AN RI STAGE IN MALAGASY? IMPLICATIONS FOR
THE ADULT GRAMMAR

NINA HYAMS

1. Introduction

Malagasy is an Austronesian language spoken on the island of Madagascar and is closely related to Philippine languages such as Tagalog. One of the most notable aspects of Malagasy grammar, which it shares with other Austronesian languages, is an elaborate voicing system. The voicing system has a distinctive morphology and it involves the promotion of an argument (actor, theme, instrument, etc.) to a referentially and syntactically prominent position, typically clause final position. Following Pearson (2001), and Schachter (1987)'s analysis of Tagalog, I refer to this prominent DP as the trigger and the promotion operation as promotion to trigger (PTT). The voice morphology on the verb identifies the grammatical function of the trigger, whether actor, theme, instrument, location, etc.

In this paper I discuss aspects of the acquisition of the Malagasy voicing system based on a longitudinal study of 3 children aged 19 to 32 months. The question I address is whether there is evidence of a stage in Malagasy corresponding to the root infinitive (RI) stage found in many other child languages. Our findings have implications for acquisition theory: Are there universal stages in grammar development? And they also bear on an important debate in the Malagasy syntax literature: Is the trigger a subject (Guilfoyle, Hung & Travis 1992) or a clause-external topic (Pearson 2001, 2005) and the related question of whether PTT involves A or A'-movement. To date, the latter issues have been looked at solely from the point of view of adult Malagasy. In this paper we examine them from a developmental perspective.
2. The Malagasy voicing system

The grammatical function of the trigger is encoded by voice morphology on the verb, as illustrated in (1) and (2) for the verb root *vidi* ‘buy’. In (1) the verb is *mividy*, the form that is used when the trigger is the actor (external argument). Following GHT and Pearson, we refer to this form as the AT (actor-trigger) form; (2) illustrates the TT (theme-trigger) form of the verb, *novedin*, i.e. the form that occurs when the trigger is the theme (internal argument).²

(1) Mividy bohy hoan’ny mpianatra ny mpampianatra.
AT.Pfx.buy books for’ Det students Det teacher
‘The teacher bought books for the students.’

(2) Novedin’ ny mpampianatra hoan’ny mpianatra ny boky.
PST.buy.TT.LNK’Det teacher for’ Det students Det books
‘The books, the teacher bought (them) for the students.’
or
‘The books were bought by the teacher for the students.’

AT morphology involves the prefix *m*-; TT morphology the suffix *-in*. In addition to *m*-, the AT form contains another verbal prefix realized as *an*- or *i*-. In coding the acquisition data we treated this prefix as a voice morpheme. There are also some verbs that occur in a root form, that is to say, without voice or tense morphology, as in (3). These are mainly TT roots.

(3) Haino-ko izy.
listen-1SG.GEN 3SG.NOM
‘As for him, I listen(ed) to (him).’

In addition to voice morphology, the Malagasy verb is inflected for tense. All the tense morphemes are prefixes. Malagasy has 3 tenses: present, past, future. Present tense is unmarked. The past tense morpheme is *n/-mo* and the future/irrealis marker is *h/ho*. However, the future/irrealis morpheme is pronounced only in very careful speech.

2.1 Competing Analyses of the Malagasy Voice System

The traditional view of the PTT process is that it involves A-movement, either passivization (e.g. Keenan 1976), or raising to subject (GHT 1992). On this view, the trigger is a subject in Spec IP. A more recent proposal is that PTT involves topicalization or A'-movement (Pearson 2001, 2005, cf. also Richards
2000), in which the trigger is a right peripheral topic. These competing analyses are reflected in the two sets of translations for the sentences in (1) and (2).

GHT propose that in Malagasy both Spec VP and Spec IP are subject positions. The actor is generated in Spec VP and the theme internal to VP. Either the actor or theme (or an oblique argument, cf. note 2) raises to Spec IP and maps onto the trigger function. As illustrated in the structure in (4), the verb adjoins to INFL, winding up at the left edge of the clause immediately preceding the actor. According to GHT, the voice affixes are case assigners that license all but one of the verb's dependents inside VP. The remaining argument raises to Spec IP to check nominative case.

