

UCLA OCCASIONAL PAPERS IN LINGUISTICS 21

# RECENT PAPERS IN AUSTRONESIAN LINGUISTICS

PROCEEDINGS OF THE THIRD AND FOURTH MEETINGS OF THE  
AUSTRONESIAN FORMAL LINGUISTICS ASSOCIATION (AFLA)

LOS ANGELES, 1996-1997

EDITED BY

MATTHEW PEARSON

UCLA DEPARTMENT OF LINGUISTICS

1998



## Table of Contents

1.	Faithfulness constraints and consonant patterns in Indonesian <i>Niken Adisasmito-Smith</i>	1
2.	Canonical types and noun phrase configuration in Fijian <i>John Alderete</i>	19
3.	Number marking in Selayarese <i>Hasan Basri</i>	45
4.	Alignment and nasal substitution strategies in Austronesian languages <i>Rita Bhandari</i>	59
5.	V to D raising in Sulawesi relatives <i>Daniel L. Finer</i>	71
6.	Chamorro: Whence and how? <i>Joseph C. Finney</i>	89
7.	Dual contradictory agent-marking: Solutions to Proto-Polynesian mysteries <i>Joseph C. Finney</i>	103
8.	The phonology-syntax interface in Rotuman <i>Mark Hale and Madelyn Kissock</i>	115
9.	Integrity in Malay VPs <i>Mark Honegger</i>	129
10.	Reduplication in Leti <i>Henry van der Hulst and Marian Klamer</i>	147
11.	Reduplication in Malagasy <i>Edward L. Keenan and Jean Paulin Razafimamonjy</i>	159
12.	The liver as the location of emotions: Noun incorporation in Kambera <i>Marian Klamer</i>	185
13.	Spelling out clitics in Kambera <i>Marian Klamer</i>	201

14.	Predicate fronting in Niuean <i>Diane Massam</i>	223
15.	The syntax and semantics of <i>olsem</i> in Bislama <i>Miriam Meyerhoff and Nancy Niedzielski</i>	235
16.	DP licensing and Spec roles in Maori <i>Elizabeth Pearce</i>	245
17.	Syntax versus semantics in Tagalog wh-extraction <i>Norvin Richards</i>	259
18.	Nominative pronouns and verb movement in Cebuano <i>Emily Sityar</i>	277
19.	Certain restrictions on A-bar movement in Malay <i>Hooi Ling Soh</i>	295



# Faithfulness constraints and consonant patterns in Indonesian\*

Niken Adisasmito-Smith, Cornell University

## 1. Introduction

Indonesian, a Western Austronesian language, is spoken in Indonesia both as the official language and as a lingua franca. From its history, there are two issues relevant to the nature of consonant clusters in Indonesian. First, Indonesian has an extensive history of borrowing lexical items from other languages that it comes in contact with. The source languages are, among others, Sanskrit, Tamil, Portuguese, Arabic, Dutch, and most recently, English. The degree to which these lexical items are nativized varies. The second issue is that, due to its status as the official language and a lingua franca, there are many varieties of Indonesian. These different varieties may be influenced by any of the languages spoken in the surrounding area. The variety that I am covering in this paper is the one which Cohn (1989) argues to be standard, that spoken by educated non-Javanese living in Java.

The goal of this paper is to analyze the distribution of consonants and consonant clusters in Indonesian, and to account for the observed patterns within the framework of Optimality Theory. Of particular interest is the interaction between faithfulness constraints, position-sensitive constraints, sonority hierarchy constraints, as well as prosodic ones. This paper is organized as follows. In section 2, I discuss the consonant inventory in Indonesian. I discuss the occurrence of consonants in the syllable in section 3, followed by the discussion of the occurrence of consonants in the root in section 4, and in affixed forms in section 5. In each section, constraints are proposed to account for the observed patterns. In section 6, I discuss relevant phonological processes at prosodic boundaries.

The data in this study are based on my own observations and a database of Indonesian lexical items, *Linguist 2.0*, compiled by Jonathan Alcantara, covering 31,000 entries. This database is based on Echols and Shadily (1990). I also draw on observations in the previous literature on Indonesian phonology, including Macdonald (1967), Lapoliwa (1981), Uhrbach (1987).

## 2. Consonant inventory

The consonants occurring in Indonesian, shown in (1), include bilabial, alveolar, palatal, velar, and glottal ones.

(1)	bilabial	alveolar	palatal	velar	glottal
	p	t	tʃ	k	ʔ
	b	d	dʒ	g	
	m	n	ɲ	ŋ	
	(f)	s			h
		l, r			
	w		j		

---

\* This research has been supported in part by NSF grant #SBR-9511185 to Abigail C. Cohn. I am indebted to her for many helpful comments and suggestions. I have also benefited from discussions with colleagues at the Phonetics Lab meeting at Cornell University and at the AFLA III. All errors are solely mine.

The fricative consonants native to Indonesian are [s] and [h]. There are no voiced fricatives. Most of the words with labio-dental fricative [f] are borrowed from Arabic, Dutch and English.

### 3. Simple consonants

In this section, I discuss the occurrence of consonants in various positions in a syllable. Indonesian allows primarily CV and CVC syllables, though some limited complex onset clusters occur. Onsetless syllables can also occur.

#### 3.1 *Onset and coda positions*

All of the consonants in the inventory shown in (1) can occur in onset position, except /ʔ/, which is not contrastive in this position. Listed in (2a) are some examples illustrating the occurrence of consonants at onset position, including voiced and voiceless obstruents, voiced and voiceless affricates, fricatives, liquids, and nasal consonants.

(2)	a.	padzak	'tax'	surat	'letter'
		buni	'sound'	rasa	'feel'
		tʃapat	'quick'	nakal	'naughty'
		dʒalur	'lane'		
	b.	ŋian	'echo'	ŋaraj	'valley'
		ŋilu	'pain'	ŋaja	'wide open'
				aŋin	'wind'

However, there is a restriction against the velar nasal occurring in root-initial onset position. Only a few indigenous lexical items occur with the root-initial velar nasal, listed exhaustively in (2b). The examples aŋin and ŋaja show that the velar nasal consonant can occur in root-internal onset position. I treat these lexical items as exceptions. In other varieties of Indonesian, particularly those influenced by Javanese, the velar nasal occurs more freely.

In coda position, most consonants can occur, except for the palatal affricates: [tʃ, dʒ], and the palatal nasal [ɲ]. As seen in (2a), they can occur both root-initially, dʒalur, and root-internally, buni and padzak. The fact that no palatal consonants are found in coda position can be accounted for by a restriction on their occurrence. The most effective position for these palatal consonants, given their status as complex segments (Keating 1991), is one where they are released, allowing them to be perceived. As discussed in Paradis and Prunet (1991), complex segments have the tendency to simplify in coda position. In Indonesian, complex segments like the palatal consonants are banned altogether from coda position; thus hypothetical \*/gatʃ/ or \*/bun/ are ruled out.

Another restriction on the occurrence of consonants in coda position is related to the voicing of obstruents. In Indonesian, voiced obstruents do not occur in coda position, as shown in (3a).

(3)	a.	saptu	'Saturday'	ribut	'noisy'
		paksa	'to force'	berkas	'bundle'
	b.	/adab/	'civilize'	/pəradaban/	'civilization'
		[p]		[b]	

The words *saptu* and *paksa* exemplify the occurrence of voiceless obstruents in coda position root-internally. The examples *ribut* and *barkas* show their occurrence root-finally. In the indigenous vocabulary, there are no final voiced obstruents. In borrowed lexical items, when a voiced obstruent occurs at word-final coda position, it is realized as voiceless. In (3b), the bilabial obstruent in the root /*adab*/ of Arabic origin is voiceless, when it occurs word-finally. However, when a vowel-initial suffix follows the root, as in /*pəradaban*/, the voicing of bilabial obstruent surfaces.

### 3.2 *OT account*

To account for the distribution of consonant clusters in Indonesian, I assume Correspondence Theory, as argued for in McCarthy and Prince (1995). Correspondence constraints determine the relationship of input and output at the segmental level, while Identity constraints regulate the input-output relationship at the featural level. The constraints considered here are MAXIO, DEPIO, and IDENT(F).

- (4) MAXIO  
Every segment of the input has a correspondent in the output  
(No phonological deletion)
- (5) DEPIO  
Every segment of the output has a correspondent in the input  
(Prohibit phonological insertion)
- (6) IDENT(F)  
Correspondent segments are identical in feature F

The interaction of Correspondence constraints with Alignment constraints (McCarthy and Prince 1993) allows us to capture the observed patterns noted above, such as the restriction against root-initial velar nasals. The constraint is formulated in (7). This constraint also plays a role in the phonotactics of other languages, such as English.

- (7) \*|ŋ            | = root edge  
No velar nasal at root-edge

While the velar nasal does not occur root-initially in the indigenous vocabulary, when it occurs in the borrowed lexical items, it is tolerated. This means that in borrowed words, the faithfulness constraint at root-edge ranks higher than \*|ŋ. This is in fact the case in the Indonesian variety with Javanese influence.

The constraint against palatal consonants from occurring in coda position is stated in (8).

- (8) IDENTCOMP[REL]  
The release feature of an input complex segment should be preserved in the output

Observing this constraint, palatal consonants always occur at a position where they can be released, i.e. onset position. Since there are no exceptions observed, this constraint must be highly ranked.

To account for the patterns of final devoicing, I adopt Lombardi's approach (1995) in which she accounts for final devoicing through the interaction of faithfulness constraints relative to the position in a syllable, and markedness constraints relative to the voicing of obstruents. To rule out any voiced obstruent in coda position, \*LAR (9) dominates IDLAR (10). Ensuring the voicing specification of obstruents at onset position, \*LAR is in conflict with IDONSLAR (11). The relative ranking of the three constraints is shown in (12). The ranking IDONSLAR >> IDLAR is argued to be universal.

(9) \*LAR

Do not have laryngeal features (= be voiceless)

(10) IDLAR

Consonants should be faithful to the underlying laryngeal specification

(11) IDONSLAR

Onsets should be faithful to underlying laryngeal specification

(12) IDONSLAR >> \*LAR >> IDLAR (Lombardi 1995)

	/adab/	IDONSLAR	*LAR	IDLAR
√	a. a.dap.		*	*
	b. a.dab.		*!*	
	c. a.tap.	*!		**

Both output candidate forms in (12a) and (12b) obey IDONSLAR. For each voiced obstruent, a candidate form incurs one violation. Even though both *adap* and *adab* in (12a) and (12b) violate \*LAR, the former is more optimal than the latter, because it incurs fewer violations. The candidate form *atap* in (12c) violates highly ranked IDONSLAR, since the input voiced obstruent in onset position corresponds to a voiceless obstruent in the output. The same input form with a vowel-initial suffix attached to it also follows from the ranking in (12). However, the result is different, shown in tableau (13).

(13)

	/pər-adab-an/	IDONSLAR	*LAR	IDLAR
	a. pə.ra.da.pan.	*!	*	*
√	b. pə.ra.da.ban.		**	

The optimal output candidate form is *pəradaban* in (13b). This candidate form incurs more violations as compared to candidate (a), with regards to \*LAR. However, it obeys IDONSLAR which is ranked higher than \*LAR. The output candidate in (13a), *pəradapan*, is ruled out, since the voiceless [p] at onset position fatally violates IDONSLAR.

#### 4. Consonant clusters

##### 4.1 Facts

The basic shape of bisyllabic roots in the indigenous vocabulary of Indonesian is (C)V(N)(C)V(C). There are thus root-medial clusters, some clusters across morpheme boundaries and to a limited extent some onset clusters, due primarily to borrowings. The data in (14) show that consonant clusters are allowed to occur in onset position, both root-initially as in *prakarsa*, etc., as well as root-internally, as in *damprat*, etc.

(14)	a. Root-initial clusters <sup>1</sup>	b. Root-internal clusters
	pra.kar.sa      ‘initiative’	dam.prat      ‘scold’
	tra.di.si      ‘tradition’	sas.tra      ‘literature’
	gra.da.si      ‘gradation’	aŋ.grek      ‘orchid’

In standard Indonesian, a sequence of consonants may be in onset position if it observes the sonority hierarchy. Otherwise the cluster is broken into different syllables, as the examples in (14b) show.

In Indonesian, complex codas are not tolerated, at least by those speakers who have not been exposed to English. So, hypothetical words like \*/rant/ or \*/damp/ would be ruled out.

In cases where a CVC syllable is followed by another consonant-initial syllable, heterosyllabic consonant clusters result. One of the most common root-internal consonant clusters in Indonesian is a nasal obstruent cluster. The data in (15) show that an obstruent following a nasal consonant can be voiceless, as in (15a): *simpan*, *pintu*, *hantŋur* and *uŋkap*. The examples in (15b): *rambut*, *landas*, *indzak*, and *baŋga* show that an obstruent can also be voiced when following a nasal. The place of articulation of the nasal in these clusters is assimilated to that of the following obstruent.

(15)	a. <i>simpan</i> ‘keep’	b. <i>rambut</i> ‘hair’
	<i>pintu</i> ‘door’	<i>landas</i> ‘base’
	<i>hantŋur</i> ‘destroyed’	<i>indzak</i> ‘to step on’
	<i>uŋkap</i> ‘to reveal’	<i>baŋga</i> ‘proud’
	c. <i>piŋsan</i> ‘to faint’	
	<i>lɔŋsɔr</i> ‘to slide down’	
	<i>aŋsur</i> ‘to do s.t. by installment’	

It should be noted also that there are cases where a consonant cluster consists of a velar nasal followed by the voiceless fricative [s], shown in (15c). In these cases, the velar nasal does not assimilate its place of articulation to that of the fricative. However, there are also cases where a root-internal cluster consists of an alveolar nasal followed by the fricative [s], such as *dansa* ‘dance’, *bɔnsaj* ‘bonsai’, etc. These consonant clusters are rare, and are usually due to borrowing. I therefore take the basic pattern in (15c) to be /ŋs/.

#### 4.2 OT constraints

To account for the restriction on consonant clusters in onset position, I follow the constraint formulation in Benua (1995). She assumes the standard sonority hierarchy as stated in (16), and the condition on syllables, SON-CON.

(16) Sonority hierarchy: | glide | > | liquid | > | nasal | > | fricative | > | stop |

SON-CON (Sonority Contour)

Complex onsets rise in sonority and complex codas fall in sonority

<sup>1</sup> In certain variety of Indonesian, a schwa is inserted to break up the tautosyllabic cluster: *pˈra.kar.sa*, *tˈra.di.si*, etc.

The relative ranking of the input-output faithfulness constraint and SON-CON is shown in tableau (17) where, for the input form *prakarsa*, candidate (a) is the most optimal output form. All segments in the input have a correspondent in the output. The consonant cluster in root-initial onset position obeys SON-CON, and the root-internal consonant cluster is divided into different syllables. Candidate (b), not realizing the /r/, is ruled out for violating MAXIO. Candidate (c) is also ruled out since a consonant cluster violates SON-CON in onset position.

(17) MAXIO      SON-CON

>>

	/prakarsa/	MAXIO	SON-CON
√	a. pra.kar.sa		
	b. p<r>a.kar.sa	*!	
	c. pra.ka.rsa		*!

The constraint regulating consonants in coda position as stated in (18) disallows tautosyllabic consonants in coda position.

(18) \*CC]<sub>σ</sub>

Be a simple coda

The tableau in (19) shows the relative ranking of the faithfulness constraint, the simple coda constraint, and SON-CON. Candidate (a) is the most optimal output form since it incurs no violation with regard to these constraints. Candidate (b) is ruled out for violating SON-CON. Candidate (c) violates \*CC]<sub>σ</sub> by allowing a sequence of consonants in coda position, even though the consonantal sequence obeys SON-CON. Candidate (d) fatally violates the faithfulness constraint for not having a correspondent of the input [p] in the output.

(19) MAXIO >> SON-CON >> \*CC]<sub>σ</sub>

	/damprat/	MAXIO	SON-CON	*CC] <sub>σ</sub>
√	a. dam.prat			
	b. da.mprat		*!	
	c. damp.rat			*!
	d. dam<p>.rat	*!		

Another possible candidate not illustrated in (19) is the case where the adjacent nasal and voiceless obstruent coalesce, resulting in \*damrat (not necessarily distinguishable from candidate (d)). In Indonesian, no coalescence occurs root-internally. For further discussion on coalescence cases, see section 5.5.1.

Goodman (1996) provides an approach to account for cases of nasal assimilation and vowel epenthesis. Based on Southeast Pomo, Ponapean, and Dahalo, she argues that nasal assimilation and vowel epenthesis in these languages result from the restriction against nasal consonants having underlying place features, as stated in (20). This constraint interacts with the markedness hierarchy of place features, following Prince and Smolenksy (1993), and Smolenksy (1993). The hierarchy is shown in (21).

(20) \*NASAL[PLACE]  
Nasal consonants do not have place features

(21) IDENT[DOR] >> IDENT[LAB] >> IDENT[COR]

Earlier in (15c), we saw cases where a root-internal nasal does not always assimilate its place of articulation to that of the following obstruent: *piŋsan*. It can be argued that the relative ranking of the constraints in (20) and (21) in Indonesian is as shown in (22), and that the underlying representation for *damprat* in (19) is /daŋprat/.

(22) IDENT[DOR] >> \*NASAL[PLACE] >> IDENT[LAB] >> IDENT[COR]

As we can see, consonant clusters may arise when root-internal syllables come in contact. They may also arise when a root co-occurs with affixes, or when an affix adjoins to another affix. This is the focus of discussion in the following section.

### 5. Consonants in affixed forms

Before discussing cases with consonant clusters in the affixed forms, we need to consider the syllable shapes of affixes. As shown in (23), an affix in Indonesian may consist of a vowel: /-i/; of an open syllable with an onset: /-ku/, /-mu/, /-ŋa/; or of a closed syllable with an onset: /pər-/ , /tər-/ , /-pun/, /-kah/, etc. or without: /-an/.

(23)	prefix +	prefix +	root +	suffix	+	suffix (clitic)	+	suffix (clitic)
	məŋ +	pər +		-i		-ku		-lah
			root +	-an	+	-mu	+	-pun
		tər +		-kan		-ŋa		-kah
		bər +				<i>poss./obj.</i>		<i>emph./quest.</i>

The schema above shows the maximal possible combination of a root and its affixes, and the order in which they appear. The maximal number of prefixes is two, and the maximal number of suffixes is three. Grammatically, some of the suffixes are clitics. The consonants in the onset position of a prefix are the bilabial nasal: /məŋ-/ , and the bilabial and alveolar obstruents: /pər-/ , /bər-/ , and /tər-/ . The only possible combination of the prefixes is that of /məŋ-/ and /pər-/ . This is further discussed in section 5.3. The possible consonants at the onset position of a suffix are the bilabial and velar obstruents: /-pun/, /-kan/, /-ku/, and /-kah/; the bilabial and palatal nasals: /-mu/ and /-ŋa/; and the liquid: /lah/. The consonant in the coda position of an affix can be a placeless nasal: /məŋ-/ , an alveolar nasal: /-pun/, /-kan/, an [h]: /-kah/, /lah/, or an [r]: /pər-/ , /bər-/ , /tər-/ .

As we saw in (15), a nasal and an obstruent (voiced or voiceless) can be adjacent root-internally. Except for the /ŋs/ cases, place assimilation takes place. In section 5.3.3, we will see that place assimilation also takes place at the prefix-prefix juncture. The following sections are organized based on phonological processes that take place. Sections 5.1 and 5.3.1 focus on the cases of nasal-obstruent clusters occurring at the root-suffix and suffix-suffix junctures, where no phonological changes take place. Section 5.3.3 focuses on cases of place assimilation at the

prefix-prefix juncture. In section 5.4, I discuss cases of degemination when identical consonants become adjacent. Nasal deletion or coalescence takes place at the prefix-root juncture, as discussed in section 5.5.

### 5.1 *Root-suffix juncture*

At the root-suffix juncture, consonant clusters result from the juxtaposition of the root-final consonant and the initial consonant of the suffix. The data in (24) show that when two consonants become adjacent at the root-suffix juncture, both consonants in the input form are present in the output.

(24)	a.	makan-pun	->	makanpun	‘even eating...’
		harum-kan	->	harumkan	‘make fragrant’
	b.	saruŋ-mu	->	saruŋmu	‘your sarong’
		saruŋ-ŋa	->	saruŋŋa	‘his sarong’
	c.	makan-lah	->	makanlah	‘please eat’
		diam-lah	->	diamlah	‘please be quiet’

In (24a), we see examples of a nasal followed by an obstruent. Neither place assimilation nor coalescence takes place at the root-suffix juncture. This is different from the case seen above in (15) where most root-internal nasal obstruent clusters consist of a placeless nasal and an obstruent, and the nasal assimilates its place of articulation to that of the following obstruent.

The data in (24b) show the occurrence of a cluster of two nasal consonants at the root-suffix juncture with the same result. These nasals maintain the place feature specification of the input form in the output form. While no nasal clusters occur root-internally or at the prefix-root juncture, they also occur at the suffix-suffix juncture. The examples in (24c) show that a nasal-liquid cluster results from the root-suffix juxtaposition. Nasal-liquid clusters can occur at the suffix-suffix juncture, but not at the prefix-root juncture.

### 5.2 *OT account*

The constraints MAXRT (25) and MAXONS (26) account for the fact that consonant clusters at the root-suffix juncture of the input forms are represented in the output forms. This assumes underlying syllabification. This is an area that warrants further investigation. The tableau in (27) shows that no relative ranking between the two constraints can be established.

(25) MAXRT  
Every segment in the input root has a correspondent in the output root

(26) MAXONS  
Onsets in the input have a correspondent in the output



(27) MAXRT &gt;&gt; MAXONS

	/makan-lah/	MAXRT	MAXONS
√	a. makan-lah		
	b. maka<n>-lah	*!	
	c. makan-<l>ah		*!

Candidate (a) is the optimal output form; it obeys both constraints MAXRT and MAXONS. Candidates (b) and (c) are both ruled out for incurring fatal violations of MAXRT and of MAXONS, respectively.

The constraint IDENT(F) (5) ensures that a consonant retains its place specification in the input form, preventing place assimilation to take place in (24a) and (24b); thus \*makampun from the input /makan-pun/ and \*sarummu from /sarun-mu/. A potential output candidate of /makan-pun/ is \*makamun, in which case the nasal and the obstruent coalesce. No coalescence takes place at the root-suffix juncture.

### 5.3 At the junctures between affixes

#### 5.3.1 Suffix-suffix juncture

The examples in (28) show the cases where a maximum of three suffixes may follow a root. As with the case at the root-suffix juncture, no coalescence or place assimilation applies to the adjacent nasal and voiceless obstruent at the suffix-suffix juncture, as shown in mæmpørtahankanku: the root-final and suffix-final alveolar nasals, tahan and -kan, are adjacent to a velar obstruent and they retain their place specification in the output form.

- (28) mæŋ-pər-tahan-kan-mu-lah → mæmpørtahankanmulah 'to defend you is....'  
 mæŋ-pər-tahan-kan-mu → mæmpørtahankanmu 'to defend you'  
 mæŋ-pər-tahan-kan-ku → mæmpørtahankanku 'to defend me'  
 pər-tahan-kan-lah → pørtahankanlah 'please, defend (it)'

As mentioned earlier, some suffixes are grammatically clitics, such as -mu in mæmpørtahankanmu 'to defend you'. This example also shows that at the suffix-suffix juncture adjacent nasals [n] and [m] can occur. Also, as is the case at the root-suffix juncture, adjacent nasal and liquid consonants can occur at the suffix-suffix juncture.

#### 5.3.2 OT account

The tableau in (30) shows that the optimal output form is the one which does not violate MAXSFX (29) and MAXONS (26).

(29) MAXSFX

Every segment in the input suffix has a correspondent in the output suffix

(30)

	/pər-tahan-kan-lah/	MAXSFX	MAXONS
√	a. pørtahankanlah		
	b. pørtahanka<n>lah	!*	
	c. pørtahankan<l>ah		!*

Candidate (a) is the optimal output form. Candidate (b) is ruled out for underparsing the coda consonant of the suffix /-kan/, in violation of MAXSFX. Candidate (c) is bad for not parsing the onset of the suffix /-lah/, violating MAXONS.

### 5.3.3 *Prefix-prefix juncture*

We have seen in the schema in (23) that two prefixes can be combined preceding a root. As mentioned earlier, only the prefixes /məŋ-/ and /pər-/ can be juxtaposed. The prefix /məŋ-/ indicates the active verbal form. Examples are shown in (31).

- |      |               |                |                  |
|------|---------------|----------------|------------------|
| (31) | məŋ-pər-cəpat | -> məmpərcəpat | ‘to make faster’ |
|      | məŋ-pər-mudah | -> məmpərmudah | ‘to make easier’ |
|      | məŋ-pər-lunak | -> məmpərlunak | ‘to make softer’ |

In the output form, a nasal and a voiceless obstruent are adjacent at the prefix-prefix boundary. The placeless nasal of the prefix /məŋ-/ assimilates its place of articulation to that of the following obstruent, i.e. a bilabial. No coalescence occurs at the prefix-prefix juncture. There are at least two possible accounts. First, one may argue that the occurrence of the two prefixes in (31) is lexicalized, and coalescence does not apply. Secondly, it is also possible to argue that the prefix-prefix boundary acts like a syllable boundary within a root, where no coalescence takes place. From the historical perspective, these two views may be related, that is, the second argument may lead to the first.

To account for the surface realization of the consonant clusters at the prefix-prefix juncture, the relative ranking of constraints presented earlier in (22), repeated in (32), is adopted.

- (32) IDENT[DOR] >> \*NASAL[PLACE] >> IDENT[LAB] >> IDENT[COR]

In the following section, I discuss the cases where identical consonants become adjacent due to affixation of a morpheme to another morpheme, i.e. an affix to a root or an affix to another affix. I then turn to the question of nasal assimilation and coalescence at the prefix-root juncture.

### 5.4 *Identical consonants at morpheme boundaries*

Generally, a sequence of identical consonants—that is, false geminates—seem to be avoided in Indonesian. The environment in which such cases may arise is as shown in the following schema:

- (33) M1[ ... C<sub>1</sub>] + [C<sub>2</sub> ... ]M2    where: C<sub>1</sub> = C<sub>2</sub>

When the final coda consonant C<sub>1</sub> of the morpheme M1, and the onset consonant C<sub>2</sub> of the second morpheme M2 are adjacent, a fake geminate results, in the case where C<sub>1</sub> and C<sub>2</sub> are identical. In Indonesian, two identical consonants do in fact become adjacent as the result of affixation, both at the prefix-root boundary as well as at the root-suffix boundary. In these cases, shortening or full degemination can occur. Careful-speech syllabification, orthography, and durational measurements can be used to diagnose the resulting patterns. The examples in (34) illustrate this phenomenon.

- |      |              |                           |                   |
|------|--------------|---------------------------|-------------------|
| (34) | a. tər-tarik | -> tərtarik               | ‘to be attracted’ |
|      | b. tər-rasa  | -> tərasa (terasa, orth.) | ‘to be felt’      |

c. *sarakah*

'greedy'

In (34a) and (34b), when the prefix coda [r] is followed by an obstruent-initial root, both consonants are present in the output form. However, when the initial consonant of the root is also an [r], only one of them survives. Preliminary acoustic duration measurements (Adisasmito-Smith, in progress) show that the duration of [r] in *tərasa* and that in root-internal intervocalic occurrences, as in *sarakah* 'greedy' (34c), are almost identical. In careful-speech syllabification, the consonant [r] may appear twice as in *tər...ra...sa*, or it may also appear once as in *tə...ra...sa*, as discussed in Cohn and McCarthy (1994). The degemination process seems to be complete at the prefix-root juncture. The evidence from the orthography supports this.

However, the situation at the root-suffix juncture is different. (35) illustrates a case where identical nasal consonants become adjacent at the root-suffix juncture.

- |      |                 |  |                 |
|------|-----------------|--|-----------------|
| (35) | <i>salam-ku</i> | -> <i>salamku</i>                          | 'my greeting'   |
|      | <i>salam-mu</i> | -> <i>salamu</i> ( <i>salammu</i> , orth.) | 'your greeting' |

Contrary to */tərasa/*, orthographically, both consonants in */salammu/* are present. Based on a preliminary durational measurement, the duration of the bilabial nasal in */salammu/* is approximately 30-50% longer than that in */salamku/*. This is consistent with the preliminary measurement observed by Cohn (p.c.). If this difference holds up, clearly, there is more to the story. This question is currently under study.

In cases where the initial consonant of the root is a nasal, as shown in (36), a sequence of two nasal consonants results. However, only one of the nasals survives: */məŋ-masak/* -> *məmasak*.

- |      |                  |                   |                  |
|------|------------------|-------------------|------------------|
| (36) | <i>məŋ-masak</i> | -> <i>məmasak</i> | 'to cook'        |
|      | <i>məŋ-nikah</i> | -> <i>mənikah</i> | 'to get married' |
|      | <i>məŋ-pala</i>  | -> <i>məpala</i>  | 'to flare up'    |

One may argue that the placeless nasals in (36) undergo place assimilation and the single surviving nasal is the result of degemination. They may also be considered to pattern together with the cases where a nasal in coda position of the prefix deletes when it follows a sonorant (see section 5.5.3).

Degemination in Indonesian may be argued as an OCP effect, as suggested in Cohn and McCarthy (1994).

- (37) OCP (OBLIGATORY CONTOUR PRINCIPLE)  
No adjacent identical segments are allowed

This constraint is ranked higher than the input-output faithfulness constraint, as shown in tableau (38). The optimal output candidate (a) is the one which avoids having a geminate consonant cluster in the output, in violation of MAXIO. Candidate (b) violates the highly ranked OCP, and is thus ruled out.

(38) OCP &gt;&gt; MAXIO

	/tər-rasa/	OCP	MAXIO
√	a. tə<r>.ra.sa		*
	b. tər.ra.sa	*!	

Note that candidate (b) can be optimal in careful-speech syllabification, as mentioned earlier, in which case the constraints in (38) are reranked: MAXIO is ranked higher than OCP.

In the following section, I discuss the cases of consonant clusters at the prefix-root juncture, where place assimilation, coalescence, and nasal deletion take place.

### 5.5 *Prefix-root juncture*

#### 5.5.1 *Nasal clusters and coalescence*

The cases of nasal assimilation and coalescence at the prefix-root juncture in Indonesian have been discussed in Uhrbach (1987) and in Pater (1995), among others. In this section, I present some of these cases.

As mentioned earlier, the possible coda consonant of prefixes is either a placeless nasal or the liquid [r]. When a vowel-initial root follows the prefix with placeless nasal: /məŋ-/ , this nasal surfaces as velar, as shown in (39). Note that when the root-initial consonant is [h], the placeless nasal also surfaces with velar articulation. This may be due to the fact that, phonetically, [h] can be characterized as the voiceless counterpart of the surrounding vowel(s).

(39)	məŋ-asuh	-> məŋasuh	'to take care of'
	məŋ-indzak	-> məŋindzak	'to step on'
	məŋ-hilaŋ	-> məŋhilaŋ	'to disappear'

When the initial consonant of the root is a voiced obstruent, the resulting cluster consists of a nasal with the same place of articulation as that of the following obstruent, followed by the voiced obstruent, shown in (40).

(40)	məŋ-buat	-> məmbuat	'to make'
	məŋ-dapat	-> məndapat	'to get'
	məŋ-dzamah	-> məndzamah	'to touch'
	məŋ-gubah	-> məŋgubah	'to compose'

Adjacent nasals and voiceless obstruents at the prefix-root juncture undergo coalescence, as shown in (41). As a result of the coalescence, the nasal consonant survives, and it assimilates its place of articulation to that of the following voiceless obstruent. The voiceless obstruent itself is deleted. However, when the root-initial consonant is the palatal affricate [tʃ], assimilation but no coalescence takes place.

(41)	məŋ-pantʃiŋ	-> məmantʃiŋ	'to fish'
	məŋ-tukar	-> mənukar	'to exchange'
	məŋ-kuruŋ	-> məŋuruŋ	'to lock s.b. in'
	məŋ-satu	-> məŋnatu	'to unite'
but:	məŋ-tʃuri	-> məntʃuri	'to steal'

### 5.5.2 OT account

It has been argued that a nasal-voiceless obstruent cluster is dispreferred in some languages. Different languages choose different strategies in the face of this dispreference. One possible strategy is by voicing the obstruent following the nasal consonant. Another possibility is to delete the nasal consonant, and so on. Pater (1995) discusses the typology of languages based on the strategy taken.

As suggested by Pater, in Indonesian, cases where nasal obstruent clusters result at the prefix-root juncture are accounted for by the input-output faithfulness constraint, MAXIO, and the constraint prohibiting nasals to have place specification, \*NASAL[PLACE]. These constraints are not ranked with respect to each other, since both of them are obeyed in the output forms, shown in (42). Note that the only occurring nasal obstruent cluster at the prefix-root juncture is a nasal-voiced obstruent sequence.

(42) MAXIO, \*NASAL[PL]

/məŋ-dapat/	MAXIO	*NASAL[PL]
məndapat	√	√

We have seen earlier that root-internally, both cases where a nasal is followed by a voiced obstruent, and where a nasal is followed by a voiceless obstruent occur. We have also seen in sections 5.1 and 5.3 that at the junctures of root-suffix, suffix-suffix, and prefix-prefix, this is also widespread. However, at the prefix-root juncture, only nasal-voiced obstruent clusters occur. On the other hand, nasal-voiceless obstruent clusters undergo coalescence.

To account for cases with coalescence, I adopt the idea of ‘fusion’ as argued for in Pater (1995). Segmentally, the process of fusion does not involve deletion and thus is not a violation of MAXIO. Instead, two segments fuse into or correspond with one segment containing at least one feature from each segment,  $C_1C_2 \rightarrow C_{1,2}$ , thus violating input-output identity of certain features. See Pater for more detailed discussion. The constraint \*NC (43) ensures that no nasal-voiceless obstruent cluster occurs at the prefix-root juncture. This constraint is in direct conflict with an ‘anti-fusion’ constraint, \*MC (44). The tableau in (45) illustrates the relative ranking of these two constraints, as well as their ranking relative to MAXIO.

- (43) \*NC (Pater 1995)  
Nasal/voiceless obstruent clusters are disallowed
- (44) \*MC (\*MULTIPLE CORRESPONDENCE)<sup>2</sup> (Lemontagne and Rice 1995)  
Elements of the input and the output must stand in one-to-one correspondence relationship with each other.

(45) MAXIO >> \*NC >> \*MC

/məŋ-tukar/	MAXIO	*NC	*MC
√ a. mənukar			*
b. məntukar		*!*	
c. mə<n>tukar	*!	*	

<sup>2</sup> The equivalent of this constraint is LIN (LINEARITY) in Pater (1995), and UNIFORMITY in McCarthy and Prince (1995).

Candidate (a) in the tableau above is the optimal output form. While this candidate violates \*MC, it is still optimal since it satisfies the higher ranked constraints, MAXIO and \*NC. Candidate (b) violates \*NC due to the nasal-voiceless obstruent sequence at the prefix-root juncture. In order to avoid the occurrence of this consonant cluster, it is also possible for an output candidate not to parse the nasal in coda position of the prefix, as shown by candidate (c). However, in Indonesian this is not an option, thus MAXIO dominates \*NC. As we can see, the evidence for the relative ranking between MAXIO and \*NC comes from the non-surviving candidates.

We have seen earlier in (15) that nasal-voiceless obstruent clusters freely occur root-internally, as in *simpan*, *unkap*, etc. This is also the case with nasal-voiceless obstruent clusters at the root-suffix and suffix-suffix junctures. The constraints MAXRT (25) and MAXSFX (29) alone will not prevent fusion from taking place: the process of fusion does not involve any deletion, and therefore is not in violation of MAXRT or MAXSFX. It seems that a constraint IDENT(F), where F refers to any features violated in the fusion process, is ranked higher at junctures other than the prefix-root juncture. See section 6 for further discussion.

### 5.5.3 Nasal deletion and coda [r] deletion

As seen above in cases of /r-r/ (34b) and /N-nasal/ (35) sequences at the prefix-root juncture, only one of the two consonants survives. The data in (46) show that when the nasal coda of the prefix /məŋ-/ and a root-initial liquid become adjacent, the liquid survives, as in *məŋ-laraŋ* -> *məlaraŋ*.

(46)	<i>məŋ-laraŋ</i>	-> <i>məlaraŋ</i>	'to forbid'
	<i>məŋ-rasa</i>	-> <i>mərasa</i>	'to feel'
	<i>məŋ-wakil-i</i>	-> <i>məwakili</i>	'to represent'
	<i>məŋ-jakin-i</i>	-> <i>məjakini</i>	'to be convinced'

Shown in (47) are cases where the consonant in coda position of the prefix is [r], all of the consonants from the input are present in the output.

(47)	<i>tər-tarik</i>	-> <i>tərtarik</i>	'to be attracted'
	<i>tər-buat</i>	-> <i>tərbuat</i>	'to be made'
	<i>tər-pata</i>	-> <i>tərpata</i>	'in fact'
	<i>tər-laraŋ</i>	-> <i>tərlaraŋ</i>	'to be forbidden'
but:	<i>tər-rasa</i>	-> <i>tərasa</i>	'to be felt'

When the initial consonant of the root is an [r], only one of them survives: /*tər-rasa*/ -> *tərasa*, as the result of degemination, as discussed above in section 5.4.

### 5.5.4 OT account

In Indonesian, it seems that there is a restriction against nasal-liquid sequences, and nasal-glides sequences. This restriction is expressed in the constraints \*NL (48) and \*NG (49) which regulate restrictions on adjacent consonants at the segmental level. I propose the constraint ranking in (50), in which MAXRT (25) and MAXONS (26) dominate \*NL and \*NG. There is no evidence for the relative ranking of MAXRT and MAXONS, nor for \*NL and \*NG.



Monosyllabic cases in Indonesian provide evidence against (52) (Cohn and McCarthy, 1994), in which case, a schwa is usually inserted to satisfy minimal word requirement, showing precisely this sort of interaction. (53) seems to suggest that there needs to be constraint reranking whenever the prosodically strong boundary is relevant. An account integrating the interaction of (53) and (54) will be most preferable, since it will refer to prosodic boundaries without necessarily triggering constraint reranking.

In the spirit of (54), I propose the following, to account for the occurrence of consonant clusters in Indonesian. The integrity of segments within the PrWd is maintained:

- (55) MAX<sub>PRWD</sub>  
Every segment in the input prosodic word has a correspondent in the output prosodic word

The constraint ranking in (56) accounts for cases of nasal deletion, shown in tableau (57).

- (56) MAX<sub>PRWD</sub> >> \*NL

(57)

	/ məŋ-[laraŋ] /	MAX <sub>PRWD</sub>	*NL	MAXIO
	a. məŋ-[laraŋ]		*!	
√	b. mə<ŋ>-[laraŋ]			*
	c. məŋ-<l>araŋ]	*!		*

Candidate (57b), məlaraŋ, is the optimal output form. Even though it violates MAXIO, it obeys MAX<sub>PRWD</sub> and \*NL. Candidate (57a) is ruled out, since it fatally violates \*NL. Candidate (57c) underparses a segment in the PrWd, thus fatally violating MAX<sub>PRWD</sub>.

While \*NC phenomenon is common across languages, as discussed in Pater, in Indonesian it can have an effect only at the periphery of the PrWd. This is captured by the constraint \*N<sub>[PrWd C]</sub> (58). However, the integrity of all features within the PrWd is maintained: IDENT<sub>PRWD(F)</sub> (59). The relative ranking of these two constraints is shown in (60). This ranking accounts for the occurrence of coalescence at the prefix-root juncture, shown in (61), and of nasal-voiceless obstruent cluster at the root-suffix juncture, which is within the PrWd, as shown in (62).

- (58) \*N<sub>[PrWd C]</sub>  
Nasal-voiceless obstruent clusters are prohibited at the left edge of the prosodic word

- (59) IDENT<sub>PRWD(F)</sub>  
Every feature specified in the input prosodic word has a correspondent in the output prosodic word

- (60) \*N<sub>[PrWd C]</sub> >> IDENT<sub>PRWD(F)</sub>

(61)

	/ məŋ-[tukar] /	*N <sub>[PrWd C]</sub>	IDENT <sub>PRWD(F)</sub>
	a. məŋ-[tukar]	*!	
√	b. mə[n <sub>1,2</sub> ukar]		*



(62)	/ [makan-kan] /	*N [ <sub>PrWd</sub> C	IDENTPRWD(F)
√	a. [makan-kan]		
	b. [makaŋ-kan]		*!
	c. [makaŋ <sub>1,2</sub> an]		*!

Candidate (61b) is optimal; even though it violates IDENTPRWD(F) by undergoing fusion, it obeys the higher ranked \*N [<sub>PrWd</sub> C. This constraint is fatally violated by candidate (61a). Candidate (62a) does not violate \*N [<sub>PrWd</sub> C or IDENTPRWD(F), and thus the optimal output form. Candidates (62b) and (62c) are ruled out because they incur fatal violations of IDENTPRWD(F). The alveolar place specification of the nasal in candidate (62b) corresponds to bilabial in the output form. Fusion takes place within the PrWd unit in candidate (62c), and thus in violation of IDENTPRWD(F).

Based on the facts of Indonesian consonant clusters, the ability to make direct reference to the edge(s) of a prosodic unit, namely PrWd, is crucial. This allows us to account for the effect of certain phonological processes that are restricted to certain positions.

In conclusion, we have seen that in Indonesian, the phonotactics of the languages at the levels of syllable, root and prosodic word play a major role in the patterning of consonant clusters. We have also seen the ways in which the constraints at different levels are distinct, yet interactive.

#### References:

- Adisasmito-Smith, N. (in progress) *Preliminary acoustic study of Indonesian consonants*. Cornell University.
- Benua, L. (1995) Identity effects in morphological truncation. *UMOP* 18: 77-136.
- Cohn, A. (1989) Stress in Indonesian and bracketing paradoxes. *NLLT* 7: 167-216.
- Cohn, A., and McCarthy, J. (1994) *Alignment and parallelism in Indonesian phonology*. Ms, Cornell University and University of Massachusetts, Amherst.
- Goodman, B. (1996) *Nasal assimilation and vowel epenthesis*. Ms, University of Wisconsin at Madison.
- Keating, P. (1991) Coronal places of articulation. In C. Paradis and J.-F. Prunet (eds.) *The special status of coronals*. (Phonetics and Phonology, vol. 2) San Diego, CA: Academic Press, 29-48.
- Lemontagne, G. and Rice, K. (1995) A correspondence account of coalescence. *UMOP* 18: 211-224.
- Lapoliwa, H. (1981) *A generative approach to the phonology of Bahasa Indonesia*. (Pacific Linguistics, Series D, no. 34) Canberra: Australian National University.
- Lombardi, L. (1995) *Position faithfulness and the phonology of voicing in Optimality Theory*. Ms, University of Maryland.
- Macdonald, R. (1976) *Indonesian reference grammar*. Washington, DC: Georgetown University Press.
- McCarthy, J. and Prince, A. (1993) Generalized alignment. *Yearbook of morphology 1993*: 79-153.
- McCarthy, J. and Prince, A. (1995) Faithfulness and reduplicative identity. *UMOP* 18: 249-384.
- Paradis, C. and Prunet, J.-F. (1991) Introduction: Asymmetry and visibility in consonant articulations. In C. Paradis and J.-F. Prunet (eds.) *The special status of coronals*. (Phonetics and Phonology, vol. 2) San Diego, CA: Academic Press, 1-28.
- Pater, J. (1995) Austronesian nasal substitution and other NC effects. To appear in the *Proceedings of the Utrecht Prosodic Morphology Workshop*.

- Prince, A. and Smolensky, P. (1993) *Optimality Theory, constraint interaction in Generative Grammar*. Technical Report #2 of the Rutgers Center for Cognitive Science, Rutgers University..
- Smolensky, P. (1993) *Harmony, markedness, and phonological activity*. Paper presented at Rutgers Optimality Workshop-1, Rutgers University, October 1993.
- Uhrbach, A. (1987) *A formal analysis of reduplication and its interaction with phonological and morphological processes*. Ph. D. thesis, University of Texas.

# Canonical Types and Noun Phrase Configuration in Fijian\*

John Alderete

University of Massachusetts, Amherst

## 1. Introduction

This paper examines a set of observations governing the distribution of nominal expressions in Fijian clause structure. Special attention is given to the semantic types of the nouns involved, especially in establishing their patterning in the positions for the grammatical functions object and subject. The object position is shown to be more restricted in Fijian, essentially limited to proper names and certain pronominals. This asymmetry is interpreted in a type-theoretic framework as a consequence of the well-motivated syntactic asymmetry, namely that subjects takes scope over objects. In sum, the two main claims put forth in this paper are:

- (i) The restrictions on possible objects and subjects in Fijian may be stated in terms of restrictions of possible semantic types;
- (ii) The observed non-uniformity in the distribution of nominals may be explained as an asymmetry inherited from Fijian clause structure.

### 1.1 Object-Subject Asymmetries

Many contemporary syntactic frameworks posit an asymmetric structural relationship between the predicate and the nominal expressions of a sentence: object NPs generally form a constituent with the predicate which excludes the subject. In the domain of noun phrase quantification, there is a related asymmetry in the way that quantifier phrases are interpreted. Quantifier phrases (QPs) are generally treated as denoting sets of sets (Barwise and Cooper 1980), hence, they are of type  $\langle\langle e\ t\rangle, t\rangle$ . Because of this type specification, subject QPs may combine directly with the intransitive verb phrase via Functional Application in any Montague-style semantics (1a). Interpreting object quantifiers, however, is more complicated, essentially because of the type mismatch between the QP and the local transitive verb of type  $\langle e, \langle e\ t\rangle\rangle$  (1b).

- (1) a.  $[QP_{\langle\langle e\ t\rangle, t\rangle} VP_{\langle e\ t\rangle}]_{IP}$       b.  $[V_{\langle e, \langle e\ t\rangle\rangle} QP_{\langle\langle e\ t\rangle, t\rangle}]_{VP}$

In order to compute the interpretation of an object QP, together with a transitive verb, additional mechanisms are called for (e.g., either Quantifier Raising as employed in Kratzer and Heim 1994, or Type-Lifting as in the approach taken in Partee and Rooth 1983 and Hendriks 1988). The important point here is that the asymmetric syntactic representations for the grammatical functions of a sentence are inherited by the semantic component, which in turn requires special effort in calculating the meaning of an object QP.

The force of this point is not to suggest a weakening of the homomorphism requirements on the mapping from the syntax to the interpretive component. Nor is it to call into question the widely believed assumption that subjects asymmetrically take scope over objects—there are very good reasons for wanting to maintain these assumptions. Rather, this observation will provide the necessary background for posing the question which will serve as the point of departure in this paper. Assuming the canonical or 'unmarked' types for nominal and predicative expressions given in (2), the leading question is, what if a given language lacked the linguistic devices necessary for

---

\*This paper has benefitted from conversations with Roger Higgins, Angelika Kratzer, Peggy Speas, and Ellen Woolford. I would also like to acknowledge Sandra Chung and Albert Schütz for their expert advice on issues related to Fijian and Austronesian syntax. All errors, both empirical and analytical, are my own.

quantifying over nominals in the object position? What would this 'quantificationally defective' language look like?

- (2) CANONICAL TYPES
- |    |                   |   |
|----|-------------------|---|
| a. | quantifier phrase | $\langle\langle e\ t\rangle, t\rangle$  |
| b. | transitive verb   | $\langle e, \langle e\ t\rangle\rangle$ |
| c. | intransitive verb | $\langle e, t\rangle$                   |
| d. | pronoun, trace    | $\langle e\rangle$                      |
| e. | name              | $\langle e\rangle$                      |

The type-theoretic framework makes the clear prediction that the distribution of nominal expressions would be more restricted in the object position than in the subject position. This is because possible objects are restricted to just those elements which may combine directly with the transitive verb expression via function-argument application. Because transitive verbs denote a function from entities of type  $\langle e\rangle$  to predicates of type  $\langle e\ t\rangle$ , only expressions like pronouns, proper names, and potentially null elements, may be interpreted in a position closely bound to the transitive verb. In contrast, the set of possible subjects in this hypothetical language will include all the expressions which are in the set of possible objects, in addition to generalized quantifiers, which may combine directly with the intransitive verb because of its higher type.

- (3) POSSIBLE GRAMMATICAL FUNCTIONS
- |    |                    |   |
|----|--------------------|---|
| a. | Possible objects:  | entities of type $\langle e\rangle$   |
| b. | Possible subjects: | entities of type $\langle e\rangle$ , or $\langle\langle e\ t\rangle, t\rangle$ |

It should be emphasized here that the restrictions on possible objects are semantic in nature, but to be more thorough, the asymmetric syntactic representations for the core grammatical functions also factor in establishing this claim. It is only because the direct object is more closely bound to the transitive verb in the syntax that the semantic restriction applies to this specific grammatical function. The general prediction, then, within this view of the syntax-semantics interface, is that nominals closely bound to transitive verb predicates will be restricted to entities denoting individuals in a language lacking certain quantificational devices. The main goal of this paper is to provide empirical evidence in support of this claim from Fijian clause structure.

## 1.2 *Pronominal Argument Languages*

Before formulating more specific research questions for this paper, it will be helpful to draw a parallel between the semantic restriction on the object position and some of the properties of so called pronominal argument languages. In an effort of provide a more restrictive theory for syntactic parameters of configurationality, a group of researchers have approached a wide range of languages with the guiding assumption that the use of pronominals in a language will be a factor in determining the structural positions for nominal expressions (Jelinek 1984, 1989, 1993, 1995a, 1995b, Baker 1994, 1995, Junker 1994, Heinholtz and Russell 1995). The languages under examination here generally have a rich pronominal system, free, or quasi-free word order, and allow some nominal expressions to freely drop. While the specific proposals differ in the details, they all share one central hypothesis, namely that the agreement markers in these languages are pronominal arguments which function in some sense as the primary arguments of the predicate. This assumption leads to the limitation of noun phrase constituents to positions adjoined to the predicate phrase because the canonical argument positions are supplanted by the agreement morphology. To summarize the results of this hypothesis, the interpretation of the agreement morphemes as pronominal arguments has implications for the syntactic behavior of other nominals in the sentence, effectively forcing the full NPs into adjoined positions where they may freely drop or permute to different positions in the sentence.

Now that some of the syntactic observations characterizing this class of languages have been sketched, we may consider the type-theoretic interpretation of these observations which will

be developed in this paper. Pronominal elements refer to individuals, and so they denote entities of type  $\langle e \rangle$ . In PA languages, the agreement morphemes are analyzed as pronominal arguments, and therefore they will be analyzed as type  $\langle e \rangle$  in the semantic component. With these assumptions in place, the main structural characteristics of pronominal arguments made stated roughly as follows.

(4) PRONOMINAL ARGUMENT LANGUAGES

- a. Nominals more closely bound to the predicate must be of type  $\langle e \rangle$ .
- b. Nominals in a more distant relationship to the predicate may be of a higher type.

The requirement in (4a) restricts the set of elements bound to the verb to pronouns (or potentially proper names), effectively excluding common nouns and quantifiers from acting as arguments of the verb. This fact may now be interpreted on a par with the observation that objects are restricted to elements of type  $\langle e \rangle$  in the hypothetical language described above: these are the only elements which may directly combine with the predicate to yield a well-formed expression for the larger clause. Furthermore, the observation in (4b) may be viewed in the same light as the possibility of interpreting subject QPs *in situ*—in both cases, the position of the noun phrase relative to the predicate enables it to combine with the predicate without the aid of any special semantic rules.

There are a few qualifications to be made before moving on. First, the semantic type restriction on possible arguments for the predicate in a PA language seems to apply only to the internal argument, that is, the argument which will be applied first to the transitive verb. Combining the transitive predicate with an agreement marker results in a meaning for the predicate phrase like that of an intransitive verb:  $\langle e, \langle e \ t \rangle \rangle + \langle e \rangle \rightarrow \langle e \ t \rangle$ . This partially saturated predicate may now combine with either an element of a generalized quantifier type meaning, or an expression of an individual type—both are permitted by the elementary rules of semantic composition. Therefore, the assumptions employed above in the type-theoretic approach would seem to predict, nothing else said, a non-uniform restriction on possible arguments which is not generally attested in PA languages. The state of affairs in Fijian, however, seems to call for such a non-uniform application of the semantic restriction. Possible objects in Fijian are more restricted than possible subjects, justifying the type-theoretic approach to PA languages which is *inherently* asymmetric.

A second clarification, to be elaborated on below, is that the type-theoretic restriction on possible arguments allows a wider class of elements than has been observed to occur in argument positions in PA languages. In particular, the requirement that predicates combine with elements of type  $\langle e \rangle$  not only allows pronominal objects, but also proper names. Names are generally treated as rigid designators, following Kripke 1972, and hence they are of type  $\langle e \rangle$  as well. While the distribution of names relative to pronominal expressions has received little attention in recent work on PA languages, a fundamental observation in Fijian syntax is that proper names pattern like pronominals in a wide range of contexts. For example, object pronouns directly follow transitive verbs (5), as do proper names when the object refers to a specific person in the utterance context (6).

- (5) e    saa    rai-ci    **ira**    a    gone    a    qase  
 3sg   ASP    see-TR   3pl    D    child    D    old person  
 'The old person saw the children' (unmarked interpretation)

- (6) a.    era    rai-ci    **Jone**    tiko    na    gone    (Bauan, Pawley 1983)  
 3pl    see-TR   Name   ASP    D    child  
 'The children are watching John'

- b.    au    aa    tu'u-ni    **Eroni**    vei    Nana Maa    (Boumaa, Dixon 1988)  
 1sg    ASP    tell-TR    Name    P    Name  
 'I told about Eroni to Nana Maa'

The importance of studying this observation alongside other observations restricting the behavior of nouns is that it evidences the generality of the semantic restriction on internal arguments, namely that they may only be expressions of type ⟨e⟩. In sum, Fijian is an interesting case to study with the type-theoretic approach to NP configurationality in mind because of the observed asymmetries in the distribution of nominals and the fact that proper names and object pronouns pattern as a class.

### 1.3 Overview

The rest of this paper has two major sections and a final section which summarizes the main results of the paper. Section 2 presents a sketch of Fijian clause structure and employs the guiding idea that the subject and object markers are pronominal arguments of the verb. This hypothesis will be shown to account for a range of characteristics of Fijian syntax, namely the primacy of the grammatical function (GF) markers, the optionality of full NPs, and the absence of overt case marking. Two problems for this hypothesis are then brought to the fore, namely the divergent phonologies of the GF markers and the semi-predictable character of word order in Fijian. Section 3 proposes to address these problems by giving the GF markers different syntactic classifications, which in turn leads to a non-uniform syntactic parsing of the subject and object. This asymmetric clause structure is finally shown to provide an adequate structure for characterizing the distribution of nominal expressions, correctly limiting possible objects to individual denoting expressions.

## 2. Fijian as a Pronominal Argument Language

In this section, a brief sketch of Fijian clause structure is given (§2.1) and subsequently the Pronominal Argument Hypothesis (PAH) is developed and employed as a means of accounting for some basic features of Fijian syntax. In §2.2, a clear statement of this hypothesis is given, and then it is applied to Fijian in §2.3. The summary in §2.4 lists the successes and failures of using the PAH.

### 2.1 A Sketch of Fijian Clause Structure

Fijian is a Malayo-Polynesian language spoken in the island group of Fiji.<sup>1</sup> Typologically, it is a verb (predicate) initial language with very little inflectional morphology and an elaborate system of pronominal marking. The major type of clause in Fijian contains a predicate head and one or two grammatical function (GF) markers which give information about the subject of the clause, and if the predicate is transitive, information about the object. The subject marker occurs clause initially, and may be separated from the predicate by one or more particles giving temporal or aspectual information about the event described by the sentence. If the information about the object of the verb comes in the form of a GF marker, it directly follows the transitive verb, and this object marker may be followed by a set of post-head modifiers or adverbials. Full NPs may also occur clause finally, and they are interpreted 'in apposition to' a grammatical function marker, meaning that they expand the reference of a GF marker.

(7) GF<sub>SUB</sub> (Tns/Asp) Pred (-TR GF<sub>OBJ</sub>) (Adv) (NP) (NP)

---

<sup>1</sup>Most of the example sentences used in this paper are drawn from the descriptions and texts given in Dixon 1988, and so they represent the western dialect of Boumaa. All of the observations supported by these examples are found in the standard Bauan dialect, and probably many other nonstandard dialects. The references consulted in making these comparisons were: Geraghty 1983, Schütz 1986, Arms 1974, Pawley 1986 [1975], Churchward 1941, and Milner 1956.

The rest of this section will test out the hypothesis that the GF markers are pronominal arguments and that the clause final NPs are in adjunct positions. It will be shown that the morphological analysis of person pronouns in Fijian supports such an analysis, and further, it will have positive consequences the fact that the appositional NPs are always optional and are not marked for morphological case. The semi-predictable word order patterns in Fijian, however, complicates the analysis, which is also discussed below.

Before studying Fijian in greater detail, however, it is necessary to establish the cluster of properties the PAH is intended to account for. This is done directly below.

## 2.2 *One Type of Nonconfigurational Language*

Beginning with Hale 1983, the term 'nonconfigurational' has been used to characterize languages with syntactic features strikingly different from English and other well-known Indo-European languages. While Hale provided a long list of syntactic traits (9), the main tests for diagnosing a language as nonconfigurational (NC) has in practice been: (a) free word order, (b) use of discontinuous expressions, (c) and null anaphora.

### (9) DIAGNOSTICS OF NONCONFIGURATIONALITY (Hale 1983)

- a. free word order
- b. the use of discontinuous expressions
- c. free or frequent pronoun drop
- d. lack of NP movement transformations
- e. lack of pleonastic NPs
- f. use of a rich Case system
- g. complex verb words of verb cum AUX systems

The 'Dual Representation' approach taken in Hale's work correlated these characteristics by defining a model which applied syntactic constraints at two distinct levels of representation, namely the level of lexical structure (LS) and the level of phrase structure (PS). The set of syntactic constraints, chiefly Chomsky's 1981 Projection Principle, is said to hold in NC languages only at the level of LS, allowing, for example, NPs to delete or permute in linear order in the mapping from LS to PS. In sum, the constellation of structural characteristics defining NC languages is the result of a very restrictive theory of the mapping of lexical structures to phrase structures.

The wholesale clustering of these features has been shown to be empirically unjustified on the grounds that languages simply do not group into two discrete groups (see Speas 1990 and references therein). For example, Italian is a language with relatively strict word order patterns and it lacks discontinuous expressions, and yet this language licenses null anaphora and has a rich inflectional system. Malayalam, on the other hand, is a language which might be characterized as a NC language in having free word order and pro-drop, but it doesn't allow discontinuous NPs and has a rather impoverished agreement system. The observation that natural languages may not be distinguished solely on the basis of the categories, configurational and nonconfigurational, has lead Speas to posit a set of distinct parameters responsible for these non-correlating syntactic features.

### (10) CONFIGURATIONALITY PARAMETERS (Speas 1990)

- a. Case Morphemes
- b. Null Topic
- c. Obligatory Agreement
- d. V-to-I in syntax
- e. Incorporated pronouns
- f. LF Focus
- g. KPs modify
- h. pronouns precede antecedents
- i. c/m command

The Pronominal Argument Hypothesis may be viewed as a middle ground between the wholesale correlation of syntactic properties predicted by the Dual Representation approach, and the Configurationality Parameters Theory, which predicts, by and large, no necessary correlations. By making the agreement morphemes the primary arguments of the predicate, obligatory agreement can be positively correlated with a limited set of configurational characteristics.

- (11) THE PRONOMINAL ARGUMENT HYPOTHESIS (PAH)  
 Agreement morphemes are pronominal, acting as the primary arguments of the predicate.

The analysis of the agreement morphemes as pronominal arguments leads to an analysis of the full NPs as adjuncts. The standard account of this analytical move is that the pronominal arguments absorb abstract morphological Case, within the theory of morphological Case outlined in Chomsky 1981, and therefore, since the Case requirements for the NPs may not be satisfied in an argument position, PA languages compel adjunction of the full NPs to the clause (Jelinek 1984, 1993, Baker 1994). One of the main issues to be taken up in section 3 is how to establish the primacy of the pronominal arguments without these Case-theoretic assumptions.

The consequences of the PAH for clause structure may be examined, however, independently from the framework within which it is implemented, so let us now become clear on how the adjunction of the NPs constituents accounts for a cluster of nonconfigurational features.

- (12) A CLUSTERING OF PROPERTIES
- a. Pronominal arguments (~ Rich agreement)
  - b. NPs may freely 'drop' (~ Null anaphora)
  - c. Free word order
  - d. Absence of overtly marked morphological case

Because adjoined NPs are interpreted as modifiers, they are not necessary components of the sentence, and so full NPs in adjoined positions may be freely dropped. Also, free word order is a consequence of NP adjunction. The set of requirements governing the sequencing of words in a sentence has always been more rigorously defined for arguments than for adjuncts. Some caveats should of course be mentioned here, particularly in the distribution of different classes of adverbs, but the central idea is clear: argument positions (A-positions) represent canonical positions for the grammatical functions, whereas, adjoined positions (A'-positions) are less fixed in their linear order (Jelinek 1984 *et seq*). Lastly, the PAH has also been argued to account for the absence of case marking on the full NPs (Baker 1995). Case morphology is usually assumed to be assigned, or checked, under strict structural conditions (e.g., Government in the framework of Government and Binding syntax, or in a Spec-Head relation employed in the Minimalist Program), and these conditions are not met between the predicate head and an adjoined NP. To summarize, the PAH predicts that if the agreement morphemes are analyzed as the primary arguments of the predicate, full NPs can only occupy A'-positions. This results in the cluster of syntactic properties given above in (12).

A few comments are in order regarding some additional patterns observed in PA languages. The PAH has been said to account for the use of discontinuous expressions in various languages, the idea being that more than one NP can refer to the same entity in adjoined positions (Jelinek 1984). This claim has been challenged in recent work, both on empirical and theoretical grounds (Baker 1994, Heinholtz and Russell 1995), and the results seem clear that discontinuous constituency is not a reliable feature of PA languages. Also, it has been claimed that PA languages lack certain types of quantification (in Jelinek 1995a for Straits Salish, and in Baker 1995 for Mohawk), a claim which has been refuted in recent work by Heinholtz and Russell 1995. In light of these findings, it seems more prudent to ignore the use of quantification in these languages for the moment. To close, the set of characteristics listed above in (12) will be the diagnostics employed below in considering the usefulness of the PAH in describing Fijian clause structure.



## 2.3 Pronouns and Configurationality in Fijian

### 2.3.1 The Pronominal System

The first step in applying the PAH to Fijian clause structure is to consider the hypothesis that the agreement morphemes, or what have been referred to above as the grammatical function (GF) markers, may be interpreted as pronominal arguments of the predicate. To this end, let us first examine the GF markers alongside the regular cardinal pronouns (i.e. pronouns which take an article when they head their own noun phrase). The general observation is that the GF markers are either homophonous, or morphologically related to, the cardinal pronouns.

#### (13) PERSONAL PRONOUNS (from Dixon 1988, ignoring some dialect mixing)

SINGULAR	1		2	3
subject	<i>au ~ u</i>		<i>o</i>	<i>e</i>
object	<i>au</i>		<i>i'o</i>	<i>e'a</i>
cardinal	<i>yau</i>		<i>i'o</i>	<i>e'a</i>
DUAL	1 incl.	1 excl.	2	3
subject	<i>(e)taru</i>	<i>'eirau</i>	<i>(o)mudrau</i>	<i>(e)rau</i>
object	<i>'eetaru</i>	<i>'eirau</i>	<i>'emudrau</i>	<i>rau</i>
cardinal	<i>'eetaru</i>	<i>'eirau</i>	<i>'emudrau</i>	<i>(i)rau</i>
PAUCAL	1 incl.	1 excl.	2	3
subject	<i>iou</i>	<i>'eitou</i>	<i>(o)mudou</i>	<i>(e)ratou</i>
object	<i>'etatou</i>	<i>'eitou</i>	<i>'emudou</i>	<i>iratou</i>
cardinal	<i>'etatou</i>	<i>'eitou</i>	<i>'emudou</i>	<i>(i)ratou</i>
PLURAL	1 incl.	1 excl.	2	3
subject	<i>(e)ta</i>	<i>'eimami</i>	<i>(o)munuu</i>	<i>(e)ra</i>
object	<i>'eta</i>	<i>'eimami</i>	<i>'emunuu</i>	<i>ira</i>
cardinal	<i>'eta</i>	<i>'eimami</i>	<i>'emunuu</i>	<i>(i)ra</i>

Ignoring the singular forms for the moment, the object markers are generally the same shape as the cardinal pronouns, with the exception of the optionality of *i* observed initially in the third person forms for the cardinal pronouns. Furthermore, a morphological analysis of the nonsingular forms has been advanced in which the object markers and the cardinal pronouns, on the one hand, relate to the subject markers in that they share the same basic pronominal roots.

#### (14) NONSINGULAR PRONOMINAL ROOTS (Dixon 1988)

	1 incl.	1 excl.	2	3
dual	<i>taru</i>	<i>irau</i>	<i>mudrau</i>	<i>rau</i>
paucal	<i>{ta, Ø}tou</i>	<i>itou</i>	<i>mudou</i>	<i>ratou</i>
plural	<i>ta</i>	<i>imami</i>	<i>munuu</i>	<i>ra</i>

The remainder of the pronominal forms may be analyzed as drawing from the roots in (14) and formed by prefixing the (sometimes optional) vowels shown in (15). In only one case is there a need to resort to disjunctive bracketing, marking the omission of *ta* in the subject marker for the first person inclusive paucal form. The morphological breakdown proposed below for the nonsingular pronouns is again inspired by Dixon's work, but it differs in treating the object markers and cardinal forms as a unified class.

(15) MORPHOLOGICAL ANALYSIS FOR NONSINGULAR PRONOUNS<sup>2</sup>

	1 incl.	1 excl.	2	3
Sub	(e)+R	'e+R	(o)+R	(e)+R
Obj, Card	'e+R	'e+R	'e+R	{i, (i)}+R

To summarize the analysis, while there are a few minor irregularities, the morphological relatedness of the subject markers and the object and cardinal pronouns stems from the assumption that they are formed with the same pronominal roots. The homomorphism observed between the object markers and the cardinal pronouns is reflected in the claim that they employ the same basic roots, and further, that they receive the same prefixal vowels.

Moving now to the singular forms, it is simply noted that the object and cardinal forms differ subtly in the first person pronouns, cf. *au* with *yau* from the chart in (13). The singular subject markers, on the other hand, may generally be monosyllabic, which highlights a second important difference between the subject markers and the object and cardinal pronouns: the members of the latter class always contain at least two vowels. One reflex of this observation is that the object markers and the cardinal pronouns are always self-standing phonological words, whereas the monosyllabic subject markers are either proclitic to subsequent phonological words, or bound to a preceding complementizer (Dixon 1988: §3.2.4). Section 3 will discuss the syntactic implications of the different phonological status of the subject and object markers. At this point it is sufficient to say that the singular personal pronouns do not contradict the main conclusion derived directly above, namely that the object markers and the cardinal forms form a class which excludes the subject markers.

To return to the main theme of this section, the morphological relatedness of the GF markers to the regular personal pronouns is consistent the claim that the GF markers are pronominal arguments. Because all the nonsingular pronouns share a common set of roots (a claim which can be extended to the singular forms), the pronominal status of these forms, presumably encoded in their semantic type, may be attributed directly to the roots, which has the positive consequence of providing for greater parsimony in the grammar of the pronominal system. In sum, the morphological relationships observed across the paradigms listed above supports a treatment of the GF markers as pronominals, on a par with the cardinal pronouns.

### 2.3.2 Appositional Mode

If the Pronominal Argument Hypothesis is correct for Fijian, the analysis of the GF markers as pronominal arguments should compel the adjunction of full NPs, with the effect that their use is optional. Full NPs are never necessary elements of Fijian sentences—this is a fundamental property of Fijian clause structure referred to as 'appositional mode'. The following discussion will show how this mode of sentence construction is employed by first showing how 'basic sentences' (or "predicate phrases") are formed, and then how the clause final full NPs may be added to the basic sentence via appositional mode.

Simple indicative clauses are formed in Fijian by combining a verb (or a predicate of a different word class) with GF markers giving person and number information about the subject and object of the sentence. Also, transitive verbs are formed by applying a transitive suffix (-TR) to a bare intransitive root, as shown in the examples below.

- (16) a.    *au*    *rai*  
           1sg    see  
           'I am looking'
- b.    *au*    *rai-ca*  
           1sg    see-TR  
           'I see (him/her/it)'

<sup>2</sup>Clarification on the bracket notation is in order: the parentheses "(") are used to indicate that the initial vowel is simply optional, while "{" represents the disjunction between prefixing *i* to the object pronouns or the optional prefixation of *i* with the cardinal forms.

- (17) a.    au    la'o  
          1s    go  
          'I am going'
- b.    au    la'o-va  
              1sg   go-TR  
              'I am going for (it...)

Reference to the object of a transitive predicate may be expressed by an object marker which must directly follow the transitive suffix. Object markers are distinguished from cardinal pronouns in that they do not take an article. Thus, *o ira* is unacceptable as the expression of the third person plural object within the basic sentence in (18a).

- (18) a.    au    rai-ci   ira  
          1sg   see-TR 3pl  
          'I see them'
- b.    e    rai-ci   au  
              3sg   see-TR 1sg  
              'He/she sees me'

When the object is third person singular, the transitive suffix *-Ca* may be used (16-17). Elsewhere *-Ci* is used, along with an object marker (18).<sup>3</sup> Abstracting away from this allomorphic variation, the behavior of the GF markers may be stated as follows: (i) every clause contains a subject marker, and (ii) if the predicate is transitive, pronominal reference to the object may come in the form of an object marker, or the *a*-final transitive suffix may be used with the interpretation of the object as third person singular.

These basic sentences may combine with clause final NPs, with the result of expanding the reference of GF marker to which the NP is in apposition. For example, reference to the subject marker in a basic sentence like *era la'o* 'they go', may be further expanded by adding clause final NPs.

- (19) a.    era    la'o   [a gone]  
          3pl   go    D child  
          'The children are going' (lit. They are going, the child)
- b.    era    la'o   [e    walu   a gone lalai   yai]  
              3pl   go    3sg   eight   D child little   Dem  
              'These eight little children are going'

The appositional mode of constructing larger sentences from basic sentences is not limited to the expansion of the reference of the subject marker. As shown by the following examples, a clause final NP may be interpreted in apposition to an object marker. The sentences in (20) are in fact ambiguous, with the anaphoric reference back to a GF marker determined by the context in which the sentence is used.

- (20) a.    e    rai-ca           [a gone]  
          3sg   see-TR       D child  
          'he/she/it sees the child' or 'the child sees him/her/it'
- b.    e    rai-ci   ira   [a gone]  
              3sg   see-TR 3pl   D child  
              'he/she/it sees the children' or 'the child sees them'

<sup>3</sup>The standard analysis of the different forms of transitive suffixes is that third person singular is marked by a morphologically complex form containing the TR suffix and pronominal *a*, and the suffixal vowel *i* is elided in this merging (Arms 1973, Pawley 1986 [1975]).

