach to the passive in English. *Syntax* 8.2:81 – 120.

- Corbet, Greville. 1991. *Gender*. CUP.
- 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.
- [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.
- Haspelmath, M. 1990. The grammaticization of passive morphology. In *Studies in Language* 14.1:25 – 72.
- Huang, Lillian. 2002. Nominalization in Mayrinax Atayal. *Language and Linguistics* 3.2 Special Issue: Nominalization in Formosan Languages, Elizabeth Zeitoun (ed). Academia Sinica. pp 197 – 227.
- Hyams, Nina, Dimitris Ntelitheos, and Cecile manorohanta. 2006. Acquisition of the Malagasy Voicing System: implications for the adult grammar. *NLLT* 24:1049 – 1092.
- 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. Gerds (eds). CSLI. Stanford. pp.171 – 217
- Keenan, Edward L. 1996. The semantics of determiners. In *The Handbook of Contemporary Semantic Theory* S. Lappin (ed). Blackwell pp. 41–53.
- [K&M 2001] Keenan, Edward L. and Cécile manorohanta. 2001. A quantitative study of voice in Malagasy. in *Oceanic Linguistics* vol 40:67 – 85.
- [K&M 2004] Keenan, Edward L. and Cécile manorohanta. 2004. Malagasy clause structure and language acquisition. In *Proceedings of the XIth Austronesian Formal Linguistics Association*. Paul Law editor. ZAS, Berlin. Pp. 178 – 203.
- [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, E.L. and Baholisoa Ralalaoherivony. 2000. Raising from NP in Malagasy. *Linvisticae Investigationes* 23:1:1 – 44.
- Keenan, Edward L. and Jean-Paulin Razafimamonjy. 2004. Reciprocals in Malagasy. in *Oceanic Linguistics* vol. 43.1:177 – 207.
- [K&S] Keenan, E.L. and E.P. Stabler. 2003. *Bare Grammar* CSLI, Stanford.
- Koopman, H. 2005. Malagasy imperatives. In *UCLA Working Papers in Linguistics*:12.
- Kroeger, Paul. 1988. Verbal focus in Kimaragang. In *Papers in Western Austronesian Linguistics* No. 3:217 – 40. Pacific Linguistics, A-78.
- Kroeger, Paul. 1993. *Phrase Structure and Grammatical Relations in Tagalog*. CSLI, Stanford.
- Mahajan, Anoop. 1994. ACTIVE passives. WCCFL XIII. R. Aranovich et al. (Eds). pp. 286 – 301.
- Malzac, S.J. 1926. *Grammaire malgache* Paris, Société d'éditions géographiques, Maritimes et Coloniales, Paris.

- Ntelitheos, Dimitrios. 2005a. Variation in the distribution of genitive subjects in Malagasy. 35th Michigan Linguistic Society, Lansing.
- Ntelitheos, Dimitris. 2006. *The Morpho-Syntax of Nominalizations: A Case Study*. PhD diss. UCLA Dept of Linguistics.
- Paul, Ileana. 1996. The Malagasy genitive. In Pearson & Paul, *The Structure of Malagasy* Vol I, UCLA Occasional Papers in Linguistics 17. Pp 76 – 92.
- Paul, Ileana. 1999. *Malagasy Clause Structure*. PhD Diss. Dept. of Linguistics, McGill University
- Paul, Ileana. 2001. Concealed pseudo-clefts. *Lingua* 111:707 – 727.
- 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. 2005. The Malagasy subject/topic as an A' element. in *Natural Language and Linguistic Theory* 23:381 – 457.
- Phillips, V. 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, R-B. 1993. Le verbe malgache – Etude des formes predicatives. Ambozontany: Fianarantsoa, Madagascar.
- [R&T] Rackowski, Andrea and Lisa Travis. 2000. V-Initial Languages: X or XP movement and adverbial placement. In *The Syntax of Verb Initial Languages*. Andrew Carnie and Eithne Guilfoyle (eds) OUP Pp. 117 – 143.
- Radford, Andrew. 1997. *Syntax: A Minimalist Introduction*. CUP.
- Rahajarizafy, Antoine. 1960. *Essai sur la grammaire malgache* Imprimerie Catholique, Antanimena, Tananarive
- [RR] Rajemisa-Raolison, Regis. 1971. *Grammaire malgache*. Fianarantsoa.
- Randriamasimanana, Charles. 1986. *The Causatives of Malagasy* Univ. of Hawaii Press.
- Richards, N. 2000. Another look at Tagalog subjects. In I. Paul, V. Phillips, and L. Travis (eds). *Formal Issues in Austronesian Linguistics* Pp 105 – 116. Kluwer
- Ross, M. 2006. The argument structure of undergoer voice clauses in Formosan and other Philippine-type languages. Presented at AFLA 13, Tsing Hua Univ. Taiwan
- Sabel, Joachim. 2003. Malagasy as an optional multiple wh-fronting language. In Cedric Boeckx and Kleanthes Grohman (eds). *Multiple wh-Fronting* 229 – 254. Benjamins.
- Schachter, Paul. 1973. Focus and relativization. *Language* 49(1):19 – 46.
- Schachter, Paul. 1976. The subject in Philippine languages: topic, actor, actor-topic, or none of the above. in C. Li (ed) *Subject and Topic* 1976:491 – 518. Academic Press.
- Schachter, P. and Fe Otanes. 1972. *Tagalog Reference Grammar* UC Press.
- Wechsler, Stephen and I.Wayan Arka. 1998. Syntactic Ergativity in Balinese. In *Natural Language and Linguistic Theory* 16:387 – 441.