(4) \[ \text{IP} \]
   \[ \text{\textit{trigger}} \]
   \[ \text{\textit{I}} \]
   \[ \text{\textit{VP}} \]
   \[ \text{\textit{I}} \]
   \[ \text{\textit{V}} \textsubscript{1} \text{\textit{(actor)}} \]
   \[ \text{\textit{V'}} \]
   \[ \text{\textit{t}} \textsubscript{i} \text{\textit{(theme)}} \]

Pearson (2005) proposes instead that the trigger is a topic located in the C-domain. The trigger is base-generated in Spec TopP where it is licensed through coindexation with a null operator \((O_p)\) that raises to the specifier of WhP (also in the C domain) from an argument position inside VP, either actor, theme (or oblique) position. Thus under Pearson's analysis, PTT involves A' movement. A simplified version of his structure is as in (5).

(5) \[ \text{TopP} \]
   \[ \text{\textit{Top}} \textsubscript{i} \text{\textit{trigger}} \]
   \[ \text{\textit{Top}} \]
   \[ \text{\textit{WhP}} \]
   \[ \text{\textit{Wh'}} \]
   \[ \text{\textit{O_p}} \]
   \[ \text{\textit{Wh}} \]
   \[ \text{\textit{TP}} \]
   \[ \text{T} \]
   \[ \text{\textit{VP}} \]
Like GHT, Pearson assumes that the base order is SVO, and he also takes voice morphology to be case related. He proposes that the voice morphemes spell out the case of the position from which the operator has moved: AT morphology is realized on the verb when the operator raises from the nominative case position to Spec WhP and TT morphology when it raises from the accusative case position.

Pearson proposes that Malagasy is a V2-like language in that (virtually) every clause contains a predicate external A'-position that must be filled. The trigger, like the Germanic topic, must be definite and is associated with 'aboutness'. An interesting property that Malagasy shares with the V2 languages is an optional rule of trigger drop, analogous to topic drop. In informal registers in the Germanic languages a discourse salient pronoun may be omitted when it occurs in topic position (and not otherwise), as illustrated in the German sentences in (6).

(6) a. (Ich) hab' *(ihn) schon gesehen.  b. (Ihn) hab' *(ich) schon gesehen.
   (I) have (him) already seen            (him) have (I) already seen
   'I already saw him'                    'Him, I already saw'

Pearson notes that a comparable pattern exists in Malagasy. In informal conversation a referential pronoun can be dropped, but only if it occurs in trigger position, as illustrated in (7).

(7) a. Mamangy an'i Tenda (izy)      c. Vangian i Naivo (izy)
   AT.visit obj-Det Tenda (3rd per.)  TT.visit Det Naivo (3rd per.)
   'He is visiting Tenda.'            (Him) Naivo is visiting.'

   b. Mamangy *(azy) i Naivo.  d. Vangian' *(ny) i Tenda.
   AT.visit 3rd per. Det Naivo        TT.visit (3rd per.) Det Tenda
   'Naivo is visiting (him).'          'Tenda, (he) is visiting.'

GHT do not discuss trigger drop. However, because their analysis places the overt trigger in Spec IP (cf. 4), the omitted trigger would on their account presumably have to be analyzed as a null subject similar to the null subjects found in Romance and other pro drop languages.

The trigger drop phenomenon will be central to our analysis of the acquisition results presented in the following sections.
3. A Developmental Perspective

Various properties distinguish the acquisition of V2/topic drop languages from null subject languages. Perhaps the most significant difference in the developmental trajectory is that children acquiring V2/topic drop languages show a root infinitive (RI) stage, while children acquiring null subject languages do not (Sano and Hyams 1994; Rhee and Wexler 1995). Some examples of RIs from the Germanic languages are provided in (8).