The ambiguity of reference of the clause final NPs is also evident in rare sentences with two NPs.

- (21) e        rai-ca    [a gone]    [a cauravou]  
       3sg    see-TR    D child     D youth  
       'The youth sees the child' unmarked interpretation  
       'The child sees the youth'

While the clause final NP is typically interpreted as expanding the meaning of the subject marker, the reverse interpretation is also found in texts, indicating that both VOS and VSO interpretations are possible.

This lesson on the appositional mode of sentence construction has summarized two fundamental characteristics of Fijian clause structure, namely that the clause final NPs are optional, and they are interpreted as expanding the reference of obligatory GF markers. These characteristics are positive diagnostics for PA languages because if the GF markers are analyzed as pronominal arguments, the clause final NPs must occur in adjoined positions. This analysis therefore makes the correct prediction that the full NPs should be optional.

### 2.3.3 Word Order

At this point, it seems quite valid to think of Fijian as a PA language: analyzing the GF markers on a par with the regular pronouns leads to an insightful characterization of the pronominal system. Furthermore, this analysis provides an avenue for explaining the appositional process of combining clause final NPs with a basic sentence. The PAH makes the additional prediction that the (adjoined) NP constituents will be freely ordered with respect to each other, and with respect to the predicate. As will be shown directly below, this is not entirely true in Fijian, and will constitute one of the complicating factors in the analysis of its clause structure.

It is difficult to postulate a basic word order pattern for Fijian, at least in examining the distribution of the full NPs. One problem is that sentences with more than one NP constituent are extremely rare. In the textual study conducted by Dixon, only about 2 or 3 percent of the clauses in texts had explicit object and subject NPs. And of those rare cases, roughly one in four sentences had fronted NPs, yielding the pattern [ NP Pred NP ]. Some conclusions may tentatively be drawn, however, from these sentences and the elicitation order of the NP constituents.

When a clause contains both an appositional subject and object NP, one observed pattern is for the appositional subject NP to be clause final. The following examples show this VOS word order pattern.

- (22) a. T4.170: 323  
       saa    'ila-a            sara    gaa    [ a 'aa]        [ o mata.ni.vanua ]  
       ASP   know-TR        MODIF MODIF    D thing        D messenger  
       'Then the messenger knew the thing'
- b. T4.19: 334  
       Ia,    ni            saa    rogo-ca    [ o 'ea a+i- rogo qoo ]    [ o Raatuu-i    ca'au ],  
       Well, WHEN    ASP   hear-TR    D 3sg D+    news THIS     D Title-POSS reef  
       'Well, when the Raatuu of the reef heard the news, ...'
- c. T6.89a: 345  
       ... saa tu'u-nati'o    [ a o-na        vosa ]    [ o 'ea ]  
       ASP tell-TR ASP     D CLASSIF-3sg word D 3sg  
       '...(at this time), he spoke the following words'

Also, in elicitation sessions, Dixon found that his consultants most often gave VOS word order for sentences with more than one NP.

VSO word orders are also found, as with the following textual examples.

- (23) a. T6.32: 336  
 ...saa mani soli-a vuaa [ o Paatere Lorosio ] [ o 'ea a kuruse qoo ] ...  
 ASP THEN give-TR TO+3sg D Father Name D 3sg D cross THIS  
 '...and Father Lorenzo then gave this cross to him'
- b. T4.199: 326  
 Saa+qei faele.-ta'ina [ o 'ea ] [ a o-na moto ] ...  
 ASP+THEN file-TR D 3sg D CLASSIF-3sg spear  
 'Then he filed his spear...'
- c. T4.183: 325  
 Saa la'o mai, saa mai tu'u-na [ o 'ea ] [ a 'aa  
 ASP come HERE ASP COME tell-TR D 3sg D thing  
 e tu'u-na mai [ o Tui Waini'eli ] ] ...  
 3sg tell-TR HERE D King Place  
 'He has come here, he has come and reported the things that the kind of Waini'eli  
 has said...'

Dixon concludes that a basic word order may not be determined on the basis of pattern frequency—sentences with two NPs are too infrequent, and sentences with VOS orders occurred with roughly the same frequency as those with VSO patterns. The elicitation order is suggestive, however, of a basic VOS pattern, with the VSO order being derived by means of a rule of rightward NP shift. While there doesn't seem to be firm empirical ground in which to base a conclusion about the underlying word order, Dixon's tentative conclusion seems reasonable.

The distribution of NP constituents relative to so called 'peripheral constituents' (i.e., NPs introduced by prepositions that specify semantic roles like goal, addressee, instrument, etc.), is less restricted, however, as shown by the following examples. In the sentences directly below, a peripheral constituent (PC) precedes a NP in apposition to a subject marker.

- (24) a. T4.43: 310  
 Saa soli sara [ vei rau ] [ a 'ee-drau i'a ],  
 ASP give MODIF TO 3du D CLASSIF-3du fish  
 'amu sa la'o yadi.-va'i maa  
 THAT KIND ASP go choose-PASS THAT  
 'Their fish, which had been selected for them, was given to the two of them.'
- b. T4.44: 310  
 ...[ o ra ] viro wale mai [ i+na 'oro ] [ o ra a marama ] ...  
 D 3pl return MODIF HERE TO+D village D 3pl D woman  
 '...and they the women had to return home to (their) village...'
- c. T4.113: 318  
 Io, 'oto [ i qere ] [ o i'o tagane ]  
 Yes, lie AT THERE D 2sg man  
 'Yes, just keep lying there, you, fellow.'

The opposite order is attested in the next set of examples, however, showing that the position of the appositional subject NP relative to PCs is not strictly governed.

- (25) a. T4.45: 310  
 Saa tu'u-na ti'o [ o Raaluve.ni.Waini'eli ] [ vei Raavouvou.ni.Boumaa ]  
 ASP tell-TR ASP D Title TO Title  
 me 'ua ni va'a-.bera.bera.-ta'i i yaa,  
 SHOULD ASP THAT MAKE-slow-PASS AT THAT

a oo-drau                    va'a-.wati        ee?  
 D CLASSIF-3du            MAKE-spouse TAG  
 'The Raaluve of Waini'eli told the Raavouvou of Boumaa that it should not be  
 delayed, their marriage, eh?'

b. T4.72: 314

La'o yane o rau saa    yaco ti'o    [ a maarau ]        [ mai Narova ]  
 go    THERE D 3du ASP    happen ASP    D celebrations        AT Place  
 'The two (youths) went there, and the festivities were in full swing at Narova.'

c. T4.114: 318

Rubica sobu [ o 'ea ]        [ i+na sautabu ]..  
 go fast    DOWN            D 3sg            TO+D chiefly graveyard  
 'He hurried down to the chiefly graveyard, ...'

While the texts in Dixon 1988 do not provide sufficient data to show the same word order behavior for appositional object NPs relative to PCs (obviously due to the rare cooccurrence of both constituents in the same sentence), Dixon clearly states (p. 244) that the same free word order is attested in this case, "Peripheral NPs may come in any order—both among themselves and in relation to the core constituents, subject and object."

Returning to the predictions of the PAH, the semi-predictable word order patterns fleshed out in the above discussion are only partially consistent with the expectations of this hypothesis. While word order permutations were observed in deriving a VSO pattern from the basic VOS order, and the patterning of the peripheral constituents relative to clause-final NPs was shown to be relatively free, Fijian still shows a strong preference for predicate initial clauses. These results should make one cautious in the applying the PAH to Fijian clause structure, as the assumptions inherent to the PAH predict free word order for the core constituents of the sentence.<sup>4</sup>

#### 2.3.4 Absence of Case Marking and Agreement

One last observation relevant to the adjunction of NPs predicted by the PAH is the absence of morphological case, or any kind of inflectional marking for that matter, on the appositional NPs. Fijian has no case morphemes and so the clause final NPs are not marked overtly for morphological case.

Neither person nor number markings are present on the appositional NPs either. As shown by the following examples, the number of the entities referred to by the GFs are determined solely by the pronominal markers closely bound to the predicate.

<sup>4</sup>It seems relevant to mention a rule of leftward NP shift, 'Fronting', because of its potential for word order permutation. Fronting is a syntactic operation which may topicalize any NP constituent to a clause initial position. This process typically involves rendering the fronted constituent more rhematic, effectively bringing the NP into the foreground of the discourse context, as Topicalization does in many languages. Consider the following example in which the appositional object in (i) is fronted in (ii).

- i.    au        via        talanoa.-ta'ina        [ a+i- talanoa lailai ] [ vei 'emudrau ]  
 1sg    WANT tell-TR            D    story    little    TO    2du  
 'I want to tell a little story to you two.'
- ii.    [ a+i- talanoa lailai ] au        via        talanoa.-ta'ina        [ vei 'emudrau ]  
 D    story    little    1sg    WANT tell-TR            TO    2du  
 'There is a little story that I want to tell to you.'

Despite the potential Fronting may have for creating predicate initial word orders, it seems best to ignore this process in our study of configurationality in Fijian. First, Dixon's textual study shows that the patterns resulting from this process are relatively marked. Moreover, Fronting is clearly a discourse rule with observable results in the topichood of the NPs involved, which is clearly different than the object shifting rule yielding VSO orders from an

(26)	e	la'o	[ a gone ]	'the child is going'
	erau	la'o	[ a gone ]	'the two children are going'
	eratou	la'o	[ a gone ]	'the (few) children are going'
	era	la'o	[ a gone ]	'the (many) children are going'
	SUB	go	D child	

The absence of case marking and agreement morphemes on the appositional NPs is relevant because, in the predicted adjunction structures, the necessary conditions for said marking (as defined above in §2.2) are not met. This observation is therefore consistent with the PAH.

#### 2.4 Summary

The following chart summarizes the results arrived at in the previous section, and compares the characteristics of Fijian clauses to those found in Mohawk (Baker 1995).

(27)	<u>DIAGNOSTICS FOR PA LANGUAGES</u>	<u>FIJIAN</u>	<u>MOHAWK</u>
a.	Rich (Pronominal) Agreement +	+	+
b.	Optional full NPs	+	+
c.	Free Word Order	?	+
d.	No Case Morphology	+	+

The obligatory GF markers are interpreted as pronominal arguments of the predicate (§2.3.1), on a par with the regular cardinal pronouns, warranting a "+" for the first diagnostic of a PA language. The appositional mode of sentence construction described in §2.3.2 showed that the clause final NPs are optional, again positively diagnosing Fijian as a PA language. The discussion of word order in §2.3.3 was less conclusive, but it seems clear that Fijian has quasi-free word order, permitting VOS and VSO patterns. The order relative to the appositional NPs and the peripheral constituents was totally free, however, suggesting that an alternative analysis in which the subject and object NPs occupy A-positions is unlikely to be successful. The quasi-free character of these word order patterns is reflected in the "?" given in (27c). Lastly, the absence of case morphology further identifies Fijian as a PA language. This cluster of properties permits a comparison with the Iroquian language Mohawk, characterized in Baker 1994, 1995 as a pronominal argument language.

The general conclusion of this section is that the application of the Pronominal Argument Hypothesis to Fijian successfully explains a set of structural characteristics found in this language. There is a residue, however, which remains unaccounted for by the PAH, and these concerns will be addressed in the next section. First, Fijian word order is not totally free: the major type of clause examined in this section strongly prefers verb initial sentences, and further, the elicitation order is subject final (VOS). Second, there were two differences observed between the subject markers and the object markers. The object markers formed a morphological class which excluded the subject markers. Also, the phonological size requirements of the object markers do not apply to the subject markers. The next section will deal with these differences, and the quasi-free word order, by giving the GF markers a different syntactic classification, making an interesting set of predictions about their semantic and syntactic behavior.

### 3. The Role of Canonical Types in Deriving Appositional Mode

The analysis of Fijian clauses has been so far, to a certain extent, pre-theoretical. The PAH was informally introduced as a means of accounting for a clustering of syntactic features, and at the same time, identifying certain problems that this hypothesis raises. In this section, a formal framework is provided for deriving the results established in the characterization of Fijian as a PA

---

underlying VOS structure.

language. This framework, which employs a type-theoretic vocabulary within a Montague-style semantics, will be shown to have a range of empirical consequences which distinguish it from the syntactic framework in which the PAH is traditionally implemented.

### 3.1 *Type-driven Appositional Mode*

#### 3.1.1 Formal Semantics

To begin, recall that the position directly following a transitive verb is reserved for the object pronouns. While subsequent discussion will show that certain other individual type expressions are also found in the object slot Verb+TR \_\_\_\_\_, this position is more restricted than the clause final positions for nominals. In particular, common nouns never directly follow transitives.<sup>5</sup>

#### (30) THE OBJECT SLOT

Verb + TR \_\_\_\_\_  
                   Object Pronoun  
                   \*Common Noun

This restriction is accounted for by the PAH by claiming that the object markers are pronominals and that they constitute the primary (internal) arguments of the predicate. Within a Case-theoretic frameworks the PAH is typically worked out, however, a fundamental question remains unanswered, namely, why *agreement markers* function as the primary arguments, and not, for example, some other piece of inflectional morphology. The main thesis to developed directly below addresses this question by paying special attention to the semantic types of the nominal and predicative expressions involved.

Maintaining the assumption inherent to the PAH that the GF markers are pronominal arguments, the unmarked semantic value for these pronouns will be variables ranging over the domain of individuals, hence they will be of type  $\langle e \rangle$ . In denoting individual type variables, the GF markers are distinguished from common nouns, which, on their canonical interpretation, denote simple intransitive predicates, i.e., functions from individuals to truth values:  $\langle e \ t \rangle$ . Transitive predicates, which in the ensuing discussion will be equated with transitive verbs, denote functions from individuals to intransitive predicates. These semantic types are the widely assumed unmarked semantic values for pronouns, transitives, and intransitives (Heim and Kratzer 1994, Partee and Rooth 1983, Partee 1987).

#### (31) CANONICAL TYPES

a.	Transitive Verb	$\langle e, \langle e \ t \rangle \rangle$
b.	Common Noun	$\langle e \ t \rangle$
c.	Pronoun (and Traces)	$\langle e \rangle$

These canonical types are reflected in the following lexical entries for the relevant word classes involved.

---

<sup>5</sup>Fijian does have a process of Object Incorporation which produces lexical compounds resembling this verb phrase construction, but Object Incorporation is only productive for a very small set of predicates. Further, when a common noun forms a compound with the predicate root, the transitive suffix is conspicuously absent, e.g. *ta'i-va* 'to fetch (something) in a container', *ta'i.wai* 'fetch water'. The absence of the transitive suffix and the weak productiveness of this rule suggests that it is an earlier syntactic process, quite different from the positioning of proper names in the object slot, which is fully productive, and involves the transitive morphology.



## (32) A SAMPLE LEXICON

## A. Transitive Verbs

- i.  $[[ \text{rai-ci} ]] = f: D_e \rightarrow D_{\langle e, t \rangle}$   
For all  $a, b \in D_e$ ,  $f(a)(b) = 1$  iff  $b$  looks for  $a$ .
- ii.  $[[ \text{'aci-vi} ]] = f: D_e \rightarrow D_{\langle e, t \rangle}$   
For all  $a, b \in D_e$ ,  $f(a)(b) = 1$  iff  $b$  calls  $a$ .

## B. Common Nouns

- i.  $[[ \text{gone} ]] = f: D_e \rightarrow D_t$   
For all  $a \in D_e$ ,  $f(a) = 1$  iff  $a$  is a child.
- ii.  $[[ \text{qase} ]] = f: D_e \rightarrow D_t$   
For all  $a \in D_e$ ,  $f(a) = 1$  iff  $a$  is an elder.

## C. Pronouns

- i.  $[[ \text{ira}_x ]]^g = \text{them}$
- ii.  $[[ \text{au}_y ]]^g = \text{me}$

The relevance of these entries to formulating the restrictions on nominals in the object slot is that the domain of the function denoted by transitives is type  $\langle e \rangle$ , and only pronouns, not common nouns, match this in semantic type. This distinguishes the two nominals in the formal semantics of verb phrase interpretation, as demonstrated in the following computations.

## (33) COMPUTATION FOR TRANSITIVE VERB PHRASES

A.  $[[ \text{Verb+TR GF}_{\text{OBJ}} ]]$ 

1.  $\text{rai-ci}^* = \lambda x_e \lambda y_e [ \text{look for } (x) (y) ]$
2.  $\text{ira}^* = \text{them}$
3.  $(\text{rai-ci ira})^* = \lambda y_e [ \text{look for } (\text{them}) (y) ]$   
From (1) and (2) by Functional Application.

B.  $[[ \text{Verb+TR Common Noun} ]]$ 

1.  $\text{rai-ci}^* = \lambda x_e \lambda y_e [ \text{look for } (x) (y) ]$
2.  $\text{gone}^* = \lambda x_e [ \text{child } (x) ]$
3.  $(\text{rai-ci gone})^*$  is undefined.

Object pronouns combine straightforwardly with transitive verbs by Functional Application (see for example Kratzer and Heim 1994 for the formulation of this rule) because these word classes translate into expressions of the right semantic type (33A). If a verb phrase, however, composed of a transitive verb plus a bare common noun is translated into the semantic component, this constituent yields a type mismatch, requiring special mechanisms to yield the right results (33B). The assumption, therefore, that Fijian lacks the syntactic or semantic rules necessary for interpreting these constructions (e.g., Quantifier Raising or Type Shifting, as discussed in the introduction) explains the lack of such verb phrases: the semantic rules cannot operate on these structures, and so they cannot receive a coherent interpretation in the semantic component of the grammar. Only pronominals may combine with transitives to form a verb phrase because the verb and its object must combine directly via function-argument application, and only pronouns are of the correct semantic value to achieve this. Thus, the assumed semantic types derive the restriction

characteristic of PA languages, namely that nominals closely bound to the verb must be pronominal. In the next subsection, the observation that proper names pattern like the pronouns in the object position will establish the generality of the restriction that the object marker be of type  $\langle e \rangle$ .

The semantic framework being sketched here also correctly characterizes the behavior of nominals which occur in full noun phrases. Thus, the semantics of clause final NPs which refer back to a GF marker is essentially the approach to NP quantification taken in Kratzer and Heim's work. In considering the following sentence, the object pronoun 'ea s/he' and the NP *a gone* 'the child' are co-indexed to indicate the appropriate anaphoric relation in the interpretation of this sentence.

- (34) a.     e       rai-ci       'ea     [ a gone ]  
           3sg    look-TR    3sg     D child  
           'He looks for the child'
- b.     [ e rai-ci 'ea<sub>1</sub> ]<sub>1</sub> [ a gone ]

Now, indices act like pronouns, they are variables which range over individuals.<sup>6</sup> The index on the clause final NP acts as a lambda abstractor, binding the index on the object pronoun, very much like semantic binding of pronouns by co-indexed quantifier phrases. Consider the following computation for the 'quantifying in' of the appositional NPs.

(35) COMPUTATION OF APPOSITIONAL NOUN PHRASES

1.     ( e rai-ci 'ea<sub>1</sub> )<sup>\*</sup> = [ look for (x<sub>1</sub>) (he) ]
2.     <sub>1</sub> = x<sub>e</sub>
3.     ( e rai-ci 'ea<sub>1</sub> )<sub>1</sub><sup>\*</sup> = λx<sub>e</sub> [ look for (x<sub>1</sub>) (he) ]  
        From (1) and (2) by Lambda Abstraction.
4.     a<sup>\*</sup> = λx<sub>\langle e t \rangle</sub> [ f<sub>\langle e t \rangle</sub> (x) ]
5.     gone<sup>\*</sup> = λx<sub>e</sub> [ child (x) ]
6.     a gone<sup>\*</sup> = λx<sub>\langle e t \rangle</sub> [ child (x) ]  
        From (4) and (5) by Functional Application.
7.     ( [ e rai-ci 'ea<sub>1</sub> ]<sub>1</sub> a gone )<sup>\*</sup> = [ look for (the child) (he) ] = 1 iff  
        he looked for the child.  
        From (3) and (6) by Functional Application.

The basic sentence in (34b) *e rai-ci 'ea* 'he looks for him' denotes the proposition specified in (35.1), which is made into a predicative expression by means of Lambda Abstraction in (35.3). Assuming the meaning for the clause final NP given in (35.6), this nominal can be quantified in via Functional Application, yielding the right interpretation for this complex sentence formed by appositional mode. Indeed, the structures resulting from the appositional mode mirror those resulting from Quantifier Raising in the interpretation of quantifier-variable binding.

Now that the semantic framework has been developed for characterizing the necessary restrictions on the object slot, and for interpreting the appositional NPs, the questions raised at the close of section 2 may be addressed in tandem. First, the problems identified above concerning the

<sup>6</sup>Indices act like pronouns in the semantics for a very obvious reason which is somewhat obscured by the discussion above. Pronouns are referential, and so all pronouns bear indices. It's these indices which translate into variables ranging over the domain of individuals, not the pronouns themselves. This is the sense in which traces and pronouns have the same semantic function.

different morphological behavior of the GF markers will be addressed, and the solution to this set of problems will suggest a way of accounting for the semi-predictable word order patterns. The first problem was that the object markers seemed to form a class with the cardinal pronouns, with regard to their morphological shape, and that this morphological class excluded the subject markers (§2.3.1). The null hypothesis for the word class of cardinal pronouns is that they are nouns (N) because they head NPs beginning with a determiner (D). And so, because the object markers are homophonous with the cardinal pronouns, it seems sensible to assume that the object markers are also nouns. This of course accords nicely with the syntactic assumptions tacitly made in the above account of the behavior of the object markers: they are assumed to be pronominal arguments which form a constituent with the verb stem. Both object pronouns and cardinal pronouns may receive the same syntactic category, N, in their respective positions in the clause.

### 3.1.2 The Thesis of NonUniformity

If the object markers and cardinal pronouns are characterized as Ns, how are they to be distinguished from the subject markers, which systematically assume a different form? The proposal to be defended immediately below is that the subject markers are the outputs of different word formation processes than those generating the object and cardinal pronouns. Recall from section 2.3.1 that both the GF markers shared the same pronominal roots, given in (14), but they differ in the prefixal vowels which they receive. These results are summarized in the following morphological analysis, repeated from (15).

#### (36) MORPHOLOGICAL ANALYSIS FOR PERSONAL PRONOUNS

	1 incl.	1 excl.	2	3
GF <sub>SUB</sub>	(e)+R	'e+R	(o)+R	(e)+R
GF <sub>OBJ</sub>	'e+R	'e+R	'e+R	i+R

The morphological constituents which form the GF markers are argued to be the output of the following lexical rules, which render the object markers as Ns, and the subject markers as the functional category Agr.

#### (37) LEXICAL RULES OF PRONOMINAL SYSTEM

A. GF <sub>SUB</sub> as Agr	B. GF <sub>OBJ</sub> as N
(i) [ (e)+R ] → Agr[ 1 incl. ]	(i) [ 'e+R ] → N[ 1 incl. ]
(ii) [ 'e+R ] → Agr[ 1 excl. ]	(ii) [ 'e+R ] → N[ 1 excl. ]
(iii) [ (o)+R ] → Agr[ 2 ](iii)	[ 'e+R ] → N[ 2 ]
(iv) [ (e)+R ] → Agr[ 3 ](iv)	[ {i}+R ] → N[ 3 ]

The choice of distinguishing the GF markers in terms of syntactic categories is initially plausible on the basis of their phonological behavior. Recall from §2.3.1 that the object markers always form independent phonological words, whereas the subject markers may be monosyllabic and are usually proclitic to the following particles or verbal stems or merged with a preceding complementizer. If the subject markers are distinguished from the pronouns by being members of a closed inflectional class, their apparently aberrant phonological behavior can be adequately accounted for. (The next subsection will develop this argument more explicitly).

#### (38) SYNTACTIC CLASSIFICATION FOR PERSONAL PRONOUNS

a. GF <sub>SUB</sub> :	Agr
b. GF <sub>OBJ</sub> :	N
c. Cardinal Pronoun:	N

Furthermore, the classification of the object markers and the cardinal pronouns as noun heads, and the subject markers as Agr heads, provides an avenue for addressing the problem concerning the semi-predictable word order patterns in Fijian clauses. If the subject marker is analyzed as the head of an inflectional category—while it may function semantically as an argument of the verb—it will not be a syntactic argument because it does not occupy an A-position, e.g., the specifier position of Agr. The proposal given above therefore has the consequence of giving a *non-uniform* treatment of the configurational features of nominals. As shown in the schematic representation below, the object markers are pronominal arguments, occupying the argument position directly following the verb, which has the effect of forcing the appositional NP into an adjoined position within the VP. The subject markers are not pronominal arguments, however, and so an appositional NP referring back to the Agr head must occupy a fixed final position.

(39) Fijian Clause Structure

$$\text{Agr}_k \ [ \ [ \text{V+TR NP}_i \ ] \ (\text{NP}_i) \ ] \ \text{NP}_k$$

The Agr head occurs clause initially and is bound by the clause final subject. Further, the N head occurs post-verbally and may be bound by an adjoined appositional NP. Thus, the asymmetric approach sketched here has the consequence that the subject position is an A-position, while the position occupied by an appositional object NP is an A'-position.

The non-uniform treatment of the GF markers therefore, in one stroke, addresses the morphological differences observed among the personal pronouns and posits a canonical clause final position for subjects. Thus, the representation above will yield the basic VOS pattern for the core sentence constituents, and it is assumed that Object Shift may permute the adjoined object to a final position to yield the VSO pattern. In sum, the mixed approach to the configurationality of nominal expressions distinguishes the GF markers morphologically, and at the same time, derives the semi-predictable word order of Fijian clauses.

### 3.2 Further Implications of NonUniformity

The syntactic characterization of the person pronouns in Fijian whereby the GF markers are treated as different word classes has further implications for the distribution of proper names, interrogative pronouns, and the phonological word status of the person pronouns. These are taken up directly below, providing further empirical support of the non-uniform approach to pronouns and NP configurationality in Fijian.

#### 3.2.1 Pronouns and Proper Names

The semantic restriction on the object slot proposed above was that only nominal expressions of type ⟨e⟩ may combine directly with a transitive verb. This restriction was responsible for ruling out bare common nouns from occurring in this position, while correctly allowing the object pronouns, which are of this semantic type. Proper names are often assumed to be rigid designators, meaning essentially that they pick out the same individual across all various possible worlds (Kripke 1972). Rigid designators are of type ⟨e⟩, and so if the object position is truly open to expressions of this semantic class, the prediction is that proper names should pattern with pronouns in the object slot. That is, they should directly follow the *i*-final transitive suffixes within the verb phrase. This prediction is in fact borne out in Fijian, which is part of a general pattern in which proper names behave in similar ways to the person pronouns.

First, proper names and person pronouns, on the one hand, are distinguished from common nouns by the kind of article they take when they head a full noun phrase. Common nouns, like *oro* 'village' take the article *a*, as shown in the following examples.

- (40) a oro                      a gone  
 D village                  D child  
 '(the) village'            '(the) child'

On the other hand, pronouns and proper names, both place names and person names, take the article *o* when they head their own NP.

- (41) a.    o Waitabu                      b.    o yau  
           D Place                            D 1sg  
           'Waitabu'                           'me!'  
           o Jone                              o ira  
           D Name                              D 3pl  
           'John'                                'them!'

The fact that pronouns and names take the same article, which is different from the one used with common nouns, supports the standard characterization of these nominals as taking the same semantic type. If the article *o* will apply to both proper names and pronouns, the domain of the function denoted by *o* will necessarily characterize the type of both names and pronouns (without, of course, writing two different lexical entries for *o*).

#### (42) LEXICAL ENTRIES FOR ARTICLES

##### a. Pronoun and Name Article

[[ *o* ]] = f:  $D_e \rightarrow D_e$   
 For all  $a, b \in D_e$ ,  $f(a) = b$  iff  $a$  is  $b$ .

##### b. Common Noun Article

[[ *a* ]] = f:  $D_{\langle e, t \rangle} \rightarrow D_{\langle e, t \rangle}$   
 For all  $a, b \in D_{\langle e, t \rangle}$ ,  $f(a) = b$  iff  $a$  is  $b$ .

The semantics of the two articles here will not be discussed any further, as their meanings are still somewhat controversial (Dixon 1988, Schütz 1985, cf. Arms 1974). The essential point here is that the assumption that names and pronouns have the same semantic value permits a unitary formulation of the interpretation of the article *o*.

The assumption that pronouns and names have the same canonical type also has the consequence that names and pronouns should have similar syntactic behavior. In this context, a repair to the observation sketched in the introduction is in order. Recall that it was noted that proper names, referring both to persons and locations, may occur directly following the transitive verb when the verb root is marked with an *i*-final transitive suffix. The sentences below, drawn from Dixon's grammar and texts, give ample exemplification of this fact, which is widely attested in the standard dialect, as well as in other Fijian dialects (Schütz 1985, Pawley 1986 [1975], Geraghty 1983).

#### (43) a. Dixon 1988 : 267 (23.1)

au aa tu'u-ni            Eroni vei Nana Maa  
 1sg ASP tell-TR        Name P Name  
 'I told Nana Maa about Eroni'

#### b. T4.88:315

Saa la'o yane e dua me            la'i 'aci-vi Raaluve.ni.Waini'eli  
 ASP go THERE 3sg one SHOULD GO call-TR Title  
 'One (person) went to call the Raaluve of Waini'eli'

c. T4.53:311  
 O i'o 'ila-i Narova  
 D 2sg. know-TR Place  
 'Do you know Narova?'

d. T4.175:324  
 la, sa+na mai 'aba-ti Boumaa o Waini'eli  
 WELL ASP+FUT COME besiege-TR Place D Place  
 'And finally, that Waini'eli would come and fight (lit: besiege) Boumaa'

e. T4.183:325  
 ni o ira sa+na mai 'aba-ti Boumaa  
 THAT D 3pl ASP+FUT COME invade-TR Place  
 '... that they would come and fight (lit: invade) Boumaa'

f. T4.192:326  
 Ra saa la'o mai me+ra mai 'aba-ti Boumaa  
 3pl ASP go HERE SHOULD+3pl COME besiege-TR Place  
 'They were to come here, come and fight Boumaa'

g. T6.21:335  
 Saa la'o mai, me la'o mai me mai rai-ci Tui.Ca'au  
 ASP go HERE SHOULD go HERE SHOULD COME see-TR Title  
 'He came here, came here to come and see Tui Ca'au (title of Raatuu Golea)'

h. T4.40:310  
 Tala.talanoa to'a sa—o Raavouvou.ni.Boumaa saa taro-gi Raaluve.ni.Waini'eli—  
 Redup-chat ASP ASP—Art Title ASP ask-TR Title  
 Raaluve.ni.Waini'eli me saa wati-na.  
 Title SHOULD ASP spouse-3sg  
 'They chatted away—and then the Raavouvou of Boumaa asked the Raaluve of Waini'eli to be his wife.'

i. T6.74:342  
 [...] qoo saa cabe-ti Viti mai, e aa cabe mai  
 THIS ASP come ashore-TR Place HERE 3sg PAST come ashore HERE  
 i+na ucu-na qoo  
 AT+D peak of land-3sg THIS  
 '... he had come ashore on Fiji, he had come ashore at this peak of land'

This observation is in fact one of the motivating factors in Arms 1973 and Pawley 1986 [1975] for treating the *i*-final transitive suffixes as the basic form, and the *a*-final ones as the result of combining the suffix with a third person pronoun *a*. Because the *i*-final form is used in a wider range of environments, namely with proper names, all other object markers, in passive constructions, and with reflexive verbs, this form is assumed to be underlying.

The implications of this observation for the issues at hand is that it attests the full range of possible objects predicted by the type-theoretic approach. The interpretation of basic sentences like the one in (44) are computed on a par with the meanings of sentences where the object pronoun directly follows the verb.

(44) e rai-ci Jone  
 3sg look-TR Name  
 'He is looking for John'

Because names denote rigid designators they are of the semantic type  $\langle e \rangle$ , and so they can combine directly with the transitive verb without the need of any semantic rules other than function-argument application.

(45) COMPUTATION OF PROPER NAME IN OBJECT POSITION

1.  $\text{rai-ci}^* = \lambda x_e \lambda y_e [\text{look for } (x) (y) ]$
2.  $\text{Jone}^* = \text{John}$
3.  $(\text{rai-ci Jone})^* = \lambda y_e [\text{look for } (\text{John}) (y) ]$   
From (1) and (2) by Functional Application.
4.  $e^* = \text{he}$
5.  $(e \text{ rai-ci Jone})^* = [\text{look for } (\text{John}) (\text{he}) ] = 1$  iff  
he is looking for John.  
From (3) and (4) by Functional Application.

It will be noted in the summary of this analysis that this consequence of the type-theoretic approach to restricting the object slot is not shared by plausible alternatives to it, in particular Case-theoretic approaches to PA languages, or syntactic theories of null anaphora.

Proper names only occur within the predicate phrase in the object slot, thus names do not substitute for the subject markers as they do for the object markers. This represents one additional reason for distinguishing the GF markers in syntactic category. A name in sentence initial position would yield a perfectly coherent interpretation: just as the agreement marker *e* combines with the VP in (45), the intransitive predicate denoted by the VP may combine with a name to yield a proposition. The failure of names to pattern with the subject marker must therefore be accomplished in the syntax, as it is, by restricting this position to Agr heads. Names are Ns, and so they are predicted not to occur clause initially.

### 3.2.2 Interrogative *cei*

The interrogative pronoun *cei* is employed in questions which call for an answer that is a name, a pronoun, or a common noun with human reference. Like pronouns and proper names, when *cei* heads a NP it follows the article *o*. If the lexical entries given above are correct, *cei* is predicted to behave like other pronouns of type  $\langle e \rangle$ . In particular, the expectation is for *cei* to substitute for proper names and pronouns in the object slot, and what is more, *cei* should not pattern with the subject markers because they are of a different syntactic category. This is in fact true: *cei* may directly follow *i*-final transitive suffixes in interrogatives where the object is being questioned (46), but when the subject GF is questioned, a subject marker is employed, and an independent NP headed by *cei* is fronted to a clause initial position (47).

- (46) o aa rai-ci cei  
D 2sg look-TR WHO  
'Who did you see?'

- (47) a. [ o cei ] e sabi-ci i'o  
D WHO 3sg hit-TR 2sg  
'Who hit you?'

- b. T4.119:318  
Saa taro-ga [ o ira ] se [ o cei ] e ca'a-va ti'o,  
ASP ask-TR D 3pl COMPL D WHO 3sg do-TR ASP

[ o cei ]	e	ca'a-va	ti'o	[ a	'aa	maa ]
D WHO	3sg	do-TR	ASP	D	thing	that

'They asked who had done it, who had done that thing.'

This final observation exhausts the survey of nominals which take the article *o*, and furthermore, it is exactly these elements which substitute for each other in the position directly following transitive verb. Thus, other interrogatives, for example *cava* 'what', always form NPs with the article *a*, and do not substitute with pronouns and names in the object slot. The semantic framework developed above predicts precisely this kind of behavior, both in the failure of *cei* to pattern with the subject Agr head (they are of different syntactic categories), and in the failure of other interrogatives which pattern with the common nouns to occur in the object slot (they are of different semantic type).

(48) CORRELATIONS WITH INTERROGATIVE PRONOUNS

- |    |                                 |  |
|----|---------------------------------|--|
| a. | <i>cei</i> <sub>(e)</sub> :     | Behaves like object pronouns and names |
| b. | <i>cava</i> <sub>(e, t)</sub> : | Behaves like common nouns              |

While a full account of the semantics of questions in Fijian would stray too far from the main line of argumentation here, the observed correlations given above are consistent with the proposed restriction on nominal semantic types. Interrogatives with *cei* in the object slot will be treated on a par with the rest of the pronouns: *cei* will saturate the internal argument of the predicate denoted by the transitive verb. Conversely, *cava* will not combine with transitives because their predictive type meaning is not of the right type to combine directly with transitives of type  $\langle e \langle e t \rangle \rangle$ . In sum, the interrogative pronouns provide further evidence for the semantic and syntactic restrictions employed in describing the distribution of nominals in basic sentences.

3.2.3 Differential Phonological Word Status of the Person Pronouns

The GF markers are outputs of different morphological rules which are responsible for the different shapes they assume, and also for associating the person pronouns with different syntactic categories. The subject agreement marker is classified as the inflectional head Agr, while the object pronouns are categorized as Ns, on a par with the regular pronouns. These different syntactic classes also make possible an important distinction which will play a role in deriving the different phonological behavior of these pronominals.

The major differences lay in the behavior of the singular pronouns, which are repeated from (13) below.

(49) SINGULAR PERSONAL PRONOUNS

	1	2	3
subject	<i>au ~ u</i>	<i>o</i>	<i>e</i>
object	<i>au</i>	<i>i'o</i>	<i>e'a</i>

The first observation distinguishing the two classes is that only the subject forms may be composed of less than two vowels. This is related to a second important observation, namely that the object and cardinal forms always constitute an independent phonological phrase, receiving stress on the penultimate syllable. The nonsingular subject personal pronouns usually have this status, but the singular forms are aberrant in their phonological behavior, being omitted or latching on to a neighboring word in a range of contexts (Dixon §3.2.4). For example, the third person subject pronoun *e* can never be more than a clitic to the next word, and it is dropped after a monosyllabic complementizer, e.g., *me* or *ni*. The second person subject marker exhibits a slightly different pattern of behavior, merging with the complementizers, e.g., [*me o*] → *mo*, and being proclitic to subsequent words in the elsewhere case. Lastly, the first person subject marker is *au* underlyingly, but drops the first vowel after a monosyllabic complementizer and coheres to it. Thus, [*ni au*] → *niu*, and [*me au*] → *meu*.



The differences in the minimal size of the pronominals, and consequently, their divergent phonological phrasing, may be accounted for by subjecting the open class forms to a word minima requirement commonly used in phonological analyses of this kind of phenomenon. Because all words are composed of at least one metrical foot, the phonological size constraint, Foot Binariness, will require that every word of the relevant class be composed of at least two moras (where this pair is either distributed over two syllables, or contained in one syllable composed of a long vowel or a diphthong).

- (50) Foot Binariness (McCarthy and Prince 1986, Hayes 1987)  
Feet must be binary at the syllabic or moraic level of analysis.

The cardinal and object pronouns are analyzed as nouns, and are thus categorized as members of an open word class. Foot Binariness will therefore apply to these pronominals, accounting for the observation that their lexical forms are always composed of at least two vowels (moras). This in turn accounts for the uniform status of the object and cardinal pronouns as independent phonological words: because they are all binary at some level of prosodic analysis they may support a well-formed foot, and hence, they satisfy the minimal requirement for prosodic word status (phonological word in Dixon's parlance).

The subject pronouns, on the other hand, are analyzed as Agr heads, and are thus inflectional morphemes of a closed word class. Foot Binariness is not, by hypothesis, operative in this class, predicting the observed subminimal lexical forms. All of the behavior of the monosyllabic forms described by Dixon, deletion, proclisis, cohesion with preceding complementizers, may now be motivated by the need for the monosyllabic pronouns to be parsed as well-formed phonological words in the phrasal phonology. It is common cross-linguistically for closed class and open class items to diverge in this way (Hayes 1995), providing outside evidence for this classification based on word minima observations. To conclude, the syntactic categories employed above in restricting the distribution of pronouns and names to the object slot provides an adequate word classification in accounting for their differential phonological status.

#### 4. Summary of Results

This section summarizes the results established in the analysis developed above, and briefly compares this analysis with a Case-theoretic treatment of configurationality in Fijian.

In the previous section, the proposal was advanced that the GF markers are the result of different morphological rules. These rules are responsible for the different morphological shapes observed in contrasting these two classes of pronominals, and also they classify the GF markers as different syntactic categories. This syntactic classification was shown to have a wide range of consequences, influencing both the phonological behavior of the different pronoun classes and their syntactic distribution when compared and contrasted with other nominals with a similar semantic type. The following list summarizes the assumptions made above regarding the word class and semantic type of the relevant nominal expressions.

- (51) SYNTACTIC CATEGORIES AND CANONICAL TYPES FOR NOMINALS
- |    |                           |     |       |
|----|---------------------------|-----|-------|
| a. | Common Noun               | N   | <e t> |
|    | Interrogative <i>cava</i> | N   | <e t> |
| b. | GF <sub>OBJ</sub>         | N   | <e>   |
|    | Names                     | N   | <e>   |
|    | Interrogative <i>cei</i>  | N   | <e>   |
| c. | GF <sub>SUB</sub>         | Agr | <e>   |

The observation that the object markers, proper names, and *cei* all patterned as a class post-verbally is accounted for in this analysis by their syntactic and semantic characterization (51b). They are exactly those nominals which may directly follow an *i*-final transitive verb and occur in

NPs headed by the article *o* because they refer to individuals, i.e. Ns of type ⟨e⟩. The individual type nominals are distinguished formally in this system from the common nouns and *cava* by means of their canonical type (51a). This correctly predicted that they are excluded from occurring in the object slot, and that they use a different article when they head their own NPs, i.e., *a* not *o*. Only the nominals in (51b) are licensed in the post-verbal position in transitive constructions because only they can be applied directly to the denotation of the transitive verb in the interpretative component of the grammar and yield a well-formed expression for the larger verb phrase.

In contrast to this semantic restriction on the denotation of post-verbal nominals, the distinction between the subject marker and the object markers is established in the syntax. For example, the account of the failure of names and interrogative *cei* to substitute for the subject marker inside a basic sentence is due to their different syntactic categorizations. The subject markers are analyzed as closed class functional categories (51c). This distinction in terms of word class had two further consequences: (i) it provides a coherent classification for distinguishing the phonological status of the GF markers, and (ii) it played a role in deriving the predictable component of the word order patterns observed for the core NPs constituents of the sentence.

An apparent drawback of this non-uniform syntactic classification is the related asymmetry with respect to null anaphora: the subject GF should be an obligatory expression (it occupies an A-position), while the object GF is optional (because it occupies a A'-position). This prediction, however, is actually not true, once the semantics of appositional NPs is carefully defined. Succinctly, the optionality of the appositional NPs is not derived from their structural positions; rather, NPs are optional because of their semantic function, and it is their semantics which derive their configurational properties. Thus, in previous analyses of this parameter, adjuncts are analyzed as restrictive modifiers, and so they are not necessary elements of the sentence. The semantics of the appositional NPs defined in section 3.1 is clearly compatible with this result. Hence, the constituent *Agr'* is a well-formed expression, denoting a proposition because the pronominal *Agr* head saturates the external argument of the verb. Further, the appositional subject NP co-indexed to the *Agr* head is interpreted exactly on a par with the appositional object NP: they both function to give more information about the GF marker (whether this is defined as restrictive modification, or some other semantic process, is not clear at present). The important point here is that if the optionality of NPs in the clause stems from its semantic function, as has always been the assumed, then the results in the case of Fijian appositional NPs are in fact quite uniform. To conclude, the syntactic classification necessary in the account of the distribution of nominals will not lead to non-uniformity in the configurational traits of the appositional NPs, if these traits are governed in the semantic framework like the one developed here.

It should be clear at this point how the results derived above distinguish the type-theoretic approach to the distribution of nominal expressions from a Case-theoretic account. The type-theoretic approach to the distribution of nominal expressions provides a genuine explanation of the patterning of pronouns and proper names post-verbally, stemming from the fundamental categories necessary for any compositional semantics. In Case-theoretic frameworks for implementing the Pronominal Argument Hypothesis, this syntactic patterning is completely unexplained. The syntactic categories at the disposal in this theory do not provide the right classification for characterizing the observation that pronouns and proper names pattern as a class in the object position. Further, an appeal to an incorporation analysis of names closely bound to the transitive verb will presumably admit common nouns in this position as well, incorrectly extending the set of nouns to this semantic class. Indeed, this observation, which is rather fundamental in Fijian syntax, seems to call unmistakably for special attention to the semantic types of these nominals.

A second component of the canonical types approach which makes it different from the alternatives is its reliance on object-subject asymmetries in governing the distribution of nouns. The syntactic structures mapped in the interpretive component provide the semantics with transitive verb phrase constituents which do not include subjects. This representational asymmetry is indeed essential to characterizing the semantic restriction on post-verbal nominals, as different syntactic constituency for the grammatical functions without this property would surely have different consequences. Thus the semantic framework employed here derives a strong prediction related to this syntactic asymmetry, namely that the class of nominals permitted as object GFs should be

more restricted than the set of possible subject GFs. As emphasized in the above discussion, the set of possible objects in Fijian is clearly more narrowly defined than the set of possible subjects. Possible objects, that is nominal complements to V+TR stems, are restricted to pronouns, proper names, and traces. Possible subjects on the other hand, i.e., NPs occurring in the specifier position of Agr, are all of the possible object nouns, NPs formed with common nouns, and potentially generalized quantifier type expressions.<sup>7</sup>

In comparing this result with the predictions of the standard syntactic approach to describing PA languages it is fair to ask, why do the agreement morphemes uniformly absorb Case, and consequently predict a set of configurational results for both the subject and object? Nothing inherent to the syntactic approach derives this result, and so, in the context of the observed asymmetries in Fijian, one is left wondering how to proceed. This is not to say that syntactic principles will not emerge as a way of accounting for the unbalanced distribution of nominal expressions in Fijian. For example, the bifurcation of the pronominal system in terms of their syntactic category proposed in this paper may provide an avenue for accounting for the observed object-subject asymmetries. On the semantic approach, however, the asymmetric distribution of nouns comes as a direct consequence of the type-theoretic denotations fundamental to the property of compositionality embedded in this framework.

## REFERENCES

- Arms, David G. (1973). Whence the Fijian Transitive Endings. *Oceanic Linguistics* 12: 503-58. Papers of the First International Conference on Comparative Austronesian Linguistics.
- Arms, David G. (1974). Transitivity in Standard Fijian. Doctoral Dissertation, University of Michigan.
- Baker, Mark (1991). On Some Subject/Object Non-Asymmetries in Mohawk. *Natural Language and Linguistic Theory* 9: 537-576.
- Baker, Mark (1994). The Polysynthesis Parameter. Ms. McGill University.
- Baker, Mark (1995). On the Absence of Certain Quantifiers in Mohawk. In Emmon Bach et al, eds., *Papers in Quantification*. Dordrecht: Kluwer.
- Bonvillain, Nancy (1973). A Grammar of Akwesasne Mohawk. *Ethnology Division Paper No. 8*. Ottawa: National Museum of Man.
- Bresnan, Joan and Sam Mchombo (1986). Grammatical and Anaphoric Agreement. *CLS* 22, part 2.
- Cammack, Floyd M. (1962). Bauan Grammar. Doctoral Dissertation, Cornell University.
- Chomsky, Noam. (1981). *Lectures in Government and Binding*. Dordrecht: Foris.
- Churchward, C. Maxwell (1941). *A New Fijian Grammar*. Sydney: Australasian edical Publishing Company.
- Dixon, R.M.W. (1988). *A Grammar of Boumaa Fijian*. Chicago: The University of Chicago Press.
- Geraghty, Paul. (1983). *The History of the Fijian Languages*. *Oceanic Linguistics Special Publications no.19*. Honolulu: University of Hawaii Press.
- Gordon, Lynn (1980). Raising in Bauan Fijian. Paper read to the LSA annual meeting, San Antonio, Texas.
- Hale, Ken (1983). Warpiri and the Grammar of Non-Configurational Languages. *Natural Language and Linguistic Theory* 1: 5-47.

---

<sup>7</sup>Noun phrase quantification in Fijian has not been studied carefully within the formal framework employed in this paper, so the relevance of this point for Fijian has yet to be established. Preliminary research on the language seems to indicate an avoidance of using determiners like *few* or *many*, these English expressions being translated into Fijian with the use relative clauses. Thus, for example, *few villages* would be translated in Fijian as *e dua a 'oro*, which means literally, 'the village which is few'. Perhaps this mode of quantifying over nouns reflects a general pattern in the language, but at this point not a lot is known on this topic.

- Hayes, Bruce (1987). A Revised Parametric Metrical Theory. *Northeastern Linguistic Society* 17: 274-89.
- Hayes, Bruce (1995). *Metrical Stress Theory. Principles and Case Studies*. Chicago: The University of Chicago Press.
- Hazlewood, David (1850). *A Feejeean and English and an English and Feejeean dictionary [...], and A Compendious Grammar of the Feejeean Language*. Vewa: Wesleyan Missionary Press.
- Heinholtz, Charlotte and Kevin Russell (1995). Quantified NPs in Pronominal Argument Languages: Evidence from Swampy Cree. In Jill Beckman, ed., *Northeastern Linguistic Society*. Amherst: Graduate Linguistic Student Association.
- Hendriks, H. (1988). "Type Change in Semantics: The Scope of Quantification and Coordination," In E. Klein and J. van Benthem (eds.), *Categories, Polymorphism and Unification*. Edinburgh/Amsterdam: CCS/TTLI.
- Jelinek, Eloise (1984). "Empty Categories, Case, and Configurationality. *Natural Language and Linguistic Theory* 2: 39-76.
- Jelinek, Eloise (1993). Ergative "Splits" and Argument Type. In Jonathan Bobaljik and Colin Phillips, eds., *MIT Working Papers in Linguistics* 18: 15-42.
- Jelinek, Eloise (1995a). Quantification in Straits Salish. In Emmon Bach et al, eds., *Papers in Quantification*. Dordrecht: Kluwer.
- Jelinek, Eloise (1995b). Predicates and Pronominal Arguments in Straits Salish. *Language* 70: 697-736.
- Junker, Marie-Odile (1994). Syntaxe du Quantifieur universel en Algonquin. *Canadian Journal of Linguistics* 39: 1-13.
- Kratzer, Angelika and Irene Heim (1994). *Introduction to Semantics in Generative Grammar*. [Forthcoming MIT Press].
- Kripke, S. (1972). Naming and Necessity. In D. Davidson and G. Harman, eds., *Semantics of Natural Language*. Dordrecht: Reidel.
- Lounsbury, Floyd (1953). *Oneida Verb Morphology*. New Haven: Yale University Press.
- Massam, Diane (1985). *Case Theory and The Projection Principle*. Doctoral Dissertation, MIT.
- McCarthy, John and Alan Prince (1986). *Prosodic Morphology*. Ms. University of Massachusetts, Amherst and Brandeis University.
- Michelson, Gunther (1973). *A Thousand Words of Mohawk*. Ethnology Division Paper No. 5. Ottawa: National Museum of Man.
- Milner, George (1956). *Fijian Grammar*. Suva: Fijian Government Press.
- Partee, Barbara and Mats Rooth (1983). Generalized Conjunction and Type Ambiguity," In R. Bäuerle, C. Schwarze, and A. v. Stechow (eds.), *Meaning, Use and Interpretation of Language*. Berlin: de Gruyter.
- Partee, Barbara Hall (1987). Noun Phrase Interpretation and Type-Shifting Principles. In J. Groenendijk, D. de Jongh, and M. Stokho, eds., *Studies in Discourse Representation and the Theory of Generalized Quantifiers*. Dordrecht: Foris.
- Pawley, Andrew K. (1986 [1975]). A Reanalysis of Fijian Transitive Constructions. *Te Reo* 29: 81-112.
- Schütz, Albert (1986). *The Fijian Language*. Honolulu: University of Hawaii Press.
- Speas, Margaret (1990). *Phrase Structure in Natural Language*. Dordrecht: Kluwer.

# NUMBER MARKING IN SELAYARESE

Hasan Basri

SUNY at Stony Brook &  
Tadulako University Indonesia

## 1. Introduction

The nature of number marking is usually induced from the kind of base the number marker is morphologically attached to. If the base is a noun, then it is considered to be nominal number marking. This is the case in most Indo-European languages, where nouns inflect to show number. English belongs to this group. Mithun (1988) points out that number can also be the property of verbs, judging from morphological attachment. Using data from various North American languages she demonstrates how verbs can have their own number marking, which does not always agree with the number marking on the subjects. Based on this she claims that number marking is not limited to nominals, as has been traditionally assumed, but can be extended to verbs as well, serving as a kind of aspectual marker, indicating multiplicity of events or states.

I will show that Selayarese has another type of number marking. The main issue that I will address is the nature of a formative *-i*, which, I claim, serves as a nominal number marker. Therefore I will call it PLURAL *-i*. I will demonstrate that the plural *-i* never attaches to nouns morphologically, though it normally quantifies them semantically. I will also discuss briefly another formative *-i* which seems to behave differently from the plural *-i* in terms of the type of base it attaches to and in its semantic contribution to the base. I will argue that, in spite of their identical phonological form, this formative *-i* performs a different function in that it quantifies an event or state, showing an iterative or durative aspect of some sort. I will call the verbal number marker the ITERATIVE *-i*, covering both iterative and durative aspects. Thus we are dealing with two types of number marking, a nominal number marking and a verbal number marking.

The paper will focus primarily on the morphological nature of the plural-*i*, its morphological attachment to the base, the kinds of bases it can attach to and its position on the base with respect to other formatives. More precisely, I will argue that the plural *-i* belongs to the clitic group syntactically, but it has its own idiosyncracies which make it different from clitics in the most usual sense. It functions like a clitic syntactically in that it moves from one host to another, rather than sticking to a particular lexical category. However, the plural *-i* is a well-behaved affix phonologically in that it attracts the stress and changes the word-final glottal stop [ʔ] to [k]. Therefore, prosodically, it becomes part of the base, like stress-shifting suffixes.

In terms of its morphological attachment to the base, the plural *-i* can be accounted for by the relevance criterion of Bybee's (1985) in that it follows the formative whose meaning is more relevant to the base, and normally obeys the Mirror Principle of Baker (1985) in the sense that it follows a formative which reflects an

earlier syntactic operation. However, when phonological consideration is in conflict with semantic justification with respect to its position on the verbal complex, it is the phonological consideration that always wins, resulting in violation of the relevance criterion or the Mirror Principle. Such a phenomenon can be accounted for by postulating that phonological considerations outweigh semantics in the ordering of morphological material on the verbal complex.

## 2. Selayarese Plural Marking

### 2.1. Plural Marker on Verb Bases

Let me make clear several things that might cause some confusion. First, whenever the occurrence of the plural marker *-i* creates a sequence of two identical vowels, a glottal stop [ʔ] is automatically inserted to break up the sequence. This glottal stop insertion applies to any sequence of identical vowels (not limited to the plural *-i*). Second, only future tense is marked morphologically in the tense system of the language. Thus any verb unmarked for future can as well be interpreted as present or past. Third, for practical reason, the male pronoun *he/his/him* is used to refer to a third person singular on the English gloss, which of course can also be *she/her* or *it/its*, so long as pragmatics is not violated. Fourth, for ease of reference, I use two types of boundary symbol, a [-] before stress-shifting formatives, and a [#] before stress neutral ones. These boundary symbols, however, do not suggest any tendency to a particular theoretical affiliation. Let us now look at the data in (1):

- (1)
- |    |                       |                              |
|----|-----------------------|------------------------------|
| a. | <i>as-sa'ssa#ko</i>   |                              |
|    | itr-wash#2A           | 'You did your laundry.'      |
| b. | <i>as-sassa'-i#ko</i> |                              |
|    | itr-wash-pl#2A        | 'You (pl) did your laundry.' |
| c. | <i>ap-pa'llu#i</i>    |                              |
|    | itr-cook#3A           | 'He cooked.'                 |
| d. | <i>ap-pallu'-i#ʔi</i> |                              |
|    | itr-cook-pl#3A        | 'They cooked.'               |

Data in (1) indicate that the plural *-i* on the verb does not contribute any semantic content to the base, rather it shows that the subject is plural. This might lead us to assume that the plural *-i* is a kind of inflectional agreement, similar to the English third singular agreement *-s*. However, if the *-i* is an inflectional agreement, then it must be obligatory when its context is appropriate. By appropriate context I mean the context where the participant the plural marker is construed with is unambiguously plural. Yet, this does not seem to be the case, for the plural *-i* is not required if the absolutive argument is overtly plural. Consider the data in (2).

- (2) a. *a?-li'ngka#?i*  
itr-walk#3A 'He (they) walked.'
- b. *al?-li'ngka'-i#?i*  
itr-walk-pl#3A 'They walked.'
- c. *a?-li'ngka#i i-Baso? na i-Ali*  
itr-walk#3A clas-Baso? and clas-Ali  
'Baso and Ali walked.'
- d. *a?-lingka'-i#?i i-Baso? na i-Ali*  
itr-walk-pl#3A clas-Baso? and clas-Ali  
'Baso and Ali walked.'
- e. \* *a?-lingka'-i#kang*  
itr-walk-pl#1plA 'We walked.'

We can see in (2a) that the plural marker is absent, and as such the subject is unspecified for plurality (i.e. ambiguous). In (2b) the plural marker is needed to disambiguate the unspecified number of the subject. In (2c) the plural marker is absent, for the subject referred to by the absolutive agreement *-i* is overtly plural already. In (2d) the presence of plural *-i* does not create any ungrammaticality (though it is slightly marginal). In (2e), on the other hand, it causes ungrammaticality. The ambiguity of (2a), the grammaticality of (2c) and (2d), and the ungrammaticality of (2e) give an indication that the plural *-i* is not obligatory, and as such, it is unlikely to belong to inflectional agreement.

If the plural marker is not a kind of inflectional agreement, then we need to investigate if it is a kind of plural inflection like the English plural *-s/-es*. For this we need to see how the plural marker behaves with respect to quantified nouns. Consider the data in (3):

- (3) a. *sa?a ba'kka?*  
snake big 'a big snake'
- b. *sa?a bakka'k-i*  
snake big-pl 'big snakes'
- c. *tallu sa?a ba'kka?*  
three snake big 'three big snakes'
- d. *tallu sa?a bakka'k-i*  
three snake big-pl 'three big snakes'
- e. \* *tallu sa?a'-i*  
three snake-pl 'three snakes'
- f. \* *tallu sa?a'-i ba'kka?*  
three snake-pl big

The data in (3) tell us two important things. The first is that the plural marker is optional when the nominal construed with it is already quantified, as exemplified by the grammaticality of both (3c) and (3d). If the *-i* were a plural inflection, then (3c) would be predicted unacceptable. As a matter of fact (3c) is not only grammatical but even

sounds better than (3d), which is somewhat redundant. The second thing that we can see in the data in (3) is that the plural marker is disallowed to attach to an  $N^{\circ}$ , as shown by the ungrammaticality of (3e) and (3f). This evidence makes it clear that the Selayarese plural marker is dissimilar to the English plural inflection in that the latter is not only obligatorily present on quantified nouns (at least on well-behaved noun stems) but is also obligatorily attached to the nouns. In short, the English plural marker is required, hence an inflection, while Selayarese plural marker is unobligatory, hence not an inflection by definition.

## 2.2. The position of the plural marker on the verbal complex

Bybee's theory of relevance implies that a formative whose meaning affects or modifies the meaning of the base is more likely to be closer to the base than a formative whose meaning does not affect the meaning of the base. This is more or less in line with The Mirror Principle of Baker (1985) which says: "Morphological derivations must directly reflect syntactic derivations (and vice versa)." In light of these two principles let us see how the plural marker places itself on the base with respect to other formatives. Consider the data in (4)

- (4)
- a. *ge'ge#ko*  
laugh#2A 'You laughed.'
  - b. *gege'-i#ko*  
laugh-pl#2A 'You (pl) laughed.'
  - c. *ku-keke'k-i#ko*  
1sgE-laugh-trn#2A 'I laughed at you.'
  - d. \* *ku-gege'-i#ko* 'I laughed at you.'
  - e. *ku-kekek-i'-?i#ko*  
1sgE-laugh-trn-pl#2A 'I laughed at you (pl).'

We can see in (4) that there are two different forms of the verb *laugh*, the intransitive *gege* and the transitive *keke?*. The *-i* in (4b) is a plural marker that quantifies the subject represented by the absolutive *#ko*. In (4c) on the other hand, the *-i* is a transitivizer which licenses the object represented by the absolutive *#ko*, while the subject is represented by the ergative agreement *ku-* 'I.' (4d) is unacceptable because the verb is intransitive. (4e) shows that the inner *-i* must be the transitivizer and the outer *-i* is the plural marker. The idea behind this is that transitivizer *-i* affects the meaning of the verb (i.e., add more valence to the verb), while the plural *-i* does not contribute any semantic element to the verb. By virtue of the relevance principle it is then reasonable to assume that the transitive marker should occupy a slot closer to the verb than the plural marker does. It is still unclear, though, why the plural marker precedes the absolutive agreement if semantic contribution is the primary justification for the ordering. Both formatives are semantically irrelevant to the base, though. However, following logic, it will be ideal morphologically if a formative which refers to an entity precedes a formative which quantifies the entity. Or could it be because the



plural marker is stress-shifting while the absolutive agreement is stress-neutral? I will claim that this is indeed the case. As we can see, this is the first instance where phonological consideration outranks semantic justification.

Let us look at the following data where the plural marker co-occurs with the benefactive *-ang* and the inchoative *#mo*. Consider the data in (5):

- (5) a. *ku-ha'lli#?i pao'-njo*  
1sgE-buy#3A mango-the  
'I bought the mango(es).'
- b. *ku-halli'-ang#ko pao'-njo*  
1sgE-buy-ben#2A mango-the  
'I bought the mango(es) for you.'
- c. *ku-halli-a'ng-i#ko pao'-njo*  
1sgE-buy-ben-pl#2A mango-the  
'I bought the mango(es) for you (pl).'
- d. *ku-halli-a'ng-i#mo#ko pao'-njo*  
1sgE-buy-ben-pl#inch#2A mango-the  
'I am already buying the mango(es) for you (pl)'

The benefactive *-ang* in (5b) licenses a new argument represented by the absolutive agreement *#ko* 'you,' while the other absolutive agreement, the *#i*, which refers to the old argument, does not show up. As expected, the benefactive *-ang* precedes the plural marker in (5c). This phenomenon follows from the Mirror Principle in that the new argument must first be licensed by the benefactive *-ang* before it can be quantified by the plural marker. However, (5d) cannot be accounted for by the relevance principle. If semantic contribution is the primary criterion for closeness to the base, then the inchoative *#mo* should be the closest to the base, for *#mo* is a kind of aspectual marker, whose meaning is the most relevant to the meaning of the base. The fact that *#mo* is two layers apart from the base is an indication that the relevance principle is violated. Or, again, could it be due to the fact that *#mo* is stress-neutral, therefore it must yield its place to the stress-shifting formatives irrespective of the semantic element those formatives contribute to the base? As I claimed before, this is indeed the case. The data in (5d) provide a second piece of evidence of the superiority of phonological consideration to semantic justification with respect to the ordering of morphological materials on the verbal complex.

From the data in (4) and (5) we can draw a generalization about the position of the plural marker *-i* on the verbal complex: it follows the other stress-shifting formatives, and precedes the stress-neutral ones. Data in (4) and (5) also exhibit the superiority of phonological criterion to the semantic consideration in the ordering of formatives on the verbal complex.

The following data present a problem to the above analysis, however. Consider the forms below:

- (6)
- |    |   |  |
|----|---|--|
| a. | <i>la'ngkasa</i>                            | 'tall'                                       |
| b. | <i>langka's-ang</i>                         |  |
|    | tall-comp                                   | 'taller'                                     |
| c. | <i>langkas-i'-ang</i>                       |  |
|    | tall-pl-comp                                | 'taller' (with plural subjects)              |
| d. | <i>ba'llo</i>                               | 'beautiful'                                  |
| e. | <i>ballo'-ang</i>                           |  |
|    | beautiful-comp                              | 'more beautiful.'                            |
| f. | <i>ballo-i'-ang</i>                         |  |
|    | beautiful-pl-comp                           | 'more beautiful' (with plural subjects)      |
| g. | <i>langkas-i'-ang ana'k-ku na ana'?'-mu</i> |  |
|    | tall-pl-comp                                | child-1pos and child-2pos                    |
|    |   | 'My children are taller than your children.' |

What is peculiar about (6c) and (6f) is the fact that the plural marker, which gives no semantic contribution whatsoever to the base, precedes the comparative marker, which does modify the base. Here both formatives are stress-shifting, so phonological justification cannot be used to account for the unusual ordering. Or could it be the case that there is a kind of phonological rule requiring a formative which ends in a consonant to occupy the final slot on the formative ordering irrespective of the semantic modification it makes for the base? This does not seem to be a very promising morphophonological solution, for in the case of benefactive *-ang* on the verbal complex (5d), the benefactive *-ang* does precede the plural *-i*. I will leave this issue open for further study.

### 2.3. Plural marker with an adjective base

The following data show how the plural *-i* attaches to adjectives and where it stands on the adjectival complex with respect to other formatives. Consider the data in (7)

- (7)
- |    |                     |                 |
|----|---------------------|-----------------|
| a. | <i>la'ba?#i</i>     |                 |
|    | wide-3A             | 'It is wide'    |
| b. | <i>la'ba'k-i#?i</i> |                 |
|    | wide-pl#3A          | 'They are wide' |
| c. | <i>la'mbere#i</i>   |                 |
|    | long#3A             | 'It is long'    |
| d. | <i>lambe'r-i#?i</i> |                 |
|    | long-pl#3A          | 'They are long' |

As we can see in (7), the plural marker precedes the absolutive agreement (7b) and (7d), changes the word final glottal stop [ʔ] to [k] (7b), and the epenthetic vowel does not show up in (7d). In the data below we can see that the plural marker precedes the definite article *-njo*, which is also stress-shifting. Consider the following forms:

- (8)
- a. *sapo'-njo*  
house-the 'the house'
  - b. *sa'po bakka?-i'njo*  
house big-the 'the big house'
  - c. *sa'po bakkak-i'-njo*  
house big-pl-the 'the big houses'
  - d. \* *sapo-i'-njo*  
house-pl-the

Notice also how the plural marker *-i* behaves differently from the epenthetic [i] in (8b and 8c) with respect to the word-final glottal stop [ʔ]. The [i] preceding the definite article *-njo* in (8b) is epenthetic. It is needed to break up the potential sequence of triple consonants, which is disallowed by the phonotactics. As we can see it does not affect the preceding glottal stop. Compare (8b) with (8a), where an epenthetic [i] is not necessary, for the derivation does not produce triple consonants. The *-i* in (8c), on the other hand, is a plural marker; and as such it not only attracts the stress but also changes the word-final glottal stop [ʔ] to [k]. The ungrammaticality of (8d) again provides good support for the claim that although the plural marker normally quantifies nouns semantically, it does not attach to the nouns morphologically.

Based on the facts that (1) the plural marker does not contribute any semantic element to the base, (2) it is not obligatory even if the appropriate context is met, and (3) it does not always obey the relevance principle with respect to its occurrence on the base, I will assume that the plural marker is unlikely to belong to inflectional agreement, and more unlikely to be a derivational expression.

If this is the case, then we have to explore further to see if the plural marker can be classified as a free grammatical expression. For this we have to look at other kinds of bases the plural marker can attach to.

#### 2.4. Plural Marker with Modal Auxiliaries

In the following data we will see that the plural marker can also attach to modal auxiliaries *a?ra?* 'willing' or *untu?* 'unwilling.' For the sake of simplicity, only *a?ra?* is presented in the following data. Consider the data in (9):

- (9)
- a. *at-to'long#i*  
itr-sit#3A 'He sat.'
  - b. *a?ra?#i at-to'long*  
willing#3A itr-sit 'He was willing to sit.'
  - c. *a?ra?#i at-tolo'ng-i*  
willing#3A itr-sit-pl 'They were willing to sit.'
  - d. *a?ra'k-i#?i at-to'long*  
willing-pl#3A itr-walk 'They were willing to sit.'
  - e. \* *a?ra'k-i at-to'long#i*  
willing-pl itr-sit-3A

We can see in (9) the interplay between the plural *-i* and the absolutive agreement. The absolutive agreement always attaches to the modal auxiliary as shown by the ungrammaticality of (9e). The plural marker may stay on the verb as in (9c) and may also raise to the modal auxiliary together with absolutive agreement as in (9d). There is no significant difference in meaning between (9c) and (9d). To my personal intuition, (9d) sounds more natural.

## 2.5. Plural Marker with Adverbs

The plural marker can also attach to adverbs, particularly adverbs of manner and adverbs of place. As in the case of modal auxiliaries above, the absolutive agreement attaches to the verb when the adverb is postverbal (10a). When the adverb is preverbal, the absolutive agreement raises to the adverb (10b), while the plural marker can stay on the verb (10b) or follows the absolutive to raise to the adverb (10c). Neither the absolutive agreement nor the plural marker attaches to a postverbal adverb (10d and 10e). This phenomenon shows a tendency for the absolutive agreements to occupy the second position (P2) (Wackernagel's law). Consider the data in (10):

- (10) a. *ap-pa'llu#ko ri'nni*  
itr-cook#2A here 'You cooked here.'
- b. *ri'nni#ko ap-pallu'-i*  
here#2A itr-cook-pl 'You (pl) cooked here.'
- c. *rinni'-?i#ko ap-pa'llu*  
here-pl#2A itr-cook 'You (pl) cooked here.'
- d. *ap-pallu'-i#ko rinni*  
itr-cook-pl#2A here 'You (pl) cooked here.'
- e. \* *ap-pa'llu rinni'-?i#ko*  
itr-cook here-pl#2A 'You (pl) cooked here.'

In the following examples we will see how the plural marker attaches to adverb of manner. I should mention that adverbs of manner always occur preverbally. Thus in terms of predication, they behave like adjectives. As in the data in (10) above, the plural marker can stay with the verbs (11b) or raise to the adverbs together with the absolutive agreement (11c).

- (11) a. *la'ssiri#?i a?-li'ngka*  
fast#3A itr-walk 'He walked fast.'
- b. *la'ssiri#?i a?-lingka'-i*  
fast#3A itr-walk-pl 'They walked fast.'
- c. *lassi'r-i#?i a?-li'ngka*  
fast-pl#3A itr-walk 'They walked fast.'

## 2.6. Plural Marker with Pronominals and *wh*-words

The plural marker can also attach to pronominals which are neutral for plurality. The following are several examples.

- (12)
- |    |                                    |            |                         |
|----|------------------------------------|------------|-------------------------|
| a. | <i>ak-kuta'?</i> nang#i ri ka'u    |            |                         |
|    | itr-ask#3A                         | at you     | 'He asked you.'         |
| b. | <i>ak-kuta?</i> na'ng-i#?i ri ka'u |            |                         |
|    | itr-ask-pl#3A                      | at you     | 'They asked you.'       |
| c. | ri ka'u#i <i>ak-kuta?</i> na'ng-i  |            |                         |
|    | at you#3A                          | itr-ask-pl | 'They asked you.'       |
| d. | ri kau'-i#?i <i>ak-kuta'?</i> nang |            |                         |
|    | at you-pl#3A                       | itr-ask    | 1. 'They asked you.'    |
|    |                                    |            | 2. 'He asked you (pl).' |
| e. | <i>ak-kuta?</i> nang#i ri kau'-i   |            |                         |
|    | itr-ask#3A                         | at you-pl  | 'He asked you (pl)'     |

As in the case of adverbs and modal auxiliaries above, the plural marker may stay on the verb (12b and 12c) and may also raise to the pronominal together with the absolutive agreement (12d). The ambiguity of (12d) is due to the fact that there are two arguments that can be quantified by the plural marker, the subject and the object. (12c) is not ambiguous because the plural marker cannot quantify the pronominal if it is attached to the verb. Thus the pronominal *ka'u* in (12c) is unspecified for plurality. It is not so clear why. One possible reason could be that pronominals in Selayarese are a kind of emphatic forms. And as such, the plural marker must attach to the pronominals in order to be able to quantify them. This might not be a very elegant analysis, but that is the only solution I could come up with at this point.