Footnotes

¹ In addition to Malagasy (Madagascar) we include here Tagalog (Schachter and Otanes 1972, Kroeger 1993) and Philippine languages generally, Timugon (Prentice 1971), Kimaragang Dusun (Kroeger 1988), Balinese (Wechsler and Arka 1998), Atayal (Huang, 2002) and Seediq (Aldridge 2002).

² Erwin (1996) and Pearson (2005) show that many roots are consonant final, in distinction to their standard dictionary form (Abinal & Malzac 1963). If a vowel is supplied via a suffix the final consonant, possibly mutated, is pronounced. With no suffix the consonant drops. For example, from the root *taov-* ‘do’ we derive *atao* ‘does’, with loss of *-v*, but imperative *atovy*. We are not concerned here with the morphophonology of affixation and tend to give roots in their dictionary form.

³ When *F* is a two place function on *A* and α is an element of *A*, we can define a one place function F_α from *A* to *A* which maps each β in *A* to $F(\alpha, \beta)$. Here we think of Merge as taking some pairs (*man*, *r*) as arguments, and we treat **man** as an abbreviation for the one place function $\text{Merge}_{\text{man}}$.

⁴ **man** maps pairs (*s*, *C*) to pairs (*s'*, *C'*). In our example it maps the ‘*s*’ coordinate *enjika* to *manenjika*, the concatenation of *man* with *enjika*, in the tree (10b). But often the derived string is not simply concatenation with *man*: the initial consonant of a root may drop or mutate: **man**(*tolotra*) = *manolotra*, **man**(*halatra*) = *mangalatra* ‘steals’, etc. We ignore these spell out problems here. This complexity is typical of W. Austronesian.

⁵ *P*Os denote elements of {True, False}. Ultimately we need proper intensions, but for now extensional simplicity rules.

⁶ In addition six **man** verbs, recognized as exceptional in traditional grammars, are in $\text{Dom}(\mathbf{ina})$: Ex: **man**(*halatra*) = *mangalatra* ‘steals’; **ina**(*mangalatra*) = *angalarina*.

⁷ Mostly with *-(v)na* suffixes, the choice of vowel is conditioned by the root. But Rahajarizafy notes a few exceptions. We call the affixing function **ina** after its most common exponent.

⁸ We intend the 20, listable, affixes introduced by PBFs. $\Theta(p)$ is $\{\theta_1, \dots, \theta_n\}$ when the category of *p* is $S[C_1, \theta_1, \dots, C_n, \theta_n]$ or $RT_{\{\theta_1, \dots, \theta_n\}}$.

⁹ We are not free to decide that two expressions *d*, *d'* are or are not isomorphic. Once the lexicon and the structure building functions are given then either there is an automorphism mapping *d* to *d'* or there isn’t. It is a matter of proof, not introspection.

¹⁰ Mathematically we may treat a single expression as unary relation with one member.

¹¹ suppose nuclear *Ss* Φ , Ψ are theta equivalent by *f*. Let *h* an arbitrary automorphism. Then in effect $h \circ f \circ h^{-1}$ bijectively maps the arguments of $h(\Phi)$ to those of $h(\Psi)$ preserving theta role in each case, whence $h(\Phi) \equiv_\theta h(\Psi)$

¹² *mi-* verbs: *mierika ny andro* ‘It is drizzling’, *mikotraka ny orana* ‘The rain thunders’; adjectives: *ratsy/tsara/mafana ny andro* ‘the weather is bad/good/hot’; one free root verb: *ho avy ny orana* ‘the rain will come’; even nouns: *alina ny andro* ‘It is night’.

¹³ We choose freedom not *m* marking as definitional in the domain of **ana** since there are two *m* prefixes – *miha* ‘become’ and *miaraka* ‘do together’ which occur outside tense marking and do not feed **ana**:

mihamangatsika aho
I am getting cold

nihanangatsiaka aho
I was getting cold

niara-nangalatra *ny omby izy roa lahy*
past+together-past+man+steal the cow 3 two man
The two men together stole the cow

¹⁴ $\mathbf{p} \neq \theta i$ means $\mathbf{p}(z)[\mathbf{v}] \neq \theta i(z, [\mathbf{v}])$

¹⁵ The *P*1 of a *RC* may be preceded by *izay* (morphologically constant) interpreted as:

$IZAY(P)(B) = B \cap \{x|P(x) = \text{True}\}$. B is a Noun denotation, P a P1 one. Thus RCs with or without *izay* are both interpret the subordinate clause as a P1. And the denotation of $N+(izay)+P1$ is the same whether *izay* is present or not (it usually is not).

¹⁶ “any structure formed by the computation ... is constituted of elements already present in the lexical items...; no new objects are added in the course of computation apart from rearrangements of lexical properties...” (Chomsky 1995:228).

¹⁷ The only fact about **self** that we use here is that it is not a generalized quantifier which lifts a function from properties to truth values.

¹⁸ There is one well known exception. Locative and temporal adjuncts can cleft in three ways, one of which leaves the post-*no* expression an apparent complete S. Thus *It is in Betafo that Rabe resides* can be rendered as (a), (b), or (c):

- a. Any Betafo no [itoeran-dRabe.poss]_{ana-P1}
- b. Any Betafo Rabe. δ no [mitoetra]_{mi-P1}
- c. Any Betafo no [mitoetra Rabe. δ]_{p0}

What follows *no* in (a) and (b) is a P1, but in (c) it is a **mi** verb with a DP_{δ} . Still, despite this one exception the post-*no* position in clefts selects for P1s, and all clear P1s are able to occur there.