(8)

A
a. Papa schoenen wassen.  d. Ook koek hebben. (Dutch)
daddy shoes wash-INF also biscuit have-INF
b. Auf Teddy fenster gucken.  e. Wasser holen. (German)
Teddy window look-INF water get-INF
c. Jag också hoppa där å där. f. Bygge tåg. (Swedish)
I also hop-INF there and there build-INF train

In languages such as Italian, Spanish, Catalan and Portuguese, we find few examples of the sort in (8) (Guasti 1993/1994; Berger-Morales, Salustri and Gilkerson, 2005; Santos p.c.). The rate of RIs in these languages is typically under 10%, while in the V2 languages, RIs comprise between 40% and 50% of verbal utterances (Dutch: Wijnen 1997; Blom 2003; German: Poeppel & Wexler 1993; Becker and Hyams 2000; Swedish: Platzack 1996; Santelmann 1995; Josefsson 2002; Norwegian: Plunkett & Strömqvist 1990). The generalization that null subject languages do not show an RI stage is not based solely on the Romance languages. Children acquiring null subject languages as typologically diverse as Hungarian and Slovenian also fail to show an RI stage. The percentage of RIs in these two languages is under 2% (Londe 2004; Rus and Chandra 2005).

The distribution of RIs differs from that of finite verbs in child language. Particularly relevant to the present discussion is the fact that RIs typically occur with null subjects. The RIs in (8A) contain overt subjects, but much more common are examples such as those in (8B). Table 1 reports the rate of subject omission in finite clauses vs. RIs in several of the V2 languages. Across children and languages, subject omission in finite clauses (i.e. null topics) averages around 23% while the average rate of subject omission in RIs is around 70%.6
Table 1: Percentage of subject omission with finite verbs and RIs

<table>
<thead>
<tr>
<th>Language</th>
<th>Child</th>
<th>Finite verb</th>
<th>RI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flemish (Krämer 1993)</td>
<td>Maarten</td>
<td>25% (23/92)</td>
<td>89% (89/100)</td>
</tr>
<tr>
<td>Dutch (Haegeman 1995)</td>
<td>Hein</td>
<td>32% (1199/3768)</td>
<td>85% (615/721)</td>
</tr>
<tr>
<td></td>
<td>Thomas</td>
<td>28% (165/596)</td>
<td>92% (246/267)</td>
</tr>
<tr>
<td>Danish (Hamann &amp; Plunkett 1998)</td>
<td>Anne</td>
<td>11% (366/3379)</td>
<td>59% (394/667)</td>
</tr>
<tr>
<td></td>
<td>Jens</td>
<td>23% (742/3173)</td>
<td>58% (539/937)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>23% (2495/11008)</td>
<td>70% (1883/2692)</td>
</tr>
</tbody>
</table>

3.1 Predictions for Malagasy

RIs are not specifically associated with V2/topic drop languages. They also occur in non-V2 languages such as English, French, and Russian. However, every V2/topic drop language studied thus far shows an RI stage. Conversely, there is no null subject language that shows an RI stage (cf. note 5). Given that Malagasy exhibits argument drop and cannot therefore be like English, French or Russian, there are two available options for the language. The first is that it is a null subject language, in which case we do not expect an RI stage and so we do not expect any relationship between trigger omission and verb finiteness. The second is that Malagasy is a V2-like language, in which case we expect it to show an RI stage, and the rate of null subjects in non-finite contexts should be significantly higher than in finite contexts, as is the case for other V2/topic drop languages. The developmental facts can therefore help to evaluate the competing analyses of adult Malagasy. Malagasy does not have an infinitival form but the omission of verbal inflection results in a bare verb (root) which is arguably an RI analogue.

Before turning to our results we provide some relevant information about our subjects and coding.

3.2. Subjects and coding conventions

The children in this study are from families that speak the Merina dialect spoken in and around the capital city, Antananarivo. Merina is also the basis for standard written Malagasy and has been the focus of much of the linguistic research on Malagasy. The tapes were coded and transcribed by Cécile Manorchanta, a native speaker of Malagasy, and independently reviewed by two non-native speakers of Malagasy. The 3 children are: Tsirisoa 2;0–2;8 (MLU 1.68–4.5), Sonia 1;6–2;2 (MLU 2.84–3.46) and Ninie 1;10–2;6 (MLU 3.09–4.09). The total number of coded utterances is 1618.
We coded for trigger (null or overt), voice and tense morphology, and case (on pronouns). Where an argument was null, we inferred identity from context, as is standard practice. When voice was not specified (a form we refer to as a "bare verb") we used case marking, word order, context, and plausibility to determine type of voice. When tense morphology was specified we noted the tense. However, because present tense is not overtly marked\(^7\) and the future morpheme is not pronounced in colloquial speech, past is the only tense that can be coded solely based on morphology in both AT and TT forms. However, children use past very infrequently (9\%) of 1444 verbs -- about 6\% and therefore systematic coding of tense was not possible.