Finally, we will see that the plural marker can also attach to *wh*-words. Like in the case of pronominals, the plural marker must attach to the *wh*-words in order to be able to quantify them. Consider the data in (13):

- (13)
- |    |                            |  |   |
|----|----------------------------|--|---|
| a. | <i>a'pa mu-ha'lli</i>      |  |   |
|    | what 2E-buy                |  | 'What did you buy?'                           |
| b. | <i>a'pa mu-halli'-'?i</i>  |  |   |
|    | what 2E-buy-pl             |  | 'What did you (pl) buy?'                      |
| c. | <i>apa'-i mu-halli</i>     |  |   |
|    | what-pl 2E-buy             |  | 'What did you buy?' (implies plural objects)  |
| d. | <i>i-na'i la-ke'o?</i>     |  |   |
|    | clas-who 3E-invite         |  | 'Who did he invite?'                          |
| e. | <i>i-na'i la-keo'k-i</i>   |  |   |
|    | clas-who 3E-invite-pl      |  | 'Who did they invite?'                        |
| f. | <i>i-nai'-'?i la-ke'o?</i> |  |   |
|    | clas-who-pl 3E-invite      |  | 'Who did he invite?' (implies plural objects) |

We can see in (13b) and (13e) that the subjects are plural, for the plural marker is attached to the verb, while the objects (represented by the *wh*-words) are unspecified for plurality. In (13c) and (13f), on the other hand, it is the objects that are plural, for the plural marker is attached to the *wh*-words, which represent the objects.

Based on the facts that (1) the plural marker can attach to practically any lexical category so long as it is not a noun, (2) it never affects the meaning of the base it attaches to, and (3) it is not obligatory, I conclude that the plural marker is a kind of free grammatical expression, more precisely a clitic, rather than a type of inflection or derivation. The fact that it changes the phonological shape of the base does not necessarily mean that it must be excluded from the clitic groups. Rather, morphological theory should recognize the existence of a category which possesses the combined characteristics of affixes on the one hand and clitics on the other. For this we do not need to coin a new name, since there has been a familiar name in the literature which fits nicely into this category, affixal clitic (Selkirk 1995, though not exactly in the sense that Selkirk uses the name).

### 3. Verbal number Marking

As I mentioned in the introduction, besides the plural *-i* there is a homophonous *-i* which functions as a kind of verbal number marker, showing iteration or durability of an event or state, depending upon the semantics of the verb. I will discuss this type of formative briefly.

The first question to ask is whether the iterative marker is really different from the plural marker, or if they are just two names for the same category, i.e. number marker, which may quantify participants as well as events, depending on context? The crucial point here is whether there are some empirical data that can support the claim that they indeed belong to two different categories. Consider the data in (13):

- (14)
- |      |  |  |
|------|--|--|
| a.   | <i>as-sa'ssa#kang</i><br>itr-wash#1plA             | 'We did our laundry.'                              |
| b. * | <i>as-sassa'-i#kang</i><br>itr-wash-pl#1plA        |  |
| c.   | <i>ak-kuta?na'ng-i#kang</i><br>itr-ask-itt/pl#1plA | 'We asked repeatedly.'                             |
| d.   | <i>ak-kuta?na'ng-i#ko</i><br>itr-ask-itt/pl#2A     | 1. 'You asked repeatedly.'<br>2. 'You (pl) asked.' |
| e.   | <i>ak-kuta?nang-i'-?i#ko</i><br>itr-ask-itt-pl#2A  | 'You (pl) asked repeatedly.'                       |

- f. *u'ntuk#ko ak-kuta?na'ng-i*  
 unwilling#2A itr-ask-pl  
 1. 'You were unwilling to ask repeatedly'  
 2. 'You (pl) were unwilling to ask.'
- g. *untu'k-i#ko ak-kuta'?nang*  
 unwilling-pl#2A itr-ask  
 1. 'You (pl) were unwilling to ask.'  
 2. \* 'You were unwilling to ask repeatedly.'

The ungrammaticality of (14b) is due to the fact that the absolutive agreement is semantically plural, thus no plural marker is needed. The *-i* in (14b) cannot be an iterative marker because the verb is a change of state verb, which cannot be performed repeatedly. That is why (14b) is ungrammatical. The verb in (14c) is semelfactive and the *-i* is an iterative marker. As we can see, (14c) is out with plural reading because the subject is semantically plural. The ambiguity of (14d) is due to the fact the *-i* can quantify the subject as in reading (1) and can also multiply the event as in reading (2). In short, in that position the *-i* can serve as a nominal number marker as well as a verbal number marker. (14e) tells us that the iterative *-i* can co-occur with the plural *-i*. This is strong support for the claim that they belong to two different categories irrespective of their identical phonological form. Should they belong to the same category, (14d) would be unacceptable. Failure to get reading (2) in (14g) gives further support to the claim that we are dealing with two kinds of number markings. If the plural marker could also function freely as an iterative marker, reading (2) in (14b) should be possible, for we could as well say that the *-i* can quantify the event denoted by the verb even if it is attached to the modal auxiliary. The fact that reading (2) is out suggests that the iterative marker does not affect the meaning of the verb if it is not attached to the verb. In short, the iterative *-i* does not behave like a clitic, but rather it serves as an affix which sticks to the verb.

Another characteristic of the iterative *-i* which differentiates it from the plural *-i* is that it attaches only to a limited set of verbs. Thus while the plural *-i* is highly mobile morphologically, the iterative *-i* places a high degree of selectional restriction on the types of verbs it attaches to. The verbs must be semelfactive or durative.

Finally, let us see the position of iterative *-i* on the verbal complex with respect to the plural *-i* and the benefactive *-ang*. Consider the data in (15)

- (15) a. *ku-kuta'?nang#ko*  
 1sgE-ask#2A 'I asked you.'
- b. *ku-kuta?na'ng-i#ko*  
 1sgE-ask-itt/pl#2A 1. 'I asked you repeatedly.'  
 2. 'I asked you (pl).'
- c. *ku-kuta?na'ng-ang#ko*  
 1sgE-ask-ben#2A 'I asked (somebody) for you.'
- d. *ku-kuta?nang-i'-ang#ko*  
 1sgE-ask-itt-ben#2A 'I asked (somebody) repeatedly for you.'

- e. *ku-kuta?nang-a'ng-i#ko*  
 1sgE-ak-ben-pl#2A 'I asked (some one) for you (pl).'
- f. *ku-kuta?nang-i-a'ng-i#ko*  
 1sgE-ask-itt-ben-pl#2A 'I asked (someone) repeatedly for you (pl).'

The data in (15) exhibit neatly how the relevance principle and Mirror Principle are obeyed. In previous discussions we stated that the relevance principle implies that a formative whose meaning affects the meaning of the base is supposed to be closer to the base than the formatives whose meanings do not affect the base. We also stated that benefactive *-ang* precedes the plural *-i* by virtue of the Mirror Principle, where the benefactive argument must be licensed first before it is quantified by the plural marker. If this line of analysis is correct, then we can expect that the *-i* preceding the benefactive *-ang* in (15d and 15f) must be the iterative marker, while the *-i* following benefactive *-ang* in (15e and 15f) must be the plural marker. As it turns out, this is exactly the case.

Based on the facts that (1) the iterative *-i* attaches only to a limited set of verbs, (2) it co-occurs with the plural *-i*, and (3) it precedes the plural *-i* when they co-occur on the same base, I will assume that the plural *-i* and the iterative *-i* belong to two different categories. Put differently, irrespective of their being homophonous and stress-shifting, the plural *-i* is a nominal number marker, whereas the iterative *-i* is a verbal number marker. As such they appear in two different morphological realizations, the former is an affixal clitic, while the latter is an affix.

#### 4. Conclusion

To sum up, in the course of the discussion we see that there is a formative *-i* in Selayarese which behaves like an affix in that it changes the phonological shape of the base it attaches to. At first glance it looks like an inflectional agreement marker. A closer look at the nature of the formative reveals that it does not behave exactly like an inflectional agreement marker, for it is not obligatorily present when the argument it is construed with is overtly plural. In addition, it attaches to various different bases, not only to verbs but also to adjectives, modal auxiliaries, adverbs, pronouns, and *wh*-words. Irrespective of this wide range of possible bases, it does not contribute any semantic element to the base, but rather it gives a clue that the argument is plural. With respect to ordering with other formatives on the base, it normally follows the stress-shifting formatives and precedes the stress-neutral ones. The ordering of formatives indicates that phonological consideration is superior to semantic justification, which sometimes results in violation of the Relevance Criterion and the Mirror Principle. Based on the fact that (1) the plural marker attaches to variously different lexical categories, (2) it does not affect the meaning of the base it attaches to, (3) it quantifies nominals irrespective of the base it attaches to, (4) it changes the phonological shape of the base it attaches to, and (5) it is not obligatory, I conclude that the formative *-i* belongs to the category of nominal number marker morphologically realized as an affixal clitic.



Besides the nominal number marker, there is another homophonous formative *-i* which quantifies the event denoted by the verb, showing an iterative or durative aspect of some sort. This formative *-i* is morphologically different from the nominal number marker *-i* in that it attaches to a limited set of verbs only, i.e. verbs showing iteration or duration. Therefore, irrespective of homophony, I conclude that this second formative belongs to the category of verbal number marker, morphologically realized as an affix.

#### REFERENCES

- Aronoff, Mark (1976) *Word Formation in Generative Grammar*, MIT Press.
- Aronoff, M., A. Arsyad, H. Basri. and E. Broselow (1987) 'Tier configuration in Makassarese Reduplication' in *Papers from the 23rd Annual Regional Meeting of Chicago Linguistic Society*, Part 2. pp. 1-15.
- Baker, M. (1985) 'The Mirror Principle and Morphosyntactic Explanation,' *Linguistic Inquiry* 16, 373 - 416.
- Bybee, Joan (1985) *Morphology: A Study of the Relation between Meaning and Form*, John Benjamins, Amsterdam.
- Imran, Indiyah (1984) *Proses Morphologi dalam Kelas Kata Bahasa Makassar*, (Morphological Processes in Makassarese Word Class), doctoral dissertation, Hasanuddin University, Indonesia.
- Klavanz, J. (1985) 'The Independence of Syntax and Phonology in Clitization,' *Language*, 61, pp. 95 - 120
- Mithun, M. and H. Basri (1986) 'The Phonology of Selayarese' *Oceanic Linguistics*, 25. 1-2: 210 - 254.
- Mithun, M. (1988) 'Lexical Categories and the Evolution of Number Marking' in M. Hammond and M. Noonan (eds) *Theoretical Morphology*, Academic Press. 211-34.
- Selkirk, Elizabeth (1995) 'The Prosodic Structure of Function Words' in *Papers in Optimality Theory*, GLSA, University of Massachusetts at Amherst, pp. 339-466.
- Zwicky, A. (1985) 'Clitics and Particles', *Language*, 61, pp. 283-305.
- Zwicky, A. and G. Pullum (1983) 'Clitization vs Inflection: English *n't*', *Language*, 59, pp. 502 - 523.



# Alignment and nasal substitution strategies in Austronesian languages\*

Rita Bhandari  
State University of New York at Stony Brook

## INTRODUCTION

The ban on sequences of a nasal followed by a voiceless consonant appears to be widespread across Austronesian languages and different strategies are employed to get rid of them. The data in (1) provides examples from two languages. In Makassarese (1a), a nasal/voiceless consonant sequence surfaces as a geminate nasal while in Indonesian (1b), it is eliminated by coalescence.

1)	Prefix + verb stem	Surface form	Gloss
a.	Makassarese <sup>1</sup>		
	aN + kanɾ	aŋaŋaɾɛ	'to eat'
	aN + tunruŋ	annaŋruŋ	'to beat'
	aN + poloŋ	ammaŋloŋ	'to cut'
b.	Indonesian		
	məN + kail	məŋail	'to fish'
	məN + tulis	məŋulis	'to write'
	məN + pilih	məmilih	'to choose, to vote'

Pater (1996) accounts for nasal substitution in Indonesian in terms of fusion of a nasal and a voiceless consonant. He also discusses gemination in Konjo prefixation, which follows the same pattern as Makassarese. Both fusion and gemination are motivated by a high ranking phonotactic constraint that disallows nasal/voiceless consonant sequences (\*NC<sub>v-less</sub>). He analyzes the difference between the Indonesian and Konjo prefixation as a difference between coalescence and total assimilation of the voiceless consonant to the nasal. He accounts for the different patterns by varying the rankings of **Linearity**, an "anti-fusion constraint", with respect to the faithfulness constraint that requires identity of the nasal feature between output and input forms (**Ident O-I [Nas]**).

---

\* The ideas developed in this paper took shape during a seminar course on Austronesian phonology and syntax taught by Ellen Broselow and Dan Finer. I thank both of them for their help and encouragement. I would also like to thank Mark Aronoff for his comments and suggestions. I also owe special thanks to Hasan Basri for his help with the data.

<sup>1</sup> In some cases the Makassarese prefix *aŋ* does not trigger assimilation and a voiceless consonant is allowed to follow a nasal. Thus the form *ammaŋloŋ* alternates with *ampoloŋ*. But there is a syntactic difference between the two. The prefix that triggers assimilation occurs on transitive verbs with indefinite objects; the one that doesn't trigger assimilation occurs with definite objects (Aronoff 1986). In this paper, I concentrate on the indefinite prefix only.

2)	N + C	*NC <sub>v-less</sub>	Linearity	Ident O-I [Nas]
a.	N <sub>1</sub> N <sub>2</sub> (Indonesian)		*	
b.	N <sub>1</sub> N <sub>2</sub> (Konjo)			*

In this paper, I show that in Makassarese, nasal gemination cannot be accounted for merely in terms of a linearity constraint driven by a \*NC<sub>v-less</sub> effect. The main evidence for this comes from the fact that there is nasal gemination even in vowel initial stems where \*NC<sub>v-less</sub> is not in effect overtly and linearity is redundant.

3)	Prefix + verb stem	Surface form	Gloss
a.	aŋ + alle	aŋaalle	'to take'
b.	aŋ + inuŋ	aŋiinuŋ	'to drink'

In the data in (3), gemination is not triggered by the need to avoid a nasal/voiceless consonant sequence. Thus, it cannot be the constraint on segmental linearity alone that bans fusion and derives geminates in Makassarese.

In view of these facts, I propose that nasal gemination is driven by the need to satisfy certain alignment constraints on the stem and the prefix without violating syllable structure constraints. Given a limited set of constraints, I will show that the differences between the Makassarese and Indonesian patterns can be accounted for. In addition to Pater's \*NC<sub>v-less</sub>, I argue that there is a high-ranking constraint on prefixes to keep their syllables distinct from the stem; this leads to gemination. The constraint on prefix edges may be sacrificed if the language bans double linking of segments; this leads to coalescence. I will then extend my analysis to account for prefixation and reduplication patterns in another Austronesian language Selayarese to demonstrate the crosslinguistic potential of these constraints.

#### THE CONSTRAINTS

In (4)-(6), I introduce the key phonological constraints that play a role in Indonesian and Makassarese prefixation. \*NC<sub>v-less</sub> ranks very high.

4) **Prefix=σ**: This constraint states that the syllable edges of the prefix in the output must be identical to its syllable edges in the input.

The constraint in (4) can be motivated on psycholinguistic evidence. According to Bloomfield, prefixes are less common across languages than suffixes. Mark Aronoff (p.c.) points out that it has been suggested that prefixes are harder to process than suffixes so prefix integrity, which demands that prefixes be identifiable, is stronger than suffix integrity. Thus, a language like

English allows nonsyllabic suffixes (the plural marker *-s* for example) but prefixes are all syllabic.

- 5) **\*Double link**: This is a constraint against gemination.
- 6) **Align Word**: The left edge of the root must be aligned with the left edge of the prosodic word (Cohn & McCarthy 1994 as cited in Pater 1996).

In the rest of the paper, I will show that in Makassarese the antifusion constraint is not **Linearity** but **Prefix=σ**. Different rankings of the constraints in (4)-(6) will account for prefixation and reduplication in Makassarese and Indonesian.

#### PREFIXATION WITH CONSONANT INITIAL STEMS

- 7) **Makassarese**      aN + poloŋ    →    ammoloŋ

8) aN + poloŋ	*NC <sub>v-less</sub>	Align word	Prefix=σ	*Double link
a. am.mo.loŋ				*
b. a.mo.loŋ			*!	
c. am.o.loŋ		*!		
d. am.po.loŋ	*!			

\*NC<sub>v-less</sub> rules out (8d) where there is an illicit sequence of a nasal followed by a voiceless consonant. The contrast between (8b) and (8c) is one of deletion as opposed to fusion. The prosodic word excludes the prefix and (8c) violates **Align word** as the root initial nasal is deleted. (8b) does no better as the edges of the prefix is not aligned with the edges of a syllable. In Makassarese, the fusion option is not available to undo \*NC<sub>v-less</sub> effects because **Prefix=σ** is ranked high (**Prefix=σ** >> **\*Double link**). Thus, candidate (8a) with a geminate nasal emerges the winner.

- 9) **Indonesian**      mƏN + pilih    →    mƏmilih

10) mƏN + pilih	*NC <sub>v-less</sub>	Align word	*Double link	Prefix=σ
a. mƏ.mi.lih				*
b. mƏm.mi.lih			*!	
c. mƏm.i.lih		*!		
d. mƏm.pi.lih	*!			

In (10), the interesting contrast is between the (a) and (b) candidates. In Indonesian, the constraint on prefixes is subordinate to the constraint against geminates (**\*Double link** >> **Prefix=σ**).

It must also be noted that although the candidates (10a) and (10c) are identical in terms of surface form, reduplication data shows that (10a) is the more faithful output form. It will be seen that **Align word** ranks high. I will return to this point in my discussion of Indonesian reduplication.

At this point the ranking between **Align word** and **Prefix=σ** in Makassarese and **Align word** and **\*Double link** in Indonesian and is left undetermined.

There are a few other possibilities in both languages, which are not shown in tableaux (8) and (10). These have been discussed by Pater and I summarize them below.

i) Voicing of the post nasal voiceless consonant ( $NC_{v-less} \rightarrow NC$ ).

This is ruled out in both languages by the constraint **Ident [obs voice]** that requires that the voice feature of an obstruent in the output be identical to the voice feature of that obstruent in the input.

ii) Another possibility is total assimilation of the nasal to the voiceless consonant resulting in gemination of the voiceless consonant ( $NC_{v-less} \rightarrow C_{v-less} C_{v-less}$ ). However, the constraint **Ident I-O [nas]** which requires that a nasal in the input have the same feature in its output correspondent disallows this type of assimilation.

The tableaux in (11) and (12) illustrate the relative ranking of these constraints in Makassarese and Indonesian respectively with those mentioned earlier.

#### Makassarese

11) aN+poloŋ	*NC <sub>v-less</sub>	Align word	Ident [obs voice]	Ident I-O nasal	Prefix=σ	*Double link
a. am.mo.loŋ						*
b. a.mo.loŋ					*!	
c. am.o.loŋ		*!				
d. am.po.loŋ	*!					
e. am.bo.loŋ			*!			
f. ap.po.loŋ				*!		

#### Indonesian

12) məN+pilih	*NC <sub>v-less</sub>	Align word	Ident [obs voice]	Ident I-O [nas]	*Double link	Prefix=σ
a. mə.mi.lih						*
b. məm.mi.lih					*!	
c. məm.i.lih		*!				
d. məm.pi.lih	*!					
e. məm.bi.lih			*!			
f. məp.pi.lih				*!		

Candidates (11e) and (12e) where the root-initial voiceless consonant emerges as voiced are ruled out input-output identity constraint on voicing. Candidates (11f) and (12f) show that the nasal feature may not be done away with as the constraint on maintaining identity of the nasal feature is ranked rather high.

The crucial difference between Indonesian and Makassarese is with respect to the ordering of the **Prefix=σ** and **\*Double link**. In Indonesian **Prefix=σ** is ranked below **\*Double link** but in Makassarese it is the other way round. This ranking of constraints leads to nasal gemination in Makassarese and nasal fusion in Indonesian to avoid **\*NC<sub>v-less</sub>** effects. These constraints however have a wider application outside the **\*NC<sub>v-less</sub>** cases as was seen with vowel initial stems in Makassarese (in (3) above).

#### PREFIXATION WITH VOWEL INITIAL STEMS

In Makassarese, there is gemination of the final consonant of the prefix even if the root is vowel initial. The relevant data from (3) is repeated in (13).

13)	Prefix + verb stem	Surface form	Gloss
	a. aη + alle	aηηalle	'to take'
	b. aη + inuη	aηηinuη	'to drink'

I propose that gemination outside the **\*NC<sub>v-less</sub>** cases in Makassarese follows from the need to satisfy the **Onset** requirement without violating the syllable constraint on prefixes, **Prefix=σ**. The constraint **Onset** requires all syllables to have onsets (McCarthy & Prince 1995). Tableau (14) illustrates the point.

14) aη + alle	Onset	Prefix=σ	Align word	*Double link
a. aη.ηal.le			*	*
b. aη.al.le	*!			
c. a.ηal.le		*!	*	

The onset requirement eliminates (14b) and the constraint on prefix syllables rules out (14c). The winning candidate (14a) violates both **Align word** and **\*Double link** but these constraints are ranked relatively low. Earlier it was established that in Makassarese **\*Double link** was subordinate to **Prefix=σ** (see tableau 8), but the ranking between **Prefix=σ** and **Align word** had been undecided. Tableau (14) shows that the ranking between the two is not crucial but **Onset** dominates **Align word**. Thus, the ranking of constraints in Makassarese, so far, is as in (15).

15) Makassarese: **Onset, Prefix=σ >> Align word >> \*Double link**

In Indonesian, there is no gemination in vowel initial stems because **Onset** is ranked lower than **Align word**. It was seen earlier that **Align word** and **\*Double link** were ranked higher

than **Prefix=σ** in Indonesian (see tableau (10)). At this point **Onset** is left unranked with respect to **Prefix=σ**.

16) mən+ atur	Align word	*Double link	Prefix=σ	Onset
a. mən.atur				*
b. mən.ηatur	*!	*		
c. mən.ηa.tur	*!		*	

17) Indonesian: \***Double link**, **Align word** >> **Prefix=σ**, **Onset**

Although the surface forms in (16a) and (16c) are identical, reduplication data shows that (16a) and not its identical surface form (16c) is the winning candidate. It will be seen that in Indonesian **Align word** ranks very high and **Onset** ranks very low. This ordering of constraints makes the right predictions for the different patterns of reduplication that are witnessed in the two languages. Reduplication data offers important insights into the constraints and their ranking proposed and therefore must be considered.

REDUPLICATION WITH CONSONANT-INITIAL ROOTS

The data in (18) and (19) show the different reduplication patterns in Makassarese and Indonesian respectively.

18) Makassarese

	Prefix + verb stem	Surface form	Gloss
a.	aN + kanrɛ	aŋaŋanrɛŋanrɛ	'to eat'
b.	aN + tunruŋ	annunruŋnunruŋ	'to beat'
c.	aN + polon	ammolonmolon	'to cut'

19) Indonesian

	Prefix + verb stem	Surface form	Gloss
a.	məN + kail	məŋailŋail	'to fish repeatedly'
b.	məN + tulis	mənulisnulis	'to write'
c.	məN + pilih	məmilihmilih	'to chose, to vote'

Both languages have a straightforward and simple form of reduplication as far as consonant-initial stems are concerned: the entire stem is copied in reduplication. The important point to note is that the Indonesian data shows that the initial voiceless consonant of the root is not deleted but indeed fuses with the nasal of the prefix, and it is faithfully reduplicated.



## REDUPLICATION WITH VOWEL-INITIAL STEMS

Vowel-initial reduplication offers a more interesting and insightful contrast as far as constraint ranking is concerned and I will analyze them in some detail.

## 20) Makassarese

	Prefix + verb stem		Reduplication	Gloss	
a.	aŋ + alle	→	aŋalle	aŋalleŋalle	'to take for fun'
b.	aŋ + inuŋ	→	aŋinuŋ	aŋinuŋŋinuŋ	'to drink for fun'

21) aŋ + RED + alle	B-R faith	Onset	Prefix=σ	Align word	*Double link
a. aŋ.ŋalle.ŋalle				*	*
b. a.ŋalle.ŋalle			*!	*	
c. aŋ.alle.alle		**!			
d. a.ŋalle.alle	*!	*	*	*	

The presence of the "epenthetic" nasal in both the reduplicant and the base confirms the earlier high ranking of **Onset** and **Prefix=σ** with respect to **Align word** and **\*Double link**. Consider now the Indonesian data in (22).

## 22) Indonesian

	Prefix + verb stem		Reduplication	Gloss	
a.	məN + əlu	→	məŋəlu	məŋəluəlu	'to cheer'
b.	məN + atur	→	məŋatur	məŋaturatur	'to arrange'

In Indonesian, unlike in Makassarese, there is no nasal either in the base or the reduplicant. This could be seen to follow from two factors: **Align word** is ranked very high and **Onset** ranks low.

## Indonesian

23) məŋ + RED + atur	B-R faith	*Double link	Align word	Prefix=σ	Onset
a. məŋ.atur.atur					**
b. məŋ.atur.atur		*!	*		
c. mə.atur.atur			*!	*	
d. məŋ.atur.atur	*!		*		*

The winning candidate (23a) has two violations of **Onset** but it does not violate any of the higher ranked constraints.

To sum up, the ranking of constraints in Indonesian and Makassarese is as follows:

- 24) Makassarese: \*NC<sub>v-less</sub>, **Onset, Prefix=σ** >> **Align word** >> \***Double link**  
 25) Indonesian: \*NC<sub>v-less</sub>, \***Double link, Align word** >> **Prefix=σ** >> **Onset**

These constraints not only provide a unified account of the \*NC<sub>v-less</sub> effects in different languages, they can also account for the crosslinguistic variation in reduplication patterns witnessed in Indonesian, Makassarese, and another Austronesian language Selayarese.

#### SELAYARESE

The data in (26) shows reduplication in Selayarese with consonant initial (a-c) and vowel initial roots (d-e). In these examples, the intransitivizer *aN* (Mithun & Basri 1986) is prefixed to the verb stem.

26)	Prefix + verb stem		Reduplication	Gloss	
a.	aN + kanɾe	→	aŋanɾe	aŋanɾeŋanɾe	'to eat to fun'
b.	aN + tunruŋ	→	anunruŋ	anunruŋunruŋ	'to beat for fun'
c.	aN + poloŋ	→	amolonoŋ	amolonoŋmoloŋ	'to cut for fun'
d.	aŋ + alle	→	aŋalle	aŋalleŋalle	'to take for fun'
e.	aŋ + inuŋ	→	aŋinuŋ	aŋinuŋŋinuŋ	'to drink for fun'

With consonant initial stems, the Selayarese pattern of reduplication is the same as that of Indonesian. However, with vowel initial stems a different pattern emerges in all three languages. Unlike in Indonesian, Selayarese reduplicates the nasal of the prefix but unlike in Makassarese the nasal does not geminate.

27)	aN + RED + kanɾe	<b>B-R faith</b>	<b>Onset</b>	<b>*Double link</b>	<b>Align word</b>	<b>Prefix=σ</b>
a.	a.ŋanɾe.ŋanɾe					*
b.	aŋ.ŋanɾe.ŋanɾe			*!		
c.	aŋ.anɾe.ŋanɾe	*!	*			
d.	a.ŋanɾe.anɾe	*!	*		*	
e.	aŋ.anɾe.anɾe		**!		*	

Candidate (27b), which is the Makassarese like form, is eliminated because \***Double link** is ranked higher than **Prefix=σ**. The Selayarese facts can be accounted for if both **Onset** and \***Double link** are ranked above **Prefix=σ**. Although it is not possible to tell the ranking between **Onset** and **Align word** from tableau (27), the vowel initial data shows clearly that the former is ranked above the latter. Consider tableau (28).

24) aη + RED + allε	B-R faith	Onset	*Double link	Align word	Prefix=σ
a. a.ηallε.ηallε				*	*
b. aη.ηallε.ηallε			*!	*	
c. aη.allε.allε		**!			
d. aη.allε.ηallε	*!	*		*	

Again, the Makassarese like candidate (28b) loses out as it violates the high ranking **\*Double link**. Candidate (28c), which is the Indonesian like form, is also ruled out because **Onset** is ranked high. While in Indonesian **Onset** ranks below **Align word**, in Selayarese the ordering of the two constraints is reversed. The relative ranking of the key constraints of all three languages are given in (29)-(31).

- 29) Makassarese: **\*NC<sub>v-less</sub>, Onset, Prefix=σ >> Align word >> \*Double link**
- 30) Indonesian: **\*NC<sub>v-less</sub>, \*Double link, Align word >> Prefix=σ >> Onset**
- 31) Selayarese: **\*NC<sub>v-less</sub>, Onset, \*Double link >> Align word, Prefix = σ**

The analysis presented here highlights the ability of Optimality Theory to capture the phonology-morphology interface across languages with the help of a limited set of constraints. Finally, I would like to point out certain advantages in terms of explanatory power that Optimality Theory has compared to other approaches. Within derivational frameworks, which have recourse to different levels of rule application, it is difficult to provide a uniform explanation of the different reduplication patterns in the three languages.

#### OPTIMALITY THEORY VERSUS DERIVATIONAL ANALYSES

In SPE (Chomsky & Halle 1968), a distinction is made between two types of affixes *stem-level* and *word-level* on the basis of whether or not they undergo lexical phonological rules. While stem-level affixes undergo cyclic phonological rules, word-level affixes do not. Aronoff (1988) extends this level distinction to the application of morphological rules. Thus reduplication may be either a stem level or a word level process applying to either the *head* (excluding the affix) or the *whole* (including the affix). Word-level reduplication “reveals prior application of phonology” and respects the syllable structure of the stem. He also states that Makassarese reduplication is word-level applying to the head (excluding the affix).

In order to derive the Makassarese nasal gemination in vowel initial stems let us consider the possibilities of whether the prefix *aη* is stem-level or word-level. Following Aronoff, I will assume that Makassarese reduplication is word-level head reduplication.

- 32) a. Stem-level prefix + stem  
 $a\eta + all\epsilon \rightarrow a.\eta all\epsilon$
- b. Word-level head reduplication  
 $*a.\eta all\epsilon.\eta all\epsilon$

If *aŋ* is a stem-level prefix, it must be able to undergo lexical phonological rules. It must resyllabify in order to provide an onset as indicated in (32a). Word-level head reduplication should deliver the form in (32b). This form is the correct output form for Selayarese but not for Makassarese.

However, if the prefix is word level there should be no need to fulfill the onset requirement since word initially syllables are allowed to be onsetless in this language.

- 33) a. Word-level prefix + stem  
           *aŋ* + *allɛ* → *aŋ.allɛ*
- b. Word-level head reduplication  
           \**aŋ.allɛ.allɛ*

If *aŋ* is a word-level prefix, it should resist application of phonological rules such as resyllabification with the base, as in (33a). If this form were to undergo word-level head reduplication, the output would be as in (33b). Here, reduplication excludes the affix, which does not interfere with the syllable structure of the stem. Again, although this is the correct form for Indonesian reduplication, it is not consistent with the Makassarese pattern where the reduplicated form is *aŋŋallɛŋallɛ*.

Thus, a simple derivational analysis based on ordering of phonological rules fails to account for the reduplication patterns in Makassarese. In this respect, Optimality Theory is more successful in accounting for the reduplication patterns of three Austronesian languages by means of constraint ranking.

## CONCLUSION

I demonstrated that **Prefix=σ** and not **Linearity** (as proposed by Pater) is the antifusion constraint in Makassarese. Fusion is not available as an option to undo \*NC<sub>v-less</sub> effects due to this constraint on syllable integrity. The constraint **Prefix=σ** was seen to have a wider application outside the \*NC<sub>v-less</sub> cases and could account for cases with vowel initial stems where there was no potential nasal/voiceless consonant sequence. Varying the rankings of a limited set of constraints **Prefix = σ**, **Align Word**, **\*Double linking**, and **Onset** provides a unified account crosslinguistically of prefixing and reduplication patterns in at least three Austronesian languages. Finally, I also showed that Optimality Theory has an advantage over derivational analyses in capturing the reduplication patterns in these languages.

## References

- Aronoff, Mark. 1988. *Head operations and strata in reduplication: A linear treatment*. Yearbook of Morphology, ed. By Geert Booij and Jaap van Marle, 1-15. Dordrecht: Foris Publications.
- Chomsky, Noam and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.
- Cohn , Abigail C. and John J. McCarthy. 1994. *Alignment and parallelism in Indonesian phonology*. Cornell University & University of Massachusetts at Amherst ms.
- McCarthy, John J. and Alan Prince. 1995. Faithfulness and reduplicative identity. *Papers in Optimality Theory: University of Massachusetts Occasional Papers 18*, ed. by Jill Beckman, Laura Walsh, and Suzanne Urbanczyk, 249-384. University of Massachusetts, Amherst: GLSA.
- Mithun, Marriane and Hasan Basri. 1986. The phonology of Selayarese. *Oceanic Linguistics*, 25.1-2: 210-54
- Pater, Joe. (To appear). Austronesian nasal substitution and other \*NC effects. *The prosody morphology interface*, ed. by Rene Kager, Harry van der Hulst, and Wim Zonneveld. Cambridge University Press.

Department of Linguistics  
 SUNY at Stony Brook  
 Stony Brook, NY 11794-4376

[bhandari@phonlab.sbs.sunysb.edu]



## V to D Raising in Sulawesi Relatives

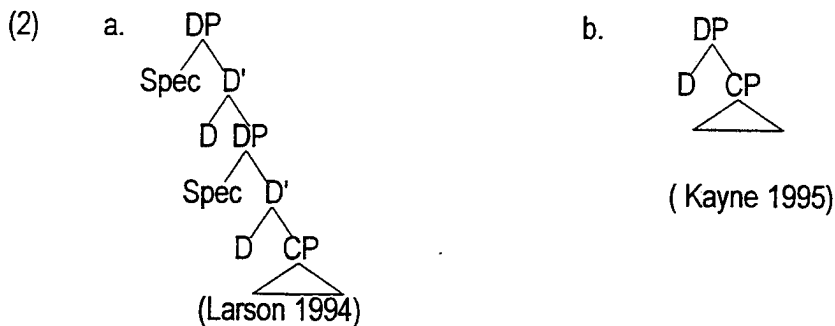
Daniel L. Finer  
 SUNY at Stony Brook  
 dfiner@ccvm.sunysb.edu

### 1. Introduction

The theoretical interpretation of relative clause data from Selayarese, Buginese, and Makassarese<sup>1</sup> (three verb-initial Austronesian languages from southwestern Sulawesi, Indonesia) presents a challenge to the classical widely-adopted CP-adjoined structures<sup>2</sup> for relative clauses, such as those shown in (1a-b).



As will be shown below, the determiner in relative clauses in these languages is an enclitic on the verb of the relative clause, and I argue that this attachment results from syntactic head movement; the verb raises from the VP of the relative clause through IP and CP and left-adjoins to D. The structures in (1), however, share the problematic feature that CP is integrated into the structure via right-adjunction, and so head-movement is not licensed by the Head-Movement Constraint (HMC, see Travis 1984 and Baker 1988).<sup>3</sup> Similarly, the Extension Condition of Chomsky 1995, which requires that overt movement extend the structure, also precludes movement out of CP to D, since the adjunction would be internal to NP in the structures in (1).<sup>4</sup> As a solution to these problems, I suggest below that a more appropriate structure for these relative clauses is a projection of DP in which CP is a complement of D rather than an adjunct of some projection of N, as proposed on independent grounds by Larson (1994) ((2a)), and Kayne (1995) ((2b)).



The common feature in (2a-b), that CP is a complement of D,<sup>5</sup> is structurally compatible with head movement to D from C. This relation thus establishes the sufficient conditions for the analysis of the V+D patterns in the data, and I propose below that the strength of the formal features of V, C and D determine the affixation patterns at spellout. The features are strong in the relevant sense, and head-raising from V (to I to C) to D takes place overtly. In addition to accounting for the Sulawesi data, the movement via feature-strength hypothesis also makes certain typological predictions regarding the form that relative and other subordinate clauses take cross-linguistically, and these predictions will be discussed briefly below.

The discussion is organized as follows: First I illustrate simple sentences which contain simple DPs, couching the discussion in loosely Minimalist terms. This is followed by a discussion of the relative clause data and candidate structures. I then present an analysis that satisfies the constraints on movement mentioned above, and finally I discuss the typological implications of the analysis.

## 2. Simple Clauses, simple DPs and the determiner

The languages are very similar in overall structure, differing mainly in lexical items and the realization of the morphology, and Selayarese will be the major source of data throughout. Declarative sentences are verb-initial, with VOS having a slight edge in preference over VSO word order. Some sample sentences and structures are provided below.

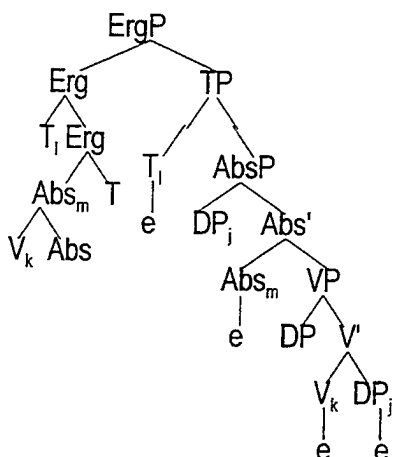
- (3) a. ku-alle-i doe?-iñjo  
*1sE-take-3A money-DEF*  
 'I took the money'
- b. la-?alle-i doe?-iñjo palopi-ñjo  
*3E-take-3A money-DEF sailor-DEF*  
 'The sailor took the money'
- c. la-?alle-i palopi-ñjo doe?-iñjo  
*3E-take-3A sailor-DEF money-DEF*  
 'The sailor took the money'
- d. la-taro-i doe?-iñjo palopi-ñjo ri lamari  
*3E-put-3A money-DEF sailor-DEF in cupboard*  
 'The sailor put the money in a cupboard'
- e. ku-keo?-ko  
*1sE-call-2famA*  
 'I called you'
- f. ak-kelong-ko  
*int-sing-2famA*  
 'You sang'

I analyze the above word order patterns as deriving from a VP-internal SVO order. V raises to the highest Inflectional projection, and the object shifts to the Spec position of an Agr ("AbsP") projection due to a strong D feature on the V+Agr complex, thus deriving VOS order. I assume that there is also a version of Agr with a weak D feature, and if this element enters the derivation, the object does not raise. As before, the verb raises across the subject, and VSO order is derived.<sup>6</sup> In each case, the subject raises to the higher Agr Projection ("ErgP") covertly (the object raises covertly as well in overtly VS sentences). While the agreement pattern is ergative, typical subject-object c-command asymmetric familiar from, e.g. English, hold, and I take this as evidence that the position to which the subject ultimately raises asymmetrically c-commands the position to which the object raises. The languages are thus ergative in their morphology only; their syntax is by and large accusative (see Finer 1994, in pre for further discussion). I adopt Bobaljik's (1993) case parameter in order to account for the ergative agreement patterns (note the data in (3e-f) in particular).

- (4) Case Parameter (Bobaljik 1992 (cf. Levin & Massam 1985, Massam 1996)):
- (i) Case X is obligatorily assigned/checked
  - (ii) a. In Nominative/Accusative languages, X is nominative
  - b. In Ergative/Absolutive languages, X is absolutive.



## (5) VOS order



In addition, the left periphery of the clause contains a position to which focused XPs move, and in all three languages, the absolutive marker is incompatible with a focused absolutive argument. No special morphological reflex occurs when the ergative argument is focused in Selayarese, but in Buginese and Makassarese, a special verbal prefix displaces the ergative agreement prefix. This same morphology recurs in Wh-questions and, as we will see below, in relative clauses, two other typical environments of A' movement.

## Selayarese

- (6) a. *doeʔ-iñjo la-taro(\*-i) palopi-ñjo ri lamari*  
*money-DEF 3E-put(-3A) sailor-DEF in cupboard*  
 'The sailor put THE MONEY in a cupboard'
- b. *palopi-ñjo la-taro-i doeʔ-iñjo ri lamari*  
*sailor-DEF 3E-put-3A money-DEF in cupboard*  
 'THE SAILOR put the money in a cupboard'
- c. *apa la-taro(\*-i) ri lamari palopi-ñjo*  
*what 3E-put(-3A) in cupboard sailor-DEF*  
 'What did the sailor put in a cupboard'

## Makassarese

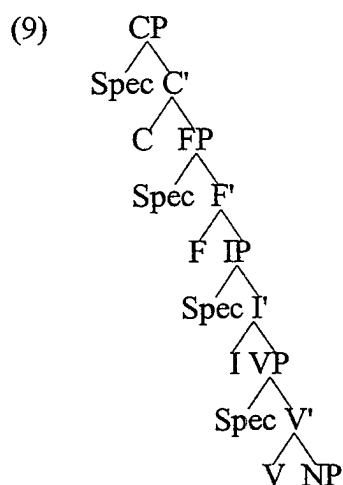
- (7) a. *na-pallu-i berasak-a i Ali*  
*3E-cook-3A rice-DEF cl Ali*  
 Ali cooks the rice
- b. *berasak-a na-pallu i Ali*  
*rice-DEF 3E-cook cl A*  
 'Ali cooks THE RICE'
- c. *i Ali am-pallu-i berasak-a*  
*cl A PREF-cook-3A rice-DEF*  
 'ALI cooks the rice'

## Buginese

- (8) a. *na-olli-wi bola-e i Abu*  
*3E-buy-3A house-DEF cl Abu*  
 'Abu bought the house.'

- b. i Abu m-molli-wi bola-e  
*cl Abu PREF-buy-3A house-DEF*  
 'ABU bought the house'
- c. bola-e na-olli i Abu  
*house-DEF 3E-buy cl Abu*  
 'Abu bought THE HOUSE'

In *Finer (1994, in press)* I argue that the focus position is what can be called a 'strong A<sub>r</sub>' position; weak crossover can be triggered from here, reflexives reconstruct from this position, filling this position creates an island for extraction of a lower item, etc. As discussed further there, this position occurs below (i.e., to the right of) the complementizer, so it is unlikely that the focus position is SpecCP. Instead, I assume that the focus position is Spec of FP, for want of a better term, as illustrated in (9), which shows the inflectional projections collapsed to I.



With this brief discussion of clausal structure concluded, I turn to the DPs.

The definite determiners in the three languages follow the N (D is boldfaced in the following examples).

- |      |             |    |  |    |   |
|------|-------------|----|--|----|---|
| (10) | Selayarese  | a. | tedong- <b>iñjo</b><br><i>buffalo-DEF</i><br>'the buffalo' | b. | palopi- <b>ñjo</b><br><i>sailor-DEF</i><br>'the sailor' |
| (11) | Makassarese | a. | ballak- <b>a</b><br><i>house-DEF</i><br>'the house'        | b. | bawi-ya<br><i>pig-DEF</i><br>'the pig'                  |
| (12) | Buginese    | a. | buaja-e<br><i>croc-DEF</i><br>'the crocodile'              | b. | anana?-e<br><i>child-DEF</i><br>'the child'             |

Indefinite DPs, on the other hand, show no overt determiner (13), and in possessive DPs, a pronominal form cliticizes to the nominal head (14) (see below for other examples).



languages.

### 3. Relative clauses

In a Selayarese relative clause, the overt head occurs to the left of the modifying clause, and there is a gap in the position that is normally occupied in an ordinary clause ((18a), (19a)) by the argument that is relativized. When the head is [+human], the relativizing prefix *to-* is affixed to the verbal complex, as in (18b) and (19b), and *nu-* is the relativizer used in other cases, such as (19c, d). In addition, free relatives such as (19b-c) are common.

- (18) a. *mu-lajañjang-i palop-iñjo*  
*2famE-see-3A sailor-DEF*  
 'You saw the sailor'
- b. *palopi to-mu-lajañjang-iñjo*  
*sailor REL-2famE-see-DEF*  
 'The sailor that you saw'
- (19) a. *la-pallu-i juku?-na i Ali*  
*3E-cook-3A fish-3P clas Ali*  
 'Ali cooked his fish'
- b. *to-la-pallu-ñjo-i juku?-na*  
*REL-3E-cook-DEF-3A fish-3P*  
 'the one who cooked his fish'
- c. *nu-la-pallu-ñjo i Ali*  
*REL-3E-cook-DEF cls A*  
 'the one Ali cooked'
- d. *tedong nu-ak-kelom-mu ri sapo*  
*buffalo REL-Int-sing-2P in house*  
 'your buffalo that sang in a house'

In addition to the presence of the relativizer, the verbal morphology of the relative clauses in (18) and (19) differs from that of the ordinary clause in other ways. The absolutive marker, for example, is absent when the absolutive argument (definite direct object or subject an intransitive verb) is relativized, and this parallels other constructions formed by A' movement (see Finer 1994 and Finer in press for discussion).<sup>10</sup> Central to the present discussion, however, is the placement of the definite determiner *ñjo* and the pronominal possessive clitic: they are attached to right edge of the verb of the relative clause. This placement is especially surprising in (19b,d) where the determiner intervenes between the direct object or PP adjunct of the relative clause and the verb, two elements from the modifying subordinate clause. Part of the analytic task is thus to provide a means to combine V with a clause-external D in the presence of V - O order.

This type of relative clause appears to be a feature of southwestern Sulawesi languages in general. The determiner is affixed to the verb in Makassarese (20b-c) as well, as shown below (see also Kahler 1974 for examples from Buginese and Makassarese showing V+D cliticization). The relativizing

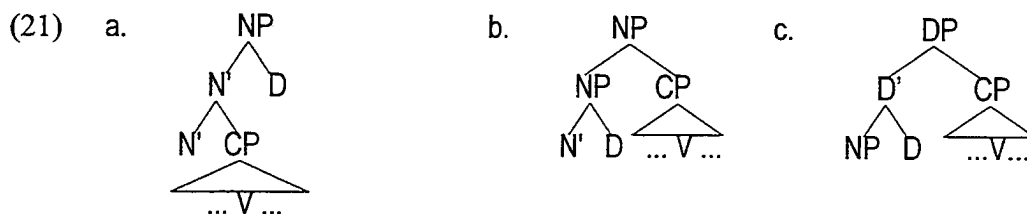
morpheme *tu* occurs in free relatives that denote humans and less frequently with headed relatives such as (20b). There is no relativizer when the nominal head denotes a non-human entity. Note that the Makassarese example in (20b) presents the same *V+D-XP* order as does (19b and 19d).<sup>11</sup>

Makassarese

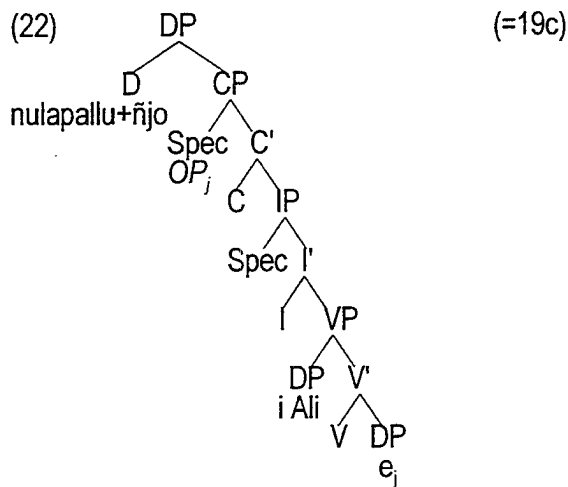
- (20) a.      *na-bunoi bawi-<sup>y</sup>a anaʔanak-a*  
                  *3E-kill+3A pig-DEF boy-DEF*  
                  'The boy killed the pig'
- b.      *anaʔanaʔ (tu-)am-buno-<sup>w</sup>a-i              bawi-<sup>y</sup>a*  
                  *boy              (REL-)PREF-kill-DEF+3A pig-DEF*  
                  'The boy who killed the pig'
- c.      *bawi na-buno-<sup>w</sup>a anaʔ anak-a*  
                  *pig 3E-kill-DEF boy-DEF*  
                  'the pig that the boy killed'

#### 4. D-CP

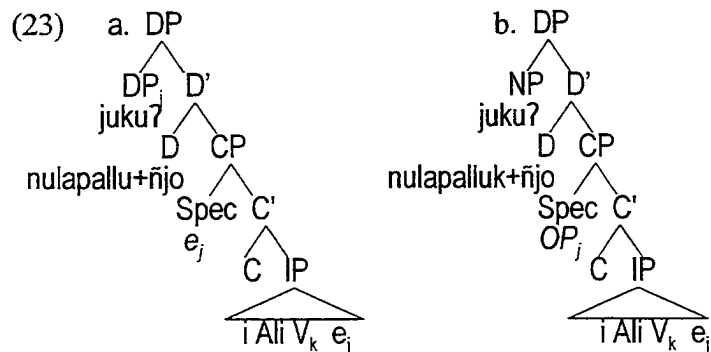
The determiner-final NP structure in (6), revised in (21a) to accommodate relative clauses becomes problematic once the *V+D* pattern in the data is considered. In order to derive the correct word order from this structure, D-lowering appears to be necessary, and such processes are in general, even if motivated, fairly rare in the theory of grammar. Similarly, while (21b) does not particularly invite a lowering analysis, it shares the adjunction structure with (21a), and so V-raising as an alternative to D-lowering fares no better here. (21c), a relabeled version of (21b), has similar shortcomings, side-to-side movement not being theoretically countenanced..



What is clearly needed for these cases is a structure in which the verb may licitly raise out of its containing CP and adjoin to D.<sup>12</sup> Each structure shown in (2), where CP is a complement of D, fulfills this requirement. As a preliminary to the analysis, I will assume, following Minimalist assumptions, that the relativizer and the subject agreement affix are both attached to the verb in the lexicon, and that the appropriate features on Agr and C are checked via overt movement of the verb. Since the determiner can occur independently of the verb (cf. (3)-(5)), the determiner is represented below as heading its own projection, and the *V+I+C* complex combines with D by left-adjunction. Let us now consider the examples in (18)-(19) in light of the D-CP hypothesis. For simplicity and concreteness, let us first take the free relative example (19c). I assume that a null operator moves from object position to a case-checking position and then to Spec,CP, thereby forming the open sentence characteristic of relative clauses.<sup>13</sup> Driven by strong V features on I, the verb raises through IP. It then moves to C, and the strong feature carried by the relativizing affix *nu-* is checked. Finally, the complex adjoins to D, checking a strong C feature on D..



Since CP is a complement of D in (22), head-raising from C to D is allowed and the categories can be licitly combined, in contrast to the derivations discussed for the examples in (21). Relative clauses with overt external heads, such as (8b), can be derived according to Kayne's (1995) proposals by raising the head from inside the clause to Spec, CP and from there to Spec, DP as in (23a). Alternately, under Larson's DP-shell approach, the head can be initially generated externally to the CP in the Spec position of DP (23b).<sup>14</sup>



As stated above, the possessive morpheme has a distribution similar to that of the definite determiner, and in line with this, I treat the possessive as a syntactic head analogous to D. As the following Selayarese examples show, the possessive (**boldfaced**) is a clitic on the nominal head if the DP contains no modifying material, and it is a clitic on the modifier otherwise.

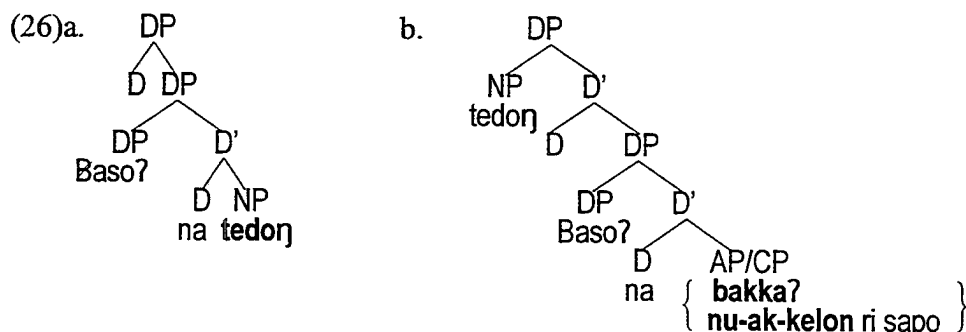
- (24) a. tedon-**na** i Baso?  
*buffalo-3P cl*  
 'Baso?'s buffalo'
- b. tedon bakka?-**na** i Baso?  
*buffalo big-3P cl*  
 'Baso?'s big buffalo'

As with the definite determiner and relative clauses, the possessive and relative clauses show further complexities. In the Selayarese examples in (25a-b), below, the possessive cliticizes to the verb of the relative clause, as would be expected if it has D-like status. What is particularly striking about (25b) is that the possessor DP occurs between the C+V complex verb of the relative clause and an adjunct PP that belongs to the lower clause. This placement strongly implicates V raising of the sort seen earlier,

indicating that V has raised out of CP into higher DP structure.

- (25) a. *tedoŋ nu-ak-kelon-na ri sapo*  
*buffalo REL-INT-sing-3P in house*  
 'his buffalo that sang in a house'
- b. *tedoŋ nu-ak-kelon-na i Baso? ri sapo*  
*buffalo REL-INT-sing-3P cl B. in house*  
 'Baso?'s buffalo that sang in a house'

In order to account for the patterns seen above, I propose a version of the DP-variant of Larson's (1988) VP-shell. A given XP within the DP, however, will not be projected into its position uniformly. This runs somewhat against some current theories of DP structure (see, e.g., Zamparelli 1995 and references cited therein for discussion of the DP hypothesis), where the NP and modifiers are uniformly projected within the DP structure. The scheme that the deployment seems to follow in the Sulawesi languages discussed here can be characterized as "ergative" in its essentials. That is, assume that the NP head of the construction is analogous to subject, the modifier is analogous to the object, and the possessive is analogous to an oblique argument. Further, assume that complement of D position is mapped to absolutive. Then the phrases within DP will be deployed in the structure as follows: An unmodified NP is projected into complement position, and if a possessor is present, it is projected into Spec position. These two constructions correspond to intransitive structures at the sentential level. The presence of a modifier (object) creates a 'transitive' structure, and the NP is then projected into Spec position, while the modifier is placed into complement position (absolutive). When there are three arguments to D, NP occupies the higher Spec position, the possessor the lower Spec position, and the modifier occupies the low complement position (cf. the projection of arguments in Larson's (1988) treatment of the double object construction). These structures are illustrated below.



The head (boldfaced) of the complement of the lowest D raises and adjoins to the possessor D, and this complex raises to the upper D. Thus, for (26b), the complex adjunction in C (*nu-ak-kelon*) raises to the possessive *na*, and the result *nuakkelonna* then raises to the highest D in the structure.<sup>15</sup>

Although there may be alternative analyses to pursue, this approach to southwestern Sulawesi DP structure looks the most promising from a position internal to the languages. Note, as perhaps part of the ergativity, the primacy of complements over specifiers, a notion exploited by Chomsky (1995) for VP structure. Following the VP analogy, the DPs are either unaccusative or transitive; there are no unergative DP structures.

### 5. Head movement and adjacency

In the previous sections, it was argued that the V raises to D. The combination of constraints that were assumed also yielded the conclusion that the V moved through intervening functional projections, and for purposes of illustration, adjunction rather than substitution was utilized. Recently, however, van Riemsdijk (1995) has proposed that adjunction is an allowable way of terminating head movement only when the affected categories are adjacent. Where they are non-adjacent, head movement applies with substitution as the result. Van Riemsdijk's definition of adjacency is given below:

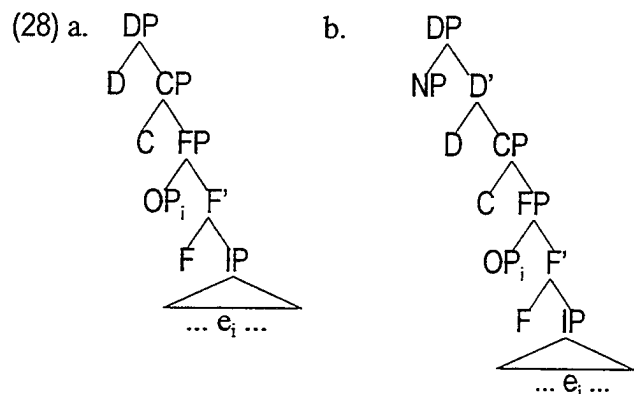
- (27) From van Riemsdijk 1995:
- a. Head-Movement takes place by adjunction only when the heads are adjacent. Otherwise by substitution.
  - b. Adjacency:  $\alpha$  and  $\beta$  are said to be systematically adjacent iff there is no element  $\gamma$  such that  $\alpha$  precedes  $\gamma$  and  $\gamma$  precedes  $\beta$  and  $\gamma$  is phonetically realized or is an obligatory X'-position.

In order to see the relation between the set of proposals here and van Riemsdijk's proposals, let us observe the assumed course of the verb's movements as it goes from V to D. For the sake of argument, let us assume that all of the bound morphology is attached in the lexicon. The verb is thus fully inflected, containing the ergative marker and the relativizer (everything except the determiner, which has independent existence, as illustrated above). The movement of the verb takes place, therefore, for simple feature-checking requirements, not for assembly. In each of the movements up through the penultimate movement to C, a plausible case can be made for substitution, for as van Riemsdijk notes, there is little difference between substitution and adjunction if what is being adjoined to is a feature bundle that disappears after the adjunction takes place. So, let us assume that substitution takes place in these instances.

The problem now becomes one of legitimizing the last move, from C to D across Spec,CP. The relevant question here is whether Spec,CP is an obligatory X'-position, and another is in what sense should we take 'obligatory' to apply at this level. Consider first the case of covert movement. As was sketched above, the VSO order that Selayarese may also exhibit comes from an SVO structure inside VP in which only the V raises to higher functional projections. The subject and object raise in the covert component to check the relevant Agr and D features on the verb. Since it can be argued that covert movement involves only the movement of formal features, not full XPs, it is plausible to assume that the specifiers of the Agr projections are simply not projected at any level of representation in the VSO cases. Further, if Spec positions are not created until they are filled, there will be no empty Spec positions. Adjunction of C to D will therefore be compatible with van Riemsdijk's constraint only if Spec, CP is not present at the time that adjunction of C to D takes place. Given this, the prediction is made that the features of the C in the relative clause that enforce operator movement are weak, not being checked until the covert component. This need not mean, however, that OP remains in situ; it is plausible that it raises first to Spec of FP, the projection between IP and CP illustrated above, creating an A' dependency. As we saw above, Selayarese focus structures are derived by overt movement, and so any island-type effects observable in relativization at spellout can be attributed to movement to Spec of FP, not necessarily to Spec of CP, and this is compatible with van Riemsdijk's constraint, since head movement from F to C (across OP, which fills the Spec position) could be viewed as substitution rather than adjunction. At the point at which C raises to D, then, the two positions are adjacent in the relevant sense.<sup>16</sup> A free relative would therefore have the structure shown below in (25), while a headed structure would receive the structure in (28b) with the NP head introduced via 'merge'. If an operator were raised to, or the head of the



construction were raised from CP, as in (23), then head movement would have to apply across a filled Spec position, in violation of the adjacency condition given above.



Summarizing to this point, this analysis of Sulawesi relatives relies crucially on two assumptions. The first is that the verb of the relative clause raises to D, and the second is that this movement is governed by general constraints governing head movement. Given these hypotheses, the CP-complement relative clause structures are supported.

## 6. Cross-linguistic implications

While the type of relative clause discussed here involves a rare form of head-movement, it can perhaps be understood as arising from the joint effect of two conditions affecting heads in the grammar, one concerning V and another concerning C. First, since the relativizing morphemes *to-* and *nu-* are attached to the verb, the verb must raise so that the relevant features can be checked. Second, raising of this C+V complex to D is also dependent upon the nature of C. Since D and CP are in a head-complement relation, D selects CP. Furthermore, the predicative nature of the CP derives from the nature of its head, which triggers operator movement. We can think of the selection relation as being enforced by feature checking between the D and the relativizing C. As a cross-linguistic consequence, it is not unreasonable to assume that the relativizing C always checks its features against D, either overtly or covertly. In the languages discussed here, this checking takes place overtly by head-adjunction.<sup>17</sup> Also in these languages, C and V are lexically attached for completely independent reasons of morphology, and so at the appropriate level, here in the overt syntax, V raises to C.

Pursuing this a bit further, if the relativizing C is affixal in nature, whatever it is an affix on will raise to C, and universally, the relativizing C raises to D at the appropriate level. Overt raising is dependent upon feature strength, and so typologically, the possibilities shown in the table below are expected, assuming that parametric variation can be attributed at least in part to variability in feature strength.

Affixal relativizer checked against C	Strong REL features on C (V raises to C)	Strong X features on D (X raises to D)	overt effect
a.	yes	yes	V raises to C, C+V raises to D
b.	yes	no	V raises to C, C+V stays within CP
c.	no	no	V stays within IP, C stays within CP
d.	no	yes	V stays within IP, C raises to D

Two varieties of languages in which the relativizing morpheme is attached to the verb of the relative are predicted. The Sulawesi type in which the verb raises overtly to C, and the type in which the verb raises covertly to C. This typology has yet to be justified fully empirically, but some of the predicted patterns seem to be attested.

Malayalam appears to be an example of the group with overt raising. Consider the data in (27), in which the word order of main clauses is very free. Note in particular the various positions in which the verb can occur with respect to its arguments. The verb may appear finally, medially, or initially. Malayalam (26b)

- (29) a. kutti aanaye **talli**  
*child elephant-ACC/DEF pushed*  
 ‘The child pushed the elephant’  
 b. **Talli** kutti aanaye  
 c. kutti **talli** aanaye  
 d-f. all other permutations are well formed.

In relative clauses, however, the verb can only appear finally, even though scrambling of the arguments is possible.

- (30) a. Gopu [[puuccaye **talliya**] aanaye] kandu  
*G cat-ACC/DEF push-REL elephant-ACC see*  
 ‘Gopu saw the elephant that pushed the cat’  
 b. \*... **talliya** puccaye aanaye ...
- (31) a. aana [[Aniline vyaakaranam **padippikkyunna**] aaline] talli  
*elephant A grammar teach-REL man-ACC push-PST*  
 ‘The elephant pushed the man who teaches grammar to Anil’  
 b. aana [[vyaakaranam Aniline **padippikkyunna**] aaline] talli  
 c. \* aana [[vyaakaranam **padippikkyunna** Aniline] aaline] talli  
 d. \* aana [[**padippikkyunna** vyaakaranam Aniline] aaline] talli

Similar restrictions are found in Malayalam conditionals. In one form, a conditional morpheme is affixed to the verb, and only the arguments can scramble. The position of the verb is fixed.

- (32) a. aana puccaye **tallunnaal**, Gopu nilavilikkyum  
*elephant cat-ACC push-CND G. shout-FUT*  
 ‘If the elephant pushes the cat, Gopu will shout’  
 b. puccaye aana **tallunnaal**, Gopu nilavilikkyum  
 c. \*aana **tallunnaal** puccaye, Gopu nilavilikkyum  
 d. \*aana **tallunn** pucceye-aal, Gopu nilavilikkyum

In another form of this construction, the conditional element is a free morpheme, and here, the verb of the conditional clause is much more mobile.

- (33) a. aana puccaye **tallunn** engil, Gopu nilavilikkyum  
*elph cat-ACC push CND G shout-FUT*  
 ‘If the elephant pushes the cat, Gopu will shout’
- b. ? aana **tallunn** puccaye engil, Gopu nilavilikkyum

The examples in (31)- (33) can be accounted for straightforwardly with the assumption of V to C raising. That is, when V raises to C, it is moved outside the domain of scrambling, and is thus immune to permutation. For the examples in (34), on the other hand, we can hypothesize that the status of the conditional element as a free morpheme does not require that a checking relation be overtly established between V and C. The verb therefore remains within the subordinate clause and it may therefore participate in word order variation with its arguments.

Internally-headed relative clauses from Cuzco Quechua (Lefebvre and Muysken 1988) provide another relevant case. Here, the relativizing element is bound to the verb, and there is word order variation within the relative clause. This is expected, given independent availability of scrambling along with covert V to C raising. In (35a) the V is initial, and the head of the relative clause ‘girl’ is on the right periphery, internal to the clause; it is not a right-hand external head, Lefebvre and Muysken argue, since it is marked with accusative case, which is incompatible with the intransitivity of the main verb. In (35b), on the other hand, the V in the relative clause is in final position, and the relative clause as a whole is topicalized.

Cuzco Quechua (Lefebvre & Muysken 1988) (27c)

- (34) a. **Riku-sqa-y** warma-ta hamu-nqa  
*see-NOM I girl-ACC come fut*  
 ‘the girl I saw will come’
- b. qayna warma **riku-sqa -y -ta-qa** hamu-nqu  
*yesterday girl see-NOM I ACC TOP come fut*  
 ‘the girl I saw yesterday will come’

If the verb does not raise to check its nominalization/relativization feature against C but remains as a constituent of IP, then it will be able to participate in the word order variations illustrated above.

## 5. Concluding remarks

It has been argued that the formation of relative clauses in three Sulawesi languages involves a form of head movement in which V adjoins to D. The only structure in which such movement is licensed, according to current assumptions, is one in which CP is a complement to D, a departure from formerly widely-adopted structures. In addition, the deployment of phrasal categories within the DP seemed to follow an ergative pattern, even though the arrangement of arguments in these languages is essentially accusative at the sentential level, despite the agreement patterns shown by the verb. The relation between these two forms of ergativity deserves further attention. It was further proposed that the mechanism that provides the motivation for the movement is feature strength. C has a strong feature that attracts V, and D has a strong feature that attracts C. If there is cross-linguistic variability in feature-strength, other forms of relative clauses are expected, and some examples were provided that conform to the predicted typological range.

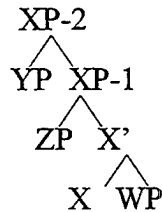
## NOTES

1. Thanks to members of the audience at AFLA IV for their comments and questions. I am particularly indebted to Hasan Basri for sharing his knowledge of Selayarese with me, and to Anthony Jukes for pointing out some errors in the Makassarese data in earlier drafts. This research was supported by grant SBR95-11171 from the National Science Foundation. Morpheme glosses should be fairly transparent; *E*=ergative, *A*=absolutive, *TR*=transitivizer, etc.

2. See, e.g., Ross 1967 and Stockwell, Schachter, and Partee 1973, for discussion.

3. According to HMC, head movement is possible only when a head/head-of-a-complement relation exists between the affected syntactic categories. In Kayne's framework, right adjunction is ruled out independently, and so the facts discussed here could in fact be viewed as evidence in favor of the principles which derive the ban on rightward adjunction.

4. Strictly speaking, even legitimate head movement does not follow the Extension Condition, as the adjunction is internal to the topmost category (see, e.g., Marantz 1995 for discussion). Instead of creating an exception clause for head movement, I tentatively propose that the condition be relativized to "reverse projections" of the current root node. Thus, movement may adjoin to the root node or any of its reverse projections. A reverse projection of XP is any member of the set of nodes that projects to XP. Thus in the structure below, XP-1, X', and X are reverse projections of XP-2.

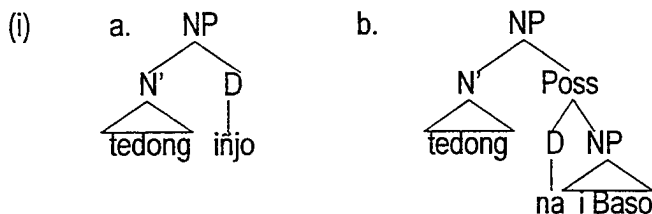


Internal adjunction is allowed to a degree by this amendment, although its full consequences have yet to be explored. Also, the implications for 'merge' are unclear at this point as well.

5. See Smith 1964 for an early proposal in which the clause is a complement to D, and Chomsky (1965) states "...restrictive relatives belong to the determiner system." (p. 217, note 26; see also Chapter 3). Stockwell, Schachter, and Partee 1973 discuss such structures alongside adjunctions such as those in (1).

6. One way to implement the apparent optionality here is to assume that the lexicon contains two instances of Abs, one with a strong D-feature, the other with a weak feature. If the Abs with the strong feature enters the derivation, VOS order results. If the other occurrence of Abs is present, then VSO order results. See Deprez 1997 for a fuller discussion of optionality in this sense.

7. An NP structure of the following sort is also possible:



Aside from the labelling, there is little relevant difference between (15) and (i).

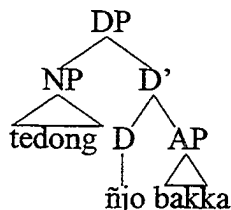
8. The adjunction structure is then interpreted by the phonology as a prosodic word, and

word-internal phonological effects are observed (see Basri et al 1997).

9. The DP in (i) shows adjectival modification:

- (i) tedong bakka-ñjo  
*buffalo big-the*  
 'the big buffalo'

This word order is compatible with the head-raising analysis in the text if it is assumed that the nominal head is projected as Spec and the additional modifier is projected as the complement of D.



See also work by Abney (1987) and Radford (1990) for proposals which inspire an alternative of the following form:

- (ii)  $[_{DP} D [_{AP} A [_{NP} N ]]] \Rightarrow [_{DP} N+A+D [ e [ e ] ]]$

The structure in (ii), however, affords no obvious extension to an analysis of the relative clauses.

10. Observe that the object agreement element is attached to the verbal complex to the right of D in these examples. This pattern suggests that the verb and determiner combine before the agreement element is attached, in turn suggesting that the agreement marker is probably neither the head of an agreement projection above VP nor attached in the lexicon (as assumed in Finer in press). If it has phrasal status instead, it can be skipped by head movement. Other data concerning the distribution of this agreement element discussed in Finer 1995 suggests that it is a free clitic, in the sense of Selkirk 1995. See also Basri et al 1997 for a discussion of the prosodic properties of this and other clitic-like elements in these languages.

11. The Makassarese form of the *to*-relativizer is *tu-*. Anthony Jukes (p.c.) informs me that it occurs primarily in free relatives, and that Makassarese speakers prefer the zero form in headed relatives.

12. V+D cliticization in and of itself, however, is compatible with either N'-adjunction or D-complementation structures. Amharic, for example, shows V+D patterns as well as SOV word order in its relative clauses, and so here, as long as the V-final CP/IP ends up to the left of D, by either direct generation or movement (Kayne 1995), simple phonological adjacency is sufficient to produce the observed data. The V+D-DP order in (9b) and (10c), however, shows that something more than phonological adjacency is required for the Sulawesi cases.

13. The choice of this particular example sidesteps the placement of the object agreement marker (see note 10). Briefly, I assume that the verb in examples such as (9b) and (10c) raises across the object marker, which then cliticizes to the derived verbal complex under phonetic adjacency (see Basri et al 1997 for more discussion). (10) would correspond more or less to the D-structure of (9c) within earlier frameworks; under Minimalist assumptions (e.g., Chomsky 1995), however, structure-building and movement proceed hand in hand, so at the time that D and CP are joined, V would already be adjoined to C.

14. The larger structure shown in (2a) will be exploited in subsequent sections where more complex DPs are considered. As discussed below, possessive constructions appear to provide evidence for introducing the nominal head of the construction via merge rather than move.