4. An "RI stage" in Malagasy?

The children in our study used voice morphology productively (cf. Hyams et al. (2005)). In addition to verbs marked with voice morphology (both AT and TT), illustrated in (10),\(^8\) the children also produced a high rate of bare verbs, exemplified in (11).

(10) a. M -i -ants Tsoso Tsorisoa, 2;3 (Adult: m-i-anatra,)
    AT-Pfx-study Tsorisoa
    'Tsorisoa studies/is studying.'
    b. Ari-na zondrina le manga (Tsorisoa 2;7) (Adult:ari-an(a))
    throw away TT dustbin Dem mango
    'As for this mango, (I) threw (it) in the dustbin.'

(11) a. Lomano za (Sonnia, 2;2) (Adult: m-i-lomano)
    swim 1sg. str.
    'As for me, I swim.'
    b. Kapo mama (Tsorisoa, 2;7) (Adult: kapoh-an(a)
    beat mommy
    'As for (me), mommy beat (me).

The rate of bare verbs used by the 3 children is given in Table 2. Imperatives are excluded from this analysis because the trigger (addressee) is unspecified in adult AT imperatives. Root verbs, that is, forms that are uninflated in the adult language (cf. 3), are also excluded.

Table 2. Percentage of bare verbs of all verbal utterances
The 43% bare verb rate in our data falls within the range of RI rates found in the Germanic languages (between 40% and 50%), and is far higher than the <10% typically found in the null subject languages (cf. Hockstra and Hyams (1998) for overview).

To confirm our hypothesis that the bare verb is an RI analogue we examined the relation between verb finiteness, viz. the specification of voice morphology, and two subject properties known to correlate with RIs in other child languages: null subjects (cf. Table 1) and strong subject pronouns. We first discuss strong subject pronouns.

Malagasy-speaking children use the strong form of the 1st person pronoun, *tsa ho*, as trigger (not possible in the adult grammar). Ntelitheos and Manorohanta (2005) observe that the strong pronoun occurs predominantly in bare verb contexts and the nominative form of the pronoun, *a ho*, occurs predominately with finite verbs, as shown in Table 3.

**Table 3. Distribution of tsaho and aho with bare and finite verbs (from Ntelitheos and Manorohanta (2005))**

<table>
<thead>
<tr>
<th></th>
<th>Finite</th>
<th>Bare</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aho</td>
<td>40 (73%)</td>
<td>15 (27%)</td>
</tr>
<tr>
<td>Izaho</td>
<td>5 (24%)</td>
<td>16 (76%)</td>
</tr>
</tbody>
</table>

This result is in line with those reported by Schütze (1997) for English and French speaking children, whose use of the default pronouns (*me, moi*) in subject position is largely restricted to non-finite contexts. Examples are given in (12). (In the Malagasy sentence in (12c/11a) *za* is a short form of *tsaho*).

(12) a. Me got bean.
    b. Moi mettre ça comme Pol.
       me put that like Pol
       ‘I (want to) put it like Pol.’
    c. Lomano za (=tsaho) (Sonnia 26)
       swim 1SG.STR
       ‘I (want to) swim’
The high percentage of bare verbs in our data and the bare verb-strong pronoun contingency provide prima facie evidence of an RI stage in Malagasy, consistent with the V2/topic drop analysis of the language, but inconsistent with the hypothesis that Malagasy is a null subject language.

Our next prediction concerns the distribution of null triggers. If bare verbs are RI analogues then null triggers will substantially outnumber overt triggers. Table 4 shows that most triggers in bare verb sentences are null while in finite clauses most triggers are overt.