15. There may in fact be a higher functional head to which D raises so that the phi-features of the D head can be checked against the possessor, which raises to Spec position. Such movements

would appear to be necessary if features can be checked only as a result of movement, not by merge (a position argued for in Finer in press for a different set of constructions). If the phi features of *na* could be checked immediately by the possessor merged into the local spec position, there is no need for the additional structure.

16. The analysis of the possessive structures offered in (17) above is in conflict with the constraint on adjunction, however, since N raises across a filled Spec. The second proposal, in which D raises across a filled Spec to a higher empty D position which is filled by substitution, however, is unobjectionable in this respect.

17. As discussed above, overt C to D raising would be expected only when Spec, CP is not projected overtly. Such raising then should correlate with either no clause internal movement to Spec of CP, or, as suggested here, OP movement to a focus position below C.

### References

- Abney, Steven. 1987. *The English Noun Phrase in its Sentential Aspect*, unpublished dissertation, MIT, Cambridge, Mass.
- Baker, Mark. 1988. *Incorporation* Chicago: University of Chicago Press.
- Basri, Hasan, Ellen Broselow, Daniel Finer, and Lisa Selkirk. 1997 "Prosodic Adjuncts in Makassarese," paper presented at the Conference on the Phonological Word, Berlin.
- Bobaljik, Jonathon: 1993, "Nominally Absolutive is not Absolutely Nominative," in J. Mead, ed., *The Proceedings of WCCFL XI*, CSLI publications, pp. 44-60.
- Brochie, Amanda. 1992. *Verbal Agreement and Constituent Order in Bugis*, unpublished Honours Thesis, University of Melbourne.
- Chomsky, Noam. 1995. *The Minimalist Program* Cambridge, Mass: MIT Press.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass: MIT Press.
- Deprez, Vivienne 1997. ms., Rutgers University, New Brunswick NJ.
- Finer, Daniel. 1994. "On the Nature of Two A' Positions in Selayarese." in N. Corver and H. van Riemsdijk, eds. *Studies on Scrambling*, J. Benjamins.
- Finer, Daniel 1995. "Covert Movement and the Distribution of the Selayarese Absolutive Marker," ms. SUNY at Stony Brook.
- Finer, Daniel. In press. "Contrasting A'-dependencies in Selayarese," to appear in *Natural Language and Linguistic Theory*.
- Kahler, H. 1974. "Relative clause formation in some Indonesian languages." *Oceanic Linguistics* 13, pp. 257-277.
- Kayne, Richard S. 1995. *The Antisymmetry of Syntax* Cambridge, Mass: MIT Press.
- Larson, Richard K. 1994. "The projection of DP and DegP structure." Ms. SUNY Stony Brook. Stony Brook NY.
- Levin, Juliette and Diane Massam: 1985, "Surface Ergativity: Case/Theta Relations Reexamined," in Steven Berman, et al. eds., *Proceedings of NELS XV*. GLSA Publications, University of Massachusetts at Amherst.
- Longobardi, Giuseppe. 1994. "Reference and proper names." *Linguistic Inquiry* 25: 609-665.
- Massam, Diane: 1985, *Case Theory and the Projection Principle* PhD dissertation, MIT.
- Massam, Diane: 1996, "Clause Structure and Case in Niuean," presented at the 3rd meeting of the Austronesian Formal Linguistics Association, UCLA.
- Marantz, Alec. 1995 "The Minimalist Program," in Gert Webelhuth, ed., *Government Binding Theory and The Minimalist Program*, Blackwell. pp. 349-381.

- Radford, Andrew. 1990. *Syntactic theory and the acquisition of English syntax* Oxford, UK: Basil Blackwell.
- Riemsdijk, Henk van 1995. "Head movement and adjacency." Ms. Tilburg University, Tilburg NL.
- Ross, John R. 1967. *Constraints on variables in syntax*. unpublished MIT dissertation. Cambridge, Mass.
- Selkirk, Elisabeth O. 1995 "The Prosodic Structure of Function Words" in *Papers in Optimality Theory*, GLSA, University of Massachusetts at Amherst, pp.439-466.
- Smith, Carlota. 1964. Determiners and relative clauses in a generative grammar of English, " *Language* 40:37-52, and in *Modern studies in English* eds. David Reibel and Sanford Schane. pp. 247-263. New York: Prentice-Hall.
- Stockwell, Robert, Paul Schachter, and Barbara Partee. 1973. *The major syntactic structures of English* New York: Holt, Reinhart, and Winston.
- Travis, Lisa. 1984. *Parameters and effects of word-order variation* MIT dissertation. Cambridge, Mass.





## Chamorro: Whence and How?

### (An Inquiry in Diachronic Syntax)

Joseph C. Finney

#### Introductory Remarks

This study is part of an ongoing program in Austronesian (AN) diachronic syntax, with emphasis on prepositions and casemarkers (henceforth PCM). In Indo-European are found both prepositions and casemarking suffixes; the latter are an older generation of former postpositions surviving in a highly lenited and grammaticalized form. Many Austronesian languages today have monosyllabic particles that can be called either prepositions or casemarkers, and are often identical with (or close to) their likely Proto-Austronesian (PAN) form.

Two major sources on Chamorro are Topping et al. (1969 textbook, 1973 reference grammar, 1975 dictionary), and Chung (1981 syntactic study): two authors with different strengths. Topping has long and deep experience with the language, while Chung has expertise in comparative syntax. The two drew quite different conclusions about PCMs and casemarking in Chamorro.

Chamorro is an Austronesian (AN) language spoken in Guam and the recently established Commonwealth of the Northern Mariana Islands. To its South are two (nuclear) Micronesian Oceanic AN languages (Ulithi and then Woleai). Southwest of Ulithi lie two other AN areas: Yap and then Palau, both rather idiosyncratic. Chamorro's syntax is of the focus type, found in Formosan languages such as Atayal and Malayo-Polynesian (henceforth MP) such as Tagalog.

Chamorro was severely impacted by centuries of Spanish, till 1899 (and since then, by German, Japanese, and American English). It is saturated with Spanish words, including even prepositions, a drawback for the present morphological-syntactic study in which PCMs take a key role. Topping (1973, 166) says, "I have never met any Chamorro speaker with any knowledge of the pre-Spanish numeral system" [which had been reported struggling for survival in the 1905-1913 period]. He gives Safford's (1909) list of the old Chamorro numbers: 1 hacha, 2 hugua, 3 tulu, 4 fatfat, 5 lima, 6 gunum, 7 fiti, 8 gualu, 9 sigua.

I am grateful to David Zorc for his generosity with his time, advice, and correction, both in letters and by telephone, and for giving me sources I had not known of. Presented with the problem, how the PAN/PMP preposition or casemarker ki can be gi in Chamorro, where ancestral \*k is normally h, he replied that it is not uncommon to see function words make their own phonology. In Waray, Central Philippines, where \*g is unchanged in major words, it becomes h in the particles, as sa > ha, and san > han. I dub this the Zorc Syntactic Phonological Principle, and shall abbreviate it ZSPP. Of course, he is not responsible for any of

my errors that remain.

In this paper, when I speak of a "definite" article, I really mean "specific" article, as it glosses not only "the" but also "a certain".

Throughout the paper, square brackets mark my comments or amendments.

For simplification, in the representation of PAN, Philippine and Formosan words, as well as Chamorro, I omit the laryngeals, mostly glottal stops, none of which are relevant to the matters of this paper. Experts on Atayal do not agree whether initial glottals are phonemic or not.

Topping (1973) considered Chamorro to be a Philippine language. More recently, Starosta (1995), on the basis of syntactic theory that will not be discussed here, proposed that Chamorro broke off on its own path fairly close to PAN, leaving a common descent path from which, several branchings farther down, Proto-MP broke off. We may somewhat inaccurately describe Starosta's theory as putting Chamorro among the Formosan languages; at any rate, definitely not among the MP languages. Malcolm Ross, however (personal communication) remarks that Chamorro shows the MP phonological changes.

### Diachronic Phonological Issues

#### Chamorro words with PAN \*C

The languages grouped as Malayo-Polynesian, as well as some in Formosa, merge PAN \*C and \*t as t, in the traditional theory. Wolff (1991), however, has proposed that C derived from t positionally, depending on the location of stress accent in PAN. Ross has pointed out that such a change is not so common as to have occurred independently in different languages. That may be read to imply that, if such an unlikely split occurred, it happened only once, and may have been the primary division within Austronesian.

Chamorro shows t in words accepted as PAN \*C, and which have \*t in MP languages: talanga "ear", mata "eye", hutu "louse", gugat "vein", and langet "sky". That set tends to support the idea that Chamorro is MP. It separates Chamorro from most Formosan languages, and fails to fit Starosta's proposed order of branching. It must be noted that PAN \*C is "t" also in some languages in southeast Formosa: Bunun, Amis and Puyama.

#### The interesting word "sa" / "ta" : "because (of)"

Chamorro has the word sa' (Topping, 1975, 179), glossed: "Because, because of, on account of, due to the fact." It has an alternate form sis (189). Examples show it used both as a preposition and as a conjunction.

MP languages have the equivalent word, with much the same range of uses, but have it as ta. It is glossed "because" in Pangasinan, and both "because" and "so" in Ilokano, Isneg and Bikol. It was not found in Hiligaynon, Tagalog, Sebuano, nor Manobo. If Chamorro is MP, the word should be ta also in Chamorro. It occurs as te in the Belu dialect of Tetun (Morris, 1984, 183). The dialect is spoken in south-west areas of East Timor and south-east areas of West Timor. It is called a conjunction, and glossed "because, so that".

### Casemarking in Some Other Archaic AN Tongues

Before examining casemarking in Chamorro, let's look at marking for the three basic cases in four archaic AN languages: Amis, Old Javanese, Atayal, and Tagalog, so that we can see where Chamorro fits into the picture.

For Nataoran Amis, a Formosan language, Chen (1987, 127) gives this table of PCMs:

	Topic	Nominative	Genitive	Accusative	Locative
-demonstrative:	o	ko	no	to	i
+demonstrative:	ia	kia	nia	tia	itia

Becker and Hunter (1988) give this table of "deictic prepositions" in Old Javanese:

	[Case 1] [Nominative] direct	[Case 2] [Genitive] oblique	[Case 3] [Dative-Locative] directional
indefinite, nonspecific	i	ni	ri
definite	ing	ning	ring
definite & specific	ikang	nikang	rikang

Victoria Rau, a Chinese who is fluent in both Atayal and Tagalog, asserts that the two have striking similarity. In those two reference languages and others closely kin to them, we find:

	Case 1	Case 2	Case 3
Atayal			
impers spec	(a) ku	na ku	sa ku, te, i
impers nonsp	a, zero	na	sa, te
pers	i, zero	ni	ki, te
Tagalog or related dialects:			
impers spec	a(ng)	na(ng)	sa(ng)
impers nonsp	i(ng)	ni(ng)	si(ng), ki(ng)
pers	si	ni	kay, ki
dem pron	i-	ni-	di-

The paradigm above is slightly idealized for simplification. In some Philippine languages the s markers have come to be used for Case 2. Note that the case 2 marker /nang/ for impersonal nouns in Tagalog is customarily written in the abbreviated form "ng".

### Casemarkers and Prepositions in Chamorro

Topping scarcely recognizes the concept of casemarkers; he often calls "article" or "preposition" the words that others may call casemarkers. He says (1973,119), "Prepositions in Chamorro can be divided into two types: Spanish and non-Spanish." The Spanish are more numerous. "The non-Spanish, or Chamorro, prepositions are presented first. They are gi, giya, para, ginen or gini, and as." Of course, para is a Spanish preposition, but it is used as if it were non-Spanish, and Topping even speculates that Chamorro may have had a homophonous preposition before the Spanish came. He illustrates (119) Spanish and non-Spanish ways of saying the same sentence, "He cleaned the house for his mother." As a "non-Spanish" preposition, para means "to" in contrast to ginen "from". (The "Spanish" forms are asta and desde respectively.)

He also lists common articles (132) i, ni, and nu and notes that ni is a contraction of nu i. He says (135), "Safford describes nu as a preposition; it seems to follow more closely the pattern of articles." He lists (130) proper articles for persons (si, as) and places (iya). He recognizes (122) that giya is a contraction of [PCM] gi and [article] iya, and suggests that gi itself is a contraction of [PCM] gi and [article] i. He fails to note that i is also a contraction of [PCM] i and [article] i, though a glimmer of it may be inferred from the confused dictionary entry (1975,93):

i The (definite article). Focus marking particle. Cf. Ni.

Topping's great corpus of sentences show gi as the main preposition in Chamorro, with common meanings "to", "for", "from", "at" and "in". It is the all-purpose space-relation preposition comparable to sa in Central Philippine languages, di in Indonesian/Malayan, and i in Wolio (Anceaux 1987). While Chamorro embedded prepositional phrases (such as "on top of") have gi as the left and i as the right particle, note that in Bikol (Zorc, 1992) the construction has sa as its left preposition and kan as its right.

Examples where i is a contraction of PCM i and article i are in Topping (1973, 120) :

gi hilo i ( = i + i ) lamasa  
on top of the table

(gi= gi + i ) fi'on (i = i + i ) lapas  
at the side of the pencil

The PCM i is from a PAN PCM, while the article i, like the form

iya, is from the PAN demonstrative \*i(y)a.

He might also appropriately regard "article" ni as PCM plus article, as he gives sentences in which ni is clearly a PCM for agent or instrument, consistently with its use in four archaic languages cited above as a marker for Case 2, the genitive/ergative case.

The ancestry of most of the PCMs and articles is clear. The i and iya articles are from PAN i(y)a. The PCMs are easily recognizable as forms shown in the four archaic languages. The preposition ginen is gi with a suffix, and is sometimes interchangeable with simple gi. The nen suffix is comparable to suffixes found on prepositions in other Philippine languages and Oceanic languages such as Tolo (Crowley, 1986, 14) where the preposition hi < ki appears as hinana and hinia. The only PCM of diachronically uncertain form is as. One can only surmise that it is the PCM sa (found in Tagalog and many other languages), with a vowel prefixed and the final vowel subsequently lost. Or, less likely, a particle or prefix a + (article/casemarker) si with elision of the final -i. Something similar happened in Palau, where the only ancient particles are er and el, created by prefixing a vowel and then losing the final vowel, from the PMP \*(d,D,Z)i and \*ni particles, Cases 3 and 2 respectively.

Topping's examples of "article" as include use as ergative agent marker (as Tomas in "Thomas saw me", as Juan in "Maria was kissed by Juan"); and as personal locative "at Pedro's [house or place]".

Chung (1981) listed the casemarkers for definite NPs (indefinites show no casemarking):

	Case 1 Unmarked	Case 2 Oblique	Case 3 Local
common noun	-- [i]	ni [nu i]	gi
proper noun	si	as	gias [gi as]
pronoun	-- [--, gu-]	nu	giya [gi iya]

(My suggested changes, in brackets, are affected by Topping's huge mass of sentences with glosses. Chung spells gias one word, no hyphen, while Topping makes it two: gi as.) Substitute nu i for ni in the table, from Topping's (not Chung's) analysis. I would fill the "unmarked, common noun" slot with i. Likeness of uses shows that whatever i and si are, it's the same. Topping calls both articles (also ni); I call all three combined casemarker+article.

If, indeed, the Chamorro particle sa comes from earlier sa as in Tagalog, it's of interest in that part of what is Case 3 is merged into Case 2 in Chamorro, while the ki-marked part, lenited to gi, remains as the surviving Case 3. Note also the curiosity

of successive PCMs "gi as". The alternative still remains that as is a prefixed si; if so, "gi as" is analogous to "gi iya", with the first word a PCM and the second an article, diachronically.

In languages in which prepositions and casemarkers are separate classes of words, the relation is often many prepositions to few casemarkers. Chamorro seems to be in a transition to that state, and hence the ambiguity of status of gi. Chung considers it a casemarkers comparable to "si", "as" and "ni", while Topping considers it a preposition comparable to other native and to the numerous Spanish prepositions.

It is notable that in Chamorro ni (or nu-i) marks case form 2 for impersonals but not personals. That seems odd because it is well-known that the reverse is true of archaic Austronesian languages in general (and presumably of PAN) where na marks impersonals.

#### More on the Diachronic Context of Chamorro Casemarking

Topping basically collapses ni and nu, pointing out (correctly) that ni is a contraction of nu + i, and noting that pronouns take the simple nu. When we note that in Paiwan, the PCM na is a contraction of nu (PCM) and a (presumably article), it appears that in this one instance, Chamorro (like Paiwan) is arrested at a development stage many thousands of years old, in which the contraction of nu + article (i, a) is in progress (an unstable state?) and not completed: an amazing and fascinating fact. It's not clear just what that means in its prehistory and its place on the family tree. The simple genitive marker nu is still used with nouns in some Western MP languages.

In most Philippine languages, such as Tagalog (Schachter 1972: 88), and Formosan language such as Amis, Saisiyat, Tsou (Starosta 1974: 301, 326, 349) the particle "si" can be defined either as a case 1 marker used only for personal nouns and pronouns or as a personal article used only in case 1 (sometimes also accusative case, if any) and usually only in third person. Ross (personal communication) mentions that in three Formosan languages si (or ti) is "a personal honorific which follows the common CM", and thinks it was also so in PAN.

Topping considers that Chamorro has two words as, one a preposition, the other a proper article. He lists the proper articles as si, as, and iya. In our table above, the proper articles are zero for pronouns for case 1 (with zero casemarkers) and case 2 (with casemarkers nu); and the si, as and iya forms fill the slots for all three cases for proper names and the case 3 slot for pronouns (where giya can be recognized as preposition gi + article/casemarkers iya, which appears to be the PAN demonstrative with or without a locative PCM i). In fact, the "article" iya is used with proper names of places, as in:

falak	iya	Honolulu
go (to)	(to) ART	Honolulu

Topping's separation of article as from preposition as seems unsuccessful, as on page 125 as an example of preposition as he gives:

gi as Bernardo "at Bernardo's (place)"

but on page 131 as an example of article as he gives:

gi as Juan "at Juan's"

using what is surely the same construction.

The case 2 (genitive) particle na, reconstructed for PAN for impersonal nouns, also occurs in Chamorro, but is not part of this paradigm. It is called a linking particle, like its cognate el in Palau. In linking nouns, the order is the expected one: possession NA possessor, and na can be glossed "of". But with an adjective the order is: adjective NA noun, which appears to make the adjective the head of the noun phrase. These statements are also true of Philippine languages such as Tagalog and Formosan languages such as Atayal. Tagalog also has: noun NANG noun, used to make the second noun specific.

To put the Chamorro case system in context, let's compare it with that of Philippine and Formosan languages.

This study works within the "focus" syntax which is found in Philippine and Formosan languages and almost certainly was the syntax of PAN. In her grammar of Atayal, a Formosan language, which was her Ph.D. dissertation under John Wolff at Cornell, Rau (1992) remarks that Atayal is very much like Tagalog, a language that she has also studied and taught. Rau found that in Atayal exactly three cases are distinguished in form, and called them nominative, genitive, and dative. Finney (1995) showed in a comparative study that the casemarking assignable to the common ancestor of Atayal and Central Philippine (possibly PAN) seems to be:

	Case 1	Case 2	Case 3
impersonals	a, zero	na	(C,s)a, ka, ti, di, i
personals	i, zero, si	ni	ki, ti, di

I had considered that forms beginning d- were found only in MP languages, not in Formosans. I suggested that Proto Malayo Polynesian (PMP) may have made doublets ti and di from PAN \*ti. Now I am grateful to Ross (personal communication) for the information that locative marker di occurred in the dead Formosan language Paze, and possibly di and even da in others.

Case 1 is the focus case, the syntactic subject. NPs of any formal case and any semantic role can be promoted to case 1. (The concept of promotion is Perlmutter's.) Case 2 can hang NP from NP (thus "genitive"). Its functions as argument of the verb can be many, but most typically ergative (marked agent) or (if impersonal) instrument. In some languages, case 2 can also be the non-focus patient (accusative object). Case 3 is most centrally allative-dative-locative, very often for motion to or toward;

like allatives everywhere, it can come to be used as direct (accusative) object or indirect (dative) object. Case 3 also has miscellaneous oblique uses (time, manner, company, cause, reason), and from locative it easily becomes quasi-genitive.

### The Accreted Initial "gu" [gw] in Chamorro

Chamorro has a number of major words (i.e., non-particles) beginning with "gu-", [gw] before vowels]. In citing words from Topping-Ogo-Dunca's (1975) dictionary, I'll insert a hyphen to make the prefix stand out. Among its occurrences are:

Personal pronouns:

gu-ahu	I (emphatic), me (emphatic)
gu-e'	he, she (subject pronouns with no definite object); him, her (transitive object pronouns)
gu-iyá	he, she, it (emphatic)

Demonstrative locative adverbs:

gu-atu	there, in that direction, away from you and me
gu-ini	here, in this place
gu-ihi	there, away from you and me
gu-enao	there, in that place toward you
gu-enao gu-atu	over there, in your direction
gu-ihi gu-atu	over there, away from you and me
gu-atu gu-enao	to there, toward you

Unclassified:

gu-aha	have, there is, there are, exists
--------	-----------------------------------

Nouns:

gu-ahi	fire (PAN *hapuy) also verb: reduce heat by taking food off the fire
Gu-ahan	placename Guam
Gu-guan	placename, third island north of Saipan

### Blust's Study of Apparent Obstruent Epenthesis

In an interesting paper, Blust (1994) discussed what resembles epenthesis of obstruents (g and d) in Chamorro. Other languages are central to his interest, and his abstract does not even mention Chamorro. But he argues that what happened in pre-Chamorro was not epenthesis of the obstruents but first epenthesis of glide w initially before all initial vowels (of which Chamorro had only a, i, and u) and then fortition of w to gw, and y to dz. In the first step, also, some glides were "predictably" added. An example Blust gives is the numeral two, PAN \*duSa > \*duha > \*dua > \*duwa > \*dugwa > hugwa. It is the addition of w in the environment u<sub>V</sub>, and of y in i<sub>V</sub>, that is considered predictable ("phonemically null but phonetically motivated"); "the unavoidable transition glide between a high vowel and a following unlike vowel".

Blust (113) states, "Before all reflexes of PAN forms that began with \*i, \*a, or the semivowel \*w, Chamorro now has /gw/, while before reflexes of PAN \*u and \*e Chamorro now has /gw/." The result should be that Chamorro would have no vowel-initial words. But Topping's dictionary (1975) shows many of them,



including words than cannot be loans. Among them are the particles i and iyān, which can be explained by the ZSP.

PAN \*ku as the motivator that may have begun it all

I suggest a causative motivation for the early apparent epenthesis before i- and a-, which, as Blust says, is not phonetically motivated. The explanation I suggest is that the process began when the PAN particle ku was melded with nouns and pronouns. The process was exactly analogous to that described by Bender (1981) in (nuclear) Micronesian languages.

Bender's thorough study showed that the singular article, which he reconstructed as te, accounts for the left element in many nouns in Micronesian languages. The only addition I would make to Bender's analysis is that his examples show that the appended element was often ta or ti; that Polynesian languages also show the article in three forms te/ta/ti; and so the article must be reconstructed as those triplets in Proto Eastern Oceanic.

The ku particle may have begun as a Case 1 marker used only for definite nouns; or (the same in different words) a definite article used only for Case 1 nouns. Note that in many Philippine languages Case 1 NPs cannot be nonspecific or indefinite. In the next few paragraphs we demonstrate that ku generalizes in some Formosan languages to a definite specific article for all cases, while in other Formosan languages it generalizes to a Case 1 marker regardless of definiteness.

Here are the specific articles in Mayrinax Atayal as shown by Li (1995) and Huang (1995).

Nom	Gen	Ben	Ins	Acc	Dat
ku	nku	nku	nku	cku	cku

The case form headings shown are those of Starosta's Lexicase. On first glance, Mayrinax seems to have a whole set of unexplained casemarkers, ending in -u. But this is an illusion. Note that true consonant clusters don't occur in any Formosan language. Between adjacently written consonants in Atayal are "schwa" vowels whose historical values are known. So, using information from Rau, we can show here markings for impersonals in Atayal.

Rau's (1992) three-case analysis of Atayal Formosan shows essentially these basic markers for impersonal nouns:

	case 1	case 2	case 3
specific	ku	na ku	ca ku
nonspecific	a / zero	na	ca

In the writing system, the specific markers look different ("nku" for "na ku" and "cku" for "ca ku"). This is because the a vowels,

being unstressed, are reduced to schwas and, being predictable, are considered non-phonemic and so not written. As expanded here, we see that ku, appearing most obviously as the case 1 specific marker, is paradigmatically the specific marker for all cases.

Now we see that, for Atayal, the pure casemarkers are those found on the nonspecific line, as the nonspecific article is zero. The specific line shows that ku is the specific article, and that the apparent case 1 marker "ku" is simply a zero casemarkers plus the specific article.

A somewhat different conclusion would be drawn from Starosta's (1974) comparison of six Formosan languages. In Starosta's texts, no examples of ku appeared in four languages (Saisiyat, Seediq, Bunun, Tsou). In Rukai, ku appeared as the case 1 marker for personal nouns. In Amis, where there is no specific/nonspecific distinction, ko < ku was seen as the case 1 marker for impersonal nouns. So for those languages, ku looks like a marker for case 1 only.

The particle ku (gu in Wulai squliq dialect) serves to introduce substantive clauses, usually serving as subject of a sentence. It resembles English "that" in marking the beginning of a noun clause, serving as a conjunction, complementizer or nominalizer. Examples are easily found throughout the grammars by Rao (1992) and by Huang (1994).

A reasonable interpretation is that ku at some very early time was the marker of specificity for Case 1. Thence it generalized in some Formosan languages to a marker of specificity regardless of case, and in others to a marker of case 1 regardless of specificity. (Zorc, personal communication, says that in PCMs in Central Philippine languages, vowel u is specific or definite, i is nonspecific or indefinite, and a is neutral.) So, ku seems a good candidate for a cognate of the gu- prefix in Chamorro. The phonology is reasonable, as Chamorro has gi for the PCM that is ki in Formosan, Philippine and Oceanic languages. Major words with PAN initial \*k have "h-" in Chamorro: hita for \*kita "we, inclusive"; hu for \*aku, "I". The ZSP allows PCMs to follow a different diachronic phonology.

Once the morphological epenthesis of gu took place, it was subject to reanalysis as a phonological process of fortition. In accordance with Kiparsky's principle that languages change by reordering of rules so as to feed their application, fortition came to be applied in steps to broader classes of occurrence. (1) The process was applied not only initially but also medially. (2) It spread from apparent w > gw to y > dz [still written "y" as in the particle "iya"]. (3) It came to be available not only to nouns and pronouns but also to all classes of major words, though not to particles.

#### Accusative or Accusative-like Objects

The "focus" system of syntax is found generally in languages believed to be most archaic in AN, such as Formosan and

Philippine tongues, and almost certainly was the basic system of Proto AN. When the (patient, undergoer, semantic object) is the focus, Case 1, the (agent, doer, semantic subject) appears in Case 2. Such sentences may be called passive or ergative. Chamorro follows that pattern, too, as Topping shows:

Lini'e' i palao'an as Pedro  
 see Case 1 woman Case 2 (of, by) Pedro  
 (The woman was seen by Pedro; or: Pedro saw the woman.)

Now, what happens in "Actor-focus", i.e., when the (agent, doer, semantic subject) appears in focus, in Case 1, the syntactic subject? What Case and what casemarking are assigned to the (patient, undergoer, semantic object)? In purely Nominative-Accusative languages an accusative case is found; but there is no such case in many AN languages that are considered archaic. Let's call the non-focus patient the accusative object or, if you prefer, accusative-like object.

In Tagalog and perhaps most such languages, the accusative object is in Case 2, at least for nouns (Schachter & Otnes, 1972, 69). Zorc (personal communication) says that in the Central Philippine languages, the accusative object, if a pronoun, is put not in Case 2, but in Case 3. Perhaps that is what Schachter & Otnes mean in saying (75) that the "object complement" must be an unmarked noun or a deictic pronoun; it cannot be a personal pronoun nor a personal noun.

In Wulai Atayal, Rau (1992, 144) reports that the accusative object is in Case 3, the sa case. She phrases it, "The direct object of an active verb is also dative, as illustrated by the following sentences with dative pronoun direct objects... The dative of other phrases [nouns] is also used as a direct object if the object can be viewed as referring to a location." The example she gives is glossed "When he saw the woman".

In Chamorro actor focus, which looks NomAc, the doer nominative agent is in case 1, and the accusative undergoer is also marked Case 1, as shown (Topping 1973, 246):

Si Pedro lumi'e' i palao'an  
 Case 1 Pedro see Case 1 woman

Some of Starosta's (1974) Formosan languages likewise casemark the accusative the same as the nominative. For Amis and Saisiyat, that marker, for personal nouns only, is cognate with Chamorro's si: ci in Amis and hi in Saisiyat. Three others (Saisiyat, Rukai, Seediq) have marker ka common to nominative and accusative for impersonal nouns. Tsou can use ta for both those cases, but has other markers for both that are not in common. For Amis [-pers] Starosta shows Ac to [ ? < \*Cu] against Nom ko [ < \*ku]. See Chen (1987), paradigm given above.

#### Other Interesting Points in Chamorro Syntax

Beside the Philippine-like system of putting affixes to verbs, (the "old" system), the Chamorro language also has, as an alternative, the "new" syntax resembling that of Proto-Oceanic, based on the "hu" series of inner nominative pronouns. The hu series of unstressed pronouns is named for its first person singular which descends from PAN aku, "I". In the stressed series, the first person singular Spanish pronoun yo has replaced the native AN form, leaving the other persons and numbers intact.

The Chamorro form ni also serves as a negative and as a relative pronoun. In neither of those uses, of course, can nu be substituted. The Chamorro negative ni is characteristic of Formosan (Atayal ini'), while Philippine languages characteristically use other words for negatives.

The Chamorro relative pronoun ni, glossed "which (relative), who (relative)", is interesting in that it is what we expect if PCM ni has merged with ia to form prefixed pronoun nia and then lenited with loss of the final vowel. This process is obvious in Central Philippine languages, and shows traces in Atayal (nya, kya). The other consonant+i-prefixed demonstratives are missing in Chamorro, such as the Oceanic \*tia which I reconstruct as the source of the \*te/\*ta/\*ti article.

#### Summary and Conclusion

A study has been done of the casemarking system of Chamorro, based in part on the previous work of two scholars, Chung and Topping, who differed considerably in their approaches and in their resulting classificatory systems. An attempt has been made to fit the Chamorro casemarking system into a general framework, based on studies done by many scholars on Philippine and Formosan languages. One goal of the study was to test Starosta's theory that Chamorro is not MP, and that Chamorro broke off very early from the line that led down to MP and most non-MP AN languages.

Chamorro shows the same three basic cases that are found in Tagalog (Schachter-Otanes 1972) and Atayal Formosan (Rau 1992). But Chamorro has accusative objects in case 1, along with nominative or absolutive subjects; and thus differs from the Tagalog-Atayal system, in which the accusative-like objects must be expressed either in case 2 (genitive-ergative) or in case 3 (allative-dative-locative, etc.). In this regard, Chamorro not only differs from the Tagalog-Atayal consensus, but also agrees with some other Formosan languages reported by Starosta (Saisiyat, Seediq, Amis and Rukai).

Chamorro has a prefix that may be a fossilized article and/or casemarker, gu < \*ku, which in Formosan languages marked case 1 (nominative-absolutive-focus-subject) at least for specific (definite) nouns. Blust has recently proposed a theory of fortition of certain phonemes that seeks to account for the gu- words as well as other classes. The two explanations are compatible and may have worked in succession, i.e., the gu- words with the fossilized prefix from \*ku may have been reanalyzed as

fortitions and motivated the fortition of other consonants.

Examination of other AN languages that are geographically near Chamorro today shows no evidence of close relation nor of substantial syntactic borrowing. In the area, only Palau shares the focus syntax. Chamorro and Palau share some interesting parallels in the phonological development of casemarkers, and in the divergence of the "linking particles" from their genitive origin. Even so, it is not certain that Palau and Chamorro shared a long common path not shared by other MP languages.

The predominance of the phonological evidence supports the conclusion that Chamorro is either MP, or from a proto-dialect close to PMP, and not one whose family tree is the one suggested by Starosta.

### References

- Anceaux, J. C., 1987. Wolio dictionary. Dordrecht: Foris.
- Becker, Alton L. and Thomas Hunter, 1988. Cohesion in Old Javanese prose. pp 475-488 in Richard McGinn (ed) Studies in Austronesian linguistics. Ohio University P. (distributed by U. of Chicago P.)
- Bender, Byron, 1981. A fossilized article in Marshallese. In: Hollyman, J. and Andrew Pawley (Eds.), Studies in Pacific languages and cultures in honour of Bruce Biggs. Auckland: Linguistic Society of New Zealand.
- Blust, Robert, 1994. Obstruent epenthesis and the unity of phonological features. Lingua 93: 111-139.
- Chen, Teresa M., 1987. Verbal constructions and verbal classification in Nataoran-Amis. PL C-85.
- Chung, Sandra, 1981. Transitivity & surface filters in Chamorro. Studies in Pacific languages & cultures in honour of Bruce Biggs (Jim Hollyman & Andrew Pawley, Eds.) 311-332. Auckland: Linguistic Society of New Zealand.
- Crowley, Susan Smith, 1986. Tolo dictionary. PL C-91.
- Finney, Joseph C., 1995. The \*iya Demonstratives. Proceedings of the 7th (1994) International Conference on Austronesian Linguistics (in press).
- Huang, Lillian M., 1994. Ergativity in Atayal. Oceanic Linguistics 33: 129-143.
- Huang, Lillian M., 1995 The syntactic structure of Wulai and Mayrinax Atayal: a comparison. Proceedings of the 7th (1994) International Conference on Austronesian Linguistics (in press).
- Josephs, Lewis S., 1975. Palauan Reference Grammar. Honolulu:

Univ of Hawaii Pr.

Josephs, Lewis S., 1990. New Palauan Dictionary. Honolulu: Univ of Hawaii Pr.

Josephs, Lewis S., 1994. Review Article. Oceanic Linguistics 33: 317-256.

Kaehler, Hans, 1974. Relative clause formation in some Austronesian languages. Oceanic Linguistics 13: 257-277.

Li, Paul Jen-kuei, 1995. The case-marking system in Mayrinax, Atayal. Proceedings of the 7th (1994) International Conference on Austronesian Linguistics (in press).

Morris, Cliff, 1984. Tetun-English dictionary. PL C-83.

Rau, Der-Hwa Victoria, 1992. A Grammar of Atayal. Taipei: Crane.

Rau, Der-Hwa Victoria, 1995. Transitivity and discourse grounding in Atayal. Proceedings of the 7th International Conference on Austronesian Linguistics (in press).

Ross, Malcolm, 1988. Proto Oceanic and the Austronesian languages of Western Melanesia. PL C-98.

Safford, William E., 1909. The Chamorro Language of Guam. Washington: Lowdermilk.

Schachter, Paul and Fe T. Otones, 1972. Tagalog Reference Grammar. Berkeley: Univ. of California Pr.

Starosta, Stanley, 1974. Causative verbs in Formosan languages. Oceanic Linguistics 13: 279-369.

Starosta, Stanley, 1995. Rukai-Tsouic: subgroup or treetop? Proceedings of the 7th International Conference on Austronesian Linguistics (in press).

Topping, Donald M. and Bernadita C. Dungca, 1973. Chamorro Reference Grammar. Honolulu: Univ. of Hawaii Pr.

Topping, Donald M. and Pedro Ogo, 1969. Spoken Chamorro. Honolulu: Univ. of Hawaii Pr.

Topping, Donald M, Pedro M. Ogo, and Bernadita C. Dungca, 1975. Chamorro-English Dictionary. Honolulu: Univ. of Hawaii Pr.

Wolff, John U., 1991. The Proto Austronesian phoneme \*t and the grouping of the Austronesian languages. In Currents in Pacific linguistics, Robert Blust (Ed.), PL C-117, 534-549.

Zorc, R. David, 1992. Grammatical introduction, C. A. Belchez & P. J. Moguet, Bikol Newspaper Reader. Kensington, MD: Dunwoody Press.

# Dual Contradictory Agent-Marking: Solutions to Proto-Polynesian Mysteries

Joseph C. Finney

## Introduction and Overview

Abbreviations are given in a Note<sup>1</sup>.

Two recent works that describe some of the processes through which languages can move from AbErg to NomAc or in the reverse are Dixon (1994, 182-206) and Manning (1996, 20, 70, 184). The present work gives evidence for a theory of how Proto Polynesian (PPN) moved from NomAc to a mixed but predominantly AbErg system through a stage of dual contradictory marking of agents. The implications for the broader theory of ergativity are reserved for a later study.

Dual contradictory agent-marking has been found in another AN language: the Central MalayoPolynesian language Kambera (Klamer, 1994, 162-166). There it survives in fading relics in the intransitive sentences, in which the Pawley-Reid (1979) move from AbErg to NomAc syntax is not quite fully completed. An example of the transition (Klamer, op. cit. 165) is:

Hi na- laku -ya na maramba  
CNJ 3sN go 3sAc ART king

And the king went.

(The Nominative proclitic pronoun was formerly Ergative, and, before that, Genitive. The n- onset of the proclitic pronoun is the remnant of a PAN ablative and genitive preposition **ni**. The Accusative enclitic pronoun was formerly Absolutive, and before that, Topic or Nominative case. Its y- onset is the remnant of a PAN topic marker **i**. See Pawley and Reed, op. cit.)

That's the situation that makes possible the dual contradictory marking shown in Klamer (op.cit. 140) as follows:

**da-** meti **-ha**  
3pN die 3pAc [they die, they/them]

"They would have been dead." (The left pronoun is the old Ergative, new Nominative, and the right one is the old Nominative, then Absolutive, new Accusative.)

The present paper shows evidence that a similar dual agent-marking occurred in PPN. This time it occurred in the transitive sentences in the move from NomAc to AbErg, and it is attested in relics today in a few languages.

Some mysteries in PPN that have not been addressed are: How did the construction marked by -Cia suffix on verbs, ancestrally the active transitive, in a nominative-accusative system, come to be a passive construction in PN languages? What must have been the intermediate stages between the NomAc syntax of Proto Central Pacific (PCP) and the Ab-Erg syntax of the Tongic-Samoic consensus? And the intermediate stages between PCP and the NomAc syntax of East Polynesian, altogether different from the ancestral NomAc syntax of PCP?

A related mystery is the origin of the new particle e, which is the preposition or casemarkers for passive and ergative agents in Polynesian.

The present study presents a reasonable account that solves those mysteries. Differing from others, which assume that PPN's syntax was identical with that of one of the existing PN groups, this study begins with the PCP syntax that was ancestral to Fijian and Rotuman as well as to PPN. One key to the solutions lies in the Polynesian subset of unaccusative intransitive verbs that have been called neuter verbs or verbs with stative agents or involuntary causes.

A subset of unaccusative intransitive verbs in PPN had patient-like subjects, and an oblique PP (with preposition *i*) denoting inanimate or involuntary causes (Chung 1978, Hooper 1984). When human agents came to be allowed in the cause slot, the preposition *i* melded with personal article *a* (required right of prepositions, before personal pronouns and names of persons). The pair of particles became *ia* and then *e*, which then became reanalyzed as ergative or passive agent. That particular construction moved through lexical passive into an ergative.

The *e* agent phrase was borrowed into the ancestral active transitive clause with fossilized *-Cia*, and temporarily the *Cia* transitive marked its agent both in the nominative and in the oblique lexical passive "by" phrases. An example is the following sentence still accepted by some speakers of Nanumanga Tuvalu Polynesian (Finney, fieldwork 1970, 1971, 1995):

<i>e</i>	<b>kau</b>	matea	ngi(n)a	koe	<b>ne</b>	<b>aku.</b>
ASPECT	I	see	Cia	thee	by	me

"I see you."

The *e* construction became ergative in the old intransitive (no *Cia*), i.e., the former neuter verbs which had had involuntary causes as "agents". That became the common Ergative transitive construction in languages such as Tongan and Samoan. But the *e* agent construction became passive in the ancestral transitive verbs (verbs with *Cia* suffix), and that became the passive construction in the East Polynesian languages (and traces of that *Cia* passive construction remain in Tongan and Samoan, Cook, 1978, 1996). Remnants of the transitional stages can be found in PN languages today.

An essential role in creating the new East PN syntax was the merger of verb classes. East PN developed a new active voice, in which the direct objects are marked with the former locative preposition *i*. A similar process has taken place in English so that the locative-looking, intransitive-looking sentence "I looked at the thing" means much the same as the transitive "I saw the thing" or "I watched the thing." "You can count of me" means much the same as "You can trust me." Chung (1968) called that verb-class "middle verbs" in Samoan and regarded them as transitive with direct objects, though Cook disagrees. For East PN the point is that the "middle" *MiT* verb class merged with the *Cia* class, forming the huge class of transitive verbs with the former "middle" as its active voice, and the old *Cia* transitive as its passive.

The elements (slots, positions) within a clause are numbered as in Finney (1983) in which slot 2 is the tense-aspect, slot 3 the preverbal inner nominative pronoun, and slot 4 the verb. Chung (1978, 31) calls them 1 through 3 as she doesn't provide for the possibility of a sentence-initial noun. Slot 5 is *Cia*, the fossilized inner object pronoun [diachronically *-C + Ci + a*], slot 6 is for subject and/or object nouns right of the verb, slots 9 and 0 for prepositions and their adjunct objects (but numbered 7 and 8 for the ergative which diachronically begins as oblique PP and becomes NP).

For ease of comparison among PN languages, I have standardized their spelling. Long vowels are written as double vowels, glottal stops from PPN glottal are written as apostrophe, and glottal stops from PPN \**k* are written "q". The stop consonants in Nukuoro and Kapingamarangi will be written *p*, *pp*, *t*, *tt*, *k*, and *kk* (not *b*, *p*, *d*, *t*, *g*, *k*). The liquid will be written "l" (not "r") in all PN languages, and *h < f* written *f*.

The points that will be shown are as follows:

1. The involuntary cause paradigm (Inca) and the canonical ergative paradigm (ICep) must have had a common origin.



2. Of the two (Inca, Icep), the (intransitive) Inca is the older and the Icep the newer; so there is a presumption that the ergative transitive Icep developed from the intransitive Inca.
3. The development from Inca to Icep is phonologically plausible; so the presumption of point 2 is supported.
4. For the construction with Cia, which is ancestrally active transitive, there is no plausible direct course for the paradigm with e agent marker, which has since become the canonical passive in East PN. The reasonable conclusion is that after the e agent phrase developed in the (ancestrally unaccusative intransitive) canonical ergative paradigm, it was borrowed into Tra, the ancestrally transitive Cia paradigm which was to become the canonical passive. In this NomAc construction, the only possible role for the e agent was passive, and that's the role it took.
5. The immediate effect of the borrowing was to convert Tra into the Trdua paradigm, the Cia clause with dual contradictory marking of agents, with the old unmarked nominative and the borrowed e phrase coindexed. Indeed, it was a triple marking when the old nominative appeared both as the preverbal (potentially cliticized) conjunctive pronoun and as a coindexed noun (or disjunctive pronoun). For most languages this was followed by loss of the nominatively marked NP, while the nominative conjunctive pronoun was either reanalyzed or progressively lost.
6. When PPN divided into Proto Nuclear Polynesian (PNP, ancestral to Proto Samoic-Outlier and Proto East PN), PNP completed the split-off of a new active quasi-transitive construction (MiT) from the Mid (middle) "look-at" construction (whose ancestry was unergative intransitive). The process went on in East PN to become the new active transitive voice, while it never quite reached that status in Samoic-Outlier. This step was not taken by the Tongic group, which uses the (ancestrally intransitive) cononical ergative Icep as its transitive, and marginalizes the canonically passsive (Cia) Trp, using it as adjective (stative verb).
7. This paper will not discuss the class of verbs with **-ina** suffix, which also borrowed the e agent. It had been commonly believed that **-ina** was a metathesis of Cia. Mosel and Hovdhaugen (1992, 198-207, 741-748) show that **ina** differs from Cia in usage.
8. Proto Samoic-Outlier kept all the paradigms, and so kept closest to terminal PPN in its syntax. Its largest class of verbs uses the (ancestrally unaccusative intransitive) canonical ergative (without Cia) as the transitive (as Tongic does). A small class of verbs uses the (also ancestrally intransitive but unergative) MiT ("look at") in a way approaching a transitive construction with i object, but never quite gets there (at least not in Samoan, in Cook's analysis). In Samoan the (ancestrally transitive) Cia construction has tended to come to be used interchangeably with the other constructions that have e agents, to wit, the ergative and the **ina** paradigms, but Cook (1996) shows that the Cia still keeps its distinctive passive use. Even today, the remains of the various PPN syntactic structures are most visible in Samoic and Outlier languages.
9. The East Polynesian languages merge the MiT (new active transitive) verb class with the Trp (canonical passive) Cia verb class to produce a contrast between active and passive forms of the same verbs.
10. Each step of the development is simple and clear. The total set of steps accounts for the very complex description that is required for the combination of all contemporary PN languages. The actual diachronic development did not coincide exactly with either the "passive" to ergative" or the "ergative to passive" theory, each of which correctly accounts for parts of the complex development.
11. Fortunately for our understanding, each of the crucial steps proposed here has left traces in contemporary languages. The transition from involuntary cause to canonical ergative is visible in examples given by Mosel and Hovdhaugen (1992), Chung (1978), Hooper (1984) and Cook (1996). The transition of the Cia construction from the ancestral active transitive to the canonical passive (through dual contradictory agent-marking) is visible in Nukuoro, Kapingamarangi, and Nanumanga Tuvalu. The coexistence of three transitive systems (canonical ergative, canonical passive, and new MiT, not to mention **ina**) is seen in Samoan and (possibly) in Chung's (1978) Pukapuka. Note that in Finney and Alexander (1998) we give an alternative explanation that Pukapuka was a Samoic-Outlier island on the border of East Polynesia that was conquered by East Polynesians with resulting mixture of dialects.

### Proposed Prehistory

Space relations (locative, directional, LD) are the well spring of forms that (at first analogically and symbolically) come to be used for many other semantic and syntactic roles.

Closely related to Proto Fijian, the earliest Proto Oceanic active transitive construction was NomAc and (for pronouns) SOV. In its basic construction, sometimes called the inner clause, most if not all core NPs were pronouns with deictic or discourse reference; when needed to clear ambiguity, nouns were added at the right as afterthoughts, making the outer clause (and could occur on the left as extraclausal topics). The transitive verb (in the absence of an incorporated object) was followed by an (inner) suffix, slot 5, consisting of a former preposition Ci and an object pronoun (which in Eastern Oceanic was reduced to third person singular -a, making Cia). Genitives were also expressed as pronouns and doubled by nouns if needed (Milner's "their customs, the Fijians", 1956, 21, which would be better said "customs of them, the Fijians"). But nonspecific indefinite objects occupied slot 5, were not doubled by clitic pronouns, and came to be incorporated into the verb unit. No theory has tried to reconcile this "pronouns primary" diachronic paradigm with that of synchronic generative grammar in which nouns are primary to their anaphors.

Early in the PPN period, two intransitive constructions, both spatial in origin, each with its class of verbs, began to rival the old transitive paradigm and usurp its functions. (This is consistent with Chung's view, *op. cit.* 254, "The evidence indicates that all the markers were originally oblique.")

One class consisted of unergative verbs with ancestral locative phrases marked by preposition *i*, or ancestral allative phrases marked by preposition *ki*. Their subjects were agent-like. In a subclass, their oblique phrases eventually became (at least in East PN) the accusative objects in a new NomAc paradigm. This seems to be a common pattern in languages of the world. A subclass of space PPs become Middle object (MiD), of which a new subclass becomes fully accusative object (MiT). Thus came the new NomAc and active construction that dominates East PN languages.

### Involuntary Causes and Stative Agents

The other class of intransitives that threatened the old transitive paradigm consisted of unaccusative verbs with ancestral locative phrases marked by preposition *i*. Their subjects were patient-like. This is the class that has been called verbs with stative agents, or neuter verbs. In a subclass, their LD phrases, in steps, came to be used for inanimate causes and to have passive-like semantics; then for involuntary human doers/actors; and at last for human voluntary agents in a construction that became ergative in some Polynesian tongues (and passive in others, when borrowed into the transitive paradigm). This path is the key that unlocks the mysteries of PPN syntax.

For Samoan, Mosel and Hovdhaugen (1992, 428) list six types of clauses that are VP-ABS-LD (locative, directional). One of the six is "the local source of a movement or the cause of a state of affairs". Their two examples are:

na paqu le tama i le niu  
TAM fall ART child LD ART coconut tree  
"The boy fell from the coconut tree."

qaa muu le aliqi i le laa  
TAM burn ART chief LD ART sun  
"The high chief is about to be burnt by the sun."

Both sentences are unaccusative intransitives, with absolutive (ancestral nominative) subjects that are patient-like. Whether intentionally or not, Mosel and Hovdhaugen gave the sentences in diachronic order: first with "from" in a space sense, glossed intransitively, and then "from" in an abstract causative sense, an involuntary cause, which in this sentence they glossed

“by”, a passive gloss that makes the sun sound agent-like. Perhaps “about to burn in the sun” would reflect the ancient syntax, but the passive translation shows the direction in which the paradigm must have slowly moved over the centuries.

Following Hohepa (1967, 1969) and Hale (1968), Chung published her revised dissertation as a major work (1978) which applied generative grammar to Polynesian, and used generative terms (including Raising and Equi Deletion) as criteria to determine which component is the subject. Although “passive” and “ergative” agents look alike in the Polynesian languages, in being marked with *e*, she designates as ergative those in which the *e* agent tested out as subject, and as passive those in which it did not. Her conclusions supported those of Hohepa and Hale that PPN had passives, like the East PN languages today, and that the Tongic and Samoic-Outlier languages showed passive-to-ergative drift.

That contradicted the theory championed most vigorously by Clark (1973, 1976) that PPN had ergatives, like Tongan and Samoan, and that ergative-to-passive drift accounts for the East PN syntax. Proponents of each theory have produced evidence to support it (or at least invalidate the competitor). Both theories suffer from the assumption that PPN was syntactically identical with one of the contemporary attested languages. In contrast, the present theory begins earlier with state 6, the earliest proto stage in which the ancestor of PN languages began to differ from the ancestors of Rotuman and Fijian languages.

Chung’s PPN reconstruction has the weakness that its proposed PPN NomAc differs totally from the NomAc system found in Fijian and universally reconstructed for POC. While admitting that, she (op. cit. 254) attacked Clark’s theory for an implausibility that is avoided in this present paper.

Chung used the older term “Stative Agent” construction for the derivative of inanimate cause, but glossed it Caus. She restricted the term Stative Agent to the offshoot in which the old inanimate or involuntary cause slot, with preposition *i*, had come to be occupied by human or human-like agents. She explained, (1987, 30):

“What is important about stative agents is that they are semantically agentive but syntactically oblique; as such, they can be used to determine whether a rule refers to semantic agency or syntactic subjecthood. They will be appealed to a number of times in a later chapter.”

An example from Samoan which Chung glosses passive but shows as Caus, excluded from the “stative agent” group because it is inanimate, is the following (Chung 1978, 29):

na lavea le tamaaloa i le masini.  
TAM hurt ART man Caus ART machine  
“The man was hurt by the machine.”

An example of Chung’s stative agent from Maori (op.cit., 29) is:

ka mate te hoariri i te toa.  
TAM die ART enemy Caus ART warrior  
“The enemy died because of the warrior.”

Another of her examples (1978, 29), this one from Tongan, is interesting in that she adds a second gloss that makes the agent an active transitive subject (despite her position that it is syntactically oblique):

na’e ngalo ‘i-ate au e peepee.  
TAM forgotten Caus-ART me ART baby.  
“The baby was forgotten because of me.”  
“I forgot the baby.”

Incidentally, Hooper regarded all the involuntary causes as oblique, human or not, as she

regarded all those sentences as both semantically and syntactically passive. If so, they come under Keenan's concept of "lexical passives" in which the simple form of the verb is passive, and the active form either is non-existent or has a form more complex than that of the passive.

### Origin of the Agent Preposition (or Casemarker) "e"

While, in Fijian, particle *e* is a phonological variant of *i*, Polynesian languages have two distinguishable prepositions, *i* and *e*. The *i* comes straight down from the PAN preposition *i*, locative "in, on, at". Its range of function expands in Oceanic and some other branches of AN to invade the "from" and "of" territory, but its central meaning is still the same as in PAN. But whence the *e* preposition of agency?

Compare the following sentences in Tuvalu (or in generalized Samoic-Outlier). The particle *a* is a personal article required after prepositions before pronouns and proper names of persons. (The verb "puni", though transitive in Samoan, is in other PN languages one of the unaccusative intransitive verbs which take stative agents or involuntary causes.)

e	puni	te	nuku	i	te	moana
TAM	surrounded	ART	island	from	ART	ocean
e	puni	te	nuku	i	a	te moana
TAM	surrounded	ART	island	from	ART ART	ocean

"The island is surrounded by the (ocean, man [conqueror])."

The solution suggested here is based on sentences like these. When it became permissible to have a person in the old stative agent slot, preposition *i* and personal article *a* melded into a new preposition *ia*, which lenited to *e* and marked the (at first non-subject) human agent. Then it became reanalyzed as the ergative agent (and hence, by some criteria, the syntactic subject). The old nominative subject ("island" in this example) was reanalyzed as patient in the absolutive case. Simultaneously, one other change had to take place. The *a* article occurs before personal pronouns and proper names of persons. In the syntactic change, the *ia* or *e* ergative marker had to come to be used with ergative NPs even when they are common nouns of persons.

Note that the NP in the ergative case must be a human agent. The ergative NP must have begun, step by step, to acquire the characteristics of a subject, in terms of the tests that Chung used as criteria of subjecthood, such as its behavior in Equi deletion and in Raising. This created a new paradigm, which we'll call ICEP [Involuntary Cause] stative-sourced ergative/passive. It is the ergative construction that occurs in former intransitive verbs, and so it has no Cia element.

The step-by-step acquiring of subjecthood is still going on in Samoan. In the (very few) verbs in Samoan that have different forms for singular and non-singular, the older generation still uses the traditional practice of having the verb agree in number with the absolutive patient, but the younger generation makes the verb agree in number with the ergative agent.

This is a simple and elegant solution to a problem to which no other solution has been offered. It is consistent with Chung's (op.cit. 259) view: "Both \**i* and \**e*-- the prime candidates for non-oblique case markers in Proto-Polynesian-- can ultimately be traced to the same oblique preposition in Proto Eastern Oceanic."

The creation of this new construction is a branching, not merely a development. Another subset of involuntary Cause or Stative Agent verbs lived on, notably in East Polynesian, which lost the *e* ergative construction but kept the Cia *e* passive. A small set of verbs with *i* Causes, not restricted to human Causes, survived. In Hawaiian, Elbert calls them "loa'a verbs" after a prominent member of the set; ("loa'a" means "come into the possession of, be gotten by"; that verb occurs only in Hawaiian and Tuamotu, and may be a contraction of "loa aka"). That set of verbs in Hawaiian includes "mau" and "puni". It may be that this small set of verbs has recently become ergative in Hawaiian and perhaps some other East PN languages.

### Dual representation of agent with conflicting marking

Here is the most startling and dramatic part of the story.

The ancestral NomAc active transitive construction Tra (with Cia) borrowed the e agent phrase, and temporarily had contradictory dual marking (Trdua) of coindexed agents, both (unmarker) nominative and e oblique passive agents. In this NomAc traditional Cia construction, the oblique e agent had to be construed as a passive agent. So in most languages (East Polynesian and some relic expressions in Samoan and Tongan) the nominative expression dropped out and the passive e agent survived. Thus the borrowing made a fundamental change in the Cia construction, which was the construction that had included perhaps most of the verbs in PPN. The deletion of the nominative subject agent from the Trdua construction produced a construction we'll call Trp, passive from ancestral NomAc active transitive. It had Cia, unmarked patient surface subject, and oblique agent marked with e and it served as a true Passive voice. It barely persists in a few fossilized verbs in Tuvalu, and it has disappeared in most of the Samoic-Outlier languages.

Remnants of the dual contradictory marking are found in Kapingamarangi, Nukuoro, and Nanumanga Tuvalu.

Lieber and Dikepa (1974, xliii) show the dual representation in Kapingamarangi:

au e moinaa e au koe  
I TAM cherish by me thee [my gloss]  
"I cherish you." [their translation as well as mine]

Chung (1978, 58) gave a similar finding by quoting a sentence in Kapingamarangi from Elbert (1948,33):

au ku kitee e au a mee.  
I TAM see by me ART him  
"I saw him."

(A peculiarity of Kapingamarangi is that the PPN word **mea** "thing, person" has come to be used as a personal pronoun and even come to take the personal article **a**.)

Following Elbert, Chung glossed the e not as "by" but as "Nom", a matter that suggests that the ergative marker has been reanalyzed.

Lieber and Dikepa's dictionary lists both **moinaa** and **kitee** as intransitive verbs. Both, however, are old transitive verbs, as shown by the Cia suffix **-a** [kitee < kite-a]. The accusative objects without PCM also show the sentences as transitive.

Another example in Lieber and Dipeka, same page, is an accusative sentence in which the accusative object is deleted.

au te iloo e au  
I not know by me  
"I don't know."

As Nukuoro and Kapingamarangi are geographically nigh and obvious kin (though said not to be mutually comprehensible) it's not surprising to find the same dual contradictory marking, as in this Nukuoro passage in Carroll (1965:459):

akai Leipua ku lango -na foki e ia takoto  
and NAME TAM hear Cia also by her the matter  
"And Leibua also heard the matter."

Chung (op.cit., 314), did not propose the dual marking for PPN, but she gave a plausible explanation for its persistence in Kapingamarangi as follows: “Finally, in Kapingamarangi, the accusative and ergative case patterns were reanalyzed as derived by a single rule of the form ‘mark the subject with *e* and the direct object with *i*’, followed by another rule optionally deleting any case marker. These rules left Kapingamarangi with an accusative case system.”

As we have seen, the *e* agent marker, which developed as the ergative construction and was limited to human agents, was borrowed into the transitive *Cia* construction, where it wreaked havoc on the syntax and produced electrifying changes. The Chung passage that we have just cited shows that the *e*-marked agent was also borrowed into the “look-at” construction which had a quasi-patient marked with the former locative marker *i*. Thus arose the construction with an agent marked by *e* and a patient or quasi-patient marked by *i*. Mosel and Hovdhaugen (op.cit., 437, 438) seem to say that that construction is viewed by Samoans as of questionable grammaticality; but the missionaries (perhaps by mistake) used it as their normal and usual way of representing English transitive sentences; and perhaps that usage contributed toward increasing the degree of acceptability of the construction to Samoans. In Rapanui (Easter Island), the construction with *e* agent and *i* patient is fully acceptable (Finney, in press, 1998).

For Nanumanga Tuvalu, I had shown (1983) that the old Proto Central Pacific NomAc construction (as in Fijian) persisted in use all the way through PPN, Proto Nuclear PN, and Proto Samoic-Outlier into (optional) current use in sentences such as:

1	2	3	4	5	6
aku	e	kau	matea	ngina	koe
I	TAM	I	see	Cia	thee

“I see you.”

When I discussed that construction with Niko Besnier, he responded by offering an expansion that he had heard, that contains the dual contradictory marking, saying (oral communication), “They can also add *e aku* at the end”:

1	2	3	4	5	6	7	8
aku	e	kau	matea	ngina	koe	e	aku.”
I	TAM	I	see	Cia	thee	by	me

“I see you.” or “You are seen by me.”

The Nanumanga *ngina* is a combination of the *ina* suffix with the true *Cia ngia* (which is used on the neighboring island of Niutao).

That confirmed my view that the sentence with eight slots, as shown, is the underlying form in Nanumanga, from which both the ergative pattern (used in most verbs, created in PPN as we have seen) and the older accusative pattern (still used in a few transitive verbs and persisting from Proto Central Pacific and Proto Oceanic) can each be derived by deleting some of the elements. Besnier does not necessarily agree with that view. His synchronic view (influenced by his main work in South Tuvalu dialects) is that the ergative pattern is basic and the NomAc and the 8-slot sentence can be derived from it by fronting and other transformations; but I view that as a reanalysis that says nothing about the diachronic ancestry. It is also interesting that North Tuvalu and South Tuvalu, though easily mutually comprehensible, may differ in their basic syntactic structures.

I have reanalyzed my report (1983) which tested in 1982 five Nanumanga informants on 56 variations of the sentence. Eka and Pulei were adult males. The other three were 22-year-olds who had been pupils of mine on Nanumanga when they were 11-year-olds: girls Lita and Sisifo and boy Lapana.

Results showed that forms with dual contradictory marking were accepted by some,

though pure NomAc and pure AbErg expressions were accepted better. All five informants accepted the twelve sentences that were pure AbErg, i.e., had the unmarked patient *koe* only once, had the agent marked with *e* or *nee*, and did not have an agent marked Nominatively (as *aku* in the outer clause nor as *kau* in the inner clause). Those were sentences 33–40 and 53–56 in the 1983 paper. Likewise, all five informants accepted the four sentences (1,2,21,22) that were pure NomAc, i.e., had the Nominative agent *kau* in the inner clause, and had no *e* or *nee* agent phrase. All sentences with triple representation of the agent (*aku + kau + e/nee aku*) were rejected by 3 or 4 informants. Of the other eight sentences with dual representation (half of which had *Cia*), only one was accepted by two informants, and the other seven by only one.

Of the eight sentences that had contradictory dual representation, and in which the Nominative was a full NP in the outer clause (not a clitic), sentences 13 through 20, two were accepted by three informants, and the other six were accepted by two.

The move into the final stage of PPN consists in the branching off from *Trdua* (dual contradictory representation of agent) into a new paradigm we shall call *Trp*, the transitive construction that is passive. This is the *Cia* form that persists as the mildly anomalous *Cia* in Samoan and some other languages, and as the passive voice in East Polynesian's new NomAc. The agent noun is marked by the PCM *e*, and the patient has no overt marking, except where an article *a* or a presentive noun marker *ko* comes to be regarded as a Nominative casemarker. In PPN and PMP three constructions of differing ancestry (four if *ina*'s ancestry differs from *Cia*'s) had the new *e* agent, though they differed in usage and in verbal morphology. The verb had no suffix in the ergative construction; it had *-Cia* in the passive, and had *-ina* in the "perfective", the one of greatest transitivity according to Milner (1956, 1962).

For Samoan, Mosel and Hovdhaugen (1992) and Cook (1996) have shown that the *-ina* and the *-Cia* suffixes differ in function, a matter I'll address elsewhere. Milner (op.cit.) showed evidence that the Samoan transitive suffix is perfective, not passive. His examples had suffix *-ina*, not *-Cia*, and so they don't contradict Cook's (1978, 1996) evidence that *-Cia* has a passive function in Samoan. Chung (op.cit, 268-284) discussed Tongan uses of *Cia*, such as attributive, passive, and resultative, all of which are marginal. The ergative *e* construction with no suffix is the normal transitive in that language.

It is because the ancestral construction had both the subject noun and the accusative object noun without overt marking, that the deletion of the nominative subject could let the old accusative object be reanalyzed as the absolutive case in a new ergative construction, and as the surface subject (nominative patient) in a passive.

For Kapingamarangi and Nukuoro the key may be in Chung's gloss of the *e* as a Nominative marker, not as Ergative nor as passive agent. If, indeed, it came to be reanalyzed that way, the contradiction disappears. The expression is still anomalous in expressing the Nominative twice in the clause. Even that anomaly may be lightened by reanalyzing the initial Nominative in position 1 as a fronted topic. In Nanumanga the dual expression is useful as an underlying paradigm from which the speakers can delete some parts to make a NomAc transitive sentence, or delete other parts to make an AbErg sentence.

### Restriction of "e" to Human and Human-like Agents

According to the theory for which we have given evidence, the agent-marker *e* arose in circumstances true only for human subjects. Some recent publications, mostly dealing with Samoan, have shown that even today, two thousand years later, in at least some PN languages, the ergative *e* marker is used only for human and human-like agents.

For the language of Rapanui (Easter Island) Du Feu has shown (1996) that the *e* agent marker is used mostly under two circumstances which on the surface seem opposed to each other. One is sentences in which the agent acts with a high degree of intentional volition. The

other is with verbs of sensation and perception, in which the degree of volition is very low. I have pointed out (1998, in press) that what the two circumstances have in common is that both normally require a human subject. Du Feu (pers. comm.) has accepted that as the explanation.

Duranti (1994) discussed the politics of ergativity in the fono (village councils) of Western Samoa. In the words of a reviewer (Povinelli, 1996), "Samoan grammar provides a distinction between subjects that are agents of transitive action (ergativity) and other kinds of subjects (actor, instrument, etc.). But much to Duranti's initial surprise, the ergative form is rarely used in everyday speech; instead its use is concentrated in speech heard in the fono. Even within the fono, ergative constructions are, in the main, restricted either to praise of the Christian god as the cause of all social harmony and order, or to the blaming of a particular person for causing social conflict. This is in spite of the fact that nothing about the ergative construction in itself would restrict it to either of these two domains."

Cook (1996), in the context of the *Cia* passive in Samoan, showed contrasting sentences with *i* and *e* for what is translated as a "by" agent in English. Two sentences that he gave with each agent-marker are glossed in English:

"Rosy was hit by a car."

"The car was hit by a landslide."

When the "by" marker was *i*, the sentences were taken much like the involuntary cause sentences that we have examined. But the same sentences, with the same translations, can occur with *e* as the agent marker. Cook explains, "The inanimate initiator ... can be marked as in the ergative case if the initiator is agent-like, or if it is blamed for the event." In other words, if it is regarded as human-like.

### Recapitulation

The transitive construction with which PPN began in its earliest stage was the active transitive NomAc construction with *Cia*, that is found in Fijian. A question is: how did it change into the passive or ergative *Cia* constructions that are found in most Polynesian languages? Another question is: how did the ergative construction arise: i.e., the construction that has no *Cia* suffix? A third question is how the *e* agent (for passives and ergatives) came into being, and why. Also, hardly anybody has noticed till recently that at least in some PN languages the use of the *e* agent, in both passive and ergative constructions, is restricted to agents that are human or human-like and can be blamed for what happened. Those are the problems we solve in this paper.

Some, perceiving (not altogether falsely) some similarity between the ergative and the passive constructions in PN languages, have thought that the passive might have arisen from the ergative by insertion of the *Cia*. That suggestion is unacceptable, of course, because the *Cia* construction existed in the ancestral language, though it must have been active transitive as in Fijian. Others have thought that the ergative construction has arisen from the passive by deletion of the *Cia*. That's less implausible, but it's not true, either. Neither Clark's theory of ergative-to-passive drift, nor Chung's theory of passive-to-ergative drift, is fully true, though both have elements of truth. In fact, neither the ergative construction nor the passive construction is descended from the other. The passive construction, with its *Cia* suffix, descends from the ancestral Fiji-like active transitive *Cia* construction though with a dramatic shift in syntax. The ergative construction descends from the anomalous unaccusative intransitive construction that has been called involuntary cause, inanimate cause, and stative agent construction, in which the patient-like nominative subject is unmarked, and the involuntary cause is marked with the preposition *i*.

The scenario that solves all those problems, and successfully explains the syntactic development of all the branches of Polynesian, is as follows.

Personal pronouns and proper names of persons, occurring after prepositions such as *i*, must take the personal article *a*. At some point it became permissible to put a human agent into



the involuntary cause slot. The preposition *i* and the personal article *a* melded into the form *e*, which came to be reanalyzed as an ergative case–marker. Thus a subset of what had been the involuntary cause construction became the ergative construction, which could only be used for voluntary human agents.

Then the *e* agent was borrowed into the *Cia* construction, and there it had dramatic and explosive effects. At first, after the borrowing, the construction had dual contradictory marking of the agent, both nominative and oblique. Some languages today keep relics of that dual marking. In the mainstream, the nominative marking disappeared in the *Cia* class of verbs, and the oblique *e*-marked agent made the construction passive. The ancestral active transitive construction was no more.

In the Tongic languages, most verbs moved into the ergative class. The passive construction with *Cia* survived in marginal uses, sometimes as an adjective.

In Nuclear PN, at least, the *e*-marked agent was also borrowed into the “look-at” construction, forming clauses with *e*-marked agents and *i*-marked patients. This odd construction was lost in East PN, but came to be fully accepted in Rapanui and marginally accepted in Samoan.

The *e*-agent was also borrowed into the *ina* construction, which we don’t discuss here, but which may have an origin different from that of the *Cia*. Thus there came to be *e*-agents in four different Polyesian constructions: the one in which it had its origin (the ergative), and three more into which it was borrowed.

In East PN, the “look at” verb class became the active transitive, in which the *i*-marked phrase became an accusative object, though, of course, a homonymous *i*-marked phrase continued with its ancestral oblique use (such as locative). That verb-class merged with the *Cia* verb class, and the *Cia* construction became the passive voice of verbs of the combined class.

Languages in the Samoic-Outlier group kept available more of the options from PAN. In most languages of this group, the *AbErg* construction has become the main transitive construction, and most of the transitive verbs have moved into that class. Some languages in the Samoic-Outlier group retain relics of the dual contradictory marking, and Nanumanga Tuvalu retains (for a very small class of verbs) the option of using the Fiji-like *Nom-Ac* construction coming down unchanged from Proto Central Pacific

### SHORT SUMMARY

**Neither ergative-to-accusative drift nor accusative-to-ergative drift occurred. What happened is: 1. An ergative construction, limited to human agents, with *e*-marked agent, split off from the intransitive involuntary cause construction. 2. Borrowing of the *e* agent into two or three other constructions produced new diathesis constructions (voices). 3. Massive movement of verbs from one voice-class to another occurred, and East PN merged two major voice classes to create an active-passive contrast. 4. As the final stage of PPN had more voices than any language needs, various branches of PN let different voice constructions fall into disuse.**

### NOTE: ABBREVIATIONS

1. The following abbreviations are used for language groups: AN Austronesian, MP Malayo-Polynesian, OC Oceanic, CP Central Pacific, PN Polynesian, SO Samoic-Outlier (in order from most to least inclusive). All these may have P added initially for Proto.  
Abbreviations for sentence-elements: ART article, PCM preposition or case-marker, TAM tense-aspect marker, NAME for proper name  
Abbreviations for paradigms or constructions: *NomAc* Nominative-Accusative, *AbErg* Absolutive-Ergative, *Tr*: ancestral active transitive *NomAc*; *Tra*: ancestral active transitive *NomAc* that had been *Tr* before it required a *Cia* element; *Int*: simple intransitive with only one core NP; *InObl*: intransitive with a core oblique complement; *Mid*: middle verbs (“look-at”

construction); MiT.: as in Samoan, development from Mid toward an accusative; Inco: an incorporated object construction; Inca: inanimate cause construction descended from an InObl; ICep: an ergative construction that was ancestrally an Inca and before that an InObl; Trp: a passive construction with Cia that had ancestrally been a Tra. Trdua: transitive with dual marking (resulting from Tra borrowing the e agent from ICep), soon becoming Trp.

## REFERENCES

- Carroll, Vern, 1965. An outline of the structure of the language of Nukuoro. *J. Polynesian Soc.* 74:1965 (Polynesian Society Reprints, Series 10).
- Chung, Sandra, 1978. *Case marking and grammatical relations in Polynesian*. Austin: U. Texas Pr.
- Churchward, C. Maxwell, 1953. *Tongan dictionary*. London: Oxford U. Pr.
- Clark, D. Ross, 1973. Transitivity and case in Eastern Oceanic. *Oceanic Linguistics* 12: 559-605.
- Clark, D. Ross, 1976. *Aspects of Proto-Polynesian syntax*. Linguistic Society of New Zealand.
- Cook, Kenneth William, 1978. The mysterious Samoan transitive suffix. *Proceedings of the fourth annual meeting of the Berkeley linguistics society*. (not seen).
- Cook, Kenneth William, 1996. The Cia suffix as a passive marker in Samoan. *Oceanic Linguistics* 35: 57-76.
- Dixon, Robert M. W., 1994. *Ergativity*. London: Cambridge U. Pr.
- DuFeu, Veronica, 1996. *Rapanui*. London: Routledge.
- Duranti, Alessandro, 1994. *From grammar to politics: linguistic anthropology in a Western Samoan village*. Berkeley and Los Angeles, U. California Pr. (not seen)
- Elbert, Samuel H. and Torben Monberg, 1965. *From the two canoes*. Honolulu: U. Hawaii Pr. (not seen)
- Finney, Joseph C., 1983. Nanumanga verbal syntax: a study in variation, with implications for Proto-Polynesian. *J. Polynesian Soc.* 92: 107-115.
- Finney, Joseph C. and James D. Alexander, 1998 (in press). The position of Rapanui in the Polynesian family. *Rapa Nui Journal*.
- Hale, Kenneth, 1968. Review of Hoheps (1967). *J. Polynesian Soc.* 77: 83-89.
- Hohepa, Patrick, 1967. *A profile generative grammar of Maori*. Indiana U. pub in anthropology and linguistics, Memoir 20.
- Hohepa, Patrick, 1969. The accusative-to-ergative drift in Polynesian languages. *J. Polynesian Soc.* 78: 295-329.
- Hooper, Robin, 1984. Neuter verbs, stative aspect, and the expression of agency in Polynesian. *J. Polynesian Soc.* 93: 39-70.
- Klamer, Marian, 1994. *Kamera: a language of Eastern Indonesia*. Haag: Holland Institute of Generative Linguistics.
- Lieber, Michael D. and Kalio H. Dikepa, 1974. *Kapingamarangi lexicon*. Honolulu: U. Hawaii Pr.
- Manning, Christopher D., 1996. *Ergativity: argument structure and grammatical relations*. Stanford: CSLI.
- Milner, George B., 1962. Active, passive or perfective in Samoan: a fresh appraisal of the problem. *J. Polynesian Soc.* 71: 151-161.
- Milner, George B., 1956 (reprinted 1967). *Fijian grammar*. Suva: Government Printer.
- Mosel, Ulrike and Even Hovdhaugen, 1992. *Samoan reference grammar*. Oslo: Scandinavian U. Pr.
- Pawley, Andrew and Lawrence A. Reid, 1979. The evolution of transitive constructions in Austronesian. *Papers from the second eastern conference on Austronesian languages* (Ed: Paz B. Naylor. Michigan Papers on South and Southeast Asia 15: 103-130. Ann Arbor: U. Michigan Pr.
- Povinelli, Elizabeth A., 1996. Review of Duranti (1994). *Anthropological linguistics*. 38(1):160-163.

# The Phonology-Syntax Interface in Rotuman

Mark Hale & Madelyn Kissock  
Concordia University & Harvard University

This paper presents a reevaluation of the phenomenon of ‘metathesis’<sup>1</sup> in Rotuman and discusses the implications of the resulting analysis for the Optimality Theoretic account proposed in McCarthy (1996a). On the basis of a detailed examination of the evidence provided by the folktales in Churchward (1938), we will argue that the domain for metathesis is not syntactico-semantic, as Churchward claims and McCarthy (1996a) assumes for his analysis, but rather prosodic (thus in the post-lexical phonology). Further, because this has direct (negative) consequences for the OT analysis based on output-output correspondence proposed by McCarthy, we will discuss an alternative OT account which refers instead to the possibility of a post-lexical OT phonology.

## 1. Churchward (1940) and the Romanticism of Terminological Aesthetics

Rotuman displays interesting surface allomorphy for individual morphemes. The relevant phenomenon is well-known in the phonological literature particularly for its cases of metathesis, but it also includes cases of deletion, umlaut, diphthongization, and ‘no distinction.’ Churchward (1940) notes these contrasts in his original (and very thorough) treatment of Rotuman grammar and innovates the term ‘phase’ to describe it – characterizing the unchanged morphemes as being in the ‘complete phase’ and the altered morphemes as being in the ‘incomplete phase.’ Each morpheme consistently undergoes a particular change in the ‘incomplete phase.’<sup>2</sup> Appendix A provides a brief sketch of how the phases are formed for morphemes of various (phonologically-defined) structures.<sup>3</sup> The examples in (1-3) illustrate the basic contrast between ‘complete phase’ (in this case, unmetathesized) and ‘incomplete phase’ (metathesized) forms.

- |   |  |
|---|--|
| (1a) <i>famori</i> ‘ea<br>people <sub>comp</sub> / say<br>‘(the) people say’                | (1b) <i>famör</i> ‘ea<br>people <sub>inc</sub> / say<br>‘(some) people say’                |
| (2a) <i>vak titi</i> ‘u<br>canoe <sub>inc</sub> / big <sub>comp</sub><br>‘(the) big canoes’ | (2b) <i>vak titi</i> ‘<br>canoe <sub>inc</sub> / big <sub>inc</sub><br>‘(some) big canoes’ |
| (3a) <i>le</i> ‘ta<br>child <sub>inc</sub> / DEF.SG<br>‘the child’                          | (3b) <i>le</i> ‘e=t<br>child <sub>comp</sub> =INDEF.SG<br>‘a child’                        |

<sup>1</sup> We use the term ‘metathesis’ initially as a cover term for the several different outcomes (only one of which is metathesis) subsumed under Churchward’s ‘Incomplete Phase,’ to be described in detail in the following section.

<sup>2</sup> For the moment, we will simply follow Churchward in stating that the particular form of certain morphemes in the incomplete phase is identical to their complete phase form, as noted above.

<sup>3</sup> In transliterating Churchward’s orthography we will use *â* for his underdotted *a*, *ā* for his *ā*, and *æ* for his over-dotted *a*. The distinction between tense and lax mid vowels will be ignored (it is irrelevant for our purposes), as it is by Churchward’s orthography.

The only explicit statement of rules governing the use of the two phases is that provided by Churchward (1940). Churchward's rules are quoted in Appendix B and their effects are summarized below.

- Rule 1. Only the final morpheme of a given morphologically-complex lexeme shows phase distinctions. All non-lexeme-final morphemes are in the *incomplete* phase.
- Rule 2. A lexeme which modifies or is modified by the following word is in the *incomplete* phase.
- Rule 3. A noun or verb normally shows up in the *complete* phase when "definite" and in the *incomplete* phase when "indefinite."<sup>4</sup>
- Rule 4. In some cases the use of the *complete* phase indicates "positiveness, finality, or emphasis."
- Rule 5. In verbs with pronominal (object) suffixes the *complete* phase indicates the completive (perfective) tense.

The first two rules can be collapsed, given a sufficiently enriched theory of prosodic domains, into a constraint which states that it is only at the right edge of a phrasal domain where we find a phase distinction. Thus, all non-phrase-final elements are in the *incomplete* phase. The phase distinctions of phrase-final elements are regulated by Rules 3, 4, and 5. Given a sufficiently loose semantics, as Churchward notes, they may also be collapsable.

Churchward provides a discussion of the origin of the phases (1940:164ff.). This discussion is methodologically rather dated, but it is of some interest to examine it for our present purposes. Churchward reveals a strong tendency towards 'romanticism' in his notions of the nature of language and language change. For example, he notes that:

"... the fact that the distinction between the phases is fundamental and seemingly indispensable to the structure and laws of the language, and so ready to seize upon every new word, would seem to indicate that it belongs historically (as well as grammatically) to the foundations [of the Rotuman language]."

It is now widely recognized that *productivity* ("ready to seize upon every new word") and archaism ("belongs historically... to the foundations") by and large show an *inverse* relationship rather than the direct one implied by Churchward's assertion.

The precise mechanism posited by Churchward to account for the origin of the *incomplete* phase, which is clearly secondary in diachronic terms, does not fare much better:

"The *incomplete* phase arose, no doubt, through the same tendency as has given rise — and still gives rise — to abbreviations of various kinds in other languages, the tendency, namely, to economy of effort."

---

<sup>4</sup> Both of these terms have a non-standard interpretation in Churchward's grammar, as we will see in detail below.

There is little merit to this proposal. The rising diphthongs created by the metathesis are quite marked (when compared to the simple vowels in the complete phase), and the resultant syllable structure, with a complex nucleus and coda consonant in place of the complete phase's simple -CVCV pattern, can hardly be called more "economical." Note that the evolution of this type of metathesis did not occur in the closely related Polynesian languages, nor in Rotuman's immediate sisters, the Fijian languages.

Churchward's further speculation on the matter is unhindered by any particularly clear understanding of phonetics or semantics, as the following quotes respectively show.

"This propensity towards economy of effort may even help to explain why it is that the inversion of the final syllable takes place when the final vowel is wider than the penultimate, but never when it is narrower or the same width. For the widest vowel (*a*) is the one that requires the least effort to pronounce..."

"First, while the inc. phase, as requiring less effort, would be adopted in general usage, the com. phase—the full form of the word—would naturally be resorted to, on occasion, for the sake of expressing positiveness or emphasis, a use now enshrined in the "fourth rule"... The transition from this to the use embodied in the "fifth rule" is not difficult to conceive; while the transition from positiveness to precision or definiteness resulting in the "third rule" is surely quite natural. Thus, the com. ph. would come to convey, in many cases, the idea of definiteness, the inc. ph. comparative indefiniteness. And the indefiniteness expressed by the inc. ph. is what underlies the second and first rules..."

"...we may say merely that completeness of form came, by degrees, to express completeness of sense, while incompleteness of form came to express incompleteness of sense."

The mechanisms, either diachronic or synchronic, which give rise to the Rotuman phase distinctions are unlikely to be sought in explanations of the type offered by Churchward. To the extent that diachronic developments may elucidate the synchronic analysis, they will be discussed below. A detailed explanation of the origin of metatheses of this type can be found in Blevins & Garrett (1995, this volume).

Churchward notes an additional oddity with respect to a relationship between phase distribution and the concatenation of particular suffixes. He offers no clear explanation of this phenomenon:

"As to the reason why certain suffixes should be attached to the inc. ph. and others to the com. ph., this is perhaps a matter either of euphony or of ease of pronunciation — 'economy of effort' again..." (Churchward 1940:166)

We will see, however, that understanding this relationship is critical to the correct analysis of phase in Rotuman.

In the following section, we examine data of the type on which Churchward based his analysis in order to determine the accuracy of his assumptions (already somewhat debatable) and consequently, the validity of basing any theoretical analyses upon them.

## 2. Trouble in Paradise

Churchward's rules governing the use of complete vs. incomplete phase are equivalent, in modern theoretical terms, to describing the function of a specific morpheme. The complete phase morpheme has, minimally, the following three functions – definiteness (specificity of reference), perfective/completive aspect, and 'positiveness of assertion'. We should note first of all that, to our knowledge, there is no human language which encodes all of these (or even any pair of them) in a single morpheme. This casts a certain doubt from the start on the correctness of Churchward's analysis. However, independent of this peculiarity, we soon find that Churchward's rules run into serious empirical problems.

Consider the function of 'definiteness' ascribed to the complete phase and the corresponding 'indefiniteness' of the incomplete phase. The examples in (4a-b) were cited together in Churchward (1940:116) as illustrating different subcategorization frames for the verb 'stop'.<sup>5</sup>

(4a) *gou la tük iris*  
 I<sub>inc</sub> / FUT / stop<sub>inc</sub> / them<sub>inc</sub>  
 'I will stop them'

(4b) *gou la tük se irisa*  
 I<sub>inc</sub> / FUT / stop<sub>inc</sub> / at, to / them<sub>cmp</sub>  
 'I will stop them'

While these sentences do, indeed, seem to indicate that the verb 'stop' may take a NP or a PP complement, they also show that none of Churchward's rules can be the governing factors in determining complete vs. incomplete phase. The only two of his rules which could apply to these sentences (see our synopsis above and Appendix B) are those of 'definiteness/indefiniteness' or 'positiveness of assertion.' The definiteness/indefiniteness rule appears to break down immediately on the first word *gou*, the personal pronoun 'I', which is in the incomplete phase in both sentences. It is difficult to imagine that 'I' could ever have an indefinite reading, as the incomplete phase would indicate. This leaves only 'positiveness of assertion', but Churchward has indicated no such distinction in his translations, which we might expect him to do if, according to his own analysis, such a contrast exists. Instead, he specifically points out that these two sentences mean the same thing.

The distinction between perfective/imperfective aspects of verbs and complete/incomplete phases seems to be a false one as well. In both (5a) and (5b), the verb *noho* 'live' is non-completive (i.e. imperfective). Note, however, that in (5a) *noho* occurs in the incomplete phase whereas in (5b) it occurs in the complete phase – contrary to Churchward's theory that complete phase indicates perfective aspect. In addition, it turns out that all verbs before the locative anaphor clitic *e* are in the complete phase, thus a rigid connection between perfective and complete phase seems highly unlikely.

(5a) *ma Titof noh ma tupue 'te 'is 'e Faufano* (II.9)  
 and / Titof<sub>inc</sub> / lived<sub>inc</sub> / with / tupu'a<sub>inc</sub> / this<sub>inc</sub> / at / Faufano  
 'and Titof lived with this *tupu* 'a at Faufano'

<sup>5</sup> *Irisa* makes its incomplete phase irregularly as *iris*, rather than the expected but non-occurring \**irias*.

- (5b) *ia tā puer se hanue=t ne Rah noho e* (I.3)  
 he / TNS / rule<sub>inc</sub> / over / land=the<sub>inc</sub> / where / Raho<sub>inc</sub> / lived<sub>cmp</sub> / there(in)  
 ‘he ruled over the land in which Raho lived’

In example (6) we have another case of a complete phase verb before the locative clitic *e*. In addition, there is a clause-final personal name, *Fä‘äf*, clearly definite in reference but found in the incomplete phase, contrary to the prediction of Churchward.

- (6) *la ia la al=‘äki e Fä‘äf* (III.131)  
 so that / he / FUT / die<sub>inc</sub>=CAUS<sub>cmp</sub> / there / Fä‘äfe<sub>inc</sub>  
 ‘in order that he might kill Fä‘äfe there’

Ultimately, we find little support from the data that Churchward’s analysis is correct. An examination of the definiteness/indefiniteness parameter alone on the example sentences cited so far gives the results shown in (7) below.

- (7) ‘hit rate’ on the definiteness theory (misses are shaded)<sup>6</sup>

Word/Phase	Definite	Indefinite
(4a) <i>gou</i> / INC	X	
<i>tük</i> / INC		X
<i>iris</i> / INC	X	
(4b) <i>gou</i> / INC	X	
<i>tük</i> / INC		X
<i>irisa</i> / CMP	X	
(5) <i>Titoj</i> / INC	X	
<i>noh</i> / INC		X
<i>tipue</i> / INC	X	
<i>te‘is</i> / INC	X	
<i>Faufano</i> / CMP	X	
(5b) <i>puer</i> / INC		X
<i>hanue=</i> / INC	X	
<i>=t</i> / INC	X	
<i>Rah</i> / INC	X	
<i>noho</i> / CMP		X
(6) <i>al‘äki</i> / CMP		X
<i>Fä‘äf</i> / INC	X	

The chart in (7) illustrates that, for instance, in example sentence (4a) the word *iris* (incomplete phase) occurs on a definite reading (marked with an ‘X’). The total shaded examples indicate that

<sup>6</sup> Following Churchward (1940:91), we will assume that non-completive aspect verbs are ‘indefinite’, though the use of this terminology seems to be another case of twisting the semantics to try to get Churchward’s preconceived notions of “definiteness” and “phase” to work out.

12 out of the 18 cases listed are incorrectly predicted by Churchward. It appears, then, that we must search elsewhere for an explanation of the occurrences of incomplete phase – a task that we turn to in the following section.

### 3. The Relationship Between Phase and Prosodic Structure

A closer examination of the contexts in which the two phases are found reveals a quite different domain for phase distinctions than that posited by Churchward – one that is determined by phonological environment rather than syntactico-semantic properties. The data shows that complete phase forms are critically linked to the set of suffixes and clitics<sup>7</sup> listed in (8a). On the other hand, the suffixes and clitics in (8b) always trigger the incomplete phase. In addition, in cases where there is *no* suffix or clitic, the incomplete phase is used. Consequently, there is a well-defined and mutually-exclusive environment for complete and incomplete phases.<sup>8</sup>

#### (8a) Suffixes and clitics which trigger the complete phase<sup>9</sup> (after Churchward)

- ga* ‘nominalizer’: *pu* ‘a’ ‘to be greedy’ > *pu* ‘aga’ ‘greed’
- me* ‘hither’: *ho* ‘a’ ‘to take’ > *ho* ‘ame’ ‘to bring’
- (*a*)*fu*<sup>10</sup> ‘away (towards listener)’: *ho* ‘a’ > *ho* ‘afu’ ‘to take away (towards listener)’
- (*a*)*ge* ‘away (towards third person)’: *ho* ‘a’ > *ho* ‘age’ ‘to take away (towards third person)’
- a* ‘completive aspect’: *rofi* ‘to lose one’s head’ > *rofia* ‘to completely lose one’s head’
- a* ‘transitive suffix’: *hili* ‘to make a choice (intr.)’ > *hilia* ‘to choose something (tr.)’
- e* ‘locative anaphor’: *noho* ‘to dwell, live’ > *noho e* ‘to dwell therein’
- t* ‘indefinite singular article’: *vaka* ‘canoe’ > *vaka-t* ‘a canoe’
- s* ‘interrogative suffix’: *vaka* > *vaka-s* ‘which canoe’
- $\emptyset$  ‘definite plural suffix’: *vaka* ‘canoe’ > *vaka* ‘the canoes’
- $\emptyset$  ‘locative suffix’: *lopo* ‘lower part’ > ‘*e lopo* ‘in the lower part, below’

#### (8b) Suffixes and clitics which trigger the incomplete phase (after Churchward)

- ia* ‘ingressive’: *sumu* ‘to be hot’ > *sun* ‘ia’ ‘to become hot’
- tia* ‘completive aspect’: *ala* ‘to die’ > *altia* ‘to have died’
- âki* ‘causative’: *tole* ‘to carry on shoulder’ > *tol* ‘âki’ ‘to cause to be carried on shoulder’
- kia* ‘transitive’: *ho* ‘a’ ‘to take (intr.)’ > *hoa* ‘kia’ ‘to take (tr.)’
- ian* ‘ingressive’: *furi* ‘to turn around’ > *fur* ‘ian’ ‘to start turning around’
- ta* ‘a’ ‘that’: *vaka* ‘canoe’ > *vak ta* ‘a’ ‘that canoe’
- tei* ‘vocative particle’: *le* ‘e’ ‘child’ > *ko le* ‘tei’ ‘o child’
- tema* ‘each’: *vaka* ‘canoe’ > *vak tema* ‘each canoe’

<sup>7</sup> The contrast between ‘suffixes’ and ‘clitics’ is in need of a fuller linguistic investigation in Rotuman. For the present purposes it will be sufficient to treat those elements which are orthographically independent as clitics and those which are written together with what precedes as suffixes. We do not believe this is ultimately correct, but, as will be seen, the contrast plays no central role in our analysis.

<sup>8</sup> These lists build heavily on those provided by Churchward, however, an examination of the folktales reveals that there are additional members in each class. We will discuss some of these below.

<sup>9</sup> Churchward also includes the ‘transitive’ suffix *-na*, the status of which is, however, rather unclear (as Churchward admits, see his hedging 1940:116-17).

<sup>10</sup> The *a* of *-afu* and *-age* elides after preceding *a*, and forms a rising diphthong, like that produced via *a*-metathesis, after other vowels. Consequently, neither suffix increases the syllable count of the word by more than a single syllable.



*te* 'this': *vaka* 'canoe' > *vak te* 'this canoe'  
*ta* 'definite singular': *vaka* 'canoe' > *vak ta* 'the canoe'

The first thing to note is that the elements in (8a) and those in (8b) are in complementary distribution in terms of their prosodic structure. The elements in (8a), which trigger complete phase, are monosyllabic and those in (8b), which trigger incomplete phase, disyllabic.<sup>11</sup> There are, however, five apparent exceptions to the generalization about the syllable counts of the elements within each list: the asyllabic suffixes *-t*, *-s*, *-Ø* (definite), and *-Ø* (locative) in (8a), where we would expect monosyllables, and the suffix *-ta* 'definite singular' in (8b), where we would expect only disyllables. It is to the explication of these 'exceptions' that we now turn.

### 3.1 Rotuman *ta* 'one' and related forms

The five elements noted above are exceptional in another respect. Typically, complex morphemes formed with the elements in (8a) and (8b) can occur in either of the two phases, as in (9).

(9)	Complete	Incomplete	
	<i>pu'a ga</i>	<i>pu'a g</i>	'be greedy + nominalizer'
	<i>vak ta'a</i>	<i>vak ta'</i>	'that canoe'

However, it is precisely in the case of these five 'exceptional' forms that we find only *one* phase as in the examples in (10).

(10)	Complete	Incomplete	
	<i>vak ta</i>	-	'the canoe'
	-	<i>vaka-t</i>	'a canoe'

Why should this be so? If we maintain the assumption that the *only* time we find the complete phase in Rotuman is when one of the suffixes or clitics listed in (8a) follows the form in question, an answer can be found. Crucial evidence to resolve this question can be obtained from a more accurate evaluation of the elements which provide contrasts between indefinite/definite and singular/plural.

The lexeme for 'one' in Rotuman is *ta*. As in many languages, it seems likely that there is a connection between this lexeme, *ta*, and the indefinite article (see Churchward's dictionary, s.v. *ta* for supporting evidence). Assuming this, the most plausible analysis of *vaka-t* 'a canoe' is that it represents underlying /*vaka+ta*/ 'canoe' + 'one'. If this is correct for cases of the indefinite article, we must reconsider the existing analysis of the definite article seen in *vak ta* 'the canoe,' since we now have an overlap with indefinite *ta*. The key to resolving this matter lies in recognizing that Rotuman, like most Central Pacific languages, has explicit marking of singularity – a singular article is obligatory, usually the word for 'one'. Plurality is indicated by a zero. Thus the best analysis of elements in (8a) and (8b) which are currently glossed as 'indefinite singular', 'definite plural,' and 'definite singular' is that they are themselves complex morphemes made up of independent

<sup>11</sup> There are a few exceptions to this general statement which we will turn to at once. In addition, *te* 'this' is of course trisyllabic. The real contrast may therefore be between monosyllabic and polysyllabic. The matter is not relevant to our immediate concerns and we will ignore it, though a full account of the data must make provisions for this.

markers for definiteness and number and these markers are distinct in all cases. An example is given in (11).

- (11) *vaka* /vaka + Ø<sub>plural</sub> + Ø<sub>definite</sub>/ 'the canoes'  
*vak ta* /vaka + ta + Ø<sub>definite</sub>/ 'the one canoe' (i.e. 'the canoe')  
*vaka-t* /vaka + ta/ 'a/one canoe'

The Ø<sub>definite</sub> marker triggers only complete phase (for reasons we will turn to shortly). In addition, it is unable to precede any incomplete phase-triggering morphemes (those in (8b)) for what appear to be semantic reasons (\*the this canoe).

Under this analysis, the contrast between *vak ta* 'the canoe' and *vaka-t* 'a canoe', where the former occurs only in the complete phase and the latter in the incomplete, can be explained in a straightforward manner. Because the definite marker invariably takes the complete phase, there is no 'incomplete phase' version of *vak ta* 'the canoe.' There is no complete phase version of *vak-t* 'a canoe' simply because there is no affix or clitic which triggers the complete phase which can be coherently added to this element (\*that a canoe, \*this a canoe, \*each a canoe, etc.). A similar argument can be fashioned for *vaka-s* 'which canoe' (in which the -s appears to reflect an interrogative morpheme *se* also found in *seia* 'who' and related forms). Finally, the locative null affix (from (8a)) appears unable to precede any 'incomplete phase' triggering suffixes or affixes from (8b), because the locative is a phrasal affix, occurring after the *last* element in a phrasal group (thus 'in that canoe' is /*e vak ta 'a* < /*e vaka ta 'a* + Ø<sub>locative</sub>/, with complete phase of *ta 'a* 'that', rather than complete phase of *vaka*). Thus the null definite and null locative affixes invariably trigger complete phase.

Further support for the above analysis comes from historical sources for the null locative and definite morphemes. These are well-attested in Polynesian, where they show a strong tendency to persist as prosodic effects in spite of their lack of segmental content, i.e., they tend to remain 'moraic' in spite of their lack of substance. A few relevant comparanda can be seen in (12ab).

- (12a) definite plural < \*V<sub>def</sub>, i.e., *famori* 'the people' < \**famori* + V<sub>def</sub>, compare Tongan Definite Accent (Condux 1989) as well as Blevins/Garrett (this conference)  
 (12b) locative < \*V<sub>loc</sub>, i.e., *se goua* 'to me' < \**se goua* + V<sub>loc</sub>; cf. Samoan Locative Accent (Condux 1990):<sup>12</sup> *fāle* 'house', *i fālē*: 'in the house'

One further assumption, that in Rotuman, as in Tongan and Samoan, these null elements continue to act as if they headed their own syllable, brings the data in (8a) and (8b) to full generality. The 'indefinite singular' morpheme *-t* is from monosyllabic *-ta*, the 'interrogative' morpheme *-s* is from *-se*, and the definite and locative null morphemes are monosyllabic (i.e., moraic) entities which lack phonological content beyond their moraicity. Thus all of the elements in (8a) are, in fact, monosyllabic. On the other hand, what was originally listed in (8b) as a 'definite singular' marker (*-ta*) actually consists of two morphemes, the singular *-ta* 'one' and the null, but crucially moraic, definite marker. It is thus bimoraic, fitting perfectly with the generalization regarding the remaining complete phase-triggering elements of (8b).

<sup>12</sup> Condux (1990) believes that phonetically only lengthening, not accent shift, is involved in the Samoan case. Given that length may be a correlate of stress, the matter strikes us as rather more complex than Condux's treatment indicates.

In sharp contrast, therefore, to the analysis offered by Churchward (1940) and adopted by McCarthy (1996a), there is no evidence for an ‘incomplete phase’ morpheme in Rotuman. Instead, the phase generation appears to be conditioned phonologically in a straightforward manner. Elements which are followed by monomoraic affixes or clitics show up in the complete phase. All other elements, including those followed by polymoraic affixes or clitics and those followed by *no* affixes or clitics, show up in the incomplete phase. To posit the *deus ex machina* of an ‘incomplete phase morpheme’ to account for this data misses the most significant generalizations about the Rotuman phases.<sup>13</sup> We turn now to an additional distinction found in Rotuman which interacts with the phase alternations discussed above.

#### 4. The “broad” and “narrow” distinction in Rotuman

In addition to the two phases already discussed, Churchward notes that any major-category lexeme with the appropriate phonological properties will also show a contrast between what he calls the ‘broad’ and the ‘narrow’ versions. The phonological prerequisite to showing a contrast in ‘version’ is given by Churchward as in (13).

(13) Churchward (1940:14)

“... the final *a* of words whose next to last vowel is *i* or *u* is changed, in certain circumstances, to *e*. When the consonant between the last two vowels is the glottal stop, an *e*-form is found in both phases... But when any other consonant comes between the last two vowels, only the complete phase has an *e*-form.”

These ‘broad’ and ‘narrow’ versions also show phase distinctions. An example of the two versions, in each phase, is provided in (14), for a lexeme which has a glottal stop between its ‘last two vowels’ (*i’a* ‘fish’) and one which does not (*pija* ‘rat’).

(14)

	complete phase	Incomplete phase
broad version	<i>i’a</i> ‘the fish (def.pl.)’ <i>pija</i> ‘the rat (def.pl.)’	<i>ia</i> ‘fish (indef.pl.)’ <i>piaj</i> ‘rats (indef.pl.)’
narrow version	<i>i’e-t</i> ‘a fish (indef.sg.)’ <i>pije-t</i> ‘a rat (indef.sg.)’	<i>ie’ ta</i> ‘the fish (def.sg.)’ <i>piaj ta</i> ‘the rat (def.sg.)’

Churchward’s romantic account for the function of these two versions, as with his account for the two phases, seems to follow from a relatively antiquated notion of the relationship between phonetic substance (such as ‘broad’ and ‘narrow’) and semantics, as can be seen from the quote in (15).

<sup>13</sup> We thus strongly disagree with McCarthy’s assertion (1996:2) that “Rotuman has a contrast in major-category words between two phases, the *complete* and the *incomplete*, distributed according to syntactico-semantic principles...”

## (15) Churchward (1940:87)

“The terms ‘broad’ and ‘narrow’ refer, primarily, to the fact that *a* is a broader or wider vowel than *e*... Happily, however, these terms are found to be in harmony with the respective significations of the two versions: for, in certain of their uses, the *a*-form conveys the idea of bigness or plurality, while the *e*-form conveys the idea of smallness or singularity.”

As was the case with the phases, however, the distribution of ‘versions’ appears to be conditioned purely phonologically. A list of the suffixes and clitics triggering the ‘broad’ version is given in (16a), while those suffixes and clitics which trigger the ‘narrow’ version are given in (16b). [For the glosses for each of these suffixes, see (8ab).]

## (16a) Clitics and suffixes triggering the ‘broad’ version

Complete phase:  $-\emptyset_{\text{definite}}$ ,  $-\emptyset_{\text{locative}}$ , *-ga*, *-ge*, *-me*, *-(a)fu*, *-(a)ge*,  $-a_{\text{compl}}$   $-a_{\text{tr}}$ , *e*

Incomplete phase: *-ia*, *-tia*, *-‘aki*, *-kia*, *-‘ian*, no suffix/clitic within the clitic group

## (16b) Clitics and suffixes triggering the ‘narrow’ version

Complete phase: *-t* ‘indefinite singular’, *-s* ‘which’, *-na*<sup>14</sup> ‘transitive suffix’

Incomplete phase: *ta‘a*, *tei*, *tema*, *te‘isi*, *ta*

Descriptively, it seems clear that the critical conditioning element for the ‘raising’ of *a* to *e* to form the ‘narrow’ version is *both* a preceding high vowel *and* a following dental (with the proviso, as indicated by *-tia*, that that dental may not itself be followed by a high vowel). Intervening glottal stops appear to be transparent to this ‘assimilatory’ effect of coronals on *a*, whereas other consonants are not.<sup>14</sup>

Strong support for this analysis is provided by a number of suffixes not explicitly given by Churchward in his list of triggers for ‘narrow’ version which an examination of the folktales reveals do in fact trigger the raising. These include *-toua* ‘1sg. pro. subj. agreement’ (found with some intransitive verbs), as in 3.109 *ka gou la ho‘ietou se Toga...* ‘while I go back to Tonga’, where *ho‘ietou* is from *ho‘ia + toua*, and meets the conditions for ‘narrow’ version.

The ‘incomplete phase’ form *ie‘ta* ‘the fish’ shows the effects of narrow version because the glottal stop which intervenes between the dental and the target of the process, *a*, is ‘transparent’ to the process. On the other hand, the ‘incomplete phase’ form *piaj ta* crucially does not show the effects of narrow version formation because the intervening *j* is *not* transparent, thus the *a* is not ‘adjacent’ to the dental *t* in the relevant sense.

## 5. Obtaining the Rotuman Phase Distinction via Output-Output Correspondence

McCarthy (1996a) uses a set of output-output conditions triggered by the ‘incomplete phase morpheme’ to generate the incomplete phase. These conditions require that that the incomplete phase form of a given string be faithful to the complete phase form in certain respects (particularly regarding vowel quality, see McCarthy 1996a:42ff.), but obey higher ranked well-formedness constraints (and thus be unfaithful to the complete phase) in other respects (e.g., regarding the

<sup>14</sup> While *t*, *s*, and *n* seem to trigger this effect, *j* does not. The explanation for this ultimately lies with the historical antecedent of *j*, and, as is usual in such cases, it will have to be handled in a somewhat adhoc manner in a synchronic analysis.

metathesis and vowel deletion processes which characterize the incomplete phase). We feel that there are two critical objections to handling the phase distinctions in the manner advocated by McCarthy.

First, McCarthy bases his analysis on the flawed assumption that there is a distinct ‘incomplete phase morpheme.’ An examination of the data has shown that this is not the case. The phase distinction is purely phonological in nature. It is not parallel to processes such as reduplication which make reference to specific morphemes and as such are perhaps justifiably based on output-output correspondences within OT. The second, and perhaps more serious objection, is that the ‘broad’ and ‘narrow’ phase effects cannot be obtained under McCarthy’s analysis. The contrast seen in (14) between *ie’ ta* and *piaj ta* illustrates the problem for an analysis based on output-output correspondences. For *ie’ ta*, the ‘output’ which the incomplete phase appears to be being faithful to (except for the ‘incomplete phase’ metathesis) is clearly the complete phase, narrow version *i’e-t*. However, *piaj ta* cannot be built to the corresponding complete phase, narrow version *pije-t*. On the other hand, if we take *pija* (the complete phase, broad version) as the basis for the output-output generation of *piaj ta*, we should take the complete phase, broad version *i’α* as the basis for the generation of *ie’ ta*, which, given faithfulness to vowel quality, will wrongly generate *\*ia’ ta*. Consequently, a single generalization about which output is to be used is impossible. Aside from opportunistic appeal to different ‘output’ bases for the output-output correspondences in these two cases – based on the presence of a glottal stop or Ø (as opposed to some other consonant) between the high vowel and the *α* in the stem – there appears to be no coherent method for dealing with these alternations using output-output correspondence.

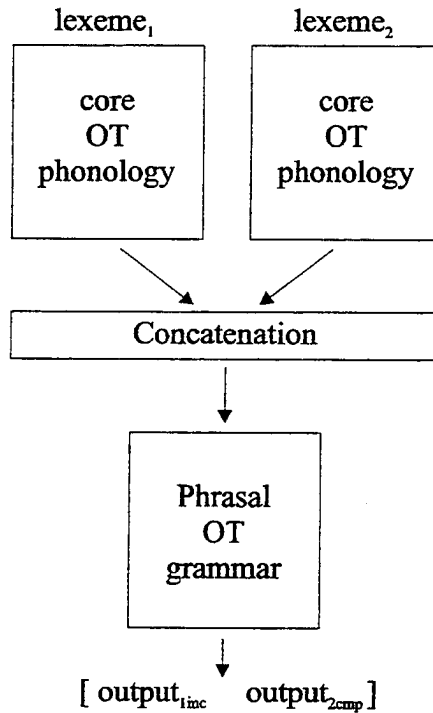
Space does not permit us to present a detailed phonological analysis of the Rotuman phases here. Having argued that an analysis based on output-output correspondences is inadequate for the reasons cited above, we would, however, like to propose an alternative analysis in which phase distinctions may still be accounted for within an OT framework.

## 6. The phonology of postlexical prosodic domains in OT

Certain predictable phonological phenomena, such as phase in Rotuman, occur in ‘postlexical prosodic domains.’ While a much more exhaustive examination of ‘external sandhi’ phenomena in a variety of the languages of the world is a necessary prerequisite to a full treatment of this issue, we propose an OT model here that might capture these sorts of facts.

Our view of the relationship between lexical and phrasal phonology is presented in (17) below. The generation of surface forms is the result of evaluation of these forms by several distinct OT-theoretic grammars, one for each (relevant) prosodic domain. For the Rotuman phenomena, for example, we would need minimally the ‘core’ or ‘lexical’ OT phonology and a phrasal OT phonology for what has usually been referred to as the ‘clitic group.’ The key distinction between the two would be that the clitic group OT phonology operates upon the output of the core OT phonology *after* the concatenation of elements within the same clitic group. While this violates the ‘surfacey’ spirit of OT, it does not, as McCarthy himself has noted (McCarthy 1996b), run afoul of any fundamental principles of this approach.

## (17) A Semi-Traditional Alternative



While we believe that our proposal, although outlined in only the briefest possible manner here, is a viable one, it raises a number of issues. For example, Faithfulness, if the theory is to be constrained at all, will have to be defined over Input-Output of the specific module (i.e., one will not be able to make reference to the underlying lexical forms in the ‘clitic group’ phonology, only to the *output* of the core lexical phonology, which serves as the *input* to the clitic group module). Other issues, such as whether the entire set of universal constraints is available at each level (the levels differing from one another only in ranking, in such a case)<sup>15</sup> or if, for instance, Well-Formedness Constraints are generally not duplicated across levels, remain open.

## 7. Conclusion

We have argued in this paper that the Rotuman phase distinctions are due to phonological processes rather than being based on syntactico-semantic principles as suggested originally by Churchward (1940) and assumed by McCarthy (1996a). Further, since they fall in the domain of predictable phonology, they should not be accounted for using output-output correspondences and, in fact, *cannot* be accounted for in that way. In addition, we have sketched an outline of how they may be handled within an Optimality Theoretic framework.

<sup>15</sup> It certainly would appear that the entire battery of Faithfulness constraints would need to be available at each level, at least in principle.

**Appendix A.**

## The Formation of the Two Phases

	COMPLETE PH	INCOMPLETE PH	
(1) Deletion			
to roll	<i>tokiri</i>	<i>tokir</i>	
big	<i>ti'u</i>	<i>ti'</i>	
coconut-spathe	<i>sulu</i>	<i>sul</i>	
to imitate	<i>rakó</i>	<i>rak</i>	
(2) Metathesis			
fish	<i>i'a</i>	<i>ia'</i>	
erroneous	<i>seseva</i>	<i>seseav</i>	
flower	<i>hosa</i>	<i>hoas</i>	
to rule	<i>pure</i>	<i>puer</i>	
prophet	<i>paraofita</i>	<i>parofiat</i>	
(3) Umlaut			
to sweep	<i>tâfi</i>	<i>taf</i>	
to sleep	<i>mose</i>	<i>mös</i>	
to pull	<i>futi</i>	<i>füt</i>	
(4) Diphthongization			
floor	<i>pupui</i>	<i>pupuy</i>	(written <i>pupui</i> )
good	<i>lelei</i>	<i>leley</i>	(written <i>lelei</i> )
to push	<i>keu</i>	<i>kew</i>	(written <i>keu</i> )
Joshua	<i>joseua</i>	<i>josewa</i>	(written <i>Joseua</i> )
(5) No distinction			
house	<i>rii</i>	<i>rii</i>	
to do	<i>ree</i>	<i>ree</i>	
cigar	<i>sikaa</i>	<i>sikaa</i>	

**Appendix B.**

Churchward's Rules (1940:88-101, summarized, the sixth, which is an ad hoc statement about pronouns, has been suppressed)

**First Rule**

1. "Except before certain suffixes, and in a few other cases, each element or component part of a composite word, other than the last, is used in its inc. phase, no matter what may be the phase of the whole word. (The phase of the word as a whole is shown, of course, by the phase of its *last* element.)"

**Second Rule**

2. "A word is used in its inc. phase when it qualifies or defines the word or group of words that follows, or (except in a few special cases) when it is qualified or defined by it."

**Third Rule**

3. "In most cases — If no defining word or group follows, a noun or verb is used in its com. phase when definite, but in its inc. phase when indefinite..."

**Fourth Rule**

4. "In some cases the use of the com. phase indicates positiveness, finality, or emphasis (or in questions) the desire to be positive or certain."

**Fifth Rule**

5. "In the case of verbs ending in a pron. suffix, the com. phase usually expresses the force of the completive tense."

**References**

- Blevins, J. 1994. The bimoraic foot in Rotuman phonology and morphology. *Oceanic Linguistics* 33:491-516.
- Blevins, J. and A. Garrett. 1995. Metathesis in Austronesian. This volume. [check title?]
- Churchward, C. M. 1938. *Tales of a lonely island*. Oceania Monographs, no. 4. Sydney.
- . 1940. *Rotuman grammar and dictionary*. Sydney: Methodist Church of Australasia.
- Conдах, I. D. 1989. Definitive accent in Tongan. *Journal of the Polynesian Society* 98 (4).
- . 1990. Locative accent in Samoan. *Oceanic Linguistics* 29:27-48.
- Hovdhaugen, E. 1990. Vowel lengthening in Samoan. In *Pacific Island Languages: Essays in honour of G. B. Milner*, J. H. C. S. Davidson (ed.). London: SOAS.
- McCarthy, J. 1996a. Extensions of Faithfulness: Rotuman revisited. ms, Rutgers Optimality Archive.
- . 1996b. Some misconceptions about Optimality Theory. ms., Rutgers Optimality Archive.



**Integrity in Malay VPs<sup>1</sup>**  
Mark Honegger  
University of Illinois, Urbana-Champaign  
honegger@cogsci.uiuc.edu

## 1.0 Introduction

This paper will examine some of the basic word order patterns in Malay. Malay is an interesting language to study with regard to word order because it is a configurational language yet appears to allow a relatively free word order compared to a configurational language like English. I say "appears" because this paper will show that the seeming flexibility of Malay word order is on par with that of English, though the parameters for Malay word order will turn out to be different than English. Malay is a good language, then, to use to test various syntactic theories and their approaches to word order. In particular, I will discuss Malay word order in light of what a Minimalist account would need to explain it. Interesting questions arise in the Minimalist Program for configurational languages that still allow relatively free word order, because all instances of the operation Move are obligatory.

Malay ostensibly appears to be an SVO language as its prototypical active sentence shows (1) and its unmarked passive follows the same SV order with the agent phrase appearing as a postverbal PP or NP (2).

- (1) Ali membaca buku itu.  
Ali / meN+read / book / that  
'Ali read the book.'
- (2) Buku itu dibaca (oleh) Ali.  
book / that / di+read / by / Ali  
'The book was read by Ali.'

However, Malay also has a number of word order patterns that it does not share with English. The task then is to identify those principles of the grammar of both Malay and English that will generate the SVO order for each but also the divergent orders as well. To do this, I will examine the word order variation of three different verbal forms in Malay with their arguments. I propose that the same principles which govern morphology govern the placement of these verbal forms with their arguments. And I will suggest that these principles are part of the phonology.

## 2.0 The basic facts

Malay has a number of verb classes that are distinguished by the type of prefix or lack thereof that appears on the verb. In this paper, I will concentrate on three classes, verbs which have the prefix /meN-/, the final segment being a nasal that assimilates in place of articulation to the following consonant, verbs which have the prefix /di-/, and verbs which have no prefix, which I will call bare verbs. The classes distinguished by the /ter-/ prefix, the /ber-/ prefix and the /ke-an/ circumfix will not be discussed in this paper, but the analysis I am proposing today could be extended to them. In this paper, I will use the terms subject and direct object to name various arguments of the verb but only in a descriptive sense. Likewise, I will use phrasal categories like NP, VP in a descriptive sense. None of these terms are intended to have any theoretical status in this analysis.

---

<sup>1</sup> This paper was originally presented at the 3rd annual meeting of the Austronesian Formal Linguistic Association. My thanks to Hiroki Koga, Mary Wu and members of the AFLA audience for their comments and criticisms.

## 2.1 /meN-/ verbs

The /meN-/ prefix marks the prototypical active transitive verb, though it does appear on selected intransitive verbs. Typically, this verb class takes an agent and a patient/theme argument. One key trait of /meN-/ verbs is that they do not allow their patient/theme argument to be extracted from appearing immediately after the verb. The patient/theme can be neither topicalized nor relativized nor extracted in a question.

- (3) Ali membaca buku itu.  
Ali / meN+read / book / det  
'Ali read the book.'
- (4) \*Buku itu Ali membaca
- (5) \*buku yang Ali membaca itu  
book / rel / Ali / meN+read / det  
'the book which Ali read'
- (6) \*Apakah Ali membaca?  
what+question / Ali / meN+read  
'What did Ali read?'

The agent phrase can be relativized though with a /meN-/ verb.

- (7) pelajar yang membaca buku itu  
student / rel / meN+read / book / det  
the student who read the book

Malay is like English in that it does not allow any phonological material between the transitive verb and its direct object.

- (8) \*Ali membaca dengan teliti buku itu.  
Ali / meN+read / with / care / book / det  
'Ali read the book carefully.'

However, /meN-/ VPs do show some variation in that the subject can appear after the /meN-/ verb plus direct object.

- (9) Membaca buku itu(lah) Ali.  
meN+read / book / det / part. / Ali  
'Read the book, Ali (did).'

The entire VP must be fronted. It is not grammatical to front only the verb.

- (10) \*Membaca Ali buku itu.  
"Ali read the book.'

## 2.2 /di-/ verbs

The /di-/ prefix marks the prototypical passive verb. The direct object of the active sentence becomes the subject of the /di-/ verb and the subject of the active sentence appears either in a PP headed by /oleh/ 'by' (11) or as an unmarked NP immediately following the verb (12).

- (11) Buku itu dibaca oleh Ali.  
book / det / di+read / by / Ali  
'The book was read by Ali.'
- (12) Buku itu dibaca Ali.  
'The book was read by Ali.'

When the /di-/ verb has its agent phrase marked with /oleh/, the following word orders are also possible in addition to (11).

- (13) Dibaca oleh Ali buku itu.
- (14) Dibaca buku itu oleh Ali.
- (15) Oleh Ali dibaca buku itu.
- (16) Oleh Ali buku itu dibaca.

The only order that is missing is (17).

- (17) \*Buku itu oleh Ali dibaca.

When the /di/ verb has its agent phrase appear as a bare NP with no preposition, then the following word order is also allowed (Dardjowidjojo 1966).

- (18) Dibaca Ali buku itu.

However, no other orders are possible. A bare agent phrase cannot occur before the verb (19-20) and the subject cannot occur between the verb and the agent phrase (21).

- (19) \*Buku itu Ali dibaca.
- (20) \*Ali buku itu dibaca.
- (21) \*Dibaca buku itu Ali.

Likewise, the subject can be relativized (22) but not the agent phrase (23).

- (22) buku yang dibaca oleh Ali itu  
'the book which was read by Ali'
- (23) \*pelajar yang buku itu dibaca itu  
'the student which the book was read by'

### 2.3 Bare verbs

The analysis of the bare verb is controversial among linguists. It is disputed whether it is a passive construction or some type of topicalization. I will not take a stand on which it is and will refer to the bare verb's arguments as the agent argument and the theme argument. The verb appears without any prefix. The agent argument must appear immediately on the left side of the verb while the patient/theme argument can appear either following the verb (24) or clause initial (25).

- (24) Ali baca buku itu  
'The book, Ali read,' or 'The book was read by Ali.'
- (25) Buku itu Ali baca.
- (26) \*Baca buku itu Ali.
- (27) \*Baca Ali buku itu.

The theme argument can be relativized (28) but not the agent argument (29).

- (28) buku yang Ali baca itu
- (29) \*pelajar yang baca buku itu

One other point to notice about this construction is that it does not allow anything to intervene between the agent phrase and the bare verb (31). Any modal or modifier must appear before the agent phrase, whether the patient/theme argument is fronted or not (32-33). Note that a modal normally appears immediately before the verb as in English (30).

- (30) Ali sudah membaca buku itu.  
 Ali / perf / meN+read / book / det  
 'Ali has read the book.'
- (31) \*Ali sudah baca buku itu.
- (32) Sudah Ali baca buku itu.
- (33) Buku itu sudah Ali baca.

The bare verb construction is the mirror image of the /meN-/ verb construction. It does not allow anything to intervene between its agent on the left and the verb. But it does allow the theme/patient to appear in situ or fronted. The /meN-/ verb construction does not allow anything to appear between the verb and its theme argument on the right, but it does allow the agent to appear postverbally.

### 3.0 Requirements for a Minimalist analysis

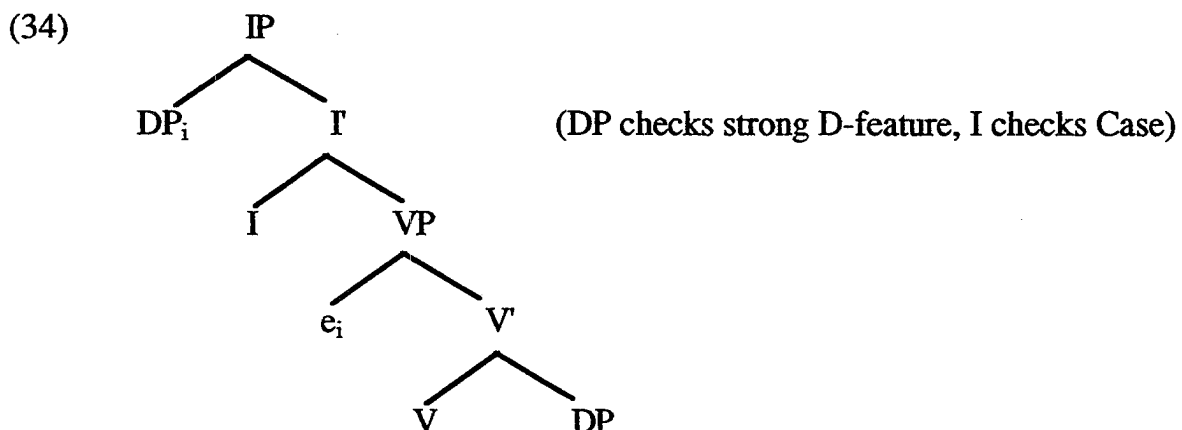
Let's examine what would be necessary for a Minimalist analysis of these phenomena. In particular, the following five questions will need to be answered.

Questions:

1. What mechanisms governs the postverbal position of the subject in the /meN-/ verb construction and the freer argument order in the /di-/ verb construction?
2. How do we account for the optionality of /oleh/ marking the agent phrase in the /di-/ verb construction?
3. Why can't the direct object move from its immediate postverbal position in the /meN-/ verb phrase?
4. What prohibits outside material from intervening in-between the /meN-/ verb and its direct object, the /di-/ verb and an agent NP, and the bare verb and its agent phrase?
5. How do we account for the placement of the patient/theme argument in the bare verb phrase?

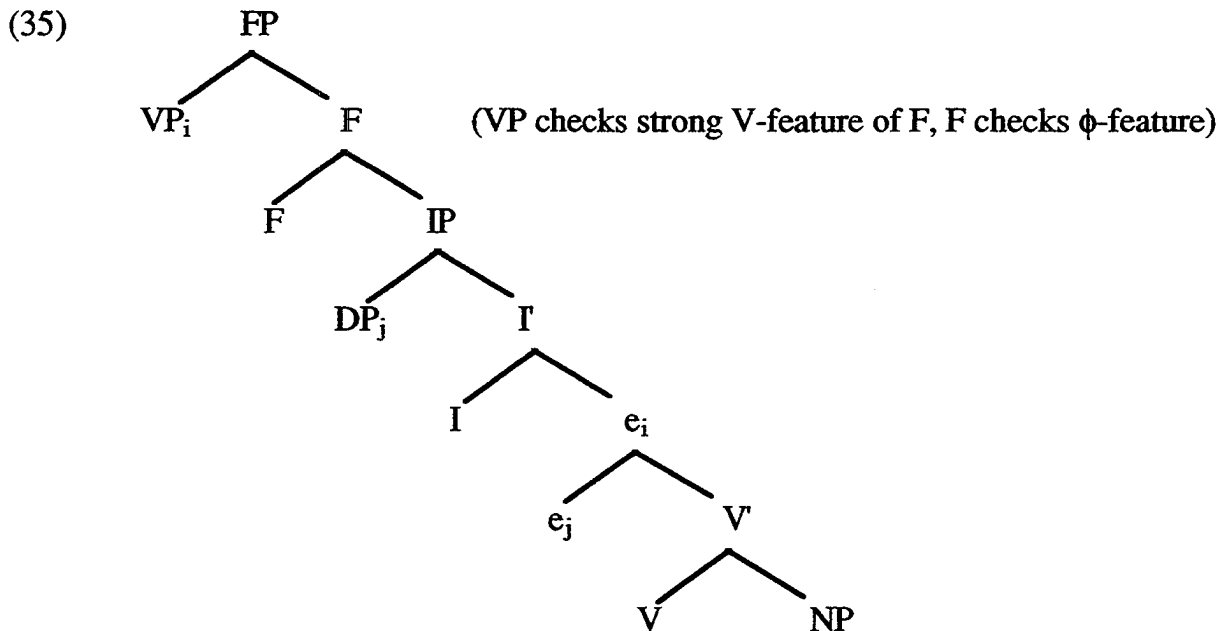
Let me discuss how such questions could be answered in a Minimalist sort of approach.

The basic SVO word order of active transitive sentences in Malay could be gotten by the same analysis used for English. Infl has a strong D-feature which must be checked and eliminated before the derivation enters Spellout. Hence, the subject must overtly raise to spec of Infl to check and eliminate the strong D-feature and to have Infl check the subject's Case feature. (I'll suppress nodes not directly relevant to the discussion.)



### 3.1 Optional/stylistic movement

The postverbal subject could conceivably be gotten by some kind of pragmatically forced movement. What obvious way to do this is to add a further functional head above Infl which has a strong V-feature cooccurring with some pragmatic phi-feature which gives a focus interpretation.



For the relatively free word order of arguments in the /di-/ verb construction, one would need whatever mechanism one would invoke for scrambling or something akin to the pragmatically forced movement I have invoked for fronting the VP in /meN-/ verb constructions.

Certainly, all of these options are technically possible within a Minimalist framework. There is no dispute on this point. The issue, rather, is how successful an analysis results from this move?

One of the clearest points of The Minimalist program (MP) is that all movement is obligatory. Optionality has been banished from the derivational component of the grammar. This leaves us with two options for accounting for empirical phenomena that look optional. One approach is to argue that such phenomena are not part of the core grammar and should not be explained on the basis of the derivational component. Such variation within a language would be accounted for by other mechanisms besides syntax, wherever prosody and pragmatics fit into the grammar. Chomsky, himself, in chapter 4 of MP, without taking an unambiguous stand seems to be heading in this direction as he argues against the adjunction of maximal categories and that order is not part of the derivational component or LF. This is a principled approach to take, but it also then becomes clear that linguists must seek non-syntactic solutions to a number of word order problems across the world's languages. However, one nagging problem remains on this approach. Under current assumptions, a minimalist approach implies that some word order is syntactically-generated and some is not. Where and how do we draw the line between these two? Is there a principled way to determine what is part of core grammar and what is not? In actual analyses, it will entail that core grammar generates some kind of basic structure which other aspects of language distort in non-

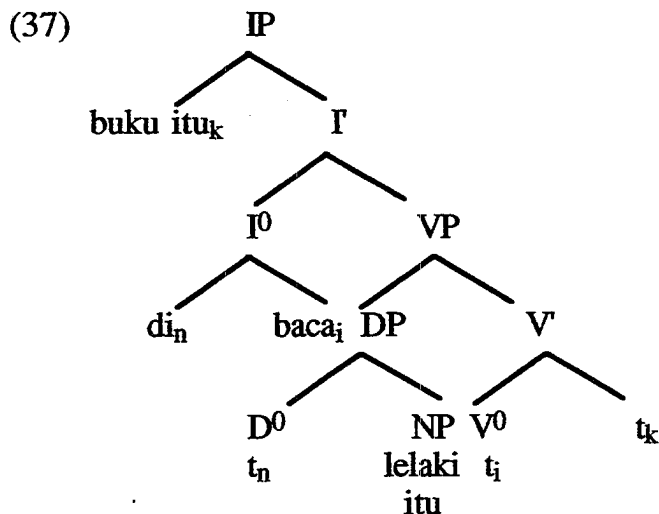
trivial ways. What limits are there on the ability of other components to distort the output of the syntax? All of these questions are unanswered at present.

The second approach is to analyze all word order as obligatory. Seeming optional word orders are forced to occur for some reason. Immediately, one can see that there will be a proliferation of functional heads and interpretable features dealing with pragmatic and interpretive issues in the derivational component. It is hard to see where such an approach would end. Syntacticians should consider how this approach would make the derivational component extremely complex and heterogeneous in an unattractive way. Right now, the derivational component does some phonology (strong features which simply put phonological material in some location), it does some semantics (raising verb constructions), and now we are going to add some pragmatic work to the mix. What links these disparate phenomena together in one component? Is this a unified domain? If one answers yes to this question, then one would subscribe to the traditional notion that syntax determines word order. But then it follows that you must analyze a number of properties that look phonological or pragmatic in nature as syntactic, a practice easily done in a feature-based system.

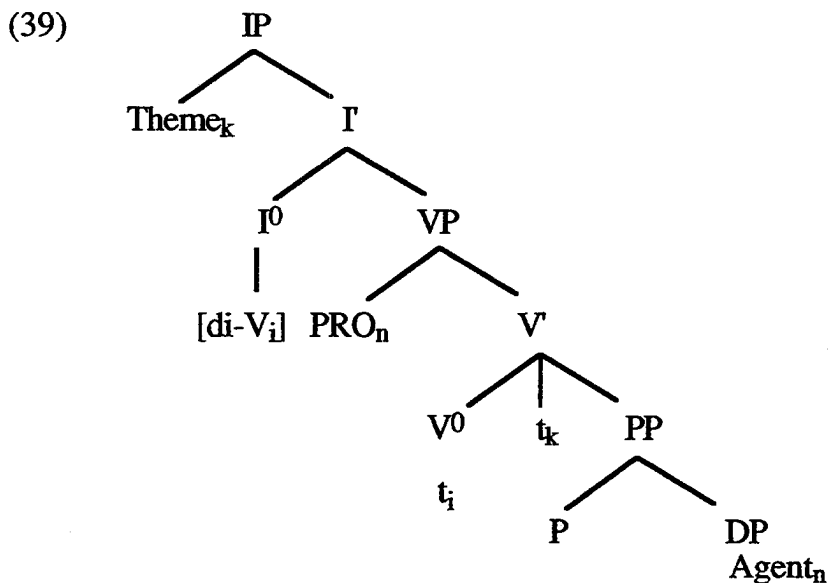
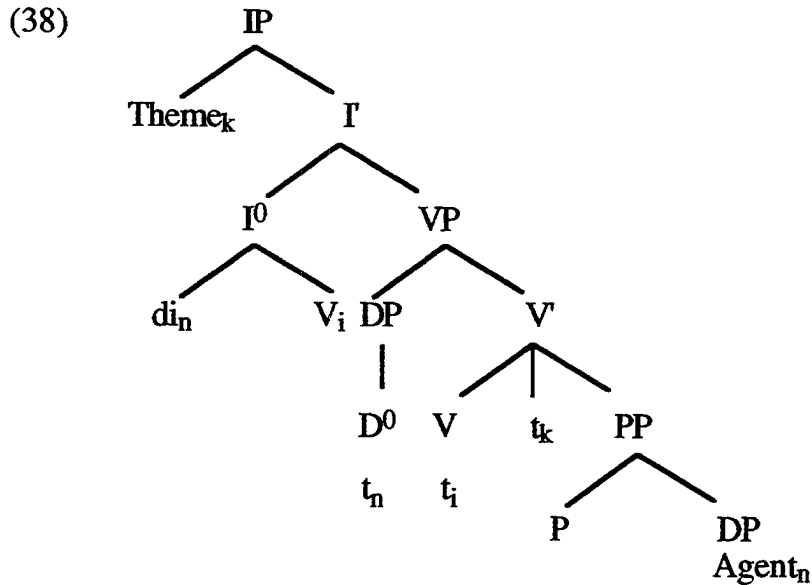
### 3.2 Optionality of /oleh/

The optionality of /oleh/ marking the agent phrase in /di-/ verb constructions does not have a simple solution. The account in Guilfoyle, Hung, and Travis (1989 and 1992) rejects analyzing this phenomenon as involving optional /oleh/ deletion or an empty preposition because this would fail to explain why outside material can intervene between the verb and its agent PP (14) but not between the verb and an agent NP (21). In their analysis of Malay word order, the Agent is generated in Spec of VP and the Theme in the complement position of V. They propose that the prefix /di-/ is generated as a passive morpheme on the verb or in  $D^0$  where it is a spellout of the formal features for the third person singular. (Note that the passive morpheme /di-/ is historically related to the third person pronoun /dia/.)  $D^0$  raises to  $I^0$  leaving behind its NP projection and  $V^0$  raises to  $I^0$  as well. This enables the Infl-V complex to Case-mark the agent phrase in Spec of VP under strict adjacency (36).

- (36) Buku itu dibaca lelaki itu.  
'The book was read by the man.'

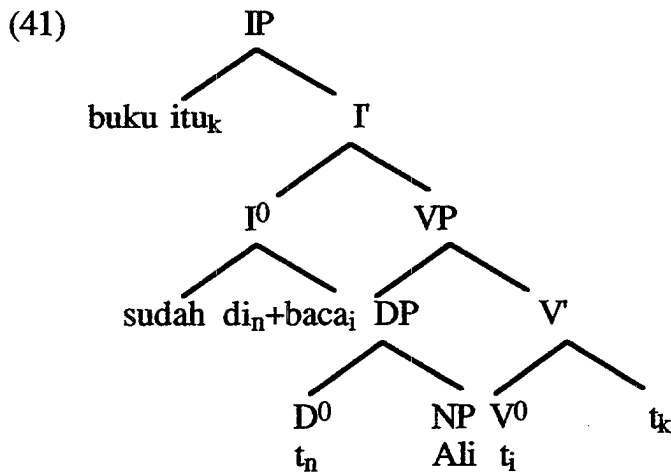


When the /oleh/ is present, the agent phrase is either coindexed with the passive morpheme /di/ (38) or with a PRO in spec of VP (39), but Case is assigned via the preposition, and so no adjacency is required between the verb and its agent phrase. (38) is the structure for what Guilfoyle et al. describe as conservative speakers who do not allow first and second person pronouns in the agent phrase of /di-/ passives while (39) is for liberal speakers who do allow first and second person there.



Two observations can be made. First, one must explain how the raising of /di-/ and the raising of the verb to INFL operate when there is an overt modal in the sentence such as (40) with the potential structure in (41).

(40) Buku itu sudah dibaca Ali.



The approach of Guifoyle et. al. strongly implies that /di-/ and the verb would incorporate into the modal. Though /di-/ is morphologically a prefix, it must incorporate after the modal because it cannot occur prefixed to the modal.

(41) \*Buku itu sudah baca Ali.

But even more telling is that the modal can appear clause-initial, not only in questions but also in indicative sentences, which means it must excorporate out of the INFL-verb complex formed by raising of the /di-/ and the verb.

(42) Boleh (pun) buku itu dibaca Ali.  
 can / particle / book / det / di+read / Ali  
 'The book CAN be read by Ali.'

Thus, it is unlikely that the passive /di-/ verb is overtly in INFL.

Second, we know that the agent PP can appear clause-initial in the /di-/ construction (15-16). Presumably, some kind of A' movement is responsible for this. But for the bare agent NP, once it is Case-marked in Spec of VP, there must be some further stipulation that prevents it from feeding A' movement. And this restriction does not appear to be semantically or pragmatically motivated as sentences like (11-12) are treated as synonymous by Malay speakers.

### 3.3 The unmovability of the direct object in /meN-/ clauses

When linguists discuss the failure of the direct object to move from its immediate postverbal position in the /meN-/ verb construction, two explanations are given. One is that the /meN-/ verb is characterized as an agent topic construction. Only topics can be relativized or topicalized. Hence, the patient/theme must stay in situ. The second explanation is that Malay obeys the Accessibility Hierarchy proposed by Keenan and Comrie 1977 which says that subjects are more relativizable than direct objects, etc.

(43) SUBJ > DO > IO > OBL > GEN > OCOMP

These explanations are not sufficient to explain what is going on in Malay. First, in regard to what is the topic and what is the focus, both the subject and the direct object in this construction can be questioned over in both wh- and yes/no questions.

(44) Siapa mencari pelajar itu di Perlis?  
 'Who was looking for the student in Perlis?'



- (45) Pelajar itu mencari siapakah di Perlis?  
 student / det / meN+seek / who+quest / prep / Perlis  
 'Who was the student looking for in Perlis?'
- (46) Alikah membaca buku?  
 'Was it Ali who was reading a book?'
- (47) Ali membaca bukukah?  
 'Was it a book that Ali read?'

Thus, as far as the pragmatics is concerned, either the subject or the direct object can be in topic or focus. But while it is possible to question the direct object, it is not possible to extract it.

- (48) \*Siapakah pelajar itu mencari di Perlis?  
 (49) \*Bukukah Ali membaca?

Second, though direct objects cannot be relativized in /meN-/ verb constructions, indirect objects can be if a different relative pronoun is used.

- (50) meja tempat saya meletakkan buku saya itu  
 table / place / I / meN+place+kan / book / I / det  
 'the table where I placed my book'

(50) does not violate the Accessibility Hierarchy (AH), because Keenan and Comrie apply the AH to language in terms of the Primary Relativization Constraint.

- (51) The Primary Relativization Constraint
1. A language must have a primary RC-forming strategy.
  2. If a primary strategy in a given language can apply to a low position on the AH, then it can apply to all higher positions.
  3. A primary strategy may cut off at any point on the AH.

The relativization of the indirect object in (50) does not utilize the primary relativization strategy in Malay which uses /yang/. Hence, Malay obeys the AH at this point, but the AH also does not explain by itself why the direct object cannot extract. (50) also argues against explaining the word order properties of /meN-/ clauses on the basis of its agent topicality, because the indirect object is the topic.

### 3.4 Adjacency requirements between verbs and arguments

The inability of outside phonological material to intervene between Malay verbs and selected arguments (8), like its counterpart in English (52) traditionally was handled by some argument to the effect that adjacency is required for the verb to assign Case to its argument.

- (52) \*Ali speaks fluently Malay.

This explanation is lost, though, in the minimalist program where adjacency is no longer a part of the theoretical apparatus and Case is checked in a spec-head position. Chomsky 1996 has suggested that (52) may be a relativized minimality violation with *fluently* blocking the attraction of *Malay* to AgrO, but this depends on the as-of-yet unproved point that the adjunct has features that are attracted to AgrO.

These sentences are precisely the kind of phenomena that make syntactic accounts of word order suspect. There is no problem in stipulating the facts in a system of syntax. The problem is motivating why such a phenomena should be in language at all.

### 3.5 Case-marking in the bare verb construction

The bare verb construction provides one of the sharpest difficulties for a syntactic account of word order that hinges on where the theme/patient argument gets its Case. The canonical position for the theme/patient is fronted before the clause (25). If this construction is analyzed as a passive, one can give the orthodox explanation for passive that the direct object of the active sentence raises to spec of Infl to get Case. This, however, leaves unexplained why the theme/patient may also stay in situ and would seem to require, on this analysis, some kind of lowering operation after Case is checked in the higher position.

The alternative is to analyze the bare verb construction as a topicalized sentence. The Case of the theme/argument would be checked in situ or at AgrO, and A' movement would put this argument in the higher position. The problem here is that the bare verb construction can feed other A movement. Thus, Chung 1976 gives examples like (53) where the theme/patient of an embedded clause can be the subject of the matrix clause.

- (53) Buku ini dianggap oleh mereka perempuan itu sudah saya baca.  
 book / det / di+believe / by / they / woman / det / perf / I / read  
 'This book is believed by them to have been read by the woman.'

Chung argues on the basis of a number of tests that the bare verb clause is a passive construction. But then one must also stipulate that the ability of the verb to assign Case to its former subject is not lost, because the agent phrase cannot be deleted in these clauses.

### 4.0 Analysis

The salient characteristic that adheres to each of these verbs is a property that we might describe as something like cliticization between one of the NP arguments and the verb itself. Specifically, the patient/theme argument "cliticizes" to the right of the /meN-/ verb, the agent argument "cliticizes" to the left of the bare verb, and an NP agent argument cliticizes to the right of the /di-/ verb. Once these properties are recognized, almost all of the word order facts observed in this paper are accounted for.

In effect, there is a constraint operating at what has been considered the level of syntax which is morphological. These cliticized arguments are acting like bound morphemes, but note that these bound morphemes need not be zero-level categories but even phrasal categories, especially in the case of /meN-/ and /di-/ verbs. I will suggest a particular optimality type of system to account for these facts, but for those who don't like this particular system, I encourage you to substitute whatever your favorite system is for generating morphology. However, these constraints are classified, they are the same ones that govern morphology. And I emphasize this point because this analysis is not adding something new to the grammar but rather applying constraints/rules already in the grammar to another empirical phenomenon. Let me describe a system that will account for Malay word order in a way that relies on these seeming "morphological" characteristics of phrases, and then discuss how the system might fit with the rest of the architecture of the grammar.

At the most basic level, where should morphemes or meaning units go in a string? If we examine nonconfigurational languages like Warlpiri, we might say that meaning units can go anywhere in a string. This is not absolutely true. Even in Warlpiri, there are second position clitics and bound morphemes whose position is restricted in the string. There is no language where there are zero restrictions on the order of meaning units, say

a language that had totally free phrasal order and no morphology. Every language has some restrictions. The basis for word order in nonconfigurational languages is phonological—realize the phonological content of the word anywhere in the string subject to a few constraints on order, such as ensuring that a verb and one NP argument do not appear before the left of the second position clitic.

Let's apply this same approach to configurational languages like Malay and say that in Malay, meaning units can go anywhere in the string subject to a set of constraints that will prohibit certain orders. At the outset, one might speculate that Malay, as a configurational language, would have a much different and larger set of constraints than Warlpiri but this does not turn out to be the case (cf. Honegger 1996).

I am adapting a proposal for SerboCroatian Clitics that Anderson (ms.) presents within an Optimality framework which relies on the most general ideas of that approach. In particular, there are constraints which hold concurrently but these constraints are ranked so that some will win out over the others. Anderson's account consists of only 3 constraint-types, **Non-Edge**, **EdgeMost**, and **Integrity**, and it turns out that these are the ones needed for the phrasal order described above

**EdgeMost** constraints come in two varieties, **EdgeMost(e,L)** and **EdgeMost(e,R)**. This constraint will be responsible for prefixes and suffixes appearing on the correct side of their hosts. So for example, the prefix *un-* in English would be subject to **EdgeMost(un,L)** while the past tense suffix *-ed* would be subject to **EdgeMost(ed,R)**.

**Non-Edge** constraints come in two varieties, **Non-Initial** and **Non-Final**. This constraint is responsible for the position of infixes, so that such a morpheme would concurrently be subject to two constraints, an **EdgeMost** and a **Non-Edge** constraint, with the **Non-Edge** constraint being ranked higher than the **EdgeMost** constraint. For example, the infix */-um-/* in Tagalog which marks the active voice is subject to **EdgeMost(um,L)** and **Non-Initial(um)**. Further phonological properties of Tagalog will then ensure that it occurs after the first consonant of its host so that */bili/* 'buy' plus */-um-/* combines to give */bumili/* 'have bought (active voice)'. **Non-Edge** constraints such as **Non-Initial** and **Non-Final** are important for analyzing second position clitics and I would also propose V2 phenomena in languages like English and other IndoEuropean languages. Significantly, it will not be needed for the Malay data discussed above.