**Table 4: Percentage of null triggers with finite and bare verbs**

<table>
<thead>
<tr>
<th></th>
<th>Finite verbs</th>
<th>Bare verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tsorisoa</td>
<td>40% (87/219)</td>
<td>52% (77/148)</td>
</tr>
<tr>
<td>Sonnia</td>
<td>56% (92/183)</td>
<td>69% (93/135)</td>
</tr>
<tr>
<td>Ninie</td>
<td>50% (100/202)</td>
<td>60% (81/134)</td>
</tr>
<tr>
<td>Total</td>
<td>46% (279/604)</td>
<td>60% (251/417)</td>
</tr>
</tbody>
</table>

The relationship between finiteness and trigger omission is marginally significant (p = 0.08 by a Friedman chi-square). The difference in percentage of null subjects in finite vs. non finite clauses is lower than in the languages in Table 1. In the next section we explore a possible explanation for the weaker effect found in Malagasy.

**4.1 Metrical effects**

We would like to propose that there are metrical effects of the sort proposed by Gerken (1994) and that this factor masks the null trigger-bare verb relation. Gerken has proposed that under production pressures, children may drop weak syllables (in iambic feet). Given the prosodic system of Malagasy we expect this kind of phonological reduction to affect prefixes but not suffixes. Suffixes form a tight unit with the verbal stem for purposes of stress assignment while prefixes form their own independent prosodic unit. Malagasy TT voice morphology is suffixal while AT voice morphology is prefixal. If prefixes are (also) being omitted for phonological reasons this would increase the number of bare AT forms relative to TT forms. Table 5 shows that this is indeed the case. There are proportionally many more bare AT verbs than bare TT verbs.

**Table 5: Percentage omission of voice morphology in AT and TT clauses**

<table>
<thead>
<tr>
<th></th>
<th>AT</th>
<th>TT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tsorisoa</td>
<td>120/309 (39%)</td>
<td>28/77 (36%)</td>
</tr>
<tr>
<td>Sonnia</td>
<td>120/241 (50%)</td>
<td>16/89 (18%)</td>
</tr>
<tr>
<td>Ninie</td>
<td>107/262 (41%)</td>
<td>40/123 (33%)</td>
</tr>
</tbody>
</table>
We thus propose that some percentage of bare AT verbs are actually finite with the voice morphology dropped in the phonology. This hypothesis leads to the prediction that the proportion of overt triggers will be higher in AT clauses (since some of these are hidden finites) than in TT clauses. This is shown to be confirmed in Table 6.

<table>
<thead>
<tr>
<th></th>
<th>Null Trigger</th>
<th>Overt trigger</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>190 (57%)</td>
<td>144 (43%)</td>
</tr>
<tr>
<td>TT</td>
<td>61 (73%)</td>
<td>22 (27%)</td>
</tr>
</tbody>
</table>

If we focus on the distribution of subjects of TT forms (controlling for the hidden finite factor), the null trigger rate is 73% and overt trigger rate 27%, very close to the (70/23%) difference found in the other RI languages (cf. Table 1).

We speculate that the 27% of bare TT verbs with overt triggers are cases in which PTT has not occurred. In other words, these are sentences in which the theme rests in situ in final position. The word order of the resulting surface string (VSO) is thus identical to one in which the theme moves to Topic position. If this hypothesis is confirmed (e.g. by spectrographic analysis showing an appropriate intonation break between the trigger and predicate), it would bring the Malagasy pattern further in line with the Germanic acquisition results, which show very few non-subject topics in RIs (which are thus overwhelmingly SOV) (cf. Poeppel & Wexler 1993; Haegeman 1995).

5. Crosslinguistic differences

One obvious difference between the Malagasy RIs and Germanic RIs is that the latter are verb final (SOV) while in Malagasy the bare verbs are clause initial (VSO). This means that in Malagasy there is V to T movement in nonfinite clauses. How do we explain this difference between Malagasy and Germanic?

Recall that on Pearson's analysis voice morphology is a reflex of null operator movement to Spec WhP. The trigger, base generated in topic position, is licensed under coindexation with the null operator (cf. 5). We suggest that the absence of voice morphology in the children's bare verbs reflects non-movement of the operator. Without operator movement the (overt) trigger is not licensed. We thus derive the bare verb-null trigger relation. In Malagasy the locus of underspecification is voice. By our hypothesis, T is specified and thus V to T
puts the verb at the left edge of the clause. In Germanic RIs, in contrast, AGR/T is underspecified and hence the verb remains in situ. This difference in the domain of underspecification in Malagasy vs. Germanic RIs may be more apparent than real, however. In both instances it is the case-related categories that are underspecified.