The final constraint we need is called **Integrity**, which ensures that outside phonological material may not intervene in the domain governed by an **Integrity** constraint. All languages have **Integrity** constraints that operate first at the word level. That is, outside phrases may not appear inside of other words. Hence, an English form like *\*wihell* for *he will* is disallowed because **Integrity(word)** is a highly ranked constraint in English, with the exception of a few lexical items that are idiosyncratically marked higher, such as the adjective *blooming* in *abso-blooming-lutely*. Notice in the Tagalog example I just gave, besides the **EdgeMost(um,L)** and the **Non-Initial(um)** constraints, it will also be necessary to rank these above the **Integrity(verb)** constraints, which will allow */-um-/* to appear as an infix on the verb.

Morphology and phrasal word order are subject to the same set of constraints, so there is no qualitative dividing line between the two. Usually, the constraints governing "morphological phenomena" will be more highly ranked than those governing phrasal order, which gets us the general inability of syntax to interfere with morphology. However, this approach to morphology and syntax predicts that it will be possible for syntactic phenomena to be interwoven with morphological phenomena, and such is the

case crosslinguistically. Hence, examples like the Pashto second position clitics that appear word internally (cited in Halpern 1995) are naturally handled in this approach by the appropriate ranking of constraints. As such, all of these constraints will make reference either to specific lexical items or non-phrasal categories such as noun, verb, etc.

Under this system, the **Integrity** constraint will figure largely for capturing the cliticization properties between the verb and one of its NP arguments. The following constraints in (54) are operative. I will assume the existence of something like a theta role hierarchy as a way of referring to the semantic relationship between a verb and its arguments. N1 will denote the noun phrase that corresponds to the highest theta role argument the verb requires. N2 will denote the noun phrase that corresponds to the second highest theta role the verb requires. Note that these constraints are stated over actual meaning units or semantic information.

- (54) **Integrity(meN-V/N2)** and **EdgeMost(N2,R)**  
**Integrity(di-V/N1)** and **EdgeMost(N1,R)**  
**Integrity(V/N1)** and **EdgeMost(N1,L)**

The first pair of constraints ensures that a noun (phrase) that functions as the second most prominent argument on a /meN-/ verb's theta role hierarchy appears cliticized to the right of the verb. The second pair of constraints ensures that a noun (phrase) that functions as the agent argument on a /di-/ verb's theta role hierarchy appears cliticized to the right of the verb. The third pair of constraints ensures that a noun (phrase) that functions as the agent argument on a bare verb's theta role hierarchy appears cliticized to the left of the verb.

Without any other rules, these constraints predict that the argument of the verb not mentioned in the **Integrity** constraint can appear either to the left or the right of the verb/argument complex, and that is exactly what we find in Malay. Subject to constraints, a meaning unit can appear anywhere in the string.

Not only can we account for the data previously given, but we can also account for a number of the "syntactic" phenomenon Chung 1976 has discussed. Chung examined a number of constructions to determine whether the bare VP was a passive construction or a topicalization. Note that I have standardized the spelling of these examples with the rest of this paper.

Her first test is that passives are subject creating, that is, a direct object in some sense becomes a subject. Then she notes that subject-to-object raising kinds of constructions can only apply to embedded subjects.

- (55) Perempuan itu dikira oleh mereka sudah membaca buku itu.  
 woman / det / di+think / by / they / perf / meN+read / book / det  
 'The woman was thought by them to have read the book.'  
 (56) \*Buku ini dianggap oleh mereka perempuan itu sudah membaca.  
 book / det / di+consider / by / they / woman / det / perf / meN+read  
 'The book was considered by them the woman to have read.'

However, when the embedded clause in this construction has an a bare verb, it is the theme/patient argument rather than the agent argument which can raise.

- (57) Buku ini dianggap oleh mereka sudah saya baca.  
 'This book was considered by them to have been read by me.'  
 (58) \*Saya dianggap oleh mereka buku ini sudah baca.

Under our analysis, (56) violates **Integrity(meN-V/N2)** and (58) violates **Integrity(V/N1)**.

Another test that Chung gives is a version of Equi involving adverbial purpose clauses. When the subject of a purpose clause is coreferential with a higher NP, it can be deleted.

- (59) Dia datang untuk bercakap-cakap dengan Ali.  
 (s)he /c ome / for / ber+**speak-speak** / with / Ali  
 'He came to speak with Ali.'

Equi can apply to the subject of a /di-/ passive.

- (60) Saya membawa surat itu untuk (dapat) dibaca oleh perempuan itu.  
 I / meN+bring / letter / det / be able / di+read / by / woman / det  
 'I brought the letter to be read by the woman.'

However, it can not apply to the direct object of a /meN-/ verb.

- (61) \*Saya membawa surat itu untuk perempuan itu dapat membaca.  
 I / meN+bring / letter / det / for / woman / det / be able / meN+read  
 'I brought the letter for the woman to read.'

(61) violates **Integrity(meN-V/N2)**. When the embedded clause has a bare verb, Equi can apply to the theme/patient but not the agent.

- (62) Mereka membeli ikan itu untuk (dapat) saya masak.  
 they / meN+buy / fish / det / for / be able / I / cook  
 'They bought the fish to be cooked by me.'

- (63) \*Saya pergi untuk mobil itu (dapat) diperbaiki.  
 I / go / for / car / that / be able / repair  
 'I went to repair the car.'

For the bare verb, the integrity constraint is with the agent phrase, so extracting the agent as in (63) is ungrammatical. But the patient/theme is not subject to an integrity constraint with the bare verb and so it can be extracted.

## 5.0 A comparison with English word order

It might appear that **Integrity(meN-V/N2)** is sufficient to derive the inability of the patient/theme argument to front, but a comparison with English shows that this story is too simple. Consider what constraints will be needed to account for English SVO word order. First, (52) shows that English is like /meN-/ verbs in Malay in that outside phonological material cannot appear between the verb and its direct object, and this suggests that English has a constraint **Integrity(V[indic]/N2)** similar to the Malay **Integrity** constraints as well as **EdgeMost(N2,R)**.<sup>2</sup>

Second, English has a kind of verb-second phenomena which shows up not only in sentences like (64) but also in *that*-trace violations.

- (64) \*Read the book Ali.

<sup>2</sup> This discussion is simplified. A fuller account will have the **Non-Initial** constraint stated over the inflectional element whether that is a separate modal or a tensed verb, and the **Integrity** constraint stated over the content verb in light of VP fronting.

(i) Eat garlic, Mary did.

I will ignore these issues for the purpose of this paper.

(65) \*What did Ali believe that e ate the cat?

One promising way to handle this is to assume that the complementizer *that* introduces a boundary with which to compute edges. If this is correct, then indicative verbs will be subject to the constraint **Non-Initial(V[indic])**.

The combination of **Integrity(V[indic]/N2)**, **Non-Initial(V[indic])** and **Edge-Most(N2,R)** gets us the SVO word order in a simple transitive sentence. (64) violates the **Non-Initial** constraint. (66) would violate the **EdgeMost** constraint.

(66) \*The book read Ali.

However, there is more to the story. These constraints do not explain why a sentence initial adjunct does not rescue a **Non-Initial(V[indic])** violation (67) nor why the extraction of the direct object in questions, relative clauses and topicalized constructions (68-70) does not violate **Integrity(V[indic]/N2)**.

(67) \*Fortunately read the book Ali.

(68) What did Ali read?

(69) the book that Ali read

(70) The book, Ali read.

The answer to this problem follows from the language specific principles needed to identify verbal arguments in a linear string. This view of word order does not depend on any syntactic structures which would contain the information needed to identify which arguments match up with which semantic requirements of the verb. For languages with no overt case marking, like Malay and English, all we would have is the position in the string for gaining that information. But that appears to be sufficient. In Malay, the parsing strategy for the /meN-/ verb is to identify the first NP on its right side as the patient/theme argument. The agent argument falls under the elsewhere condition. The patient/theme must be found on the right side of the verb. Once that is identified, the agent is found next, so that it will be correctly identified even in sentence-final position.

In English transitive verbs, it is the subject that is identified by the parsing strategy. The subject, or argument bearing the highest theta role must be found on the left side of the verb. The direct object falls under the elsewhere condition. When both arguments occur on the left side of the verb, the one closest to the verb is the subject. (71-72) summarizes the parsing strategies.

(71) English: The closest unparsed noun argument to the left of the verb is the highest argument of that verb. The second highest argument is elsewhere.<sup>3</sup>

(72) Malay: /meN-/ verbs— The closest noun argument on the right of the verb is the patient/theme argument. The agent argument is elsewhere.

What this suggests is that the failure of the patient/theme argument to topicalize in Malay is due to a failure to parse such an argument appropriately. The Malay parser is looking for an argument immediately on the right of the verb, and it can't be found. Likewise, (64) and (67) are bad in English because the English parser must identify the subject on the left side of the verb.

This further suggests that the fronted element, whether in a question or relative clause or topicalized structure is not an N2 but simply an N. That is, a fronted NP is not

---

<sup>3</sup> Again, this is simplified as the parsing principles will need to be stated in relation to the inflectional element.

identified in either Malay or English automatically by its position in the linear string. If so, **Integrity** is not violated in English because the integrity constraint is stated over a different element. It follows then that there should be semantic and pragmatic differences between a fronted N and an N2 within an **Integrity(V[indic]/N2)** domain because they are not unified as far as the grammar is concerned. The only thing that links these two positions is the ability of the English parser to identify both as "direct objects" when they fall under the elsewhere condition.

Notice, this would also suggest that clause initial adjuncts could rescue a violation of **Non-Initial(V[indic])** in English if the noun is on the left side of the verb (73), which is exactly the case in the Adverb Effect noted by Culicover 1993 (inter alia). (See Honegger 1996 for more details.)

(73) Robin met the man who Leslie said that \*(for all intents and purposes) was the mayor of the city.

For Malay /meN-/ verbs, it is disallowed because it shuts down the parser. For English, it is allowed, but its identification will depend on correctly parsing the rest of the string.

From an efficiency point of view, it is not surprising that languages have only one integrity constraint between a verb class of one argument, or that a parsing principle depends on the identification of one argument for a two argument verb. That is all that is needed. Once one argument is identified, the second one comes for free.

## 6.0 Conclusion

1. I have presented a simple set of constraints that account for the order of the arguments in three Malay verb classes. Thinking about the overall architecture of the grammar, where are these constraints located? Some might want to say that these constraints are part of the morphology, but that move presumes that there is a separate module of the grammar called morphology and we have some way of delineating its own properties and empirical phenomena from that of other modules. If this is part of the morphology, than morphology is swallowing up syntax, at least in regard to word order, so the relationship between syntax and morphology is changed at any rate.

Some might want to say that these constraints are part of the syntax, but then the question becomes how do these constraints relate to syntax as it has been traditionally conceived? These constraints do not depend on or build any kind of hierarchical structure. They draw no rigid boundary between phrasal and lexical categories. They do not work on the basis of positing some kernel structure that is distorted by movement or in some representational sense. Furthermore, these constraints are susceptible to phonological information such as stress, prosody, syllable structure etc. In Malay for example, the bare verb construction becomes increasingly degraded as the agent phrase becomes longer. An unstressed pronoun clitic is most acceptable, followed by single words. A multi-word NP is not.

- (74) Buku itu kubaca.  
 'The book, I read,' or 'the book was read by me.'
- (75) Buku itu Ali baca.
- (76) \*Buku itu pelajar yang rajin itu baca.  
 'The book was read by the industrious student.'

I propose that these constraints are best understood as a part of the phonology. It has long been known that phonology is responsible for the order of segments and

syllables. The time has come to consider whether phonology might be responsible for the order of units above the syllable as well. This is similar to the view of morphology presented in Anderson 1992 where the structure of words is primarily phonological rather than syntactic. The constraint-system presented in this paper should be viewed as a language-specific way of realizing the sound sequences in a string.

Viewing word order this way also makes provides a natural way of handling exceptions such as heavy NP shift. Exceptional cases for these word orders are phonological in nature. If syntax derives the basic order, then the phonology must distort the syntax, a major problem. But allowing phonological considerations to affect the phonology is natural and leads to a parsimonious explanation and grammar.

2. While English and Malay look similar in their SVO order, there are really different combinations of constraints determining both. The Malay order is driven by **Integrity(meN-V/N2)** while the English one is driven by the V2 phenomena, **Non-Initial(V[indic])** as well as **Integrity(V[indic]/N2)**. This accounts for the word order properties presented as well as further facts such as the lack of expletives in Malay. Expletives, in this approach, are primarily the method of specific verb-second languages to prevent **Non-Initial(V[indic])** violations.

(77) Ada buku di atas meja.

(78) There is a book on the table.

3. Phonology provides the resources of for the word order in any given language. It determines what is or isn't a possible string at the most basic level. But that doesn't mean that every string made possible by the phonology is wellformed. There will be strings that are phonologically wellformed but semantically illformed or uninterpretable, and these aspects are outside the scope of this paper.

### References

- Anderson, Stephen R. 1992. A-morphous morphology. Cambridge: Cambridge University Press.
- Anderson, Stephen R. 1995. How to put your clitics in their place. Ms., Yale University.
- Chung, Sandra. 1976. On the subject of two passives in Indonesian. In *Topic and subject*, ed. by Charles N. Li. New York: Academic Press.
- Culicover, Peter. 1993. The adverb effect: Evidence against ECP accounts of the *that*-t effect. *Proceedings of NELS* 23:97-111.
- Dardjowidjojo, Soenjono. 1966. Indonesian syntax. Georgetown University, Ph.D. dissertation in Linguistics.
- Guilfoyle, Eithne, Henrietta Hung, & Lisa Travis. 1989. SPEC of VP as thematic subject and SPEC of IP as structural subject in Malayo-Polynesian languages. *McGill Working Papers in Linguistics* 6:21-47.
- . 1992. Spec of IP and Spec of VP: Two subjects in Austronesian languages. *Natural Language and Linguistic Theory* 10(3):375-414.
- Harahap, Darwis. 1989. Struktur sintaksis ayat selapis dalam Bahasa Malaysia. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Honegger, Mark. 1996. A phonologically-driven account of English word order. Ms., University of Illinois.
- Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal hierarchy. *Linguistic Inquiry* 8(1):63-99.



Mahmood, Abdul Hamid. 1992. *Ayat pasif*. Kuala Lumpur: Dewan Bahasa dan Pustaka.  
Yeoh, Chiang Kee. 1979. *Interaction of rules in Bahasa Malaysia*. University of Illinois,  
Ph.D. dissertation in linguistics.



# Reduplication in Leti<sup>1</sup>

*Harry van der Hulst & Marian Klamer*

Leiden University & Vrije Universiteit Amsterdam

## *0. Introduction*

In this article we present an analysis of reduplication facts in Tutukeian-Letine, Leti for short.<sup>2</sup> We will show that the facts are rather straightforward once it is understood how Leti reduplication interacts with two independent phonological processes of the language: metathesis and fusion.

After some preliminary remarks about syllable and root structure in Leti in section 2, section 3 offers a brief discussion of the metathesis and fusion facts, following and somewhat modifying an analysis proposed in Van der Hulst & Van Engelenhoven 1995 (henceforth VDH and VE). In section 4 we turn to the reduplication data.

Leti is an Austronesian (Central Malayo-Polynesian) language and is spoken on the island of Leti which is situated off the easternmost tip of Timor. The language has about 600 speakers, the majority of whom are around sixty to seventy years old. The data analyzed in this article come from Van Engelenhoven (1995).

## *2. Basic phonology*

In this section we provide some basic facts about the phonological structure of Leti. For a more detailed description we refer to Van Engelenhoven (1995) and to VDH and VE.

The segmental inventory of Leti (excluding a few loan phonemes) is displayed in (1). The high vowels /i/ and /u/ occur as glides /y/ and /w/ if they do not form the syllable peak and precede a non-high vowel.

---

<sup>1</sup> This contribution to AFLA III has appeared in identical form, with identical title in: Crit Cremers and Marcel den Dikken (eds). *Linguistics in the Netherlands 1996*. Amsterdam: Benjamins. Please quote the published paper.

<sup>2</sup> We are grateful to Aone van Engelenhoven for answering some questions we had about the reduplication patterns.

(1) a	consonants	b	vowels (all long and short)
	p    β    m		i            u
	t   d   s   n   l   r		e            o
	k		ε    ɔ
			a

At the surface Leti seems to allow for the phonotactic patterns that are given in the first column of (2). These patterns suggest that Leti has branching onsets (2a), closed syllables (2b,c) and complex nuclei (2d,e):

(2) a	complex onsets	#CCV	#CØ.CV
b	intervocalic clusters	VCCV	V.CØ.CV
c	closed syllables	VC#	V.CØ#
d	long vowels	VV	V.ØV
e	consonant-glide-vowel sequences	CGV	V.CV

Despite these variable surface patterns VDH and VE argue that Leti can be analyzed as a language with syllables that are strictly CV. The second column in (2) shows the analysis of the surface patterns in the first column as proposed by VDH and VE. In their view, the deviations of the strict CV pattern are only apparent. The strict CV-analysis that they propose relies on the presence of empty syllabic positions (in 2a-d) and on the analysis of a post-consonantal glide as a pre-consonantal vowel. The pre-consonantal vowel surfaces as a post-consonantal glide through a process which we call fusion (cf. section 3.2 below).

Given the strict CV-analysis referred to in the previous section, Leti roots are minimally bisyllabic. Some have a consistent bisyllabic CVCV structure, others are trisyllabic and surface in two forms, namely CVCØCV and CVCVCØ. The trisyllabic roots are involved in a process of metathesis (cf. section 3.1). We characterize the root as forming (minimally) a trocheic Foot (CVCV) or (maximally) a trocheic FootPlus (CVCVCV); cf. Van der Hulst and Klammer (to appear). Stress falls on the first root vowel, except when this is an empty vowel position.

### 3. Metathesis & fusion

We now turn to the two processes that seem to 'obscure' the regular reduplication facts. These processes are called 'metathesis A' and 'metathesis B' in VDH and VE and 'internal' and 'external' metathesis in Van Engelenhoven 1995. This terminology suggests that these processes are variants of the same process, which is not the case. To avoid confusion we will therefore distinguish these two

## REDUPLICATION IN LETI

processes by referring to the first one as 'metathesis' (section 3.1), calling the second one 'fusion' (section 3.2).

The present section draws on Van VDH and VE, although we slightly alter the view on the lexical representation of Leti forms. In particular, we do not assume that lexical representations involve so-called 'plane segregation' (cf. McCarthy 1989), but rely on storing both allomorphs. Constraints on the output determine the selection of allomorphs.

3.1. *Metathesis*. The data presented in (3) involve metathesis. The first column, headed 'final', presents the forms as they surface at the end of a phonological phrase, in the second column their phrase-medial form is given. We will return to this below.

(3) *Metathesis*

	<i>final</i>	<i>medial</i>	
a.	penta	penat	'grass'
	kuksi	kukis	'sandwich'
	βarnu	βarun	'(kind of) pigeon'
b.	βu:ra	βuar	'mountain'
	ru:ni	ruin	'dugong'
	lo:tu	lout	'servant'
	la:ra	la:r	'Anona squamosa'
	nu:nu	nu:n	'banyan'
c.	anni	anin	'wind'
	βenna	βenan	'kill'

VDH and VE propose an analysis for these facts which is based on the idea that Leti has only CV-syllables. In addition, they suggest that the template for stems that are involved in this type of phonological alternation is fixed: /CVCVCV/. The alternating forms are a result of the fact that empty nuclei are not tolerated (or licensed) phrase-finally, so that the forms in the first column in (3) end in a full vowel. In (4) we give the representation of some of the words that are involved in this metathesis alternation. (4a) are phrase-final allomorphs, (4b) are phrase-medial allomorphs.

## HARRY VAN DER HULST AND MARIAN KLAMER

(4) a	<i>final</i>	b	<i>medial</i>
	p ε n t a		p ε n a t
	c v c v c v		c v c v c v
	n u n u		n u n
			/ \
	c v c v c v		c v c v c v
	r u n i		r u i n
	c v c v c v		c v c v c v
	β ε n n a		β ε n a n
	c v c v c v		c v c v c v

Following the theory of Government Phonology (Kaye, Lowenstamm and Vergnaud 1990), we assume that empty V-positions must be licensed. Licensing can take place via Proper Government which holds if the empty V-position is followed by a filled V position in the next syllable (5a). In other words, two empty V-positions cannot occur in sequence (5b):

(5) a	V C V	b	* V C V
	∅ α		∅ ∅
	↑		

An ungoverned, and thus unlicensed empty V-position violates the Empty Category Principle (ECP). A representation containing an unlicensed empty V-position is therefore illformed. However, the V-position can be 'saved' by phonetically realizing the empty position. Standard Government Phonology does not consider 'saving' an empty V-position as a choice that languages can make or not, but rather as 'what will automatically happen': an unlicensed V-position must always be realized. Realization may take place in various ways: by producing a 'neutral' vowel sound or inserting a vowel element. We argue that Leti uses a third strategy: it chooses a different allomorph, i.e. one that does not incur the violation.

We will now explain the column headings 'final' and 'medial' in (3). According to Van Engelenhoven (1995), the 'final' forms occur when the relevant words occur phrase-finally, whereas the 'medial' forms occur in phrase-medial position (except in a number of context that we will mention below). VDH and VE now claim that the distribution of final and medial forms can be understood if the

## REDUPLICATION IN LETI

domain of Proper Government is taken to be the phrase. In that case, final forms end in a filled V-position because within the phrase there is no filled V-position following the empty position to license it. When a word occurs phrase-medially, however, the final empty V-position is followed by a filled V-position so that it is licensed. In certain phrase-medial positions the ‘final’ form occurs, whereas the expected ‘medial’ form is considered illformed. This is, for instance the case when the following word starts with a consonant cluster (cf. (2a) above), as represented in (6):

- (6) [penØta CØC.....]<sub>phrase</sub>  
 \*[penatØ CØC.....]<sub>phrase</sub>

VDH and VE show that in such cases the following word has an empty V-position in its initial syllable. In such a configuration the empty V-position of the ‘medial’ form cannot be licensed, and thereby this form is ruled out as an illformed one and the ‘final’ form is used as the only one available.

Standard Government Phonology also allows, as a parametric option, empty positions to be licensed by being domain final. In Leti, crucially, this option does not hold at the phrase level.

Lexical items with a CVCV structure (such as *lopu* ‘dolphin’ and *koni* ‘grasshopper’) do not show metathesis. They differ from forms like those in (3) (e.g. *kuksi* ~ *kukis* ‘sandwich’) in having a bisyllabic template rather than a trisyllabic one.

3.2. *Fusion*. We now turn to fusion, which can be viewed as a type of phonological liaison. The data in (7) illustrate the phenomenon:

- (7) *Fusion*
- |                    |      |        |          |          |
|--------------------|------|--------|----------|----------|
| koni ‘grasshopper’ | + de | ‘once’ | → kondie | [kondye] |
| pipi ‘goat’        | + do | ‘then’ | → pipdio | [pipdyo] |
| asu ‘dog’          | + de | ‘once’ | → asdue  | [asdwe]  |
| lopu ‘dolphin’     | + do | ‘then’ | → lopduo | [lopdwo] |
| kai ‘wages’        | + de | ‘once’ | → kadie  | [kadye]  |
| rou ‘motive’       | + de | ‘once’ | → rodue  | [rodwe]  |

The phenomenon at issue involves the high vowels /i/ and /u/. These vowels emerge as secondary articulations on the consonant that follows them in the input forms. /u/ and /i/ are lost if the vowel in the following syllable is high (i.e. also /u/ or /i/). In that case there is no fusion effect (cf. (8a) below). The vowel /a/ completely disappears, i.e. it does not leave a trace in the neighbouring syllable (cf. 8b). If the following word starts with a vowel, the high vowel shows up as an onset glide (cf. 8c):

## HARRY VAN DER HULST AND MARIAN KLAMER

- (8) a koni 'grasshopper' + di 'now' → [kondi]  
 asu 'dog' + di → [asdi]  
 tβa:lu 'we (inc) throw it' + ti 'to' → [tβa:iti]  
 b rɔna 'pot' + de 'once' → [rɔnde]  
 c ma:nu 'bird' + enu 'turtle' → [ma:nwenu]

According to VDH and VE, fusion is triggered by the delinking of a vowel melody from its V-position if this position is metrically weak — a post-tonic environment in the examples in (7) and (8). In the analysis of glide vowel nuclei, they assume that fusion also takes place pretonically, i.e. the postconsonantal glide is analyzed as a preconsonantal vowel surfacing as secondary articulation on the preceding consonant (cf. (2e) above). An illustration of this case is (18) below. Leti fusion is illustrated in (9), the dotted line indicates fusion:

- (9) a konide → [kondye] 'grasshopper once'      b koni Te:nu → [kon tye:nu] 'Teunese grasshopper'

<p>( x . )  c v c v c v          \    k o n d i e     </p>	<p>( x . )  c v c v c v c v c v          \ \ /        k ɔ n t i e n u     </p>
--	--

- c asu to:nu → [astwo:nu] 'dog pool'

( x . )  
c v c v c v c v c v  
| | | | \ \ / | | |  
a s t u o n u  
| |

The vowel melody that fuses with the following consonant leaves behind an empty V-position that is properly governed by the next vowel. For further details about Leti fusion we refer to VDH and VE (1995). We now turn our attention to the reduplication data.

#### 4. Reduplication

Van Engelenhoven (1995) presents a somewhat complicated picture of Leti reduplication which suggests that reduplication takes place to the right, i.e., the reduplicated part is a *suffix* to its base. There are cases, however, where he has to assume that reduplication is leftward, i.e. *prefixing* to the base. Our proposal



## REDUPLICATION IN LETI

is that Leti reduplication is prefixing only — it uniformly takes place to the left of the base. The reduplicant prefixes to the main stress foot and copies segmental material from it.

Leti reduplication has various functions which are given in (10):

- (10) a Category change (V → N, V → A, N → A)  
 b Relativizations  
 c Diminutives  
 d Iterative aspect

In the remainder of this paper we will discuss representative examples of Leti reduplication. They are given in (11):

(11)	<i>Reduplication</i>		<i>Root form</i>	
a	palpyali	'raft'	pali	'to float'
	wɛrwɛra	'watery'	wɛra	'water'
	olwolu	'which is sold'	olu	'sell'
	sɔpsɔpna	'servant'	sɔpna ~ sɔpan	'order'
b	sɔsɔpna	'which is ordered'	sɔpna ~ sɔpan	(idem)
	lululi	'taboo (adj)'	luli	'taboo'
	titikli	'kick for a while'	tikli ~ tikil	'kick'
c	mtatwa:tu	'afraid'	mta:tu ~ mtaut	'to be afraid'
	kririta	'low'	kri:ta ~ kriet	'to be slow'
d	pepperta	'heavy'	ppɛrta ~ pperat	'to be heavy'
e	tuɔtɔna	'the questioned'	tuɔna (/utona/)	'to question'
f	mwomodi	'which you carry'	mu - odi	'you (sg)-carry'
	vavata	'fourth'	vo - ata	'ordinal prefix-four'

In the remainder of this paper we will show that despite surface appearances, Leti employs only two reduplicative prefixes:

- (12) a CV = syllable  
 b CVCV = foot

We will now discuss the forms in (11), starting with those in (11a). The reduplicative prefix is CVCV here and the diagrams in (13) illustrate the interaction between this CVCV reduplicative template and the independent phonological process of fusion:







## REDUPLICATION IN LETI

Though Van Engelenhoven's 'rightward infixation' analysis — i.e. suffixing to stress foot disregarding the final consonant of the base — seems a reasonable alternative, it cannot handle all the cases of Leti reduplication that we have discussed. For the form in (17a), 'rightward' (suffixing/infixing) reduplication would predict the wrong form *\*ppErperta* (cf. 17b)). This case, then, Van Engelenhoven chooses to analyze as 'leftward' (i.e. prefixing) reduplication. In his analysis, therefore, Leti reduplication is bidirectional, whereas in our analysis it is unidirectional: Leti has only prefixing reduplication. Furthermore, Van Engelenhoven's analysis also runs into problems with simple cases like *sOsOpna* (14). If this form were the result of rightward infixation, the reduplicative infix would have to ignore not only the rightmost consonant /n/, but also the consonant preceding that /p/: *sO [sO] pna*.

To save the analysis, we would have to say that in the case of *sO [sO] pna* the infixation site is after the stressed syllable rather than after the stressed foot. However, this entails a disjunction in the statement of the infixation site: reduplication is suffixing either to the stressed *syllable* or to the *foot*. In contrast, our analysis does not show a comparable complication. Thus, rather than concluding that the reduplication system has choices with respect to both direction and infixation site, we prefer to take the form in (17a) as our witness for a uniform prefixing-to-foot analysis. Crosslinguistically, reduplicative prefixation to the stress foot seems to be more common in any case.

We conclude with a brief discussion of Leti reduplications that show so called 'reduplicant-to-base copying', or 'retrograde over-copying' (McCarthy and Prince (to appear) discuss similar cases in Chumash and Kihehe). Cases of 'retrograde over-copying' are of special interest because they can be used to motivate a correspondence approach to reduplication as it is proposed within Optimality Theory. Though in this paper we are not concerned with exploring an Optimality Theoretic formalisation of our analysis, the relevant cases are given in (11f), represented in (19a), output forms are given in (19b):

(19) a.	PFX	RED	BASE		PFX	RED	BASE
	c v -	c v -	c v c v		c v -	c v -	c v c v
	m u	o	o d i		v o	a	a t a
	_____↑				_____↑		
	b. [mwomodɪ]				[vavata]		

The striking fact about the output forms is the occurrence of the consonants /m/ and /v/ in the base. When a stem begins with a vowel (like *odi* and *ata* here), the reduplicative copy will also begin with a vowel. In case there is another prefix present (*mu* and *vo* here), coalescence takes place so that the first prefix becomes the onset of the reduplicative syllable; this involves the loss of /o/ in /vo-a-vata/,

the second form of (19). Thus, we see that the reduplicative syllable acquires a property that is lacking in the base. This newly acquired property is then copied back into the stem in order to acquire maximal reduplicant-base identity. This happens systematically in Leti. The OT-analysis proposed for such cases applies to the Leti cases as well. Standard derivational accounts face serious problems in dealing with this kind of phenomenon.

### 5. Conclusions

In this article it was our intention to provide an insightful analysis of reduplication in Leti. First we argued that despite the complex surface patterns, Leti reduplication can be analyzed straightforwardly once the interaction of reduplication with two important phonological processes of the language, metathesis and fusion, is understood. It was shown that Leti has two types of reduplication, foot and syllable reduplication, and that reduplication involves prefixing to the stress foot. An alternative analysis based on rightward or suffixing reduplication was argued to be less preferred on both language-internal and crosslinguistic grounds.

Secondly, our analysis of the Leti reduplication facts crucially relied on the analysis of two other processes in the language, metathesis and fusion, given in Van der Hulst and Van Engelenhoven (1995), and thus supports that analysis.

Finally, we drew attention to cases that involve 'retrograde over-copying' which are crucial in arguing in favor of an Optimality Theoretic approach to reduplicative processes.

### References

- Engelenhoven, A. van. (1995) *A Description of the Leti Language*, PhD dissertation, Leiden University.
- Hulst, H.G. van der, and A. Van Engelenhoven (1995) 'Metathesis Effects in Leti', in H. van der Hulst and J. Van de Weijer, eds., *Leiden in Last: HIL Phonology Papers I*. [HIL Publications # 1]. Holland Academic Graphics, The Hague, 243-267.
- Hulst, H.G. van der, and M. Klammer (to appear) 'The Uneven Trochee and the Structure of Kambara Roots', in M. Nespore & N. Smith, eds., *HIL Phonology Papers II*. [HIL Publications # 2]. Holland Academic Graphics, The Hague.
- Kaye, J. J. Lowenstamm and J-R Vergnaud (1990) 'Constituents Structure and Government in Phonology', *Phonology* 7/2, 193-232.
- McCarthy, J. (1989) 'Linear Order in Phonological Representations' *Linguistic Inquiry* 20/1, 71-99.
- McCarthy, J. and A. Prince (to appear). 'Faithfulness & Reduplicative Identity', in R. Kager, H. van der Hulst and W. Zonneveld, eds., *The Prosody-Morphology Interface*.

## Reduplication in Malagasy

Edward L. Keenan  
UCLA

Jean Paulin Razafimamonjy  
DIFP, Université d'Antananrivo

Our purpose here is to provide a descriptively adequate characterization of Reduplication in Malagasy. Our primary concern is precision and comprehensiveness. We intend that our description will serve as an adequacy test for the various theoretical approaches to reduplication and to Malagasy morphology that we only touch upon here. We do conclude with some challenges Malagasy poses for an Optimality Theoretic account.

**Introduction** Malagasy, like many Austronesian languages, uses reduplicated forms extensively in everyday discourse. It is not surprising to hear sentences in which essentially every content word is reduplicated.

The primary meaning of reduplication is one of attenuation: *fotsy* 'white', *fotsifotsy* 'whitish'; *maro* 'many', *maromaro* 'somewhat many'. In some cases reduplication is frequentative: *miteny* 'speaks', *miteniteny* 'jabbbers'. Used with nouns it often has a derogatory implication: *latabatra* 'table', *latabatabatra* 'sort of a table'. It is also used optionally in forming comparative adjectives (with no weakening, frequentative, or derogatory interpretation).

### I. Defining Malagasy Reduplication

**Given:** the set of roots of Malagasy (Abinal & Malzac 1888).

**Define:** the relation "*x redup y*", read as "*x reduplicates as y*". *redup* is properly a relation as a few forms have two reduplications, but usually just one, given by a function *Dup*. Our definition takes the form:

*x redup y* iff  $y = Dup(x)$  or *y* satisfies one of four special cases given adhocly later.

**Def 1 a.** The domain of *Dup* is the set PPW of *possible prosodic words*

**b.**  $Dup(\sigma) = Basic(\sigma, \sigma_i, \dots, \sigma_n)$ , where  $\sigma_i$  carries primary stress in  $\sigma$ .

We must, obviously, define PPW, *stress marked syllable* and *Basic*, which we now do.

The **vowels** of Malagasy in standard orthography are *a, e, i, o = [u]*, with diphthongs *ai/ay, ao, oy*. Vowel length is not phonemic. Word final *i* is *y*. The Malagasy **consonants** are given by the table on the next page. Malagasy **syllables** are all of the form *cv*, *c* a consonant or the empty string, *v* a vowel. So all syllables are open and (excluding recent borrowings) there are no consonant clusters. **stress marked syllables** are represented as pairs  $(cv,k)$  where  $k = 0,1, \text{ or } 2$ .  $(cv,2)$  or *primary* stressed syllables, are abbreviated *cv'*. *Secondary* stressed syllables,  $(cv,1)$ , are abbreviated *cv'*, and unstressed syllables,  $(cv,0)$ , are abbreviated *cv*. For example, the sequence of stress marked syllables in the word *elatra* 'wing' is  $\langle (e,2), (la,0), (tra,0) \rangle$ , which abbreviates to *élatra*, just the standard orthography with stress marked. A **possible prosodic word** is a finite

non-empty sequence of stressed marked syllables exactly one of which has primary stress. PPW is the set of possible prosodic words.

**consonants:** the phoneme *dz* is orthographic *j*. It is the voiced counterpart of *ts*. *tr* and *dr* are single affricates articulated with the blade of the tongue against the alveolar ridge.

<sup>n</sup>C or <sup>m</sup>C, a prenasalized C, = orthographic *nC* or *mC*, even when separated by a hyphen indicating a morpheme boundary, as in *n-tr* and *n-dr*. The C vs <sup>n</sup>C/<sup>m</sup>C distinction is phonemic in all cases. Here is a complete set of minimal pairs: *dóbo* 'pond' vs *dómbo* 'dull'; *tápoka* 'cut/dilute' vs *tám-poka* 'suddenly'; *màndodóna* 'urge (imp)' vs *màndondóna* 'knocks (at a door)'; *éto* 'here' vs *énto* 'carry (imp)'; *majájana* 'completely separated' vs *majánjána* 'strikes hard'; *átsy* 'there' vs *ántsy* 'knife'; *sédra* 'a challenge' vs *séndra* 'meet (by accident)'; *atráno* 'be prepared (imp)' vs *an-tráno* 'at home'; *sóga* 'cotton cloth' vs *sóngá* 'pulled back, cleft'; and *máika* 'rushed' vs *máinka* 'so much the more'.

	nasals	stops		affricates	fricatives	liquids
labial	m	b	p			
		<sup>m</sup> b	<sup>m</sup> p			
labio-dental					v f	
dental-alveolar	n	d	t			l
		<sup>n</sup> d	<sup>n</sup> t			
alveolar tongue tip				j ts	z s	r
				<sup>n</sup> j <sup>n</sup> ts		
tongue blade				dr tr		
				<sup>n</sup> dr <sup>n</sup> tr		
velar		g	k			h
		<sup>n</sup> g	<sup>n</sup> k			

**Basic** is a function which combines two possible prosodic words (ppw) to form a single ppw. **Basic** is used in many morphological derivational processes (MDPs), including *incorporation* of Ns into As and Vs; of As into Ns; *noun compounding*; *genitive constructions* (Paul 1996): V<sub>[-act]</sub>+Agent, N+Possessor, Prep+NP (most Preps). Three changes take place under **Basic**:

1. nasalization of onsets (*nset*) of certain syllables, defined by: For all c,v as above,

$$\begin{aligned}
 nset(c+v) = & \begin{array}{ll} n+v & \text{if } c \text{ is empty} \\ mc+v & \text{if } c = b \text{ or } p \\ nc+v & \text{if } c = d, t, j, ts, dr, tr, g, \text{ or } k \\ c+v & \text{otherwise} \end{array} \quad \text{(Here and later + is strict concatenation)}
 \end{aligned}$$



2. **Basic** shifts each *continuant* consonant  $f, v, s, z, h, l, r$  to its homorganic stop or affricate, defined by the function *stop*:

x	<i>stop</i> (x)	x	<i>stop</i> (x)
	<i>stop</i>		<i>stop</i>
f	→ p	h	→ k
v	→ b	l	→ d
s	→ ts	r	→ dr
z	→ j	c	→ c, all other consonants c

*Naturalness*: voice is invariant under *stop*:  $\text{VOICE}(c) = \text{VOICE}(\text{stop}(c))$ ,  $\text{VOICE}(x) \in \{+, -\}$ .

3. **Basic** reduces ( $\sim$ ) primary stresses to secondary ones, defined by  $\checkmark$ :

$$\text{For all stress marked syllables } \sigma = (cv, k), \checkmark = \begin{cases} (cv, k) & \text{if } k < 2 \\ (cv, 1) & \text{if } k = 2 \end{cases}$$

Of course  $\checkmark$  extends to sequences of stress marked syllables by product lifting. That is,

if  $\sigma = \langle \sigma_1, \dots, \sigma_n \rangle$  is a sequence of stress marked syllables then  $\checkmark = \langle \checkmark_1, \dots, \checkmark_n \rangle$ .  $\square$

The value of **Basic** at  $\langle \sigma, \tau \rangle$  depends on whether  $\sigma$  is *weak* or *pseudoweak*:

**Def 2. a** A possible prosodic word  $\sigma$  is *weak* iff  $\sigma$  has primary stress on the

antepenultimate syllable and the last syllable of  $\sigma$  is weak (= *-na, -ka, -tra*).

Some roots with weak endings are treated as weak by MDPs even though stress is not antepenultimate. We call these roots *pseudo-weak*. Almost all cases are two syllable roots.

**Def 2. b**  $r$  is *pseudo-weak* iff  $r =$

*tanána* 'village', *lalána* 'law' *héna* 'diminish', *fóka* 'absorb', *zátra* 'accustomed', *trátra* 'caught', *poka* 'blow', *dóna* 'knock', *sáina* 'mind', *léna* 'wet, fresh', *fétra* 'limit', *díttra* 'naughty', *táittra* 'startled', *tsóka* 'blow', *ríttra* 'dried up', *píka* 'snap', *tratra* 'caught'...

The pseudo-weak are roots+their meanings, defined by listing; membership in this class is not predictable solely on the basis of phonological and prosodic identity. Several pseudo-weak have homophones which are not treated as weak by MDPs (and so are not in the list of pseudo-weak). E.g. *héna* 'meat', *fóka* 'idiotic', *sáina* 'flag' (<Fr. *enseigne*) and *trátra* 'chest'. Some other roots with weak endings but not pseudoweak are *rehétra* 'all', *dáka* 'a kick', *lóka* 'bet', *téna* 'body', *sétra* 'brutal', *píttra* 'sad look'. We now define **Basic**:

**Def 3**  $\text{Dom}(\mathbf{Basic}) = \{ \langle \sigma, \tau \rangle \mid \sigma, \tau \in \text{PPW} \}$ . Let  $\sigma = \sigma_1 \dots \sigma_n$  and  $\tau = \tau_1 \dots \tau_m$  be possible prosodic words:  $\mathbf{Basic}(\sigma, \tau)$  is given by cases:

*Case 1 (Vowel Elision)*  $\sigma_n = (cv, k)$  and  $\tau_1 = (v', k')$

$$\begin{array}{ccc} \sigma & \tau & \mathbf{Basic}(\sigma, \tau) \\ \sigma_1, \dots, \sigma_{n-1}, (cv, k) & (v', k'), \tau_2, \dots, \tau_m & \Rightarrow \end{array}$$

*Case 1.1*  $k = 2$  (i.e.  $\sigma_n$  carries primary stress). Then  $\mathbf{Basic}(\sigma, \tau) = \check{\sigma} + \tau$ .

$\sigma$	$\tau$	$\mathbf{Basic}(\sigma, \tau)$
mandá 'refuses'	ázy 'him'	⇒ mandà ázy (= /mandàázy/)
mandá "	ólona 'people'	mandà ólona
bé 'many'	élatra 'wing'	bè élatra
mànomé 'give'	itý 'this'	mànomè itý
mànkafý 'delect in'	itý 'this'	mànkafý itý
mànkató 'obey'	ólona 'people'	mànkató ólona
mandoká 'Praise! (imp)'	andríana 'nobles'	mandokà andríana

Malagasy accepts hiatus here. Vowel coalescence here is ungrammatical. \*/mandázy/

*Case 1.2*  $k < 2$  (and  $\sigma_n = (cv, k)$  and  $\tau_1 = (v', k')$ )

*case 1.2.1*  $v = v'$  or  $v = a$ . Then  $\mathbf{Basic}(\sigma, \tau) = \check{\sigma}_1, \dots, \check{\sigma}_{n-1}, (cv', k'), \tau_2, \dots, \tau_m$

So the final vowel of  $\sigma$  elides if it is unstressed  $a$  or it is the same as the initial vowel of  $\tau$ . Except for reduplication, Vowel Elision is not registered in the orthography of MDPs .

$\sigma$	$\tau$	$\mathbf{Basic}(\sigma, \tau)$
tápaka <i>broken</i>	élatra <i>wing</i>	⇒ /tâpakélatra/ <i>has a broken wing</i>
ólona <i>person</i>	éfatra <i>four</i>	⇒ /òlonéfatra/ <i>four people</i>
mamíta <i>accomplish</i>	íraka <i>mission</i>	⇒ /mamitíraka/ (orth: mamita iraka)
áloka <i>shade</i>	áloka	⇒ àlokáloka <i>a bit of shade</i>

ívy <i>spit</i>	ívy	⇒	ìvivy <i>spittle</i>
óva <i>change</i>	óva	⇒	òvóva <i>little changes</i>

Vowel Elision is normal in ordinary speech but failure to elide in careful speech is fully intelligible. Note also that several of the derived forms present secondary stresses adjacent to primary ones, not an attested stress pattern at the level of (affixed) roots.

*Case 1.2.2*  $v \neq v'$  and  $v \neq a$ . Then  $\mathbf{Basic}(\sigma, \tau) = \check{\sigma} + \tau$  (as in Case 1.1)

So the final vowel in  $\sigma$  remains (final /i/ may reduce a bit), but stress reduction still applies:

$\sigma$	$\tau$		$\mathbf{Basic}(\sigma, \tau)$
ántso 'call'	ántso	⇒	àntsoántso    */àntsántso/
mijéry 'sees'	ólona 'people'		mijèry ólona    */mijerólona/

*Case 2 Consonant Mutation*  $\tau_1 = (cv', k')$  for some consonant  $c$ .

$\sigma$	$\tau$		$\mathbf{Basic}(\sigma, \tau)$
$\sigma_1, \dots, \sigma_{n-1}, (cv, k)$	$(cv', k'), \tau_2, \dots, \tau_m$	⇒	

*Case 2.1*  $\sigma$  is neither weak nor pseudo-weak. Then  $\mathbf{Basic}(\sigma, \tau) = \check{\sigma} + \tau$  (as in Case 1.1)

$\sigma$	$\tau$		$\mathbf{Basic}(\sigma, \tau)$
mànantány <i>asks</i>	závatra <i>something</i>	⇒	mànantàny závatra    (*mànantànin- jávatra)
mikápa <i>cuts</i>	házo <i>wood</i>		mikàpa házo    (*mikàpa-kázo)
máro <i>many</i>	ánaka <i>children</i>		marò ánaka    (*/maránaka/)
mandá <i>refuse</i>	vóla <i>money</i>		mandà vóla    (*/mandàm-bóla/)

*Case 2.2*  $\sigma = \sigma_1 \dots \sigma_n$  is weak or pseudo-weak (and  $\tau = (cv', k'), \tau_2, \dots, \tau_m$ )

*Case 2.2.1*  $\sigma_n = (ka, k)$  or  $(tra, k)$ . Then  $\mathbf{Basic}(\sigma, \tau) = \check{\sigma}_1, \dots, \check{\sigma}_{n-1}, (\mathbf{stop}(c) + v', k'), \tau_2, \dots, \tau_m$



án(a) <i>at</i>	sáha <i>fields</i>	an-tsáha <i>in (the) fields</i>
“	hády <i>ditch</i>	an-kády <i>in (the) ditch</i>
”	ráno <i>water</i>	an-dráno <i>in (the) water</i>
”	lamósina <i>back</i>	an-damósina <i>in (the) back</i>
mánana <i>has</i>	pératra <i>ring</i>	mànam-pératra <i>has a ring</i>
mánana <i>has</i>	námana <i>friend</i>	màna-námana <i>has friends</i>
fitiávana <i>love</i>	téna <i>self</i>	fitiàvan-téna <i>love of oneself</i>
mánana <i>has</i>	dídy <i>rules</i>	mànan-dídy <i>has rules</i>

Lastly contrast the pseudoweak a-examples with their non-pseudoweak homonyms (b):

$\sigma$	$\tau$	<i>Basic</i> ( $\sigma, \tau$ )	
a. sáina 'mind'	záza 'child'	sàin-jáza	
b. sáina 'flag'	fotsy 'white'	saina fotsy	*saim-potsy
a. (mi)héna 'decrease'	vídy 'price'	mihèm-bídy	
b. héna 'meat'	léna 'fresh'	héna léna	*hèn-déna
a. (mi)fóka 'absorb'	ráno 'water'	mifò-dráno	
b. fóka 'idiotic'	fóka 'idiotic'	fòkafóka	*fòpóka

This completes the definition and illustration of *Basic*. ☺☺

**An historical explanation for the behavior of weak words** While synchronically arbitrary, this behavior of weak roots receives an historical explanation first presented and supported empirically by Dahl (1951, esp. pp 105 – 115). The languages to which Malagasy is most closely related, specifically Maanjan of the S.E. Barito group in Kalimantan (S. Borneo), present a variety of closed syllables. Dahl supports that the shift to open syllables in Malagasy took place under Bantu influence when the Malagasy began settling Madagascar (0 – 400ad). Certain

word final consonants, such as *h*, *s*, and *l* were generally dropped, but words ending in *k*, *tr*, *n*, and *r* added an *a* in conformity with the open syllable pattern of Eastern Bantu. The synchronic dropping of these sounds under MDPs then is historically illusory: the derived forms existed before the *-a* was added and simply did not change (see Keenan 1996 for the role of *Inertia* in language change). That morphological derivational processes are conservative in this sense is supported elsewhere. Erwin 1996 treats weak roots underlyingly as consonant final forms and derives our roots by a rule of *-a* epenthesis, thus, roughly, reflecting the history of these forms.

### Examples of Reduplication

#### Roots stressed on the last syllable (Oxytones)

	$\sigma$	$\sigma_i \dots \sigma_n$	$Dup(\sigma) = Basic(\sigma, \sigma_i \dots \sigma_n)$
bé	'big, many'	bé	bèbé 'fairly big, fairly many'
váo	'new'	váo	vàováo 'somewhat new; the news'
fý	'delicious'	fý	fífý 'fairly good'
ré	'heard'	ré	rèré 'heard a bit'
pý	'blink'	pý	pìpý 'blinks a little' (= mipìpý)
omé	'give'	mé	omèmé (mànomé 'gives'; mànomèmé 'gives a bit')
teté	'drip'	té	tètété (mitètété 'drips, drop by drop')
vovó	'bark'	vó	vovòvó (mivovòvó 'barks some')
ampó	'in the heart'	mpó	ampòmpó 'often in the heart'
indráy	'again'	ndráy	indràindráy 'sometimes'
lèhibé	'big'	bé	lèhibèbé 'biggish'

To see e.g. that (3) yields *omèmé* as the reduplicated form of *omé* 'give', observe:

$$(3) \quad Dup(omé) = Basic(omé, mé) \quad \text{Def } Dup; me \text{ has primary stress}$$

$$= omèmé \quad \text{Def } Basic, \text{ Case 2.1.}$$

□

#### Roots stressed on the penultimate syllable (Paroxytones)

##### two syllable roots

	$\sigma$	$\sigma_i \dots \sigma_n$	$Dup(\sigma) = Basic(\sigma, \sigma_i \dots \sigma_n)$
máro	'many'	máro	màromáro 'somewhat many'
fótsy	'white'	fótsy	fòtsifótsy 'somewhat white'
háfa	'different'	háfa	hàfaháfa 'somewhat different'
máinty	'black'	máinty	màintimáinty 'somewhat black'
máitso	'green'	máitso	màitsomáitso 'somewhat green'
máimbo	'stinky'	máimbo	màimbomáimbo 'somewhat stinky'

*Dup(máro)* = *Basic(máro,máro)* = *màromáro* by the definition of *Basic*, Case 2.1.

### roots of three or more syllables

$\sigma$		$\sigma_i \dots \sigma_n$	$Dup(\sigma) = Basic(\sigma, \sigma_i \dots \sigma_n)$
hadíno	'forget'	díno	hadinodíno (mànadinodíno 'forgets a bit')
ontány	'ask'	ntány	ontànintány (màontànintány 'asks a bit')
safáry	'feel out'	fáry	safàrifáry 'feel someone out indirectly'
salóndo	'cloudy'	lóndo	salòndolóndo 'to be a bit cloudy'
haríva	'evening'	ríva	harivaríva 'early in the evening'
àlahélo	'sadness'	hélo	àlahèlohélo 'little sadness'
pàtalóha	'pants'	lóha	(mi)pàtalòhalóha 'wear as pants'
saláma	'healthy'	láma	salàmaláma 'somewhat healthy'
tanóra	'young'	nóra	tanòranóra 'somewhat young'

### Weak words (antepenultimate stress) (Proparoxytones)

$\sigma$  is always treated as weak. We illustrate all the consonant mutations:

	$\sigma$		$\sigma_i \dots \sigma_n$	$Dup(\sigma) = Basic(\sigma, \sigma_i \dots \sigma_n)$
$h \Rightarrow k$	háingana	'quickly'	háingana	hàingankáingana 'somewhat quickly'
$l \Rightarrow d$	lávitra	'far'	lávitra	làvidávitra 'somewhat far'
$f \Rightarrow p$	fántatra	'known'	fántatra	fàntapántatra 'known a bit'
$z \Rightarrow j$	závatra	'thing'	závatra	zàvajávatra 'sth of little value'
$s \Rightarrow ts$	sítrana	'cured'	sítrana	sitrantsítrana 'a bit cured'
$v \Rightarrow b$	vélona	'alive'	vélona	vèlombélonà 'more or less alive'
$r \Rightarrow dr$	résaka	'conversation'	résaka	rèsadrésaka 'chit-chat'

Observe:  $Dup(vélona) = Basic(vélona, vélona)$  Def *Dup*  
 (4)  $= v\grave{e}lo + nset(stop(v) + \acute{e}) + lona$  Def *Basic*, Case 2.2.2  
 $= v\grave{e}lo + nset(b + \acute{e}) + lona$  Def *stop*  
 $= v\grave{e}lo + mb + \acute{e} + lona$  Def *nset*  
 $= v\grave{e}lomb\acute{e}lona.$  orthography (+stress)

### Weak roots of four or more syllables

$\sigma$		$\sigma_i \dots \sigma_n$	$Dup(\sigma) = Basic(\sigma, \sigma_i \dots \sigma_n)$
latábatra	'table'	tábatra	latàbatábatra
làvarángana	'verandah'	rángana	làvaràngandrángana
karátsaka	'rustling (leaves)'	rátsaka	karàtsadrátsaka
karétoka	'seize with teeth'	rétoka	karètodrétoka
potsiátra	'spurt suddenly'	tsiátra	potsiatsiátra

satrótroka	'swelling of face'	trótroka	satròtrotrótroka
sòmarítaka	'preoccupied'	rítaka	sòmarítadrítaka

Equally *Dup* yields correct results when the copied portion begins with a vowel:

$\sigma$	$\sigma_i \dots \sigma_n$	$Dup(\sigma) = Basic(\sigma, \sigma_i \dots \sigma_n)$	
áloka	'shade'	áloka	àlokáloka 'a bit of shade'
ívy	'spit'	ívy	ìvívý 'spittle'
óva	'change'	óva	òvóva 'little changes'
ólika	'twisting route'	ólika	òlikólika 'go indirectly'
ádana	'peace, slowness'	ádana	àdanádana 'a bit peacefully'
órotra	'pull up'	órotra	òrotrórotra 'pull up a bit'

Thus  $Dup(\acute{o}va) = Basic(\acute{o}va, \acute{o}va) = \grave{o}v\acute{o}va$  by Def *Basic*, Case 1.2.1.

Finally observe that the pseudo-weak roots behave as weak under reduplication.

$\sigma$	$\sigma_i \dots \sigma_n$	$Dup(\sigma) = Basic(\sigma, \sigma_i \dots \sigma_n)$	
héna	'diminish'	héna	hènkéna *hènahéna
fóka	'absorb'	fóka	fòpóka *fòkafóka
zátra	'accustomed'	zátra	zàjátra *zàtrazátra
trátra	'caught'	trátra	tràrátra *trátratrátra
póka	'blow'	póka	pòpóka *pòkapóka
dóna	'knock'	dóna	dòdóna *dònadóna
sáina	'mind'	sáina	sàintsáina *sàinasáina
léna	'wet, fresh'	léna	lèndéna *lènaléna
fétra	'limit'	fétra	fèpétra *fètrafétra
díttra	'naughty'	díttra	dìdítra *dìtradíttra
táitra	'startled'	táitra	tàitáitra *tàitratáitra
ríttra	'dried up'	ríttra	rìdríttra *rìtraríttra
tsóka	'blow'	tsóka	tsòtsóka *tsòkatsóka

Note the adjacent stresses:  $\langle 1,2,0 \rangle$  (= secondary + primary + unstressed).

The following roots with weak endings are not in the list of pseudo-weak:

$\sigma$	$\sigma_i \dots \sigma_n$	$Dup(\sigma) = Basic(\sigma, \sigma_i \dots \sigma_n)$	
dáka	'a kick'	dáka	dàkadáka *dàdáka
lóka	'bet'	lóka	lòkalóka *lòdóka
fóka	'idiotic'	fóka	fòkafóka *fòpóka



téna 'body'	téna	tènaténa	*tènténa
ména 'red'	ména	mènaména	*mèména
sáina 'flag'	sáina	sàinasáina	*sàintsáina
sétra 'brutal'	sétra	sètrasétra	*sésétra
pítra 'sad look'	pítra	pìtrapítra	*pìpítra
trátra 'chest'	trátra	tràtratrátra	*tràtrátra

**Three special cases and an instance of reanalysis** Here we note three cases of reduplication, all of limited extent, which yield forms in addition to those predicated by *Dup*.

**1. Problems with *h*- initial roots** Under *Basic*, a root initial *h* only mutates to *k*. And this is also the most common mutation in reduplication, (5a). But  $h \Rightarrow g$ , (5b), and  $h \Rightarrow \emptyset$ , (5c), and possibly  $h \Rightarrow tr$ , (5d), are also attested.

(5) a.  $h \Rightarrow k$

- |                                |               |                 |
|--------------------------------|---------------|-----------------|
| háingana 'quickly'             | $\Rightarrow$ | hàingankáingana |
| hénjana 'stiff, tense'         | $\Rightarrow$ | hènjankénjana   |
| hínana 'eat'                   | $\Rightarrow$ | hìnkánkana      |
| híntsana 'fall (leaves, hair)' | $\Rightarrow$ | hìntsankíntsana |
| hávana 'relative'              | $\Rightarrow$ | hàvankávana     |
- |                            |               |                |
|----------------------------|---------------|----------------|
| hántsika 'arched'          | $\Rightarrow$ | hàntsikántsika |
| hítrika 'penetrate'        | $\Rightarrow$ | hìtrikítrika   |
| hítsaka 'fouler aux pieds' | $\Rightarrow$ | hìtsakítsaka   |
- |                   |               |                            |
|-------------------|---------------|----------------------------|
| hévitra 'thought' | $\Rightarrow$ | hèvi-kévitra (one speaker) |
|-------------------|---------------|----------------------------|

b.  $h \Rightarrow g$  (only with *-na* final roots)

hírana 'ray of light'	$\Rightarrow$	hìrangírana
hílana 'oscillate'	$\Rightarrow$	hìlangílana
hívina 'shaking of head'	$\Rightarrow$	hìvingívina

A few roots – *hírana* 'blink' and *hélina* 'sudden appearance' and *hórona* 'a rolling up' accept both  $h \Rightarrow k$  and  $h \Rightarrow g$ : *hìrinkírina/hìringírina*, *hèlinkélina/hèlingélina* and *hòronkórona/hòrongórona*.

c.  $h \Rightarrow \emptyset$

hévitra 'thought'	$\Rightarrow$	hèvitrévitra
héndratra 'startled'	$\Rightarrow$	hèndratréndratra
hóditra 'skin'	$\Rightarrow$	hòditróditra
hóndratra 'tremble'	$\Rightarrow$	hòdratróndratra

d. ? h ⇒ tr

hébina	'agitation'	⇒	hèbintrébina
hálona	'sparkling'	⇒	hàlontrálona
hílona	'oscillations'	⇒	hìlontrílona
hólana	'being difficult'	⇒	hòlantrólana

The cases in (5a.2) and (5c) could also be analyzed as  $h \Rightarrow \emptyset$ . The forms in (5d) cannot be analyzed this way, but they exhaust the cases of this sort in Abinal & Malzac and they are all listed as frozen (the left hand forms above not being separately listed).

Note that *h* is typically not sounded but we cannot treat even the *h*- words in (5c) as vowel initial with orthographic *h*- a relic since other applications of *Basic* show the  $h \Rightarrow k$  shift:

filazána	'saying' + hévitra	'thought'	⇒	filazàn-kévitra	'advertising'
an(a)-	'on, at' + hóditra	'skin'	⇒	an-kóditra	'on the skin'

The *h/g* and *h/∅* alternations are independently attested in official Malagasy with (*m*)*aN* prefixation (with *h/∅* more common; Paul 1996):

(6) root r		maN(r)	
hálatra	'stolen goods'	mangálatra	'steals'
hátaka	'ask'	mangátaka	'asks'
hètahéta	'thirst'	mangètahéta	'is thirsty'
			h ⇒ g
héty	'cut hair'	manéty	'cuts hair'
hátona	'approach'	manátona	'approachs'
hídy	'lock'	manídy	'locks'
			h ⇒ ∅

The *h/tr* alternation is not otherwise attested in official Malagasy, but it is attested in various regional varieties. Thus where official Malagasy has *ravin-kazo* (*ravina* 'leaf' + *hazo* 'tree') Northern Betsileo has *ravin-trazo*. This dialect variation suggests an historical basis for the *h/tr* alternation, but we have not pursued this point.

**2. *k*-insertion** A few vowel initial weak roots ending in *-na* accept optionally a *k* inserted initially in the reduplicant (Rahajarizafy 1960;88).

(7)	σ		<i>Dup</i> (σ)	
ádana	'slowness'	àdanádana	and	àdankádana
ómana	'preparer'	omanómana		omankómana
óndana	'pillow'	òndanóndana		òdankóndana
ídina	'descend'	ìdinídina		ìdinkídina
ádina	'exam'	àdinádina		àdinkádina

This variation would receive an historical explanation if it was found that these roots began historically with *h*. Then the *k* forms are expected from the application of *Basic*, Case 2.2.1, to the historical root, and those without the *k* result from application of *Basic* once the vowel initial form is taken as the root. So as with the weak roots, reduplication here also would be built on the historically earlier forms.

**3. final vowel retention** In some cases of two syllable vowel initial roots with penultimate stress whose final syllable is not weak, the elision of the final vowel is optional. Thus both *òvóva* (already given) and *òvaóva* are acceptable reduplications of *óva* 'change' and *ìvivy* (already cited) and *ìvívvy* are acceptable reduplications of *ivy* 'spit'.

**4. a case of reanalysis** A few Ns have been relexicalized with their third person genitive ending *-ny* (which does not attract stress to the right), resulting in a form reanalyzed as underlyingly ending in *-na*. E.g. from *sásaka* 'half' we form the regular *sásany* 'its half', now relexicalized as a quantifier meaning 'some'. It reduplicates to *sàsantsásany* 'some, a few'. Similarly from *rámbo* 'tail' we have *rámmony* 'its tail' and the reduplicated form *ràmbondrámbony* 'in the last ranks, towards the end'; again a regular form if we analyze *rámmony* as *rambona+ny*. And building ultimately from *lóha* 'head' we have *vòalóhany* 'at first', which reduplicates to *vòalòhandóhany*, as though the underlying form were *vòalóhana*. To handle these cases we shall include *-ny* among the weak endings.

### Domain of reduplication

Dom(*Dup*) includes almost all contentful roots (including contentful Preps: *lavitra* 'far' ⇒ *lavidavitra* 'somewhat far', *akaiky* 'near' ⇒ *akaikikaiky* 'somewhat near', *tandrify* 'opposite' ⇒ *tandrifindrify* 'more or less opposite') and excludes in general grammatical morphemes, number names, demonstratives, and proper names.

- (8) a. Tsy faly ve izy?      "Is he not happy?"  
       not happy ? he
- b. \*Tsitsy faly ve izy?    \*Tsy faly veve izy?    \*Tsy faly ve izizy?

Number names and demonstrative do enter other sorts of copying structures: distributive numeral formation is illustrated here for *telo* 'three' ⇒ *tsitelotelo* 'in threes, three by three'. Demonstratives like *eo* 'here, visible, non-past' form adverbials like *eo ho eo* 'approximately' and *eo no ho eo* 'shortly'. Demonstrative adjectives occur as framing expressions: *ity tranon-dRabe ity* 'this house of Rabe's this' for "this house of Rabe's".

Another class of unreduplicatable roots in Abinal and Malzac (1888) are those of the form  $\sigma\tau\sigma\tau$ , such as *tàbatába* 'noise', *sàlasála* 'hesitation', *vèzivézy* 'run around'; *sèraséra* 'go back and forth, communication'. These forms are obviously frozen reduplications whose base no longer exists as an independent root. All frozen reduplications are of the form  $\sigma\tau\sigma\tau$ , there being none of the sort  $\sigma\tau\sigma\tau\tau$ , e.g. *tàho-táhotra*.

These remarks come close to defining  $\text{Dom}(\mathbf{Dup})$ . There are just two cases where morphemically complex forms reduplicate.

1. Most active verbs prefixed with  $aN$ - (forms given with the present tense  $m$ - prefix) apply  $aN$ - to reduplicated roots:

(9)	$\sigma$	$maN(\sigma)$	$Red(\sigma)$	$maN(Red(\sigma))$	$*Red(maN(\sigma))$
váky	'cut, read'	mamáky	vàkiváky	mamàkiváky	*mamàkimáky
váha	'untie'	mamáha	vàhaváha	mamàhaváha	*mamàhamáha
fótotra	'basis'	mamótotra	fòtopótotra	mamòtopótotra	*mamòtomótotra
vádika	'other side'	mamádika	vàdibádika	mamàdibádika	*mamàdimádika
vélona	'living'	mamélona	vèlombélona	mamèlombélona	*mamèlomélona
sóratra	'writing'	manóratra	sòratsóratra	manòratsóratra	*manòranóratra
vángy	'visit'	mamángy	vàngivángy	mamàngivángy	*mamàngimángy

Crucially we see that when  $maN$  applies to the roots above, or their reduplicated forms, the initial consonant is lost or modified (Paul, 1996). But that consonant appears in the reduplicant. In the case of the weak *soratra*, *fototra*, *vadika* and *velona* it is replaced by its value under *stop*, exactly the changes predicted by reduplicating the root. Had we reduplicated the  $maN$  prefixed form this consonant would not appear, an incorrect result.

But there are other cases where  $\mathbf{Dup}$  visibly applies to  $maN$  prefixed forms. Corresponding cases in Indonesian have motivated analyses of "overapplication" (for which, in Malagasy, we provide an original, if prosaic, analysis). First some examples:

(10)	root $\sigma$	$maN(\sigma)$	$\sigma_i \dots \sigma_n$	$\mathbf{Dup}(maN(\sigma))$	$maN(\mathbf{Dup}(\sigma))$
hóvitra	'shiver'	mangóvitra	ngóvitra	mangòvingóvitra	*mangòvikóvitra
lá	'refusal'	mandá	ndá	mandàndá	*mandàlá
léha	'go'	mandéha	ndéha	mandèhandéha	*mandèhaléha
lóa	'pay, vomit'	mandóa	ndóa	mandòandóha	*mandòalóa

Clearly the copied part includes the prenasalization induced by  $maN$ . So  $\mathbf{Dup}$  applies to some affixed roots which themselves must be excluded from  $\text{Dom}(\mathbf{redup})$ . The *action* of  $\mathbf{Dup}$  is as given: it copies to the right beginning with the stressed syllable, and then applies *Basic*, modified with the  $h$  alternations specific to reduplication. All that is at issue is the identity of the set of forms that the copying function applies to. And clearly this set must include some derived forms in addition to roots. As we find no conditioning factor we can do no better than list those *man*- verbs that lie in the domain of the reduplication function.

In some cases both the root and the derived *man*- verb are in the  $\text{Dom}(\mathbf{Dup})$ . So both  $\mathbf{Dup}(maN(\sigma))$  and  $maN(\mathbf{Dup}(\sigma))$  are attested:

(11) root $\sigma$	$maN(\sigma)$	$Dup(maN(\sigma))$	$maN(Dup(\sigma))$
vóno 'hit, kill'	mamóno	mamònomóno	mamònovóno
láinga 'lie'	mandáinga	mandàingandáinga	mandàingaláinga
táo 'do'	manáo	manàonáo	manàotáo

**caught in the act!** We are witnessing historical change in progress. In (11), children and teenagers are more likely to use the form that reduplicates after *maN* prefixation, and older generation speakers are more likely to use the form that reduplicates the root first and then applies *maN* prefixation. Once we think of reduplication as a function (or relation) it is easy to see that what is changing is its domain. Verbs built by *maN*- are being added to that domain, and in some cases their roots are being removed, in others the roots remain, yielding doublets as in (11). Wrt **redup** some *maN*- forms are being reinterpreted as roots. This is not too surprising: *maN* prefixation is partially non-transparent, often destroying the initial consonant of the root making retrieval of the root difficult. Thus while *manoratra* 'writes' is in fact derived from the root *soratra*, on purely phonological grounds it could also have been derived from *horatra*, *toratra*, *foratra* and *oratra*.

Commonly historical changes take the form  $A \Rightarrow A,B = B$ . See Keenan (1996) for the period in English in which both *him* and *himself* occurred locally bound.

2. The second case of verbs entering the domain of **redup** is the 20 odd  $\emptyset$ -prefix roots which exceptionally prefix tense markers (*m-* / *n-* / *h-*) directly to form active verbs. These roots (Rahajarizafy, p.47-48) are never independent words and lack a suffix passive distinct from the circumstantial form. Some examples:

(12) root $\sigma$	$act(\sigma)$	$Red(act(\sigma))$	$act(Red(\sigma))$
éty	méty 'ok, agrees'	mètiméty	*mètiéty
ísy	mísy 'there exists'	mìsimísy	*mìsísy
ódy	módy 'go home'	mòdimódy	*mòdiódy
ónina	mónina 'reside'	mònimónina	*mòninónina
ánana	mánana 'has'	mànamánana	*mànanánana
ínona	mínona 'drink a certain ritual poison'	mìnomínona	*mìnonínona

So the roots in (12) must be excluded from Dom(**redup**). Rather more often however both the root and the derived verb lie in Dom(**redup**).

(13) root $\sigma$	$act(\sigma)$	$Red(act(\sigma))$	$act(Red(\sigma))$
áka	máka 'takes'	màkamáka	màkàka
ídina	mídina 'descends'	mìdimídina	mìdinídina
íditra	míditra 'enters'	mìdimíditra	mìditríditra
íla	míla 'needs'	mìlamíla	mìlaíla

íno	míno 'believes'	mínomíno	mínoíno
índrana	míndrana 'borrows'	míndramíndrana	míndraníndrana
íta	míta 'cross (water)'	mitamíta	mítaíta

This full verb reduplication applies also in past and future tense. Thus alongside *màkamàka* 'takes a bit' we have *nàkanàka* 'took a bit' and *hàkahàka* 'will take a little'.

### Syntactic Distribution of Reduplicated Forms

In general if x *redup* y then y has the same possibilities of occurrence as x except that it cannot undergo reduplication. More formally,

$x \text{ redup } y \Rightarrow (1) \neg \exists z y \text{ redup } z$  and

(2) for all generating functions (relations)  $R \neq \text{redup}$ ,

$(...x...) \in \text{Dom}(R)$  iff  $(...y...) \in \text{Dom}(R)$

So if x reduplicates as y and x has an *-ina* or an *a-* passive so does y. If x forms active verbs with *(m)i-* or *(m)aN-*, so does y; if x accepts reciprocal or causative affixation so does y; if x forms imperatives so does y. In general then a reduplicated form has the same distribution as its unreduplicated counterpart, save that it cannot further reduplicate<sup>1</sup>.

And as most items that undergo reduplication are roots, which may fail to be words, it might seem reasonable to think that Reduplication in Malagasy is a lexical process, one that "takes place" in the lexicon. In support of this is the fact that some reduplications have idiosyncratic meanings compared to forms they are reduplications of: Thus from *tsangana* 'erect (adj)' we form the simple active verb *mitsangana* 'stands up'. But the reduplicated root *mitsangantsangana* means 'walks around'. From the root *tamby* 'salary, payment' we form the active *manamby* 'hires', but the reduplicated active *manambitamby* means 'caress, cajole'. Similarly *faka* 'cause, root' yields the reduplicated passive *fakafakaina* 'is examined'. And from the root *fana* 'heat' we have both *mafampana* 'lukewarm' and *mafanafana* 'lively'. On the other hand the fact that tense markers and the active prefix *aN-* are sometimes included in the forms that undergo Reduplication argues against this. At the moment then we must just

---

<sup>1</sup> Keenan & Polinsky (1998) note one exception (in addition to reduplication itself). Namely, *tafa-* prefixation. Thus (i) below is natural, but its phonologically well formed reduplication in (ii) is not.

- i. Tafiditro (tafa+iditra+o) ny omby  
 pass+enter+1sg.gen the cow(s)  
*I got the cows in (or The cows were gotten in by me)*

- ii. \*Tafidi(k)iditro ny omby

acknowledge that the place of Reduplication in standard organizations of grammar is unclear.

This completes our descriptive account of reduplication in Malagasy. We close with a brief and tentative consideration of an Optimality Theoretic (McCarthy & Prince 1995, henceforth M&P95) account of reduplication in Malagasy. To this end we note the following constraints on prosodic words in Malagasy:

(14) A *prosodic word* in Malagasy is a sequence  $\sigma = \sigma_1, \dots, \sigma_n$  of stress marked syllables satisfying the following *PrWd* Constraints:

1. Exactly one  $\sigma_i$  has primary stress
2. If  $\sigma_i$  has primary stress then  $i+2 \geq n$ .  
(= the primary stressed syllable in  $\sigma$  is not farther left than antepenultimate position)
3. If primary stress in  $\sigma$  is antepenultimate then  $\sigma_n = -na, -ka, -tra, -ny$ , or  $-ko$ , where  $-ny$  and  $-ko$  are the 1sg and 3 person genitive suffixes (which do not shift stress rightward). E.g. *lámba* 'clothes'  $\Rightarrow$  *lámbako* 'my clothes', *lámbany* 'his/their clothes'
4. The rightmost *e* or diphthong *ao, ai/ay* or *oi/oy* has primary stress (dominated by *PrWd*(3))
5.  $\sigma_1 \neq {}^n cv, c$  voiceless. E.g. \* ${}^n k, {}^n t, {}^n ts, {}^n tr$  or  ${}^m p$  root or word initially  
(The common agentive prefix *mp-* is heard as */p/*. Exceptionally one word does begin with  ${}^n t$ : *ntaolo* 'the ancients').
6. for  $v \neq o$ , \* $e+v$  tautomorphemically (Erwin 1996); also \* $a+a$
7. No subsequence of four contiguous  $\sigma_i$  lack consonants

□

M&P95 represent Reduplication as a pair  $\langle X, Y \rangle$  where  $X$  is a pair  $\langle \text{RED}, \text{stem} \rangle$  and  $Y$  is the reduplicated form of *stem* assumed deconcatenated(!!!) into a pair consisting of a *Base* (B) and a *Reduplicant* (R). Crucially  $Y$  is an expression in the language, an "output" form, not some sort of noumenal creature underlying the phenomenal world of audible delights. Important constraints on Reduplication are given in terms of identity conditions holding of B and R, S (*stem*) and R, and S and B, as well as language particular conditions on R, e.g.  $R = \text{PrWd}$  or  $R = \sigma$ , etc.

Distinguishing B and R is crucial for M&P, but no criteria for making the distinction are offered. They do say that in total Reduplication one can't tell which part is B and which is R, so by implicature one can tell in partial Reduplication. They do at times suggest that B+R corresponds to "base + affix" or "base + copy" on more derivationally oriented theories. Most consistent with their treatment then is the following: when one part of  $Y$  is identical to S and the rest is a proper substring of S then the identical part is the B and the substring is R.

The Malagasy order is then presumably Base+Reduplicant, since the most usual pattern with partial reduplication puts the best approximation to a full copy of the Stem leftmost:

(15) teté 'drip'	⇒	tetè + té (mitetètè 'drips, drop by drop')
vovó 'bark'	⇒	vovò + vó (mivovòvó 'barks some')
ampó 'in the heart'	⇒	ampò + mpó 'often in the heart' [= a.mpò.mpó]
indráy 'again'	⇒	indrài + ndráy 'sometimes' [= i.ndràì.ndráy]
lèhibé 'big'	⇒	lèhibè + bé 'biggish'
ontány 'ask'	⇒	ontàni + ntány (mà.no.ntà.ni.ntá.ny 'asks a bit')
haríva 'evening'	⇒	harìva + ríva 'early in the evening'
àlahélo 'sadness'	⇒	àlahèlo + hélo 'little sadness'
pàtalóha 'pants'	⇒	(mi)pàtalòha + lóha 'wear as pants'
saláma 'healthy'	⇒	salàma + láma 'somewhat healthy'
tanóra 'young'	⇒	tanòra + nóra 'somewhat young'

But notice that even analyzing e.g. salàmaláma 'somewhat healthy' as [salàma]<sub>B</sub>[láma]<sub>R</sub> the Base is not quite identical to the Stem, since it carries no main stress whereas the Stem does. And since a main stress in the Base reduces to secondary stress in the reduplicated form we will get cases (indeed many of them) where a syllable  $\sigma_i$  has greater stress than some  $\sigma_j$  in the Base but they have equal (secondary) stress in the reduplicated form. This is the case with *àlahélo* and *pàtalóha* above.

Note also that if we analyze reduplication as Reduplicant + Base then for the basic cases cited above the Base will never be identical to the Stem. We thus adopt (16):

(16) In OT terms, then, if x **redup** y then y = Base + Reduplicant

Moreover, thinking of the Reduplicant as the "copy" we see that Malagasy falls into the usual pattern that it copies to the side it copies from. Specifically it copies from the righthand side of the Stem and it copies it to the right. Perhaps Reduplication in Malagasy is slightly unusual in that it is suffixal not prefixal, as appears to be the case both for most languages and in particular for most languages genetically related to Malagasy.

(17) *Applying some OT generalizations on Reduplication to Malagasy*

1. *the Reduplicant may be phonologically less marked than the Base or than expressions in the language generally.*

This is non-trivially supported: R is never more marked than B or S, and in one case the Reduplicant assumes a less marked form. Namely, in the  $\emptyset$ -prefix verbs we may apply **Dup** to the consonant initial derived form rather than the root: *misy* ⇒ *mìsi+mísy*, \**mìsiísy*, \**mìsísy*

2. a. *General constraints in "Input-Output" relationships apply in Reduplication*  
 b. *Identity constraints applicable in Reduplication apply elsewhere (ideally)*

(17.2b) The morphophonological alternations used in **Dup** (except the marginal  $h \Rightarrow tr$ ) all occur in other MDPs e.g. ones that use **Basic** or the  $h \Rightarrow g$ ,  $h \Rightarrow \emptyset$  alternations in *maN* prefixation.



(17.2a) There are phonological changes used outside of Red that do not occur in Red. E.g. under *maN* prefixation root initial *f*-may delete: *maN+fôtotra* = *mamôtotra*; but only *fôtopôtotra* exists as a reduplicated form. Also (below) affixing induces stress patterns unknown to reduplicated forms.

### 3. *the Reduplicant bears an affix-like relationship to the Base*

This seems not to hold. I don't see much similarity between *Basic* and the suffixing processes discussed by Erwin (1996). Specifically,

#### 1. Affixing induces stress gaps, *Basic* does not:

*fináritra* 'is pleased' = (m)àhafináritra 'is pleasing' = fâhafinarétina 'circ.nom.'  
 $\langle 0,2,0,0 \rangle \Rightarrow \langle 1,0,0,2,0,0 \rangle \Rightarrow \langle 1,0,0,0,2,0,0 \rangle$

#### 2. *Basic* induces weak stress clashes, affixing does not

*filána* + *hévitra* = *filàn-kévitra* 'Advisory board'  
 $\langle 0,2,0 \rangle + \langle 2,0,0 \rangle = \langle 0,1,2,0,0 \rangle$

*fialána* + *sásatra* = *fialàn-tsásatra* 'a rest period, "removal of tiredness"  
 $\langle 0,0,2,0 \rangle + \langle 2,0,0 \rangle = \langle 0,0,1,2,0,0 \rangle$

*fétra* + *fétra* = *fêpétra*      $\langle 2,0 \rangle + \langle 2,0 \rangle = \langle 1,2,0 \rangle$

*jamóka* + *móka* = *jamòmóka* 'old name for cattle'  
 $\langle 0,2,0 \rangle + \langle 2,0 \rangle = \langle 0,1,2,0 \rangle$

#### 3. Suffixing (passive, circumstantial, imperative all voices)

##### a. triggers epenthetic consonant insertion (and some vowel changes), *Basic* MDPs do not

ex: the passive suffix *-ana/-ina*:

root	passive	
àmpy 'help'	àmpiana	(regular)
àndry 'wait'	àndrasana	insert [s]
ély 'disperse'	èlézana	insert [z]

a'. alternatively, adopting Erwin's (op cit) in which weak roots are consonant final and epenthetic *-a* is inserted late, we see that *Basic* triggers final consonant deletion for

weak words, suffixing does not.

#### 4. There are no similarities between *Basic* MDPs and prefixing or infixing.

4. *the Reduplicant is characterized templatically in prosodic terms:*

(core)(light)(heavy) *syllable, foot, prosodic word.*

But the Reduplicant does not seem to satisfy a template of any sort:

a. Were the template just a CV sequence template we would then not get the simple *moramora, salamalama*, etc. If it is CVCV we get wrong results for *lehibe* (\**lehibehibe* instead of the correct *lehibebe*) and also for e.g. *sarotra* (\**sarorotra* instead of *sarotsarotra*). If CVCVCV then get wrong results for all words with final stress and all with penultimate stress.

b. Replacing CV by  $\sigma$  (a syllable) in (a) above we see that  $R = \sigma$ ,  $R = \sigma\sigma$ , and  $R = \sigma\sigma\sigma$  all yield wrong results.

c.  $R = \text{foot}$ ? Feet are minimally and maximally binary (M&P). Given the absence of long vowels and closed syllables one expects feet in Malagasy to be disyllabic. But we can with Erwin count "moras" and accept that *cv* is bimoraic when *v* is a diphthong. This gives

correct results for *indray*  $\Rightarrow$  *indraindray* and *jay*  $\Rightarrow$  *jaijay*, as well as *salama*  $\Rightarrow$  *salamalama* and *ontany*  $\Rightarrow$  *ontanintany* assuming that non-diphthongs are short.

But then we fail to predict the existence of reduplicated forms for final stress words: *py*  $\Rightarrow$  *pipy*, *ome*  $\Rightarrow$  *omeme*, *ampo*  $\Rightarrow$  *ampompo* (as well as those ending in stressed *e*: *be*  $\Rightarrow$  *bebe*, *lehibe*  $\Rightarrow$  *lehibebe*). Also the  $n \geq 3$  syllable weak words are not cleanly generated: *tahotra*  $\Rightarrow$  *taho-tahotra*, *rafitra*  $\Rightarrow$  *rafi-drafitra*, where we seem to have copied three syllables. (But Erwin would derive the final *a* by an epenthesis rule which applies after moraification and so doesn't count for weight, so e.g. *rafitra* is "really" just the two syllable = one foot *rafet*).

d.  $R = \text{PrWd}$ ? This seems the most plausible. Crucially  $R$  must begin with a (primary) stressed syllable and thus satisfy a major condition **PrWd(1)** for being a prosodic word. It also satisfies **PrWd(2) – PrWd(4)**.

**BUT** 1. The Reduplicant clearly fails **PrWd(5)**: voiceless prenasalized consonants can initiate  $R$ .

(18)

<i>ampo</i>	$\Rightarrow$ <i>ampo + mpo</i> ;
<i>ampirina</i> 'sub. order'	$\Rightarrow$ <i>ampiri + mpirina</i> 'be put in order'
<i>fona</i> 'ask forgiveness'	$\Rightarrow$ <i>fo + mpona</i>
<i>hantona</i> 'suspension'	$\Rightarrow$ <i>hanto + nkantona</i>
<i>man[ontany]</i>	$\Rightarrow$ <i>man[ontani + ntany]</i>
<i>antara</i> 'glacial'	$\Rightarrow$ <i>antara + ntara</i>
<i>antonina</i> 'suitable'	$\Rightarrow$ <i>antoni + ntonina</i>
<i>antsambotra</i> (< <i>sambotra</i> ) 'a leap'	$\Rightarrow$ <i>antsambo + ntsambotra</i>

hatsikana 'farce, plaisanterie'	⇒ hatsika + ntsikana
havana	⇒ hava + nkavana
hinana	⇒ hina + nkinana
sasany	⇒ sasa + ntsasany

2. When R is just one syllable, is it a PrWd? Erwin claims Malagasy bans degenerate feet, but I am doubtful. Here are my counts for one syllable words, based on A&M. Grammatical words are in (19), content words in (20).

(19) *sa* 'or (in questions)'; *fa* 'that (complementizer), but'; *na* 'whether', *i* and *ry* 'proper noun articles'; *sy* 'and (phrasal)'; *ny* 'definite article', *no* 'focus particle', *ho* 'future', *ka* 'and so', *ve*, *va* 'question particle', *tsy* 'not', *mba* 'in order to', *ao* 'there+non-visible', *tao* 'there+non-visible+past', *sao* 'lest', *háy/káy* 'exclamation'

(20) *fe* 'thigh'; *fy* 'delicious'; *fo* 'heart'; *be* 'big, many, very'; *ra* 'blood'; *lo* 'rotten, spoiled'; *la* 'refusal', *mby* 'arrived'; *re* 'heard'; *ro* 'sauce'; *to* 'true, just'; *vy* 'metal'; *zo* 'rights'; *tsy* 'steel', *py* 'a blink', *ráy* 'father', *ráy* 'received', *fóy* 'abandoned', *vóy* 'action of rowing', *hóy* 'is said', *tóy* 'like', *ndre/ndry* 'interjection of surprise or pain', *táy* 'excrement', *máy* 'burnt, hurried', *láy* 'tent', *mbáy* 'step aside', *váy* 'a boil', *ndáo* 'let's go', *jáy* 'pride', *jáo* 'big; a big steer with long horns'

Thus Malagasy presents  $48 = 18 + 30$  one syllable words out of 175 possible ones (7 of the form V: 4 vowels, 3 diphthongs; the latter in need of further study) and  $24 \times 7 = 168$  of the form CV ( $24 = 29$  consonants less 5 prenasalized ones which do not begin words). So 27.4% of the possible one syllable words are actual. So let's ban the ban.

Actually these data just show that there are real words that do not contain a foot. If we assume the Prosodic Hierarchy (but see Erwin) then many of the single syllable Rs are not feet and so a fortiori not PrWds, even if they are in some other (ordinary) sense words. If we don't require PrWds to be feet then the monosyllabic Rs are not at least immediately a counterexample to the claim that  $R = \text{PrWd}$ .

But ignoring our first objection above, even if Rs are PrWds they are not minimal ones, since they can be two or more syllables long and a minimal PrWd can be just one syllable long. So *lama* in *salama-lama* is not a minimal PrWd, nor is *tahotra* in *taho-tahotra*. Whence a requirement that R be a PrWd leaves a lot of room open. (But it does commit us to something). Thus it will not distinguish the following:

(21) (a) *salàma + láma*    (b) *salà + maláma*    (c) *sa + làmaláma*

Note that these forms are the same string, they just differ wrt which parts are the Base and Reduplicant. In general,

**5. A reduplicated form deconcatenates into a part which matches the Stem and a partial copy of the Stem.**

This generalization holds for the examples in (22) where R matches the Stem in the first four cases and the Base matches it in the last case (modulo stress mismatch):

- (22) táhotra 'fear' ⇒ tàho + táhotra  
 (mi)pétraka 'sits' ⇒ (mi)pètra + pétraka  
 tápaka 'broken' ⇒ tàpa + tápaka  
 námána 'friend' ⇒ nàma + námána  
 salàma 'healthy' ⇒ salàma + láma

But in (23) and (24) neither B nor R match S mod stress (syllabification marked)

- (23) hái.nga.na ⇒ hàì.nga.+ nkái.nga.na  
 lá.vi.tra ⇒ là.vi. + dá.vi.tra  
 fá.nta.tra ⇒ fã.nta.+ pá.nta.tra  
 zá.va.tra ⇒ zà.va. + já.va.tra  
 sí.tra.na ⇒ sì.tra. + ntsí.tra.na  
 vé.lo.na ⇒ vè.lo. + mbé.lo.na  
 ré.sa.ka ⇒ rè.sa. + dré.sa.ka

- (24) latábatra 'table' ⇒ latàba + tábatra  
 làvaràngana 'verandah' ⇒ làvarànga + ndràngana  
 karátsaka 'rustling (leaves)' ⇒ karàtsa + drátsaka  
 satrótroka 'swelling of face' ⇒ satròtro + trótroka  
 sòmarítaka 'preoccupied' ⇒ sòmarita + drítaka

In short, a reduplicated form cannot in general be deconcatenated into a part that matches the Stem and another (partial) copy of itself.

6. *Stem-Base identity universally outranks Base-Reduplicant and Stem-Reduplicant identity:  $I(S,B) > I(B,R), I(S,R)$*

This generalization correctly picks (21a) over (21b) and (21c) above. But the reduplication of 3 syllable weak roots whose initial consonant is invariant under *stop* yield forms as in (22) in which R is an exact copy of Stem, Base being truncated. Also problematic here is finding what prevents reduplications like

táhotra ⇒ tàhotra-táho

The primacy of  $I(S,B)$  would seem to favor the righthand side above over the correct *tàho-táhotra*. One is inclined to say that  $\text{RightAnchor}(S,R) > \text{RightAnchor}(S,B)$ , but this is just a particular case of the violation of (6). Another natural reaction based on knowledge of the language is that adjacent unstressed syllables are unacceptable in the output of reduplications. But they are not otherwise bad. Indeed suffixing creates such stress gaps in reduplicated forms:

(m)ipètrapétraka + CIRC ⇒ ipètrapetráhana

Worth emphasizing here is that there is nothing independently bad about weak syllables occurring word internally:

- ka- alàkamísy 'Thursday'; takálo 'exchange; akánjo 'clothes
- na- (m)ánana 'has'; ánatra 'moral, lesson'
- tra- fátratra 'exceedingly'; (m)ipétraka 'sits'

*Interim conclusion* A convincing OT analysis of Red in Malagasy remains to be given.

### References

- Abinal et Malzac. 1888. *Dictionnaire malgache-français* Editions Maritime et D'outre-mer. Paris. edition of 1963.
- Carrier-Duncan, J. 1984. Some Problems with Prosodic Accounts of Reduplication. Ch. 13 in M. Aronoff and R. Oehrle eds. *Language Sound Structure*.
- Dahl, O. C. 1951. *Malgache et Maanjan* Edege-Instituttet, Oslo
- Erwin, S. 1996. Quantity and moras: an amicable separation. in M. Pearson and I. Paul (eds) *UCLA Occasional Papers in Linguistics* No. 17. pp.2–31.
- Finer, D. 1986-87. Reduplication and verbal morphology in Palauan. in *The Linguistic Review* 6. pp. 99–130.
- French, K. 1988. *Insights into Tagalog Reduplication, Infixation, and Stress from Nonlinear Phonology* Summer Institute of Linguistics and the University of Texas at Arlington.
- Gonda, J. 1950. The Functions of Word Duplication in Indonesian Languages. in *Lingua* Vol. II.2 pp. 170 – 197
- Hayes, B. and M. Abad. 1989. Reduplication and syllabification in Ilokano. in *Lingua* 77. pp. 331–374.
- Keenan, E.L. 1996. Creating Anaphors. ms. UCLA
- Keenan, E.L. and J.P. Razafimamonjy. 1996. Malagasy Morphology: Basic Rules. in M. Pearson and I. Paul (eds) *UCLA Occasional Papers in Linguistics* No. 17. pp.31–49.
- Marantz, A. 1982. Re Reduplication. in *Linguistic Inquiry* 13.3 pp. 435–482.

- Moravcsik, E. 1978. Reduplicative Constructions. in J. Greenberg, ed. *Universals of Human language*, Stanford University Press.
- Paul, I. 1996. The active marker and nasals in Malagasy. in M. Pearson and I. Paul (eds) *UCLA Occasional Papers in Linguistics* No. 17. pp 49 - 58.
- McCarthy, J. and A. Prince. 1995. Faithfulness and Dupuplicative Identity. in *Papers in Optimality Theory* U. Mass. Occasional Papers 18, J. Beckman, L. Walsh Dickey & S. Urbanczyk (eds) GLSA UMass, Amherst
- Moravcsik, E. 1978 . Reduplicative Constructions. in *Universals of Human Language* J. Greenberg (ed) Stanford University Press.
- Rahajarizafy, A. 1960. *Essai sur la Grammaire Malgache* Imprimerie catholique, Antanimena Tananarive, Madagascar
- Rajaona, S. 1977. *Problemes de Morphologie Malgache* Librairie Ambozontany, Fianarantsoa, Madagascar.
- . 1985. Le Redoublement par suffixation en malgache' in *Hiratra* No. 4, pp. 17 – 48. Département de langue, littérature et civilisation malgache, Université de Madagascar, Antananarivo.
- Rajemisa-Raolison, R. 1971. *Grammaire Malgache*, Librairie Ambozontany, Fianarantsoa, Madagascar
- Rajaona, S. 1980. L'Alternance en Malgache in *Hiratra* No. 2, Université de Madagascar, Antananarivo. pp. 12 – 92
- Rahajarizafy, A. 1960. *Essai de Grammaire Malgache*. Imprimerie Catholique, Antanimena Tananarive.
- Razafimamony, J. P. 1988. *Tsipelin-Teny: Fitsipika sy Olana*. Dingana Voalohany. Dika 10 Centre d'Etudes et de Recherches de Malagasy. Université de Madagascar, Antananarivo
- 1988 Ny Verin-droa (Reduplication). ms DIFP, Université de Madagascar, Antananarivo.
- Sch lindwein, D. 1991. Reduplication in lexical phonology: Javanese plural reduplication. in *The Linguistic Review* 8. pp.97–106
- Steriade, D. 1988. Reduplication and syllable transfer in Sanskrit and elsewhere. in *Phonology* 5 pp. 73 - 155.

Urbach, A. J. 1987. *A Formal Analysis of Reduplication and its Interaction with Phonological and Morphological Processes* PhD Diss. Univ. of Texas at Austin

Wilbur, R.B. 1973. *The Phonology of Reduplication* PhD Thesis, University of Illinois.





## The liver as the location of emotions: Noun incorporation in Kambera

Marian Klamer  
Vrije Universiteit/HIL  
Amsterdam  
klamerm@let.vu.nl

### 1. Introduction

Crosslinguistically, verbal expressions for emotions often contain a special type of noun indicating the location of emotions. In many cases this is a bodypart noun, such as *heart* in the expression *my heart bleeds*; or *nok* ‘throat, neck’ in the Chocktaw verb *noklibisa* ‘to be neck-heated: to be in a passion’. In the English example the body part noun is a separate constituent, in Chocktaw it is incorporated into the verb (Mithun 1984:879).

In the last decade there has been a debate about the analysis of noun incorporation: is it a lexical compounding process, the result of lexical word formation (Rosen 1989 and others), or a syntactic process where compounding is the result of syntactic head movement (Baker 1988, 1996 and others)? (For an overview of the various positions I refer to Spencer 1995, Mohanan 1995, Gerds 1997). Various authors have proposed various diagnostics to distinguish lexical from syntactic noun incorporation. The parallel existence of two functionally and formally similar structures, one where the noun is an independent constituent (a ‘phrasal’ construction) and one where it is ‘incorporated’, is often taken to suggest that the incorporated construction is syntactically derived from the phrasal construction by head movement of the noun into the verb.

In this paper I discuss the derivation of emotion expressions in Kambera. Kambera, classified as belonging to the Central Malayo-Polynesian subgroup of Austronesian languages, is spoken by 150,000 speakers on the island of Sumba in Eastern Indonesia. In this language, productively derived expressions for emotions contain the body part noun *eti*, as illustrated in (1):<sup>1</sup>

(1)	Kambera	literal translation <sup>2</sup>	free translation
	<i>bàrang eti</i>	‘(have a) pounding liver’	‘be worried’
	<i>mila eti</i>	‘(have a) poor liver’	‘have compassion’
	<i>hàmu eti</i>	‘(have a) good liver’	‘be good’
	<i>bàkul eti</i>	‘(have a) big liver’	‘be happy/relieved’
	<i>nggauru eti</i>	‘(have a) glad liver’	‘be glad’
	<i>jangga eti</i>	‘(have a) high liver’	‘be arrogant/haughty’
	<i>kendar eti</i>	‘(have a) humble liver’	‘be humble/meek’
	<i>kudu eti</i>	‘(have a) small liver’	‘be disappointed’
	<i>mbaha eti</i>	‘(have a) wet liver’	‘be pleased’
	<i>mbata eti</i>	‘(have a) broken liver’	‘be shattered’
	<i>mbana eti</i>	‘(have a) hot liver’	‘be fervent/very angry’
	<i>ha.nduka eti</i>	‘(have a) troubled liver’	‘be sad’
	<i>ka.lài eti</i>	‘(have a) wrong/left liver’	‘be disappointed’
	<i>ka.dua eti</i>	‘(have) two livers’	(1) ‘be in two minds’

<sup>1</sup> Notational conventions: the accent on *à* marks a phonemic contrast between short *à* and long *a*. This contrast is only valid in the first, stressed, syllable of the root; it is neutralised in the second syllable of the root, in affixes and in clitics. Morpheme boundaries within words are marked by [·], the attachment of clitics to their host is marked by [-]. Glossing conventions: ‘1s.Subj’, etc. marks the person, number and grammatical function of a pronominal clitic, ‘App’=Applicative, Art= ‘Article’, ‘Cau’= Causative, ‘Cnj’=Conjunction, ‘Dim’= Diminutive, ‘Emp’= Emphasis, ‘Impf’= Imperfective, ‘Loc’= Locational preposition, ‘Mod’= Mood, ‘Poss’= Possessive, ‘Prf’= Perfective, ‘Rel’= Relative clause marker.

<sup>2</sup> Kambera does not have a category of adjectives: adjectival concepts are expressed by intransitive verbs. The literal translation is thus not entirely literal. Note that the verb in the literal translation could just as well be the verb *be*: *bàrang eti* ‘(be) pounding-livered’, *mila eti* ‘(be) poor-livered’, *mbaha eti* ‘(be) wet-livered’, etc.

ka.leka eti	'(have a) crooked liver'	(2) 'know right from wrong'
ka.ràu eti	'(have a) dark liver'	'be untruthful/dishonest'
ma.nganga eti	'(have a) stealing liver'	'be angry'
pa.ngàndi eti	'take livers'	'be greedy'
		'be touching/sweet (said of a person)'

The emotion expressions feature in two parallel constructions: one is a phrasal construction in which the body part noun forms an independent NP, and the other is an incorporated construction where the noun has become part of the verbal constituent. Both constructions are semantically (virtually) identical and are illustrated in (2) and (3).

In (2) the body part noun *eti* 'liver' is part of the NP *na eti-na na maramba nuna* 'the liver of that king', which is crossreferenced as the subject *-nanya* of the verb. In (3) the noun *eti* 'liver' is incorporated into the verbal constituent, while its possessor, the NP *na maramba nuna* 'that king' is left 'stranded'. This possessor NP is crossreferenced as the subject of the derived verb.

- (2) Mbaha -nanya<sub>j</sub> -ka [[na eti -na] [[na maramba] [nuna]]]<sub>NP<sub>j</sub></sub>  
 be.wet -3s.Subj -Prf Art liver -3s.Poss Art king that.one  
 'That king was feeling pleased' (lit. 'The liver of that king (was) wet')
- (3) Mbaha eti -nanya<sub>j</sub> -ka [[na maramba] [nuna]]<sub>NP<sub>j</sub></sub>  
 be.wet liver -3s.Subj -Prf Art king that.one  
 'That king was feeling pleased' (lit. 'That king (was) wet-livered')

The aim of this paper is to determine how the incorporation in (3) is related to the phrasal construction in (2). I will argue that, despite the parallel synchronic occurrence of (2) and (3), the construction in (3) is not derived from (2) by syntactic head movement and incorporation but by a lexical compounding process that manipulates the lexical conceptual structure of (2).

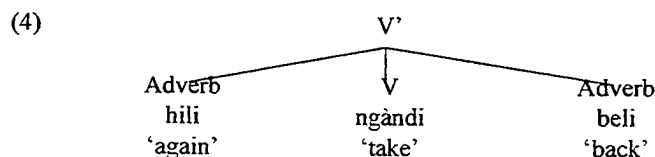
The paper is organised as follows: in order to characterise the unique position of the incorporated noun *eti* in (3), I start in section 1 with a brief discussion of the positions of argument NPs in Kambera. In section 2 the lexical properties of the incorporation of *eti* are discussed. If *eti* incorporation indeed takes place in the lexicon, the logical possessor of *eti*, *na maramba nuna* 'that king' in (3) is not a 'stranded' nominal modifier. In section 3 I present some language-internal arguments why this constituent should be analysed as a locative argument instead. Section 4 summarises the conclusions.

## 2. Incorporation in Kambera

### 2.1 Introduction: the various positions of verbal arguments in Kambera

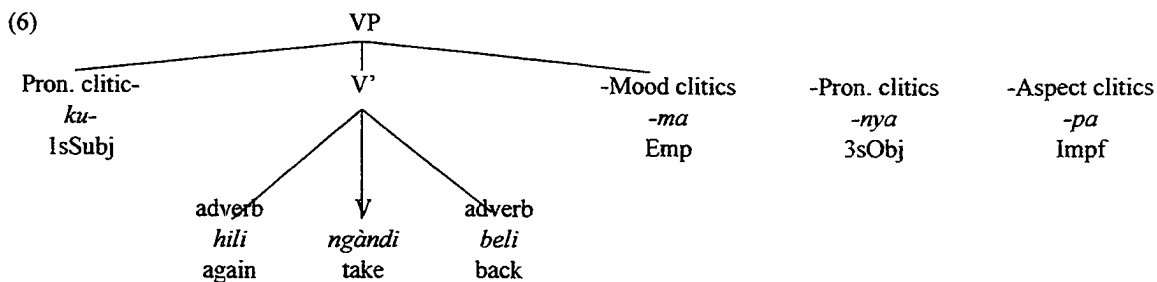
In this section I will sketch the various positions of verbal arguments in Kambera, with the aim to show that the position of incorporated *eti* in (2) -- immediately adjacent to the verb -- is not a position that is not normally available for verbal arguments. Below I assume that the pronominal clitics which are attached to the verbal projection express the verbal arguments, while the NPs that optionally double them are adjuncts (Klamer 1994, to appear). With respect to verbal complement NPs there is a difference in the grammatical status of definite and indefinite objects: definite object NPs are crossreferenced on the verb and generally resemble adjuncts, while indefinite ones are not crossreferenced and occupy an argument position. Subjects, on the other hand, are crossreferenced irrespective of their definiteness. Indefinite subject NPs only feature in very specific contexts and they are not incorporated.

Kambera is a head-marking (Nichols 1986) language, in which the clause is an endocentric projection of the verb. The core of a Kambera sentence is a verbal constituent consisting of a verb plus maximally two adverbial modifiers; all of them separate morphological and prosodic words. This is the phrasal category V', as illustrated in (4).

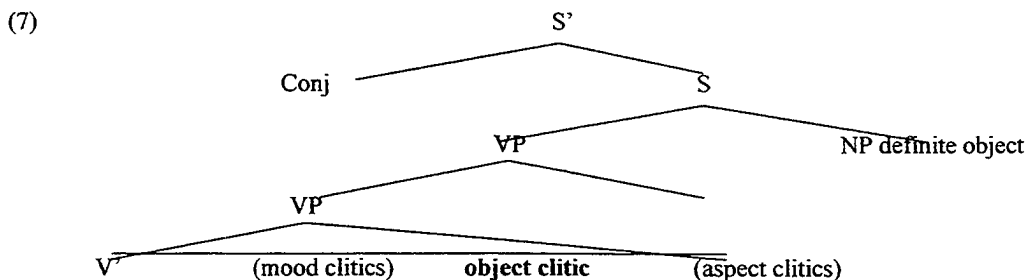


Note that the Kambera  $V'$  differs from a  $V'$  in a language like English: a Kambera verb does *not* form a constituent with its complement but rather with its modifier(s). The verb and its complement are not structurally adjacent and never behave as one constituent, in contrast to the verb and its modifier(s) (see below; and see Mohanan (1995:90-91) for a similar view on the  $V'$  in Hindi). A Kambera VP also differs from what is commonly assumed to be a VP: it consists of  $V'$  and an elaborate clitic cluster containing mood, pronominal and aspectual clitics (Klamer 1997, to appear). As this constituent can form a clause on its own, I refer to it as the 'nuclear clause':

- (5) Ku- hili ngàndi beli -ma -nya -pa  
 1sSubj- again take back -Emp- 3sObj -Perf  
 'I'll bring it back again'



The definiteness of an NP is grammatically expressed by the presence of an article (*na* singular, *da* plural, *i* proper noun), indefiniteness by the absence of an article. The pronominal clitics function to crossreference the subject NP and the definite object NPs, i.e. definite object NPs double their object clitics, indefinite object NPs do not. The NP, not the clitic, is optional and used for emphasis and/or disambiguation. In addition to this, definite NPs may scramble rather freely (see Klamer 1994, to appear). The optional presence and relatively free word order of definite NPs, combined with the obligatory presence of pronominal clitics, suggest that the pronominal clitics express the verbal arguments and the NPs, who are in apposition to the clitics, are adjoined to VP. As represented in (7), NPs are *not* part of the nuclear clause (VP, see (6)) though they *are* within the scope of clausal conjunctions, i.e. they are not topicalised or dislocated constituents. The canonical position of the definite object NP is after the nuclear clause, as in (7), but it may scramble to other positions.



In contrast to definite object NPs, indefinite object NPs are not crossreferenced on  $V'$ : they do not double an object marking clitic, as the sentences in (8) show. The object NPs in these sentences contrast in definiteness (cf. the article *na* in (8b));<sup>3</sup> the indefinite NP in (8a) is not crossreferenced, while the definite NP in (8b) is crossreferenced with the object clitic *-ya*.

- (8) a. Na- tàru nàhu angu kotak -na  
 3sSubj- watch now companion village-3sPoss  
 'He watches (a) fellow villager(s) of his'

<sup>3</sup> Note that possession and definiteness are grammatically distinct in Kambera: if a possessed noun is not preceded by an article it is indefinite.

- b. Na-      t̄aru      -ya      n̄ahu      [na    angu                  kotak    -na]  
 3sSubj-   watch    -3sObj   now      Art    companion        village-3sPoss  
 'He watches his fellow villager'

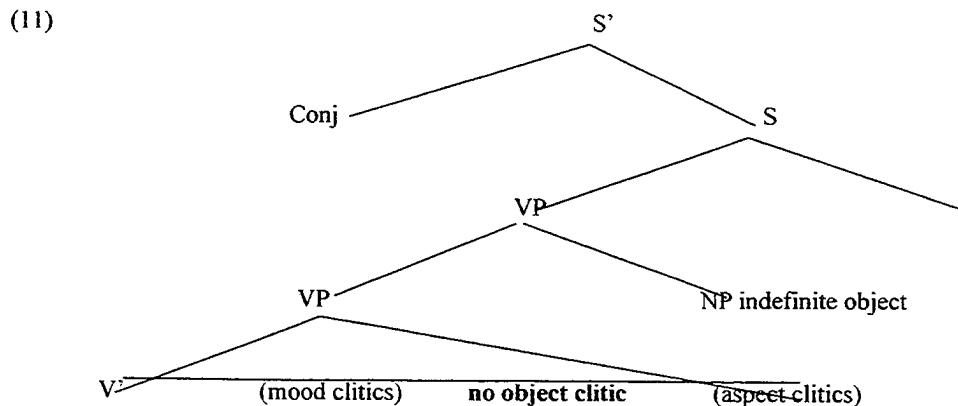
Unlike definite NPs, indefinite object NPs are obligatorily present, cf. (9a) and (9b):

- (9) a. Na-      t̄aru      -ya      n̄ahu  
 3sSubj-   watch    -3sObj   now  
 'He watches him'
- b. \* Na-      t̄aru      n̄ahu  
 3sSubj-   watch    now  
 'He watches (someone)'

Indefinite object NPs occur linearly closer to the verb than definite NPs: they must remain within the scope of the locational/directional PPs (compare 10a,b), while definite NPs may scramble over a PP (compare 10c,d):

- (10) a. T̄u    uhu      [la mbola]<sub>PP</sub>  
 put    rice      Loc    basket  
 'Put rice in the basket'
- b. \* Tu    [la    mbola]<sub>PP</sub>    uhu  
 put    Loc    basket      rice
- c. T̄u-ya                  na    uhu      [la    mbola]<sub>PP</sub>  
 put-3sObj            Art    rice      Loc    basket  
 'Put the rice in the basket'
- d. T̄u-ya      [la mbola]<sub>PP</sub>    na    uhu  
 put-3sObj    Loc basket      Art    rice  
 'Put the rice in the basket'

The distinct distribution and crossreference of definite and indefinite NPs show that there is a difference in their grammatical status: definite NPs are clausal adjuncts, while indefinite object NPs occupy an argument position within the VP at the expense of an object marking clitic, as represented in (11):



In other words, a Kambera nuclear clause or VP *minimally* consists of a verb, a subject clitic and (in case of a transitive verb) *either* a bound pronominal object --an object clitic-- *or* an indefinite object NP.

Finer (1994:162) accounts for the syntactic differences between indefinite and definite object NPs in the western Austronesian language Selayarese by analysing the indefinites as the incorporated counterparts of the definites. If we use 'incorporation' in the standard sense, as a process where the verb and the noun are combined to become one verb stem (cf. Mithun 1984), there is no evidence to apply a similar analysis to the

Kambara indefinite objects. Firstly, the indefinite object never surfaces adjacent to the verb, as illustrated in (12):

- (12) a. Na-      tàru      -bia      nàhu      angu      -na  
          3sSubj- watch   -Mod      now      companion   -3sPoss  
          ‘He just watches his companions’
- b. \* Na-      tàru      angu (-na)      -bia      nàhu  
       3sSubj- watch   companion (-3sPoss)   -Mod      now

Moreover, there is no evidence that an object NP is ever a sister of the verb because adverbs, subject pronominal clitics and mood clitics have a fixed position between the verb and its complement. An illustration is given with the mood clitic *-bia* in (13) and the subject enclitic *-na* in (14):

- (13) a. Mu-      himbu   -bia      ana      pangangu.  
          2sSubj- search   -Mod      Dim      food  
          ‘You just go find some food’
- b. \* Mu-      himbu      ana      pangangu   -bia  
       2sSubj- search      Dim      food      -Mod
- (14) a. Ngàndi      -na      -nya      ana      pangangu  
          take      -3sSubj      -3sObj      Dim      food  
          ‘He brought him some food’
- b. Ngàndi      ana      pangangu      -na      -nya  
          take      Dim      food      -3sSubj      -3sObj

Secondly, if we assume that the indefinite object is the productively incorporated counterpart of the definite one, we expect modifiers of the definite NP to be able to strand. This is not the case: the modifying article *na* of the definite NP in a sentence like (15a) cannot be stranded, as in (15b):

- (15) a. Tú      -ya      na      uhu      la      mbola  
          put      -3sObj      Art      rice      Loc      basket  
          ‘Put the rice in the basket’
- b. \* Tú      uhu<sub>j</sub>      na      t<sub>j</sub>      la      mbola  
          put      rice      Art           Loc      basket

Thirdly, the incorporation of an object into a transitive verb is a valency reducing operation, but Kambara verbs with indefinite objects are still transitive: in causative and applicative derivations they pattern with the transitive verbs rather than with the intransitives (cf. Klammer 1994, to appear: chapter 6). (Passivization cannot be used as a test because Kambara has no passive, cf. Klammer 1996).

We conclude that there is no reason to assume that indefinite object NPs are incorporated into the verb or into the verbal complex *V'*. Rather, they are part of a larger constituent, the VP. Indefinite definite object NPs in Kambara are not incorporated complements; they differ from definite object NPs in having argument rather than adjunct status.

Turning now from object NPs to the NPs marking the single argument of an intransitive verb, we find that in the few intransitive clauses where a subject NP *is* present (in most cases the NP is omitted because the referent is known from the (situational/discourse) context, this NP is postverbal and definite:

- (16) Ba      tàka      -nanya<sub>j</sub>      -ka      [na      mbapa      -nggu]<sub>NPj</sub>  
       Cnj      arrive      -3sSubj      -Prf      Art      husband      -1sPoss  
       ‘Because my husband was arriving’

Indefinite subject NPs do occur, but I have found them only with transitive verbs. Indefinite subject NPs are crossreferenced on the verb, unlike indefinite object NPs:<sup>4</sup>

- (17) Na- ngàndi -ha iwa  
3sSubj- take -3pObj flood  
'They were taken (away) by a flood' (lit. 'It took them a flood')
- (18) Lapi -na -nya tau  
cheat -3sSubj -3sObj person  
'He was cheated on by others' (lit. 'They cheated on him (some) people')

Indefinite subject NPs are never incorporated, neither as intransitive arguments (19a) nor as transitive agents (19b):

- (19) a. \* Ba jorong mbapa-nggu (-nanya<sub>j</sub>) -ka  
Cnj topple husband-1sPoss -3sSubj -Prf
- b. \* Lapi tau -na -nya  
cheat person -3sSubj -3sObj

To summarise the findings of this section, I have given a sketch of the the distinct grammatical status of Kambera object NPs: definite object NPs resemble adjuncts, while indefinite ones occupy an argument position. I also presented some evidence against the idea that the indefinite objects are incorporated counterparts of the definite ones. With respect to subjects we observed that both indefinite and definite subjects are crossreferenced, but, unlike object clitics, a subject clitic is seldom doubled by an NP. If it is, the NP is usually definite. Indefinite subject NPs only occur in very specific contexts. They are not incorporated. From this overview of Kambera sentence structure, we conclude that the noun *eti* in (2) occupies a position that is not normally available for verbal arguments. Additional evidence is presented in the next section.

## 2. 2. The incorporation of the body part noun *eti*

In this section we take a closer look at the incorporation of Kambera *eti*. First I discuss the lexical characteristics of the incorporation. Then I show that a syntactic account would account for only part of the data.

The first indication that *eti* incorporation is not a productive syntactic process is that it is the only noun that can occupy the position adjacent to V. In the previous section we saw that the incorporated position of *eti* is not available for verbal argument in general; in (20) and (21) we see that nominals occurring in idiomatic expressions cannot be incorporated either. Apart from *eti* no other body part noun can be incorporated, cf. *katiku* 'head' in (20); and neither can idiomatic objects such as *ahu* in *palu ahu* 'kill a dog' in (21).

- (20) a. Kawunggur -nanya<sub>j</sub> [na katiku -na]<sub>j</sub>  
be.dizzy -3sSubj Art head -3sPoss  
'He was becoming confused'
- b. \* Kawunggur katiku -nanya -ka  
be.dizzy head -3sSubj -Prf

<sup>4</sup> Only definite NPs are optional and used for emphasis or disambiguation. Therefore, we expect that a construction where an indefinite subject NP doubles a subject clitic is used in specific contexts and for specific purposes only. The sentences in (17) and (18) are indeed exceptional because here the subject is explicitly defocused, its status comparable to a logical subject in the English passive construction. Unlike the English passive construction, constructions like (17)-(18) are used very seldom in Kambera -- dropping or scrambling NPs are common ways to (de)focus arguments. The relative rare use of indefinite subject NPs may also explain why my corpus does not contain intransitive clauses with such an NP doubling a clitic.

- (21) a. Palu        -na            -ngga            ahu  
 hit        -3sSubj        -1sObj(App)     dog  
 'He kills a dog for me' (lit.: 'He hits a dog for me')
- b. \* Palu            ahu            -na            -ngga  
 hit            dog            -3sSubj        -1sObj(App)

Thus, the incorporation of *eti* is unique in the language, and so is its position adjacent to the verb.

The second indication of the lexical nature of this incorporation is that *eti* has no referential function in either of the constructions. That *eti* is non-referential can be inferred from the fact that its possessor cannot be questioned. Normally, Kambera question words appear *in situ*, as illustrated in (22), where the nominal possessor is questioned:<sup>5</sup>

- (22)        Tàka        -nanya<sub>j</sub>    -ka            [[na mbapa-na]            [i nggamu]]<sub>j</sub>?  
 arrive    -3sSubj    -Prf            Art husband -3sPoss        Art who  
 'Whose husband has arrived?'

When we question a possessor of incorporated *eti*, such a question renders a nonsensical interpretation:

- (23)        ?? Na-bàrang            [[na eti -na]            [i nggamu]]<sub>k</sub> ?  
 3sSubj-pound            Art liver -3sPoss        Art who  
 Intended idiomatic reading: \* 'Who was worried?'<sup>6</sup>  
 Literal interpretation: ? 'Whose liver was pounding?'

In a syntactic analysis à la Baker (1988, 1996), noun incorporation takes place for reasons of case. Such an analysis implies, among other things, (i) that the process is productive and obligatory, (ii) that the modifier(s) of an incorporated noun can be left stranded, (iii) that the incorporated noun is a verbal complement at some structural level, (iv) that a stranded possessor is 'raised to subject' and (v) that object incorporation must feed applicative formation. In what follows I will consider these implications and argue that none of them applies to Kambera *eti* incorporation.

In the previous section we have seen that the phrasal construction (2) is used alongside the incorporated one in (3), while both have the same semantics. (In fact, the phrasal construction seems to be the preferred one: in my corpus of spontaneous Kambera speech the phrasal expressions for emotions far outnumber the incorporated ones.) Thus, the incorporation of *eti* does not obligatorily take place for reasons of case. Below I discuss additional evidence from the applicative derivations of emotion verbs.

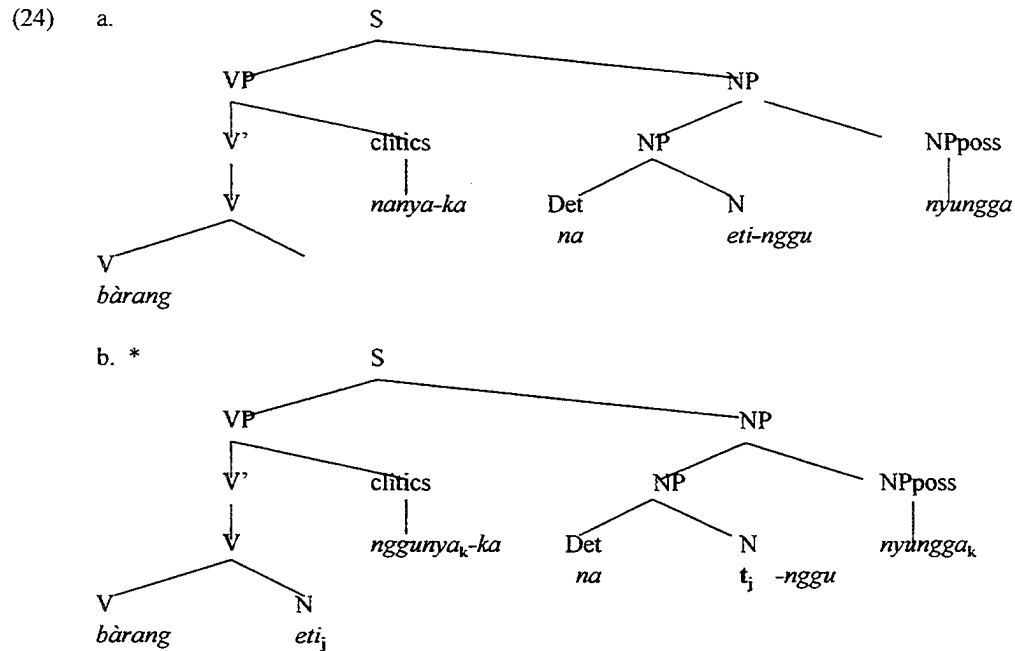
The diagrams in (24) show how *eti* incorporation by syntactic head movement would take place for the sentences in (25). Diagram (24a) represents the phrasal construction given in (25a), diagram (24b) shows the syntactic incorporation of *eti*. In the phrasal construction the subject clitic crossreferences the NP expressing the location of the emotion, *na eti-nggu nyungga* 'my liver'. In the incorporated construction *eti* is moved out of the possessed NP to form a compound with the verb and the subject clitic crossreferences the possessor of *eti*, the pronoun *nyungga* 'I'. The fact that the possessor is crossreferenced as the subject of the derived predicate is accounted for by assuming that as a result of the incorporation, the possessive modifier of *eti* is stranded and reinterpreted as a verbal argument ('possessor raising to subject').

<sup>5</sup> Kambera does not allow question word-extraction:

\*        Nggamu<sub>j</sub>        tàka-na            na        mbapa-na            t<sub>j</sub>?  
 who            arrive-3sPoss        Art        husband-3sPoss

<sup>6</sup> The argument of a verbal compound with *eti* can be questioned using a relative structure:

Nggamu        na        ma-        bàrang        eti?  
 who            Art        Rel-        pound        liver  
 'Who is/was worried?'



The syntactic incorporation in (24b) is ungrammatical. The sentence is repeated in (25b). Comparison of (25b) with the correct form in (25c) shows that Kambera does not allow the article and possessive enclitic to be stranded.

- (25) a. B̄arang      -nanya   -ka      na   eti   -nggu   nyungga  
 pound      -3sSubj   -Prf      Art   liver   -1sPoss   I  
 'I am/was worried' (lit. 'My heart is/was pounding')
- b. \* B̄arang      eti<sub>j</sub>   -nggunya<sub>k</sub>   -ka      na   t<sub>j</sub>   -nggu   nyungga<sub>k</sub>  
 pound      liver   -1sSubj   -Prf      Art        -1sPoss   I
- c. B̄arang      eti   -nggunya<sub>k</sub>   -ka      nyungga<sub>k</sub>  
 pound      liver   -1sSubj   -Prf      I  
 'I am/was worried' (lit. 'I was pounding-hearted')

Is *eti* a verbal complement at some level? The intransitive base verbs in (1) are all non-active verbs; and the thematic role of *eti* non-agentive. There is no pragmatic reason why concepts like 'be cheerful/happy' should not be expressed with a combination of an active intransitive verb and *eti*. Yet, such combinations do not occur:

- (26) \* ludu eti      Intended reading: 'be cheerful'      (lit. 'have a singing liver')  
 \* riki eti      Intended reading: 'be happy'      (lit. 'have a laughing liver')

In other words, the incorporated noun *eti* resembles a canonical object in that it cannot have an agentive interpretation. However, we cannot account for this semantic restriction by formulating a condition in structural terms such as 'external/internal argument' or 'subject/object'. The forms in (27), where *eti* is combined with a transitive verb, are evidence that it is irrelevant whether *eti* is an external argument/subject (27a), or an internal argument/object:

- (27) manganga eti      'be greedy'      (lit. 'have a stealing liver')  
 --> *eti* is external argument/subject
- pangãndi eti      'be touching/sweet'      (lit. 'capture other people's liver')  
 --> *eti* is internal argument/object



The conclusion is that because in *eti* incorporation internal and external arguments are treated identically, a syntactic analysis for these derivations is excluded. The data suggesting this are admittedly few, and we could assume a lexical derivation for *manganga eti* ‘be greedy’ in (27), while maintaining a syntactic one for *pangàndi eti* and the other derived forms. But even then, we would still have no account for the fact (discussed above) that Kambera verbal complements/arguments are not generally incorporated, while *eti* is. To account for this, *eti* would have to be marked in the lexicon as the only exception.

The syntactic account also implies that the modifier(s) of an incorporated noun can be left stranded and that a stranded possessor is ‘raised to subject’. Superficially, this seems to be the case in Kambera. However, in (24) and (25) we saw that both the possessor enclitic *-nggu* and the definite article *na* are present in the phrasal construction but absent in the incorporation: they are neither incorporated along with *eti*, nor left ‘stranded’. In addition, neither transitive objects, nor intransitive arguments undergo productive nominal incorporation followed by possessor raising. The sentences in (28) illustrate this for a possessed object:

(28) a. No incorporation with possessor stranding:

*	Palu	ana <sub>k</sub>	-nanya	t <sub>k</sub>	nyungga
	hit	child	-3sSubj		I

b. No possessor raising to object:

*	Palu	-na	-ngga <sub>j</sub>	ana-nggu	(nyungga <sub>j</sub> )
	hit	-3sSubj	-1sObj	child-1sPoss	I

Good for: ‘My child hits me’

c. No object incorporation and possessor raising:

*	Palu	ana <sub>k</sub>	-na	-ngga <sub>j</sub>	t <sub>k</sub>	(nyungga <sub>j</sub> )
	hit	child	-3sSubj	-1sObj		I

The sentences in (29) show that the same applies to the argument of a (non-active) intransitive verb:

(29) a. No incorporation with possessor stranding:

*	Tàka	mbapa <sub>k</sub>	-nanya-ka	t <sub>k</sub>	nyungga
	arrive	husband	-3sSubj-Prf		I

b. No possessor raising to (intransitive) subject:

*	Tàka	-nggunya <sub>j</sub> -ka	na	mbapa-nggu	(nyungga <sub>j</sub> )
	arrive	-1sSubj-Prf	Art	husband -1sPoss	I

c. No incorporation of intransitive argument and possessor raising:

*	Tàka	mbapa	-nggunya	-ka	t <sub>k</sub>	(nyungga <sub>j</sub> )
	arrive	husband	-1sSubj	-Prf		I

Kambera employs another construction where a possessor argument seems to be ‘raised’, illustrated in (30):

(30) a. Base form:

Na-	rasa	-a	-ya <sub>j</sub>	[na	mbotu -mu	nyumu] <sub>j</sub>
3sSubj-	feel	-Mod	-3sObj	Art	weight-2sPoss	you

‘She will notice (it) your weight’

b. Licit: possessed nominal ‘stranded’ and possessor ‘raised’:

Na-	rasa	-a	-kau <sub>k</sub>	[na	mbotu -mu	t <sub>k</sub> ]
3sSubj-	feel	-Mod	-2sObj	Art	weight-2sPoss	

‘She will notice you (by) your weight’

c. Illicit: possessed noun incorporated, possessor raised’

(construction parallel to *eti* incorporation + possessor raising):

*	Na-	rasa	mbotu <sub>j</sub>	-a	-kau <sub>k</sub>	[ t <sub>j</sub> t <sub>k</sub> ]
	3sSubj-	feel	weight	-Mod	-2sObj	

The ‘possessor raising’ construction in (27b) differs from the ‘possessor raising’ associated with the incorporation of *eti* in that the ‘stranded’ nominal is the possessee rather than the possessor. If we assume for the moment that this construction indeed involves possessor raising, it shows that in Kambera a complement noun does not need to be incorporated for its possessor to be raised.<sup>7</sup> In other words, possessor raising and noun incorporation are two formally unrelated processes in Kambera; and the possessor is not necessarily part of the complement NP. The alternative option that the possessor NP is a separate constituent is discussed in section 3.

Finally, let us look at the properties of derived emotion verbs when they form the basis of applicative and instrumental verbs. The syntactic analysis predicts that a verb undergoing applicative formation must incorporate its base object before it can be made applicative. As the object case of the derived applicative verb must be assigned to the new (applicative) object, the base object would fail to get case in an applicative construction and being unlicensed, it would remain unexpressed. Hence, in order to surface in an applicative construction it must be incorporated into the verb before applicative formation applies. In other words, noun incorporation feeds applicative formation. This prediction is not borne out in the applicative forms of Kambera emotion verbs.

Intransitive verbs can be made transitive by prefixing the causative morpheme *pa*, as illustrated in (31). This also applies to the intransitive emotion verbs. In (32) the emotion expression *hamu eti* ‘be good’ is causativised and NP containing *eti* is its object.

- (31)            *hàmu*            ‘be good’  
                   *pa.hàmu*        ‘cause X to be good: improve/restore/relieve X’
- (32)            *Na-*            *pa.hàmu*        *-ya* ;            [*na eti -nggu nyungga*];  
                   3sSubj-        Cau-be.good    -3sObj            Art liver -1sPoss I  
                   ‘He relieves my heart’ (lit.: ‘He causes my liver to be good’)

On subsequent applicative formation, the object of the causative verb *pa.hàmu* is not incorporated into the verb, as shown in (33a,b):

- (33) a.            *Na-*            *pa-hàmu*        *-ngga*            *eti*            *nyungga*  
                   3sSubj-        Cau-be.good    -1sObj(App)        liver            I  
                   ‘He makes me happy/relieved’ (lit.: ‘He relieves (for) me (my) liver?’)
- b. \*            *Na-*            *pa-hàmu*        *eti*            *-ngga*            *nyungga*  
                   3sSubj-        Cau-be.good    liver            -1sObj(App)        I

If *eti* incorporation does not feed applicative formation, as shown by (33b), there is no reason to assume that *eti* incorporation takes place for reasons of case, as is assumed in the syntactic approach.

Apart from being causativised and made applicative, emotion verbs can also be the base for an instrumental derivation. In Kambera, instrumental verbs are derived by compounding a transitive or intransitive base verb with the verb *wà(ngu)* ‘use’, as illustrated in (34) (the final syllable *ngu* is present in the infinitive and citation form of the verb, with object marking it disappears, cf. (20b)).

- (34)            *palu*                            ‘hit X’  
                   *palu wa(ngu)*                ‘hit X with/using Y’
- kamakih*                        ‘be embarrassed’  
                   *kamakih wa(ngu)*            ‘be embarrassed about/because of Y’

Verbal complements are normally not incorporated before the derivation of instrumental verbs. This is illustrated for the transitive base verb *palu* ‘hit X’ in (35). The derivation of instrumental *palu wangu* ‘hit X with/using Y’ does not involve the incorporation of the base object noun *tau* ‘person’, nor the instrumental object noun *hurung* ‘spoon’:

<sup>7</sup> The data are also circumstantial evidence that the incorporation of *eti* and the ‘possessor raising’ that subsequently takes place are not triggered by syntactic structure only because then noun incorporation would also happen in (21), i.e. (21b) would be a grammatical Kambera clause, contrary to fact.

- (35) a. Palu wàngu hurung  
eat use spoon  
'Hit (it) with a spoon'
- b. Palu wà-nya<sub>i</sub> hurung [na tau nuna]<sub>i</sub>  
eat use-3sObj spoon Art person that.one  
'Hit that person with a spoon'
- c. \* Palu hurung wàngu (tau)  
hit spoon use person
- d. \* Palu tau wàngu (hurung)  
hit person use spoon

The same is illustrated for the intransitive base verb *kamakih* 'be embarrassed' in (36), where the instrumental derivation does not involve the incorporation of the instrumental object *ana-nda* 'our child(ren)':

- (36) a. Ta- kamakih wàngu ana-nda  
1pSubj- be.embarrassed use child-1pPoss  
'We are embarrassed about our child(ren)'
- b. \* Ta- kamakih ana wàngu  
1pSubj- be.embarrassed child use

In contrast to this pattern of non-incorporation, the derivation of instrumental emotion verbs is fed by incorporation of the noun *eti*, which appears adjacent to the verb in the instrumental construction in (37a). A phrasal construction, where *eti* is part of an independent NP, is not possible for instrumental emotion verbs, as shown in (37b), and neither can *eti* appear as an indefinite NP, as in (37c):

- (37) a. Na- [[jangga eti]<sub>v</sub> wà] -nda  
3sSubj- be.high liver use -1pObj  
'He behaves arrogantly towards us' (lit. 'He has a high liver with us')
- b. \* Na- jangga wà -nda na eti -na  
3sSubj- be.high use -1pObj Art liver -3sPoss
- c. \* [Na- jangga wà -nda eti  
3sSubj- be.high use -1pObj liver

These restrictions apply to all the instrumental emotion verbs illustrated in (38):

- (38) kudu eti wà(ngu) 'be disappointed with Y' ('have a small liver with Y')  
bata eti wà(ngu) 'be shattered because of Y' ('have a broken liver with Y')  
jangga eti wà(ngu) 'be arrogant because/towards Y' ('have a high liver with Y')  
hàmu eti wà(ngu) 'be happy together with Y' ('have a good liver with Y')  
karau eti wà(ngu) 'be angry because of Y' ('have a dark liver because of Y')

The conclusion is that, again, *eti* does not behave like a common verbal complement and that *eti* incorporation is a process with lexical characteristics.

### 2.3 A lexical analysis of *eti* incorporation

At first glance, the form of the Kambera *eti* incorporations suggest a syntactic derivation because both the phrasal and the incorporated construction are used in parallel: *eti* incorporation is semantically vacuous, *eti*

is always a non-active (theme/patient) argument (of mostly inactive intransitives), and after *eti* is incorporated, its possessor may be left stranded and ‘possessor raising’ seems to apply.

However, by looking at the language in more detail and by considering other traits of the incorporated construction, we saw that *eti* incorporation is a lexical derivational process. In a lexical approach to noun incorporation (as advocated in e.g. Spencer 1995), noun incorporation involves an operation on the structure of a predicate and its argument(s) (Predicate Argument Structure) and/or the semantic content of those arguments (Lexical Conceptual Structure). PAS is linked to the appropriate semantic values at LCS by default mapping in terms of a semantic hierarchy (Jackendoff 1990, Grimshaw 1990).

The standard assumption is that PAS is structured, and distinguishes at least the following argument positions: external argument (canonically the subject though not all verbs that surface with a subject have an external argument), direct internal argument (canonically the direct object) and indirect internal argument (second object, location, oblique).

In a lexical approach to noun incorporation, a nominal stem can identify or saturate an argument of a verb, so that this position is no longer available in syntactic structure. In the Kambera case of *eti* incorporation, the argument which is saturated by *eti* is the single (theme) argument of a stative intransitive verb. There is a second location argument:

(39) Lexical conceptual structure of Kambera stative intransitive verbs

LCS:	V	THEME	LOC
------	---	-------	-----

We have seen that Kambera emotions are expressed by a phrasal and an incorporated construction that occur side by side and have the same semantics. To account for this fact, I assume the same LCS for both constructions, but the incorporated construction differs from the phrasal one in that one of its arguments is saturated by *eti* in the lexicon, leaving only one argument position available for the syntactic derivation, while in the phrasal construction both arguments are available in syntax.

The LCS and the syntactic expression of the argument in a phrasal construction is given in (40). When a stative verb combines with a theme argument *eti* to express an emotion concept, the thematic hierarchy, in which themes are higher than locations, links *eti* to the theme argument and the possessor of *eti* to the location.

(40) Emotion verbs in ‘phrasal’ construction:

LCS:	V	THEME	LOC
syntax:		 <i>eti</i>	 possessor of <i>eti</i>

The NP containing *eti* is thus the syntactic subject of the emotion verb (cf. (2) and (25a)), and the possessor of *eti* the location of the emotion.

The incorporated emotion verbs are derived from the phrasal ones by lexical compounding: in the lexicon, *eti* saturates the theme argument, so this position is no longer available in syntax. The location, the only argument available for syntax, becomes the grammatical subject of the incorporated construction, as represented in (41). This accounts for the fact that verbs with incorporated *eti* have a non-active subject that is semantically the location (possessor) of the incorporated noun *eti*.

(41) Emotion verbs with incorporated *eti*:

LCS:	V	THEME	LOC
lexical compounding:	[V	THEME <sub><i>eti</i></sub> ]	LOC
syntax:			 possessor of <i>eti</i>

Because in this analysis the incorporated construction is not syntactically derived from the phrasal construction, it is predicted that there is no modifier stranding, and this prediction is borne out by the fact that the determiner and possessive enclitic of *eti* fail to be stranded. The stranding of the logical possessor of *eti* in (3) seems a counterexample, but when we consider the logical possessor as a separate (locational) argument NP rather than a nominal modifier that is *part of* the direct internal argument, there is no ‘stranding’ of this NP. Evidence to view the logical possessor NP as an independent locational argument is discussed in the next section.



formed through a process of lexicalisation, analogous to the incorporated forms with an intransitive base.

### 3. Possession as location in Kambera

Kambera nominal modifiers such as determiners, numerals, and classifiers cannot be stranded, while at the same time *eti* incorporation does seem to leave the possessor NP stranded. When we assume that modifying constituents are part of the modified NP the asymmetrical behaviour of the possessor NPs is not accounted for. Kambera allows possession to be expressed by the simple juxtaposition of NPs:

- (48) [Iyang]        [tau        wàu]<sub>NP</sub>  
       fish        person        smell  
       ‘Fish of a smelly person’

The possessor NP is not part of the constituent it modifies, but is an independent NP juxtaposed to it. The dependent, modifying function of this NP may be marked by a genitive enclitic on the possessed noun which crossreferences the possessor NP:

- (49) [...N-genitive clitic<sub>i</sub>]<sub>NP</sub>        [ NP ]<sub>i</sub>  
       possessed                                possessor

It is a well-known observation that possessive constructions derive from conceptual schema(s), one of which is the locative one where the possessor is linguistically coded as a locative argument (Lyons 1967, Clark 1978, Jackendoff 1983, Freeze 1992, Heine 1997). Below I present evidence to view a Kambera possessor NP as an independent locative argument rather than a modifying nominal within the possessee NP. If this view is correct, the incorporation of *eti* does not involve the stranding of a modifier which undergoes possessor raising; two properties that would suggest its syntactic derivation. In addition, the asymmetrical behaviour of modifying possessors within the class of nominal modifiers is accounted for.

In Kambera, possession can be expressed in a locative construction employing the intransitive locative verb *ni(ngu)* ‘be (here)’.<sup>8</sup> This verb is derived from the deictic element *ni* ‘here’ and is used as a locational verb (existential predicates such as *he [is my son]* are expressed by a nominal predicate without a copula). Normally, the verb *ningu* has a theme argument and an (optional) oblique location argument, *la uma* in (52):

- (50) Ningu        -ka    tau        [la    uma]<sub>pp</sub>  
       be.here    -Prf    person    Loc    house  
       ‘There are people present (in the house)’

The locative verb can also be used to express possession, as in (53). Here, the theme argument of *ningu* is interpreted as the possessee; while the location is the possessor:

- (51) Ningu        kabela        [lai    nyuna]<sub>pp</sub>  
       be.here    machete    Loc    he  
       ‘He has a machete’ (Lit.: ‘A machete is on/at him’)

With respect to emotion verbs we have seen that the possessor of *eti* is the subject of the derived verb; the experiencer of the emotion. This possessor can also be expressed as a locative argument, *lai nyuna* in (54). In this sentence, the argument of the locative verb *ningu* is the relative clause as a whole. The noun *eti* is incorporated into the predicate of the embedded clause. Its possessor, *nyuna*, is expressed as an oblique locative argument in the matrix clause:

- (52) Ni -nya<sub>j</sub>                                [lai nyuna]<sub>pp</sub> [na    pa-    karau        eti        wàngu]<sub>Rel Clause j</sub>  
       be.here -3sObj        Loc    he    Art    Rel-    be.dark    liver    use  
       ‘He is the one that is angry’ (i.e. not me)

<sup>8</sup> The syllable between brackets is dropped when the argument of *ningu* is marked with a clitic, as in (54) (for further discussion of the properties of *ningu* see Klamer 1994, to appear).

(Approx. lit. translation: 'It is (located) at him (the reason why) his liver is dark')

In sum, Kambera possessors may, and often are, expressed as (oblique) locations.

In addition, they are also relativised as locations, as illustrated in (55). The base structure contains the oblique location *la luku* 'in the river' and is represented in (55a). In (55b) the locative noun heads the relative clause that is marked by the relative marker *ma-* (also used to relativise intransitive arguments and transitive subjects):

- (53) a. Ningu wuya la luku  
 be.here crocodile Loc river  
 'There is a crocodile in the river'
- b. Luku [ma- ningu wuya -ng]<sub>Rel. clause</sub>  
 river Rel be.here crocodile -ng  
 'A river with crocodiles' (lit.: 'A river that crocodiles-are')

The sentences in (56) illustrate that the same relativisation applies to possessors:

- (54) a. Na- marihak na kalembi -na na tau  
 3sSubj- be.dirty Art shirt -3sPoss Art person  
 'The person's clothes are dirty'
- b. Tau [ma- ningu kalembi marihak -ng]<sub>Rel. Clause</sub>  
 person Rel- be.here shirt be.dirty -ng  
 'Someone with dirty clothes' (lit.: 'A person that dirty clothes are')

Now consider the 'stranded' and 'raised' possessor of *eti*, the pronoun *nyungga*, crossreferenced with the clitic *-nggunya* in (57):

- (55) Barang eti -nggunya ; -ka nyungga ;  
 pound liver -1sSubj -Prf I  
 'I am feeling/getting worried'

I propose that what we witness here is not possessor stranding and raising but rather the crossreference of the argument of the derived intransitive predicate, which is *nyungga* here. This is the argument that is linked to the LOC argument (41) above. As is the case for all definite NPs that double a clitic, *nyungga* is an optional NP.

Semantic locations may be surface as *objects* of transitive predicate as well. The relevant illustrations are repeated in (58). In sentence (58a) the possessed NP *na mbotu-mu nyumu* 'your weight' is crossreferenced with the 3rd person sg. clitic *-ya* as the object of the predicate, while in sentence (58b) the possessor NP *nyumu* 'you' is marked as the object with the 2nd person sg. clitic *-kau*:

- (56) a. Na- rasa -a -ya<sub>j</sub> [[na mbotu-mu] [nyumu]]<sub>i</sub>  
 3sSubj- feel -Mod -3sObj Art weight-2sPoss you  
 'She will notice your weight'
- b. Na- rasa -a -kau<sub>k</sub> [na mbotu-mu]<sub>j</sub> [nyumu]<sub>k</sub>  
 3sSubj- feel -Mod -2sObj Art weight-2sPoss you  
 'She will notice you (by) your weight'

The evidence presented here suggests that what looks like possessor stranding and/or raising in Kambera can be analysed differently: the possessor NP is an independent NP which expresses a locational argument and may be crossreferenced as a direct verbal argument.

#### 4. Conclusions

In Kambera, expressions for emotions use both a verb and the body part noun *eti* which is the location of the emotions. The verb-*eti* combination is used in both a phrasal and an incorporated construction, which synchronically occur side by side and have minimal semantic difference.

I have argued that, despite surface appearance, the Kambera emotion verbs are not derived by syntactic head movement and do not involve syntactic noun incorporation. Rather, they are derived by a lexical compounding process that manipulates the thematic structure of the verb. The theme argument of the base verb is satisfied by the body part noun *eti*, and the remaining argument for the derived intransitive verb is the locational argument. Because in Kambera possession is expressed using a locational structure, the possessor of *eti* may grammatically be analysed as a locational argument, which explains why possessors are the only 'modifiers' that can be 'stranded' in this language.

This case study shows that noun incorporation and 'modifier stranding' and 'raising' are unrelated phenomena and/or may be only apparent. They cannot be used crosslinguistically as reliable evidence for the analysis of incorporated nouns as syntactically moved heads.