6. Conclusions

The high rate of bare verbs in Malagasy is suggestive of an RI stage. The contingencies between bare verbs and (i) strong pronouns and (ii) null subjects further support this hypothesis. This is expected if Malagasy is a V2-like language. Our data also suggest that metrical effects are a contributing factor in the omission of AT prefixes in early Malagasy grammar and that some apparent bare verbs are hidden finite forms (similar proposals have been made for English (cf. Rizzi 2000; Blom 2003)). Our acquisition results may contribute to an understanding of the adult grammar in so far as we find that the data support a topic/A’-movement analysis of Malagasy voice system.

Notes

For reasons of space many details and references are omitted. We apologize to the authors whose work we have not cited. For fuller discussions of the issues raised here see Hyams et al. (2005). Many thanks to Dimitris Ntelitheos and the GALA audience for their comments, and to Nathan Klinedinst for editorial and other assistance.

1 This study was done in collaboration with Dimitris Ntelitheos (UCLA) and Cecile Manorohanta (Université Nord, Madagascar) and is part of a larger study of the acquisition of the voicing system of Malagasy (Hyams, Ntelitheos & Manarohanta, 2005).

Malagasy also has a circumstantial voice, used when the trigger is an oblique nominal such as instrument, location, manner, etc. The children in our study produced almost no CT forms. See Hyams et al. for discussion of the acquisition of the different voice types.

3 Pearson (2001) proposes that the trigger final order is derived through the fronting of TP to a left specifier position, a phrasal movement analogue to T to C in the V2 languages. Malagasy is thus a V2 language in this sense as well.

4 Pronoun omission (7d) is grammatical only under the interpretation that Tenda is being visited by some arbitrary person. The omitted pronoun does not have the definite reading associated with the null topic.

5 Swahili, Greek and Kiché are apparent exceptions to the generalization that null subject languages do not show an RI stage. Pye (2001) claims that children acquiring Kiché exhibit an RI stage. In Kiché aspect and agreement morphology is prefixal and Pye notes that children typically produce only the last syllable of the verbal complex. This suggests
an alternative explanation for the Kiché bare verbs as involving phonological reduction rather than underspecification of a verbal functional projection. Phonological omission is discussed in more detail in section 4.1.

Swahili is a null subject language and children do produce bare verbs, illustrated in (i), which Deen (2005) analyzes as RI analogues. Following Barrett-Keach (1986) and Buell (2005), we assume that Swahili subject agreement, object agreement, and tense markers comprise a separate Aux constituent, as in (ii), which children omit during the bare verb stage. The Swahili bare verb is thus analogous to the Italian bare participle, as in (iii), and not a non-finite main verb, as is found in the Germanic languages.

(i) 0- 0- ka a hapa (child)
    a-na- ka a hapa (target)
    SA 3PRS live IND here
(ii) [a-na-] ka a hapa (Barrett-Keach 1986; Buell 2005)
(iii) Disegno cascato
     picture fallen
     'The picture has fallen.'

A null aux analysis is also available for the ‘bare subjunctive’ form used by Greek children, e.g. *Plo vrodis* ‘Spiros reads’ (cf. Varlokosta et al. (1997), but cf. also Hyams (2002) for a different analysis).

6 In these topic drop languages, subject omission in finite clauses refers to the dropping of a subject topic.

7 Traditionally, *m-* has been analyzed as the present tense morpheme (Rajaona 1972; Rajemia-Raolison 1971). However, Pearson (2001, 2005) (cf. also Builh 1988; Travis 1994) argues persuasively that it is an AT voice marker and that present tense is unmarked in Malagasy. In our analysis we follow Pearson and take *m-* to be a voice marker.

8 Malagasy-speaking children produce both AT and TT clauses just like children acquiring V2 languages produce both subject and object topics. The AT rate is higher than the TT rate for all children, also consistent with V2 languages in which subject topics are much more frequent than object topics in early language. (For discussion, see Hyams et al. (2005)).

9 Though specified tense may be inaudible because, as noted earlier, present tense is not phonologically realized, future tense is audible only in careful speech, and there are very few past tense forms at this stage of development.
References


Nina Hyams 305