#### References

- Baker, Mark (1988). *Incorporation*. Chicago: University of Chicago Press.
- Baker, Mark (1996). *The Polysynthesis Parameter*. Oxford: Oxford University Press.
- Finer, Daniel (1994). 'On the nature of two A' positions in Selarayese'. In: Corver, N & H. van Riemsdijk (eds). *Studies in scrambling: movement and non-movement approaches to free word order phenomena*. Berlin: Mouton De Gruyter, 153-183.
- Freeze, Ray (1992). 'Existentials and other locatives'. *Language* 68, 553-95.
- Gerds, Donna (1998). 'Incorporation.' In: A.J. Spencer & A.M. Zwicky (eds). *Handbook of Morphology*. Oxford: Blackwell.
- Heine, Bernd (1997). *Possession: cognitive sources, forces and grammaticalization*. Cambridge: Cambridge University Press.
- Klamer, Marian (1994). *Kambera: a language of Eastern Indonesia*. HIL dissertation, Vrije Universiteit.
- Klamer, Marian (1996). 'Kambera has no passive'. In: M. Klamer (ed). *Voice in Austronesian*. [NUSA 39], 12-30. Jakarta: Universitas Atma Jaya.
- Klamer, Marian (1997). 'Spelling out clitics in Kambera.' *Linguistics* 35, 895-927.
- Klamer, Marian (to appear). *A grammar of Kambera*. New York/Berlin: Mouton/De Gruyter.
- Lyons, John (1967). 'A note on possessive, existential, and locative sentences.' *Foundations of Language* 3, 390-396.
- Mithun, Marianne (1984). 'The evolution of noun incorporation' *Language* 60, 847-894.
- Mohanan, T. (1995). 'Wordhood and lexicality: Noun incorporation in Hindi.' *Natural Language and Linguistic Theory* 13, 75-134.
- Rosen, Sarah T. (1989). 'Two types of noun incorporation: A lexical analysis.' *Language* 65, 294-317.
- Spencer, Andrew (1995). 'Incorporation in Chuckchi.' *Language* 71, 439-89.
- Van Valin, Robert (1992). 'Incorporation in Universal Grammar: A case study in theoretical reductionism. Review of Baker 1988.' *Journal of Linguistics* 28, 199-220.

Marian Klamer, Dept. of Linguistics, Vrije Universiteit, P.O. Box 7161, 1007 MC Amsterdam, The Netherlands, email: klamerm@let.vu.nl, fax: 31+20-4446500



## SPELLING OUT CLITICS IN KAMBERA\*

Marian Klamer  
*Vrije Universiteit/HIL*  
Amsterdam  
klamerm@let.vu.nl

### Abstract

This paper presents a case study in the positional properties of a complex set of clitics: the mood, pronominal and clausal aspect clitics in the Austronesian language Kambera.

The clitics of this language may form a cluster of maximally nine clitics. The positional properties of the clitics can be distinguished into two distinct categories: (a) the position of the clitic cluster as a whole with respect to its host, and (b) the position of the clitics within the cluster relative to each other. The aim of this paper is to present an account of both these aspects of Kambera clitic placement.

The placement of the cluster as a whole will be characterized syntactically, while the ordering of the clitics within the cluster has the characteristics of inflectional morphology. I argue that the placement of the Kambera clitics with respect to each other is neither the result of the syntactic manipulation of terminal elements of functional categories, nor of lexical word formation rules or position class morphology, nor is it determined by the phonological properties of the language alone. Instead, it is the result of the morphological spell out of morpho-syntactic feature bundles (Anderson 1992) at the end of the syntactic derivation, at the interface between syntax and prosody: the postlexical level.

### 1. Introduction

This paper is a case study in the positional properties of an intriguing set of clitics: the mood, pronominal and aspectual clitics in the Austronesian language Kambera.<sup>1</sup> The clitics of this language may form a cluster of maximally nine clitics. The positional properties of the clitics can be divided into two distinct categories: (a) the position of the clitic cluster as a whole with respect to its host, and (b) the position of the clitics within the cluster relative to each other. The latter category can again be divided into two subcategories: (i) the position of three subgroups of clitics with respect to each other, and (ii) the placement of the clitics within each subgroup. The aim of this paper is to present an account of all these aspects of Kambera clitic placement.

The placement of the cluster as a whole will be characterized syntactically, while the ordering of the clitics within the cluster has the characteristics of inflectional morphology. Part of this morphology is idiosyncratic, while other aspects of it can be either functionally/semantically or historically motivated.

I argue that the placement of Kambera clitics is neither the result of the syntactic manipulation of terminal elements of functional categories, nor of lexical word formation rules or position class morphology, nor determined by the phonological properties of the language alone, but rather the result of the morphological spell out of morpho-syntactic feature bundles (Anderson 1992) at the end of the syntactic derivation, at the interface between syntax and prosody: the postlexical level.

The paper is structured as follows. In section 2, I present information on Kambera clitics and clitic placement: first I discuss in section 2.1 the potential complexities of the Kambera clitic cluster. I discuss how the clitic cluster as a whole is located with respect to its host (the (a) question above) and conclude with a list of the characteristics of Kambera clitic placement within the cluster that must be accounted for (the (b) question above). In section 2.2 I show that the Kambera clitic clusters as they are actually used in spontaneous speech are relatively small and simple compared to their potential complexities. The majority of clauses (94%) contain no more than 4 clitics: one mood clitic, one aspect clitic and (depending on the valency of the predicate) one or two pronominal clitics. In section 3 some possible accounts of Kambera clitic placement are evaluated: a syntactic (3.1), a phonological (3.2) and a morphological (3.3) account. Neither of these appears to be satisfying. Section 4 presents an account of Kambera clitic placement. The clitics are considered to be the morphological spell-out of inflectional features. Their ordering properties is discussed in section 4. Section 4.1 — 4.3 concern the ordering of the clitics within the functional subgroups (pronominal: 4.1, clausal aspect: 4.2 and mood: 4.3), section 4.4 presents an account of the ordering of these three subgroups relative to each other. Section 5 summarizes the findings.

## 2. Kambera clitics and clitic placement: the data

### 2.1. The complexities of potential Kambera clitic clusters

Following Zwicky (1977) two types of clitics are usually distinguished: 'simple' clitics which are syntactically regular but prosodically deficient, and 'special' clitics which are syntactically irregular, but not necessarily prosodically deficient. In this paper I will focus on the positional properties of the 'special' clitics in Kambera.

As shown in (1), Kambera has three general classes of such special clitics, expressing mood (emphatic, hortative, diminutive) and clausal aspect (perfective, imperfective, iterative) and marking verbal arguments (pronominal clitics marking subject, direct and indirect objects).

(1)	a. Mood <sup>2</sup> :	A.	bia 'just', mbu also/too, wa 'hortative', àru 'hortative (polite)'.			
		B.	ma 'emphasis', du/di 'emphasis', ki 'just a bit/while (diminutive)', a 'just/no more than'			
	b. Clausal aspect:		ka 'perfective', pa 'imperfective', i 'again/also (iterative)'.			
	c. Pronominal:		NOM	ACC	DAT	GEN
		1s	ku-	-ka	-ngga	-nggu
		2s	(m)u-	-kau	-nggau	-mu
		3s	na-	-ya	-nya	-na
		1p(inc)	ta-	-ta	-nda	-nda
		1p(exc)	ma-	-kama	-nggama	-ma
		2p	(m)i-	-ka(m)i	-ngga(m)i	-mi
		3p	da-	-ha	-nja	-da

Before I discuss the functional and structural properties of the Kambera clitics, I will first present some evidence for their 'clitic' (rather than affix) status. Kambera clitics are clitics because they attach to syntactic constituents and do not show selectional restrictions for a specific morphological base. As I will specify below, their syntactic attachment is to the border of a syntactic phrase, while prosodically they attach to the element that happens to linearly precede them (enclitics) or follow them (proclitics).

The examples in (2)–(4) illustrate some of the possible syntactic hosts for clitic clusters. In (2) the clitics attach to a verbal predicate which consists of a verbal projection. In (3) they attach to a nominal predicate, consisting of a nominal projection. In (4) they attach to a locative predicate, consisting of a prepositional phrase. The predicate plus the clitics make up a clause in Kambera which will be referred to as 'nuclear clause' or S and is printed in bold in (2)–(4).<sup>3</sup> (For ease of exposition, only pronominal clitics are used in these examples, but mood and aspect clitics could in principle be added as well. The properties of the pronominal clitics will be discussed below.)

(2)	a.	[Na-	[	mài]] <sub>s</sub>	na	sopir			
		3SN-		come	ART	driver			
		'The driver comes (here)'							
	b.	[Ku-	[	hili	mài]] <sub>s</sub>				
		1SN-		again	come				
		'I come again/I'll come again'							
	c.	[[Bidi	mài]	-nggu]] <sub>s</sub>					
		new	come	-1SG					
		'I have just come (here)'							
	d.	Ka	daingu <sup>4</sup>	[[ana laku]	-bia	-nggu]] <sub>s</sub>	duku		
		CNJ	surely	DIM	go	-MOD -1SG	EMP. 1s		
		'Because I really am going/will be going (lit. go a bit)'							
	e.	Napa	[ku-	[hili	beli	pàku]	-nya]] <sub>s</sub>	[na	umbuk-nggu]] <sub>j</sub>
		later	1SN-	again	return	first	-3SD	ART	grandson-1sG
		'I'll first have to go back to my grandson again'							

In Kambera, the pronominal clitics attached to the predicate generally have definite referents only.

The coreferent definite NPs are generally optional and used for emphasis and disambiguation.<sup>5</sup> In (2a) the verbal argument is marked as a nominative proclitic to the verb itself, in (2b) it is a proclitic to the preverbal adverb *hili* ‘again’, in (2c) it is a genitive enclitic to the verb, while in (2d) the same genitive enclitic attaches to the modal clitic *-bia*. In (2e), where the verb is transitive, the nominative proclitic that marks the agent attaches to the preverbal adverb *hili* ‘again’, while the clitic marking the patient attaches to the postverbal adverb *paku* ‘just, first’. Because of this pattern of attachment we can say that the clitics attach to a verbal projection. This projection consists of a verbal head plus its modifiers — the adverbs.

In (3) the phrase to which the clitics attach is a nominal projection. In (3a) the phrase is a possessed NP and the genitive enclitic attaches to a noun. (3b) is a possessed and modified NP, and the genitive clitic attaches to the modifier of the noun. In (3c,d) the NPs are used propositionally as nominal predicates. Observe that Kambera does not employ a copular verb in such constructions and that the argument of such a nominal predicate is marked with an accusative enclitic.<sup>6</sup> In (3c) the subject clitic attaches to the edge of the predicate and its prosodic host is an adverb, in (3d) the mood clitic *-bia* occurs between the subject clitic and the predicate.

- (3) a. [Na uma -nggu]<sub>NP</sub><sup>7</sup>  
 ART house -3sG  
 ‘My house’
- b. [Na uma bākul -nggu]<sub>NP</sub>  
 ART house be.big<sup>8</sup> -3sG  
 ‘My big house’
- c. [[Uma [bākul ai lulu]]<sub>NP</sub> -ya]<sub>S</sub>  
 house be.big very -3SA  
 ‘It (’s) a very big house’
- d. [[Tau mayila]<sub>NP</sub> -mbu -kai]<sub>S</sub> nyimi ná  
 person be.poor also -2pA you (pl) there  
 ‘...(moreover) you (are) also poor people’

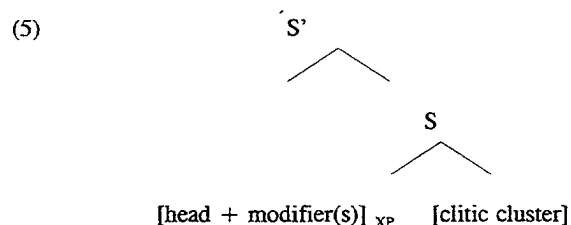
In (4) the pronominal clitics are attached to a prepositional phrase. In (4b,c) the PP in (4a) is used as a locative predicate the argument of which is marked with the clitic *-ya/-nya*. In (4a), the clitic attaches to the noun, in (4b) to the possessive clitic *-na* ‘his’ of the NP which is part of the PP *la uma-na* ‘in his house’. *-Nya* is the clitic that marks the argument of the locative predicate, *mbu ndaba-na* ‘everything’ is its coreferent subject NP.

- (4) a. [La uma]<sub>PP</sub>  
 LOC house  
 ‘At home’
- b. [La uma]<sub>PP</sub> -ya<sub>j</sub>  
 LOC house -3sA  
 ‘He<sub>j</sub> (is) at home’
- c. [Mbu ndaba -na]<sub>j</sub> [ [la uma -na]<sub>PP</sub> -nya<sub>j</sub> ]<sub>S</sub>  
 everything -3sG LOC house -3sG -3sD  
 ‘Everything is at/in his house’

Note that the argument of the locative predicate is expressed with an accusative in (4b) (the standard way to express arguments of non-verbal predicates, see (3c,d)) while a dative is used in (4c), when there is a genitive clitic preceding. This is an idiosyncratic restriction on Kambera clitic clusters, which I will return to below (see (18e, ii))

Once we view the clitics as attaching to syntactic phrases, the position of attachment is predictable. Though the particular word they attach to may change, the syntactic constituent of clitic attachment is constant: a phrasal constituent with a verbal, a nominal or a prepositional head that functions as the predicate of a clause. I call this constituent the predicate XP.<sup>9</sup> Kambera word order facts (see Klamer 1994, 1996) indicate that this phrase together with the pronominal, modal and aspectual clitics (the ‘nuclear clause’ or S) behaves as one syntactic unit. Because there is no evidence that clitics in this

language ever occur outside this constituent S,<sup>10</sup> I do not assume that they are part of a higher projection than S. The answer to question (a) in section 1 above is thus that the position of the Kambera clitic cluster with respect to its host can be formulated in simple syntactic terms, namely as ‘suffixing to XP within S’, i.e.:



Question (b) in section 1 above concerned the internal structure of the clitic cluster, which is what I will turn to now. Within the cluster, the clitics occur in two possible specified orders and combinations as given in (6):

- (6) a. Predicate XP - Mood A or {Mood B1 > Mood B2 > Mood B3 > Mood B4}  
 > Genitive > Dative 1 > Dative 2 >  
 Aspect > Aspect
- b. Nominative >  
Predicate XP > Mood A or  
 {Mood B1 > Mood B2 > Mood B3}  
 > Dative 1 or  
 Accusative > Dative 2 >  
 Aspect > Aspect 2

In the remainder of this section I illustrate some of the possibilities represented in (6) as well as the major functional and structural properties of the Kambera clitics on the basis of the illustrations (7)-(16):

- (7) Napa [da- bākul -bia -ka]<sub>S</sub> nú haromu da manu  
 later 3PN- be.big -MOD -PRF DEI tomorrow ART chicken  
 ‘The chickens just get big(ger) in time’
- (8) [Na- hili kano-nodār -ma -du]<sub>S</sub>, [na- patanda -ma -nya-ka]<sub>S</sub>  
 3SN- again RED.hesitate -MOD -MOD 3SN- think over -MOD-3SD-PRF  
 ‘He hesitates again and again, he (keeps on) thinking it over’
- (9) [Ta- tārū -ha]<sub>S</sub> [da ma- kahingir hāmu -ma-ka una]<sub>S</sub> reij  
 1pN- watch -3pA ART REL-clear nice -MOD-PRF DEI.3s  
 ‘We search for the ones that are nice and clean’
- (10) Rupu -bia -mu -nya -i -ka dumu nyumu  
 kill.chicken -MOD -2sG -3sD -ASP -ASP you.EMP you  
 ‘You just (go on and) slaughter it (i.e. the chicken)’
- (11) Lalu bākul -na<sub>i</sub> [na huru nuna]<sub>i</sub>  
 too be.big -3sG ART spoon DEI.3s  
 ‘That spoon is too big’
- (12) Māta -ma -ki -a -nja la hindi  
 leave MOD MOD MOD -3pD at attic  
 ‘Just leave them at the attic for a bit (longer)’
- (13) Daingu [wua -na -nggau -nya]<sub>S</sub> haromu, jāka [u- laku]<sub>S</sub>  
 surely give -3sG -2sD -3sD tomorrow when 2sN- go  
 ‘I’ll surely give it to you when you go’
- (14) [Njāpu -ma -đu -a -na-nya -i]<sub>S</sub> nú, na ngara ngia uhu  
 finished -MOD-MOD-MOD -3s.CONT -again DEI ART way place rice  
 ‘Thus it is finished, (the story about) the way to grow rice’

- (15) [Tau mini -du -ya -ka]<sub>s</sub>  
 person male -EMP -3SA -PRF  
 'It/he (is) a man'
- (16) [La pingi kokur -ka -i]<sub>s</sub>,  
 LOC top coconut -PRF -again/also
- [la pingi a-ài -bia -ka]<sub>s</sub> una  
 LOC top RED.tree -MOD -PRF DEL.3s  
 'Either up in the coconut (tree) or up in any other tree'

Objects are marked by an accusative enclitic, as in (9), or a dative enclitic, as in (8) and (10)–(13). In case a verb has both a patient/theme and a recipient/beneficiary object, both objects may be cliticized on the predicate simultaneously (order: recipient/goal-patient), as in (13). Agents of transitive verbs need not always be marked (i.e. zero subjects occur) but if they are, they are marked with either a nominative proclitic, as in (9), or a genitive enclitic, as in (10) and (13).

The marking of the sole argument of intransitive predicates shows a more variable marking: it can be marked by a nominative proclitic, as in (2a,b), (7) and (8), a genitive enclitic, as in (2c,d) and (11), but also by a cluster of a genitive plus third person singular dative clitic, as in (14), by an accusative clitic (that is, identical to a transitive object), as in (3c), (4b) and (15), or a dative clitic, as in (4c). In addition, it can also be doubly marked by the simultaneous attachment of a nominative and an accusative clitic, a construction not illustrated above (cf. Klamer 1994, 1997b)

Various factors determine the choice for the various different markers of transitive and intransitive subjects (Klamer 1994, 1997a,b). For present purposes it is sufficient to know that in most cases the choice for a particular clitic either depends on either the properties of the clause in which the clitic appears (such as the discourse function of the clause, aspectual properties of the sentence and the presentation/interpretation of the intransitive argument as more or less actively involved); or on co-occurrence restrictions in the clitic cluster. In other words, there is no evidence that the shape of a subject-marking clitic is determined lexically or under government from an adjacent element, but rather that it only gets shape after the derivation of the sentence is completed. This will be further discussed below.

Observe that all clitics under consideration are enclitics, i.e. they attach to the right of XP, except the nominative subject proclitic (cf. (7)–(9) above), which attaches to the left. The pronominal subject and object marking clitics at the right edge of XP follow the mood clitics, and are followed by the aspectual clitics, as represented in (6) above.

Thus, with respect to the marking of subjects (esp. transitive ones) we can say that they may surface in two shapes and two positions: one preceding the XP (the nominative) and one following the XP (the genitive). In the first case, the nominative clitic is directly adjacent to XP, in the latter case the mood clitics intervene between XP and the genitive subject marker, while the subject clitic itself occurs between the verb and its object complement. An analysis of the positional properties of Kambera clitics should take this variation into account.

The information in (1) and (6) is put together in (17) and (18) gives a summary of the characteristics of Kambera clitic placement to be accounted for.

- (17) Predicate XP -MOOD -PRONOMINAL -CLAUSAL ASPECT  
 B -ma -ki -du -a -gen -dat/acc-dat -pa/ka/i-i/ka  
 A -bia  
 -mbu  
 -wa  
 -àru

- (18) Characteristics of Kambera clitic placement to be accounted for:

- a. There are minimally zero and maximally nine post-predicate clitic positions that can be filled by members of the three ordered subgroups mood, pronominal (Dative, Accusative) and clausal aspect clitics.

- b. Within the subgroups of mood and pronominal clitics there is a strict ordering, whereas in the aspectual group at least two (of the three) clitics can take each other's position.
- c. The two subgroups of mood clitics (Mood A and B) are mutually exclusive.
- d. The marking of transitive agents and the single argument of intransitives depends on either the properties of the clause in which the clitic appears or on co-occurrence restrictions in the clitic cluster.
- e. The language has the following restrictions on sequences of clitics:
- (i) A genitive subject clitic occurs closer to the verb than an object clitic.
  - (ii) If there is a pronominal clitic following a genitive, it must be dative (compare (4b) and (4c)).
  - (iii) The two objects of ditransitive verbs can be marked in sequence. In such a sequence the inner clitic always marks the beneficiary/recipient (etc.), the outer clitic the patient/theme.
  - (iv) A double object sequence like this is subject to the restriction that they can occur in a sequence only if the first clitic is not third person while the second clitic is third person (Klamer 1994: 65-66, 76-78), stated alternatively:
    - A sequence of two object clitics marking third person (number is irrelevant) is always disallowed
    - Two object clitics can only occur in sequence if the inner clitic is first or second person and the outer clitic is third person.
- Illustrations of sequences that are allowed are given in (19), disallowed sequences are illustrated in (20):

(19)	a.	Na- wua -nya 3sN- give -3sD 'He gives (it) to him/He gives it'	b.	Na- wua -ngga 3sN- give -1sD 'He gives (it) to me/He gives me (e.g. as bride to someone)'
	c.	Na- wua -ngga -nya 3sN- give -1sD -3sD 'He gives it to me'	d.	Na- wua-nggau -nja 3sN- give -2sD -3pD 'He gives them to you (e.g. apples)'
(20)	a.*	Na- wua -nja -nya 3sN- give -3pD -3sD	b.*	Na- wua -nya -ngga 3sN- give -3sD -1sD
	c.*	Na- wua -ngga -nggau 3sN- give -1sD -2sD		

## 2.2. The relative simplicity of spontaneously used Kambera clitic clusters

In this section I will show that the Kambera clitic clusters that are actually used in spontaneous speech are relatively small and simple and never show the possible maximal expansion represented in (17). I considered 154 clauses containing clitics, taken at random from my database which consists of transcriptions of spontaneously uttered spoken texts.

Table 1 below shows that 94% of the 154 clauses under consideration contained one to four clitics, and none of them more than seven.

**Table 1.** Number of clitics used per clause

1-2	52%
3-4	42%
5-7	6%

That is, despite the fact that it is possible to have a cluster with five to nine clitics, Kambera hardly ever uses more than four clitics per clause. This means that many complexities encoded in the picture in (17) are irrelevant for 94% of the Kambera clauses containing clitics.<sup>11, 12</sup>

Table 2 shows the distribution of the three major categories of Kambera clitics (mood, pronominal

and aspectual) over the clauses. Apparently, there is no difference in how often clitics from the three subgroups are used in actual discourse. Approx. 70% of the clauses contain a mood, pronominal and aspectual clitic. They all seem to be used more or less with equal frequency; i.e. none of the subgroups has clitics that are especially frequent in a cluster.

**Table 2.** Distribution of category of clitics

Mood	70%
Pronominal	68%
Aspect	71%

Table 3 shows how many clitics per category are used in one clause. In other words: how many clauses contain one modal clitic, how many contain two modal clitics, how many have one pronominal clitic or two, etc.

**Table 3.** Number of clitics per category, used in one clause

Mood	1	83%
	2	13%
	> 2	4%
Pronominal	1	57%
	2	43%
Aspect	1	78%
	2	22%

The number of pronominal clitics depends on the number of verbal arguments that must be marked: in general we could say that a transitive predicate more often has two enclitics (marking subject and object) than it has one (marking object), while the opposite applies to an intransitive predicate. In the corpus there were no instances of a cluster with three pronominal enclitics. Table 3 shows that in four out of five clauses just one mood clitic is used rather than the two, three or four that are possible. In addition, four out of five sentences contain just one aspect clitic, rather than the two that are possible.

The question we can now ask is whether the mood, pronominal and aspect categories contain certain 'favourite' clitics that are used more often than other clitics from the same category. Table 4 presents a summary of the various frequencies of clitics per category.

**Table 4.** Frequency of clitic (paradigm) per category

Mood	group (a) ( <u>bia</u> , <u>mbu</u> , <u>wa</u> , <u>aru</u> )	15%
	group (b) ( <u>ma</u> , <u>du/di</u> , <u>ki</u> , <u>a</u> )	85% ( <u>ma</u> , <u>du</u> : 57%)
Pronominal	Genitive	35%
	Dative/Accusative	65% (Dat: 39%, Acc: 26%)
Aspect	<u>ka</u> 'perfective'	62%
	<u>pa</u> 'imperfective' <u>i</u> 'again, also'	38%

This table shows that the most frequent mood clitics are ma, du while the favourite aspect clitic is ka.

On the basis of the data presented in tables 1-4 we conclude that the clitic clusters in the majority of Kambera clauses are quite simple — much simpler than the picture in, for example (17) suggests. Of the possibilities represented in this picture only a small selection is actually used: the majority of clauses contain just one mood clitic (that one very often being either ma or du), just one aspect clitic (mostly ka) and one or two pronominal clitics — depending on the valency of the predicate. So in most cases, Kambera employs only a few of the clitic cluster possibilities available.

The functional/semantic subgrouping of the Kambera enclitics into mood, pronominal and aspect clitics seems to be a distinction that is also structurally relevant: if a clause has more than one or two clitics the various clitics come from every subgroup rather than just one. That is, rather than having, for instance, three mood clitics in a row, the language prefers to have one mood, one pronominal and one aspect clitic in a row. In other words, if we know the meaning/function of a clitic (an inherent property

of every clitic specified in the lexicon) we know to which subgroup it belongs. And given the fact that most clauses contain just one mood clitic, one aspect clitic and one pronominal clitic, we know the clitic orderings within these clauses when we know the ordering of the three subgroups mood, aspect and pronominals. I will return to this in section 4 below.

### 2.3. Overview of the paper

The facts presented in this section raise questions about the nature of the Kambera clitics during the syntactic derivation and afterwards. How can we account for their positions relative to each other? In this paper the latter question will be addressed in connection with the former. I will argue that some standard ways to account for clitic placement do not seem the most promising way to deal with the structural characteristics of Kambera clitics: First, I evaluate a possible syntactic account (3.1). Next I suggest that some prosodic factors may be relevant for the size (not the content) of the Kambera clitic cluster (3.2). Then I discuss why two traditional morphological accounts do not work for Kambera (3.3).

The alternative that I propose is to consider Kambera cliticization as inflection (rather than derivation) and view the clitics as phrasal affixes (cf. Anderson 1992). The use of the 'feature' morphology that has been proposed to account for inflectional processes (Anderson 1992, 1995, Stump 1992, Halle & Marantz 1993, Steele 1995) can thus be extended to the domain of (Kambera) clitics.

In such a view of morphology, the actual shape of the morpheme is separated from its morphosyntactic content. For instance, in Kambera, the segmental sequence [nggu] would be separated from its morphosyntactic features ([1st person], [singular], [possessive]). These features are relevant to and can be manipulated by the syntax but the actual segmental form of the morpheme cannot because it becomes only available after the sentence derivation has been completed: only then does a morphological spell-out rule link the morpho-syntactic feature bundles to their appropriate forms. Thus, only the clitics that are syntactically relevant appear at the surface and only at the very end of the derivation is the choice for their segmental form made. This ensures, for instance, that the pronominal clitic marking the subject is chosen from the right paradigm, and that the correct mood or sentential aspect clitics are used. The details of this account will be discussed in section 4.

## 3. Possible accounts of Kambera clitic placement

### 3.1. Syntactically derived?

Much current work treats clitics as the terminal elements of functional projections or as lexical heads projected and moved around by syntactic rules. This position presupposes a correlation between the position of a morpheme and its syntactic function whereby syntax is responsible for placing morphological elements relative to one another (Muysken 1986, Baker 1988, 1995, Halle and Marantz 1993). It also assumes that the morphological marking of dependency relations in syntax (e.g. by agreement, case marking and word order) is mediated by syntactic compounding, so-called 'head-to-head movement' (Baker 1988).

The empirical base for a syntactic account of word formation is the well-known cross-linguistic observation that word-internal structure often 'mirrors' clausal syntactic structure. However, this generalization is not correct for a head-marking (or 'non-configurational', or 'polysynthetic') language like Kambera, where the ordering of independent phrases is relatively free while the ordering of clitics is very much fixed. This fact alone already denies a literal interpretation of the 'mirror principle' and it questions the assumption that word formation, including clitic placement, should be syntactically analysed. Furthermore, in Kambera, only the pronominal clitics which crossreference NPs (nouns, pronouns) in the sentence may have 'full' counterparts in the sentence, while mood and clausal aspect clitics do not have such 'full' counterparts. Strictly speaking, then, only the pronominals in the cluster could tell us something about whether the clitic cluster reflects clausal structure — the part of the cluster that consists of the mood and aspectual clitics would neither confirm nor deny it.

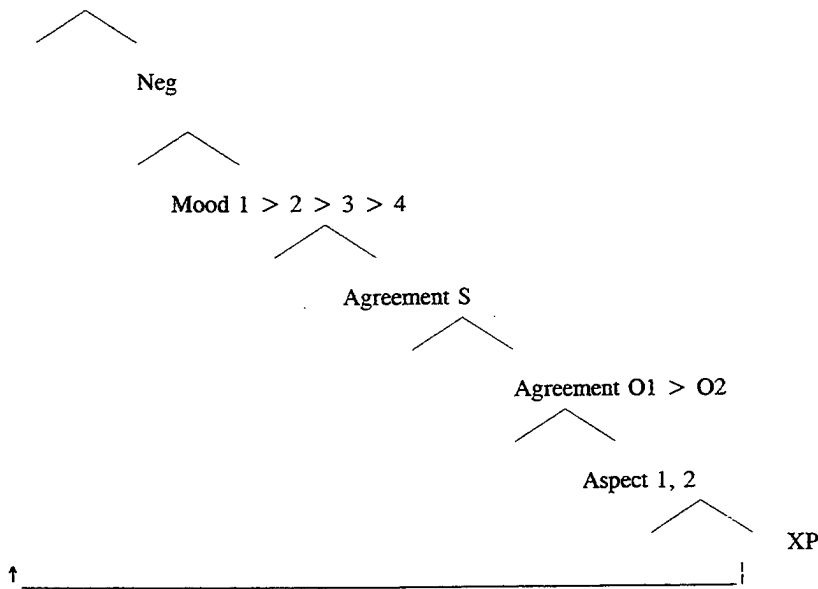
Ignoring this discrepancy for the moment, we could assume, in line with much of the syntactic literature on clitic placement, that syntactic base generation and movement in the context of an elaborated



theory of functional categories must be sufficient to locate clitics properly. The only question that should then be asked concerns the nature and motivation of that functional organization. For Kambera, it would be reasonable to assume three distinct functional projections for pronominal 'agreement': one for subject marking, one for direct object marking and another one for indirect object marking. In addition, the language would have a mood projection. We saw in (1) that this projection takes the (surface) position between the predicate and the pronominal clitics. It must have four clitic positions available (cf. (1) and (17)) which must be extrinsically ordered. An aspectual projection would be needed too, with two unordered positions. Finally, as there is no evidence that any of the clitics under discussion ever occur outside the scope of negations or conjunctions, the functional projections that we assume must be complements of the functional projections for negations and conjunctions in this language.

Above, it was mentioned that a Kambera subject (transitive/intransitive) can be marked with either a nominative proclitic or with one of the various enclitics (genitive, accusative, genitive-dative cluster) while the particular morphological case form that the subject eventually gets often depends on properties of the sentence after its derivation is completed. A syntactic analysis has to account for this. In addition, at a more basic level, such an analysis has to deal with the simple fact that subject-marking clitics surface in either preverbal or in postverbal position, in the latter case intervening between the verb and its complement. In a syntactic account that treats the clitics as terminal elements of functional projections, the structure in (21) could be a possible structure for Kambera clauses:

(21) Comp

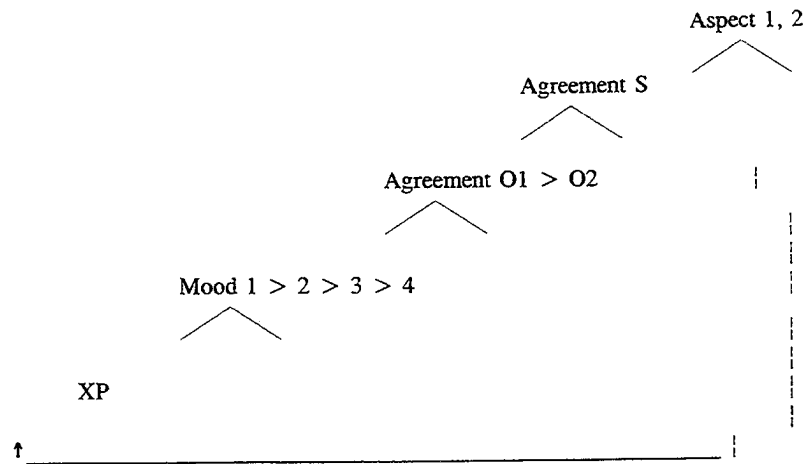


To arrive at the correct surface order on the basis of (21), XP is moved to a position above the mood projection. An analysis along these lines would derive the correct clitic order, but we would still need a motivation for the particular ordering properties of the Kambera functional projections at underlying structure, in particular: Why is the mood projection highest and aspect lowest?

We could also assume the 'standard' (European) underlying order of functional projections which differs from the one in (21), and derive the typical Kambera order from this 'universal' order by various movements. However, as long as there is no other (independent) motivation for such movements, we would still not have explained the distinct positions that mood and aspect clitics ultimately occupy in a Kambera sentence.

An alternative syntactic account would be the right-headed base-generated structure in (22).

(22)



This structure would not involve movement of the predicate XP to get the correct surface clitic order facts. However, it is at odds with the left-headed structure that is often assumed to be universally valid. In addition, in this right-headed structure we would need additional machinery to get the conjunction and negation in their correct position (always preceding the whole cluster).

Whether the underlying configuration is left-headed as in (21) or right-headed, as in (22), in both configurations we need to account for the variable position of the subject marking clitics (pre- or post-predicate). In (21), the subject clitic, base-derived as a proclitic, surfaces as an enclitic to XP when XP is moved.

But we know that subjects may also be marked by nominative proclitics and thus precede the predicate. For this alternative configuration, the subject marker would also have to move up, after XP has moved, to a position even higher up. Assuming that there is a position available in that region and that we can somehow motivate the movement, this analysis would get the word order right.

The structure in (21) where the subject marker is base-generated in post-predicate position has more problems, because the subject marker would be generated as intervening between the verb and its complement (cf. (9) and (13)). This is an unusual base generated construction, to say the least. Furthermore, to get to its pre-predicate position the subject marker would have to pass the foot of the tree; not to mention the complications that this right-headed structure causes for the correct placement of conjunctions and negations.

In addition, recall that the structure with a subject proclitic is not the mirror image of the one with the subject enclitic. While the subject proclitic is directly adjacent to the predicate, the subject enclitic is not: the mood node intervenes. However this variation is accounted for, it is clear that it involves rather complicated syntactic movement.

Apart from questions like which order should be assumed to be basic or derived and why, a syntactic movement analysis of Kambera clitic clusters raises more specific analytical questions concerning the process of movement in general, such as: If there is movement, what would be the trigger for movement in Kambera? If the movement is dictated by the morphological requirements in the form of selectional restrictions on the Kambera clitics this implies that the affixes are lexically specified for their neighbours. Such a lexical specification, however, has certain drawbacks that are discussed in section 3.3 below. Apart from the question of what would force the constituents to move, we should also ask which constituent(s) move: the XP, the individual clitics, or the functional projections? To what position(s) would they move, and what would be the evidence that those positions exist? And, last but not least, how would an account that has as its basic assumption that morphology mirrors syntax regularly take into account the different restrictions on [person] and [case] in certain clitic combinations, the fixed ordering of some (mood, pronominal) clitics versus the free ordering of others (aspect)? These are the kind of questions that should be addressed in a purely syntactic analysis of Kambera clitic placement. However, I will not deal with Kambera cliticization in this fashion, because I do not see any reason to treat the morphological marking of dependency relations in Kambera syntax such as the agreement between predicates and arguments as mediated by a process of syntactic compounding like head-to-head movement is.

Finally, recall that in Kambera the morphological form of the subject marking clitic crucially depends on properties of the sentence after derivation is completed. A syntactic account that treats the subject clitics in their actual form as the terminal elements of functional projections could not account for this fact. Rather, we need to distinguish between clitic properties that are relevant during the syntactic derivation, such as person and number, and other properties of the clitic, such as its morphological case (i.e. whether it is marked by nominative, a genitive clitic, a combination of genitive and dative, etc., see above) and its meaning properties, so that all subject markings are treated alike until a late stage in the derivation where they diverge. In other words, the morphosyntactic featural content of functional categories should be present in and accessible to syntax, but how such features are overtly realized is only decided after the sentence derivation is completed.

### 3.2. Phonologically derived?

The shape and ordering properties of the Kambera clitics do not depend on phonotactic properties (syllable/word structure), stress or phonological processes in the language (see Klamer 1994, chapter 2). However, there are some indications that the prosodic properties of the language partly determine the preferred size of the clitic cluster.

Almost all Kambera clitics can be prosodically characterized as monosyllabic (cf. the forms in (1)), unable to bear stress<sup>13</sup> and prosodically dependent on the head of a higher prosodic unit.

Kambera lexical items are generally either disyllabic roots or trisyllabic derived forms (a root with a monosyllabic prefix). The prosodic template for a morphological root/basic lexical item in Kambera is a trochaic foot (Van der Hulst & Klamer 1996). The lexical head of a predicate XP is also the prosodic head of the higher phonological unit.<sup>14</sup> Predicate and clitics together form a higher syntactic unit (S or the nuclear clause) which is also a higher prosodic unit — presumably a prosodic word.

A nuclear clause (predicate plus clitics) cannot be considered a prosodic compound.<sup>15</sup> Kambera compounds are prosodically right-headed: [‘ana ’mini] ‘son’<sup>16</sup>. In a nuclear clause the main stress remains on the left element — the lexical item in the predicate — while the clitic cluster on the right remains unstressed.

Table 1 above shows that only 6% of the 154 nuclear clauses investigated have more than 7 syllables. Most of them consist of a lexical head (a root consisting of 2 syllables plus perhaps a monosyllabic prefix), one proclitic (1 syllable) and maximally four enclitics (4 syllables). In other words, we may hypothesize that the size of Kambera clitic clusters is bound by prosodic maximality limits: Kambera prosodic words prefer to consist of maximally two feet.<sup>17</sup> However, due to the fact that my data is limited as far as prosodic information is concerned and native speakers can only be consulted locally, these observations should be taken as tendencies rather than hard facts.

### 3.3. Morphologically derived?

One way to order the clitics could be by linking them to nine sequential ‘slots’ that follow the predicate. Such ‘morpheme slots’ or ‘position classes’ are concepts that have often been used to describe the positional properties of morpheme clusters (usually affixes rather than clitics, e.g. Bloomfield 1962, Muysken 1986, Simpson and Whitgott 1986). Theoretically there is little sympathy for such an approach because its structure-specific character defies a generalization over morphological structure within and across languages. Furthermore, it has undesirable theoretical implications, one being that unfilled positions are structurally as relevant as full positions — contrary to fact in Kambera.

An alternative morphological analysis is the traditional lexical account which views clusters of morphemes as the result of morphological word formation. This approach too cannot account for the facts of Kambera. First, the placement of Kambera special clitics is not ‘regular’ word formation: because the lexical category and the morphological shape of the host to which the clitics attach may vary, the host cannot be characterised in morphological terms but must be characterized syntactically instead. Secondly, in a lexical account each clitic would have to have a specification of its potential sister in its lexical entry. In this way it would not be possible to capture some of the Kambera facts such as the free linear order of some clitics (the aspectual clitics) versus the fixed linear order of the other clitics, and the

fact that all the clitics are optionally present. In other words, the presence of a sentential aspect clitic (which is positioned at the right edge of the cluster) does not imply the presence of any other clitic: -pa can also occur as the only enclitic in a clause, as illustrated in (23):

- (23) Na- mutung -pa  
 3SN- burn -IMPF  
 'It's still on fire'

Alternatively, we could assume that individual clitics do not select one specific neighbour to attach to, but are lexically specified for the exhaustive set of clitics that it may follow. How this distributional information would look for one of the Kambera clitics, the imperfective marker -pa, is sketched in (24) (the pronominal clitics, here represented with the name of their paradigms, should be individualised too):

- (24) pa:  
 [ XP - [ {ma ki du a} Gen Dat Dat ] \_\_\_ ]  
 [ {bia/mbu/wa/aru} ]

This alternative has the following drawbacks. First of all, even if there is information like this present in the lexicon, it would still not be enough to derive the correct clitic combinations at the surface structure. For instance, the actual marking of, for instance, subjects depends on either the properties of the whole clause in which the clitic appears and/or on specific co-occurrence restrictions in the clitic cluster. Because the information on the basis of which we can choose a clitic from one of the paradigms in (1c) is only available after the syntactic derivation, a word formation rule in the lexicon cannot specify from which pronominal paradigm the clitics in a particular context should be chosen. The same applies, of course, to the mood and sentential aspect clitics. Secondly, the optionality of the neighbouring clitics still needs to be encoded in some way.

Furthermore, if clitic cluster information of this sort is specified in the lexicon, all the possible derived forms must be specified. We have seen that what attaches to the nuclear clause is a choice from a set of nine mood clitics, four paradigms of pronominal clitics (i.e. 4x7=28 pronominal clitics) and three aspectual clitics. Such an account would be enormously redundant.

Finally, note that not all the aspects of Kambera clitic order are arbitrary language-specific facts that belong in the language's lexicon. For instance, the fact that mood markers occur closer to the predicate than pronominal markers occurs crosslinguistically (Bybee 1985:35).

Another observation that has been made for many languages is that the encoding of verbal arguments interacts with their animacy according to an animacy hierarchy proposed by Silverstein (1976). According to this hierarchy, the more animate an argument is, the more it is likely to be expressed pronominally. In many languages in the configuration where an (applicative) verb has two object markers this animacy hierarchy is reflected when the pronominal marker for the beneficiary/recipient occurs closer to the verb than the marker for patients/themes. This can be explained by the fact that beneficiaries/recipients are more often animate than patients/themes (i.e. giving/handing over something to someone is the standard case, giving/handing over someone to someone else is a relatively exceptional situation). The more animate an argument is, the more relevant it is for the semantics of the predicate and the closer to the predicate it occurs (Bybee 1985). In the animacy hierarchy, 1st/2nd person pronominals are ranked at the 'top' of the hierarchy: they are the canonically animate pronominals and are followed by 3rd person pronominals. The animacy hierarchy may be an explanation for the two conditions that Kambera has on sequences of object clitics, namely that the recipient is marked closer to the verb than the patient; and that if the recipient is 3rd person, the patient/theme marker may not be 1st/2nd person, the reverse being fine.

In a lexical derivational approach crosslinguistic patterns such as these are ignored because the facts are all treated as arbitrary and coincidental facts of the lexicon of a particular language.

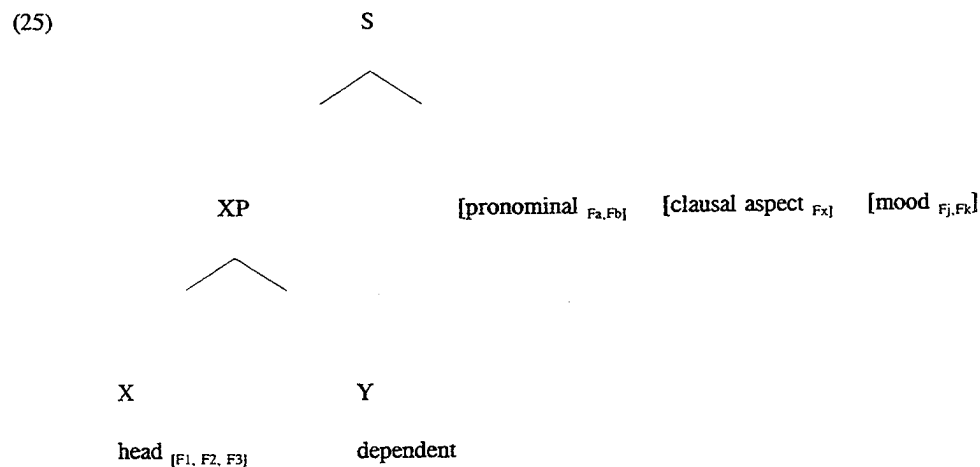
#### 4. The realization of clitics by morphological spell out rules

We have seen that Kambera clitics are 'special clitics' in Zwicky's (1977) terms because their placement with respect to each other cannot be handled by normal syntax. I also argued that for various

reasons the structure of the clitic cluster should not be described in terms of lexical word formation either. Kambera clitics are considered ‘phrasal affixes’ — they express inflectional notions and attach to a syntactic constituent (the predicate XP). At the same time as arguing for the inflectional morphological status of the Kambera clitics, we also observed that some fixed orderings and preferred combinations of Kambera clitics occur in other unrelated languages as well, hence should not be treated as language particular idiosyncratic facts.

Assuming a morphological theory like Anderson’s (1992), Kambera inflection is considered a list of morphosyntactic features cumulated during the derivation of a clause. The bundle of features becomes one or more formal objects — clitic(s)— through the application of morphological spell-out rules at the end of the derivation.

The accumulation of features is represented in (25). The XP constituent is the Kambera predicate, which consists of a head X and a dependent Y. X can be a verb, noun or preposition while Y should be interpreted as either a modifier or a complement of the head (i.e. either an adverb, noun or noun modifying a verb or a noun, or a prepositional complement NP) (see section 2.1). As the focus of this paper is to give an account of the correct surface order of the Kambera clitics, I will neither discuss the details of Kambera sentence syntax, nor the technical details of the derivation of inflectional features, nor consider how the feature copying actually takes place, nor concern myself with the internal composition of the feature bundles. I will simply assume that during the syntactic derivation of the sentence, an as yet unordered bundle of features concerning inflectional notions of mood, pronominal and aspect attach to the predicate XP within the domain of the nuclear clause S:



At the end of the derivation the feature bundle(s) are spelled out as clitics and are then linearized. This analysis implies that syntactically irrelevant clitics are not spelled out, which accounts for the optionality of all Kambera clitics. Because the optionality of the clitics is built into the system, there are no empty positions nor unused levels in the case of unused clitics, as a position class account or lexical morphological account would have.

There are two aspects to the linearization of Kambera clitics: (a) the order of the clitics within three different functional subgroups, and (b) the order of the three subgroups with respect to each other. The sections 4.1-4.3 deal with the former aspect, section 4.4 with the latter.

#### 4.1. The ordering of the pronominal clitics

Kambera pronominal clitics, being inflectional elements, are the spell out of inflectional features that are copied from independent NPs (full or empty) to the predicate XP. In section 2.1 we saw that Kambera pronominal clitics show morphological case distinctions which depend on either the properties of the entire clause or on co-occurrence restrictions within the clitic cluster (see (18d)). NPs, on the other hand, are not marked for morphological case, hence I assume that morphological case is not an inherent feature of the NPs but a feature that is added to the bundle of inflectional features in the course

of the syntactic derivation.

Kambera pronominal inflection differs from well-known 'pro-drop' languages. Clitics and full NPs are sometimes in complementary distribution but in other cases full NPs may be 'doubling' the clitics. In particular, a distinction should be made between the inflection of subject and object NPs. The cliticization of transitive objects is determined by definiteness, unlike subjects. In other words, definite objects of simple transitive verbs must be cliticized on the predicate and their coreferent NPs are always optional, while indefinite objects cannot be marked with clitics but must be expressed by full (indefinite) NPs.<sup>18</sup> Subjects allow for more options. Definite subjects are generally cliticized on the verb,<sup>19</sup> in that case the coreferent NP is optionally present. Indefinite subjects, however, are sometimes cliticized and sometimes appear as NPs. In case an indefinite subject is cliticized, the coreferent NP is optional.<sup>20</sup>

What this shows is that in general Kambera pronominal clitics have definite referents. That is, it is the feature [definite] which triggers Kambera pronominal inflection more than nominal features like [person] and [number].

Except for the nominative, all Kambera clitics attach to the right of XP. I therefore assume that this is the default value for clitic attachment in this language. (In line with the 'suffixing preference' of Greenberg 1966, Cutler et. al. 1985, Hall 1992). The nominative clitics are the only ones that have to be (lexically) specified for their attachment to the left of XP.

The following spell-out rules illustrate some of the discussion above. The features relevant for the correct spell out of the Kambera pronominal clitics include syntactic information. (26a) is the feature bundle of a definite object which is marked with a clitic from the accusative paradigm. The feature bundle for an indefinite object in (26b), however, cannot be spelled out as a clitic. The definite subject in (26c), is spelled out as given.

- (26) a. [definite, 3rd person, singular, accusative] → -ya  
 b. [indefinite, 3rd person, plural] → ∅  
 c. [definite, 1st person, nominative] → ku-

Ignoring further details of the feature (bundle) derivation and composition, let us now look at the ordering of the pronominal clitics with respect to each other. The restrictions that were summarized in (18e) are repeated in (27):

(27) Kambera has the following restrictions on sequences of clitics:

- (i) A genitive subject clitic occurs closer to the verb than an object clitic.
- (ii) If there is a pronominal clitic following a genitive, it must be dative (compare (4b) and (4c)).
- (iii) The two objects of ditransitive verbs can be marked in sequence. In such a sequence the inner clitic always marks the beneficiary/recipient (etc.), the outer clitic the patient/theme.
- (iv) A double object sequence like this is subject to the restriction that the objects can occur in a sequence only if the first clitic is not third person while the second clitic is third person (Klamer 1994: 65-66, 76-78), stated alternatively:
  - A sequence of two object clitics marking third person (number is irrelevant) is always disallowed
  - Two object clitics can only occur in sequence if the inner clitic is first or second person and the outer clitic is third person.

(See for illustrations of allowed and disallowed sequences (19) and (20) above).

According to (27i) a genitive subject enclitic should precede the object enclitic(s), as illustrated in (28):

- (28) Daingu [wua -na -nggau -nya]<sub>s</sub> haromu  
 surely give -3sG -2sD -3sD tomorrow  
 'I'll surely give it to you tomorrow'

At first sight, this seems in contradiction to the fact that in general a verb and its complement are

assumed to form a syntactic unit; the object marker should attach closer to the verb than the subject marker. Recall, however, that the canonical way to mark transitive subjects in Kambera is by using the nominative proclitic. A clause with a genitive subject is a marked construction.

There is a probable historical source for this special type of subject marking. According to Finney (1997; p.c.) the Kambera genitive subject clitic is related to the ergative marking in some of the languages related to Kambera, which would imply that the subject clitic that occurs between the verb and the object marker is a reflex of an ancient ergative pattern of the language. Though present-day Kambera morphosyntax is mainly nominative-accusative, it has some clearly absolutive-ergative properties too (Klamer 1997b). In ergative languages (in contrast to languages of the accusative type) the transitive subject is expected to be closer to the verb than object agreement (Bittner & Hale 1996:568). If the Kambera genitive is indeed a reflex of an older ergative marker, restriction (27i) becomes understandable as a restriction reflecting a ‘morphologized’ earlier syntactic pattern of the language.

Sentence (28) above also illustrates the restriction in (27iii) that the recipient/indirect object clitic always precedes the patient/direct object. Restriction (27iv) states that if in a sequence of object clitics the recipient is 3rd person, the patient/theme marker may not be 1st/2nd person, while the reverse is fine (see the illustrations in (19) and (20) above). Again, these are restrictions on surface clitic order that not only apply in Kambera, but in other unrelated languages too. Hence, I tentatively proposed the animacy hierarchy of Silverstein (1976) as a functional motivation/explanation for these two conditions (section 3.3).

The only restriction on pronominals that appears to be truly language particular is (27ii): In a transitive clause where the subject is marked with a genitive clitic, the object must be dative (see (29a,b)). (The canonical case is when the subject clitic is nominative and the object is marked with an accusative clitic, as in (29c)).

- (29) a. Mbàda      rongu-nggu    -nya  
 already      hear-1sG      -3sD  
 ‘I (have) heard it already/before’
- b. \* Mbàda      rongu-nggu    -ya  
 already      hear-1sG      -3sA
- c.    Hi      ku-      rongu      -ya  
 CNJ    1sG-    hit          -3sA  
 ‘So I’ll hear it’

If we assume that the correct linearization of the clitics is determined by the order in which the morphological spell-out rules apply, we can extrinsically state the order of the spell-out of pronominal clitics as Nominative > Genitive > Dative 1 / Accusative > Dative 2.

However, in this ordering the restrictions (27ii, iii, iv) are not taken into account. These restrictions cannot be formalized by rule ordering because they obviously have a more global reference. Hence, we must assume that the configuration that is the result of morphological spell-out is ‘almost right’ but not quite; at the final stage it is subject to surface restrictions, which include idiosyncratic restrictions like (27ii) and functionally motivated ones like (27iii, iv). I will return to this in section 4.4 below.

#### 4.2. The ordering of the clausal aspect clitics

Being part of the same clitic cluster as the pronominal clitics, the clausal aspect clitics<sup>21</sup> are assumed to be the spell-out of inflectional features too. An aspectual feature like [imperfective] is copied onto the XP from a constituent expressing imperfective clausal aspect. This constituent must be assumed to be an abstract (covert) entity always, because the Kambera aspectual clitics are never accompanied by overt aspectual phrases (unlike the pronominal clitics which may occur with ‘doubling’ NPs). The three clausal aspect clitics are repeated in (30) and illustrated in (31)–(33):

- (30) Clausal aspect: ka ‘perfective’, pa ‘imperfective’, i ‘iterative’
- (31) Mutung<sup>22</sup> -nanya -ka  
burn -3s.CONT -PRF  
‘It has been burning/It is burnt down’
- (32) Mutung -nanya -pa  
burn -3s.CONT -IMPF  
‘It is (still) burning’
- (33) Mutung -nanya -i  
burn -3s.CONT -again  
‘It is/has been burning again’

Perfective and iterative clitics have been attested in either order, i.e. either *i-ka* or *ka-i* (which are semantically distinct). Combinations of imperfective and iterative clitics have only been attested as *pa-i*. However, given the distributional similarities between *ka* and *pa*, the sequence *i-pa* is expected to be grammatical too. A combination of imperfective *pa* and perfective *ka* is predicted to be impossible because of their semantic incompatibility. In summary, in our account we do not need a statement fixing the order of the aspectual clitics.

#### 4.3. The ordering of the mood clitics

Mood clitics occur in two mutually exclusive groups: group A and group B in (34):

- (34) Mood: A. bia ‘just’, mbu ‘also/too’, wa ‘hortative’, àru ‘hortative’ (polite)  
B. ma ‘emphasis1’, du (di) ‘emphasis2’<sup>23</sup>  
ki ‘just a bit/just a while (diminutive)’, a ‘only/no more than’

Like the other clitics in the cluster of which they are a part, we assume the mood clitics to be the spell out of inflectional features. The inflectional mood features are copied onto XP from one or more (again: covert) mood phrase(s) elsewhere in the sentence structure. If the bundle of inflectional features attached to XP contains mood features, the spell out of the mood A/B clitics may be represented as in (35), (36):

- (35) a. [hortative] → [wa]  
b. [‘just’] → [bia]  
c. [‘also, too’] → [mbu]  
d. [‘hortative’ (polite)] → [àru]
- (36) a. [emphasis1] → [ma]  
b. [‘no more than’] → [a]  
c. [diminutive] → [ki]  
d. [emphasis2] → [du/i]

We do not need to state the ordering of the rules in the A-group (35), because these clitics are specified to occur on their own without any other mood clitics. The clitics of group B, however, may appear together. If they do, they must occur in a fixed order: *ma-ki-du-a* ‘emphasis1’- ‘diminutive’ - ‘emphasis2’ - ‘only/no more than’ (vs. \**ma-du-ki-a*, \**du-ki-a-ma*, etc. etc.). We account for the fixed order of the mood B clitics by the formal mechanism of extrinsically ordered spell out rules, as in (37):

- (37) [emphasis1] → [ma] >> [diminutive] → [ki] >> [emphasis2] → [du/i] >> [‘no more than’] → [a]

An alternative would be to propose spell out rules combining (unordered) features:



- (38)
- |    |  |   |                       |
|----|--|---|-----------------------|
| e. | [emphasis1], ['no more than']          | → | [ <u>ma a</u> ]       |
| f. | [emphasis1], [diminutive]              | → | [ <u>ma ki</u> ]      |
| g. | ['no more than'] [emphasis2]           | → | [ <u>du/i a</u> ]     |
| h. | [emphasis2], [emphasis1]               | → | [ <u>ma du/i</u> ]    |
| i. | [diminutive], [emphasis2]              | → | [ <u>du/i ki</u> ]    |
| j. | [diminutive], [emphasis2], [emphasis1] | → | [ <u>ma du/i ki</u> ] |
| k. | etc.                                   |   |                       |

However, this type of rule is theoretically unattractive because they imply the logical possibility of clitic orderings that do not exist (for instance, if ma-a is the spell out of two features, nothing in the formal notation of the rules prevents the spell out of a-ma, which does not occur). Furthermore, in rules like the ones in (38) the same information is repeated again and again, leading to redundancy.

Thus, extrinsically ordering the spell-out rules of (36) as is done in (37) seems to be the better option. Despite the general conceptual unattractiveness of such extrinsic rule ordering, note that in this case it does reflect the properties of the fixed ordering of the Kambera mood B clitics. After all, the idiosyncratic ordering of these Mood B clitics seems to be a language particular morphological fact without any independent structural motivation. It is a property of the language that a speaker just 'has to know'. A way to formalize this is to extrinsically order the rules by which the clitics are introduced.

#### 4.4. The ordering of the three functional subgroups of clitics

Now we have considered the spell out rules for the clitics per subgroup, we turn to the question how the three subgroups are ordered with respect to each other. As we have seen in section 2.2 the majority of the Kambera clauses that were investigated contained just one mood clitic (that one very often being either ma or du), just one aspect clitic (mostly ka) and one or two pronominal clitics — depending on the valency of the predicate. In other words, if we can account for the order of the functional subgroups, we have accounted for the clitic order in more than 90% of the actually uttered Kambera clauses.

Again I propose an extrinsic ordering for the functional subgroups of clitics because there seems to be no structural reason why the mood clitics should precede the pronominals and why the aspect clitics are attached at the end of the cluster. In this case, however, the extrinsic ordering is not completely idiosyncratic (unlike the ordering of the mood B clitics discussed in the previous subsection), because what seems to be at work here is a functional principle regulating the order of the subgroups. It is formulated as a universal principle in (39). This principle is similar to the Relevance principle formulated by Bybee (1985) to account for the crosslinguistic properties of affix ordering, but here it concerns clitics rather than affixes.

#### (39) UNIVERSAL CLITIC ORDERING PRINCIPLE : Relevance to meaning of predicate

According to this functional principle the higher the relevance of the clitic to the meaning of the predicate, the closer it occurs to the edge of the predicate XP.

From a functional point of view we find two types of clitics crosslinguistically: clitics that have relevance to the meaning of the predicate and clitics that have relevance to larger domains, like the discourse surrounding the predicate (see Anderson 1992:218-219 for a similar observation). Put differently, some clitics have a function within the phrase that contains the predicate, other clitics have phrase-external relevance.

Mood clitics like the emphatic and hortative Kambera clitics in (1a) are relevant to the meaning of the predicate — they are used as modifiers and have an adverbial function. Clitics representing basically 'grammatical' material, such as the Kambera pronominal clitics express the relation between a predicate and its arguments. They have an argument-indexing function. Both mood and pronominal clitics are therefore more relevant to the meaning of the predicate than the clausal aspect clitics because the latter express the relation between a sentence and its context. So the universal functional principle in (39) would result in the universal ordering of clitics in (40):

#### (40) Predicate-modifying clitics > Argument indexing clitics > Contextual clitics

In summary, the ordering of the functional subgroups has a functional motivation: the principle of relevance to predicate. Another functional motivation that plays a role is the 'animacy hierarchy' discussed in section 3.3. Both can be viewed as functional constraints on Kambera clitic clusters.

## 5. Summary and discussion

In this paper I have argued for a morphological feature account of Kambera clitic clusters. Many characteristics of Kambera clitic clusters were discussed as evidence for the proposal that the actual clitic morphemes are introduced late in the derivation. That is, in Kambera, inflectional clitics are not the terminal elements of syntactic projections ('heads'); functional categories like mood, person and aspect are neither morphological nor phonological 'objects', nor syntactically separate constituents with associated projections. Rather, in order to derive the correct clitic cluster, the morpho-syntactic featural content of the clitics needs to be separated from their actual (phonological) form until the end of a derivation where they are spelled out by rules. Every rule operation 'adds' a clitic to the output of the previous operation.

A morphological feature approach allows us to treat the content, but not the overt realization of functional categories as present in and accessible to syntax. This is what we need because the clitic cluster reflects some aspects of syntactic structure (inflection) while it also shows a surface order that is not syntactically motivated.

The positional properties of the Kambera clitics were described as follows. The cluster as a whole attaches to the predicate XP within S. It consists of the three ordered subgroups of inflectional clitics: mood, pronominal and clausal aspect. The order of the subgroups is functionally motivated by extending Bybee's (1985) principle of relevance to the order of clitics as well as affixes.

The ordering of the clitics *within* the three subgroups required a separate treatment. Because the order of the aspect clitics is free, we do not need to state their ordering. On the other hand, the mood B clitics that occur together must be ordered extrinsically. Finally, we saw that the ordering of the pronominal clitics reflected either a historical (ergative) property of the language, or was determined by semantic/functional restrictions (the definiteness hierarchy).

As yet it is not so clear how the 'history' and 'function' of grammatical elements could be incorporated in the synchronic model of language structure. Recall from section 3.2 that, in addition to functional and historical motivations determining the shape of the clitic cluster, prosodic constraints seem to constrain its size. Thus, both the shape and size of the Kambera clitic cluster seem to be constrained by requirements that cannot be defined in purely structural (syntactic, morphosyntactic) terms. Rather, requirements of the human computational and perception/production system seem to play an important role in shaping the Kambera clitic cluster. This is an additional indication that the cluster is formed at a late level, for example between syntax and prosody. Of course, the linearization of the clitics in a cluster like the Kambera one may also be formulated in Optimality Theoretic terms (Prince and Smolensky 1993, Prince and McCarthy 1994) employing ranked constraints on surface output, referring, for instance, to the alignment of clitics to their host. In this case, language variation (language-internally or crosslinguistically) should constitute the evidence for re-ranking the constraints. If, however, variation is (assumed to be) absent, as in the account above where a universal ordering principle was proposed, there is no need for constraint re-ranking resulting in language variation, hence the OT framework is not crucially needed to account for this type of clitic placement.

## Notes

\* This paper has appeared as: Klamer, M. (1997). 'Spelling out clitics in Kambera.' *Linguistics* 35, 895-927 and is a revised version of the paper 'Optimal clitic placement in Kambera' which was presented at AFLA III, UCLA, Spring 1996. Please cite the published version. Many thanks to Geert Booij, Karijn Helsloot, Harry van der Hulst and the two anonymous reviewers for their comments on previous versions of this paper. I also thank the audiences at AFLA3 (UCLA, April 1996) and the OT/Derivationalism Workshop (HIL, Leiden, December 1996) for their valuable questions and comments. The research for this paper was supported by a fellowship from the Royal Netherlands Academy of Arts and Sciences (KNAW).

1. Kambera is an Austronesian language of the Central-Malayo Polynesian branch, has c. 150,000 speakers and is spoken on the eastern part of the island of Sumba in Eastern Indonesia; Klamer 1994, to appear are grammars of the language.

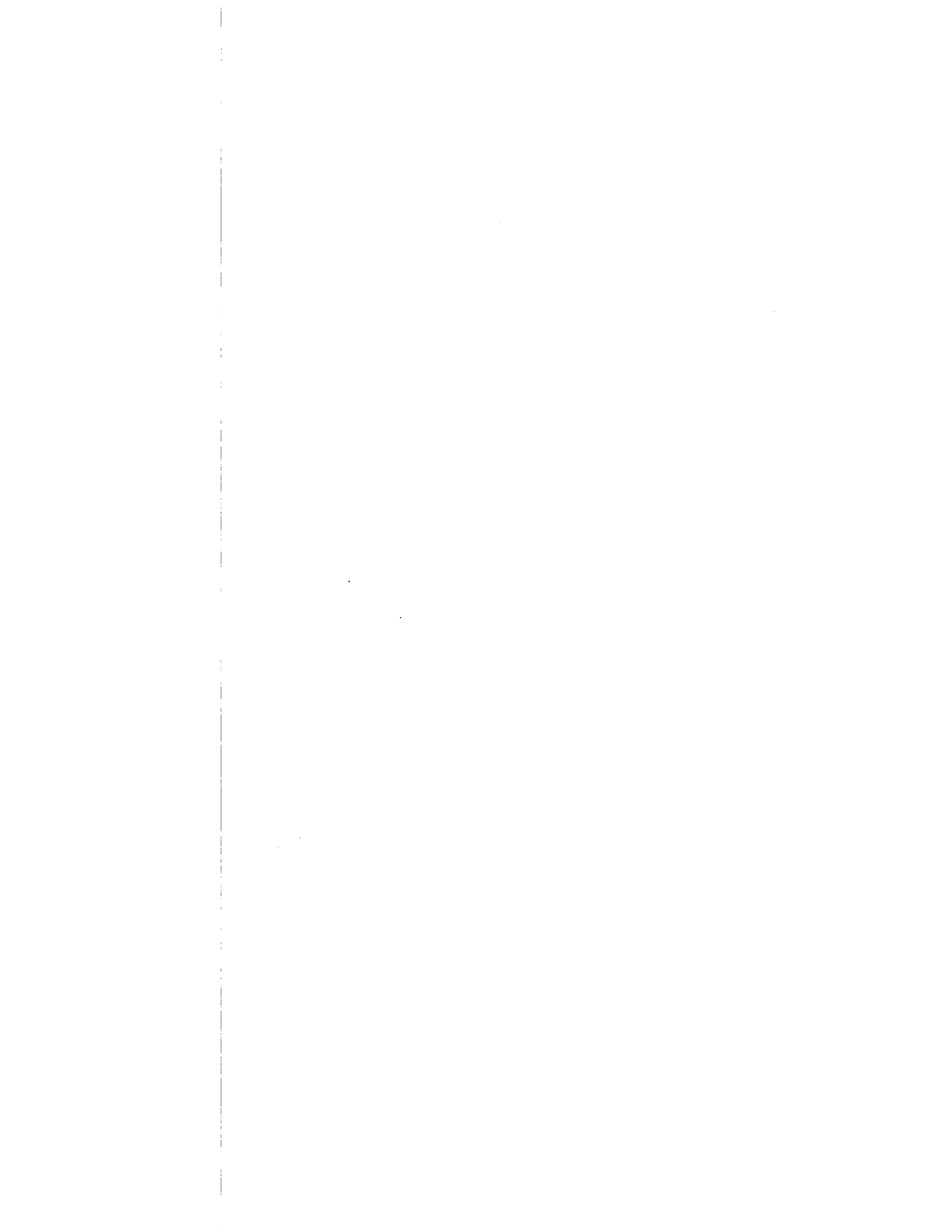
2. The term 'mood' is employed here as an (admittedly vague) cover term for the clitics that are employed to express subjective feelings of the speaker with respect to the action/state denoted by the predicate, including emphasis, hortative mood and diminutive. The group of 'mood' clitics can also be defined as follows: (i) they do not express aspectual or pronominal notions and (ii) they immediately follow the predicate XP. (For the contrast between 'mood' and 'modality', and the difficulties in defining these notions, see Bybee 1985).
3. **List of abbreviations:** A=Accusative, ART=Article (*na* = sg., *da* = plural), CAUS = Causative prefix, CNJ=Conjunction, CTR=Marker of control sentence, D=Dative, DEI=Deictic element (space/time), DEM=Demonstrative, EMP=Emphasis marker, G=Genitive, IMPF=Imperfective aspect marker, LOC=Locative preposition, MOD=Mood marker, N=Nominative, NEG=Negation, p=plural, PRF=Perfective aspect marker, REL=Relative marker, s=singular. **Notational conventions:** In the notation of the Kambera examples a clitic is separated from its (syntactic) host by a hyphen [-]. Accents on vowels mark contrastive vowel length. **Note on translations:** Third person singular pronominals in Kambera are neutral with respect to gender but are translated as 'he', 'him' or 'his', unless the context demands otherwise. Kambera verbs are not marked for tense and the tense used in the English translations was determined by the original context of the utterances.
4. *Daingu* 'surely' is a sentential adverb rather than a verbal adverb; it can be separated from the verbal projection by a conjunction, i.e. is not part of the 'nuclear clause' S (see below).
5. The relation between pronominal clitics and NPs is more complicated than this; see section 4.1 and Klamer (1994, to appear) for more discussion.
6. This marking is obligatory and is one of the ergative properties of Kambera, see Klamer 1997 for a discussion of this structure.
7. Possession and definiteness are notions that are structurally independent in Kambera. An NP (possessed or not) is indefinite if there is no article present, as in (i), versus (ii) where the NP is definite:
- (i)           Ningu       uma-nggu  
          be.here     house-1sG  
          'I have a house (lit. there (is) a house of mine)'
- (ii)           Wua-nja     na     uma-nggu  
          give-3pD   ART   house-1sG  
          'Give them my house'
8. There is no structural evidence to assume a category of Adjectives in Kambera. (Hence, there is also no adjectival projection). Adjectival notions are usually expressed by stative intransitive verbs. Bare verbs (active and stative intransitive, transitive) can be used to modify nouns.
9. Note that the structure of a Kambera verbal projection (VP?) differs from what is standardly assumed because the verb and its complement (expressed as an accusative or dative enclitic) do not form a separate constituent — adverbs, genitive enclitics and mood enclitics always intervene between the verb and the object marking clitics. See also the next note.
10. There is no evidence that Kambera clitics ever move at all. Nor do clitics ever occur before an interrogative, a conjunction or a negation. Interrogative pronouns *nggamu* 'who' and *nggara* 'what' may remain *in situ* but may also head a relative clause (i.e. 'who(m) did you see' expressed as 'who is the (one that) you saw').
11. Note that in this respect, some example sentences in (7)–(16) (chosen to illustrate the larger clitic combinations) are atypical because many of them contain four or five clitics and/or more than one aspect or mood clitic (see below).
12. This preference for smaller clitic clusters may be caused by prosodic restrictions of the language, see section 3.2 below.
13. Exceptions (1) disyllabic mood clitic *aru*, pronominal clitics *nggama*, *kama*, *nggami* (= *nggai*), *kami* (= *kai*) (2) stressed mood clitics *bia*, *aru*, *du* (cf. Klamer 1994: 30-32).
14. That is, the lexical head is the prosodic head as long as it is not modified by another lexical item (e.g. the adverb *hili* 'again' in (8) or the noun *pingi* 'top' in (16)), because then that item becomes the prosodic head.
15. Thanks to Ruben van de Vijver (p.c.) for suggesting this possibility.
16. Compounds are syntactically left-headed: *ana mini* 'child male' = 'son'.

17. Prosodic maximality is an important issue in phonological grammar. More particularly, many have argued that prosodic constituents are not n-ary branching (contra Nespor & Vogel 1986). That is, syllables, feet, prosodic words and phonological phrases are bounded by a maximal number of two or three subordinate constituents (cf. Helsloot 1995 and references cited there). A Kambera prosodic word would then consist of maximally two or three feet, i.e. approx. six or seven syllables.
18. Definite second objects of ditransitive verbs behave differently from objects of simple transitives — they may either be marked with a clitic or with an NP, depending on other factors such as discourse saliency (Klamer 1994 chapter 8).
19. An empty subject (no clitic, no NP) frequently occurs when it is known from the context.
20. Inflected indefinite subjects are used very rarely. An indefinite subject is usually expressed by the full indefinite NP only.
21. Note that the aspect expressed by the imperfective and perfective clitics in Kambera is not lexical aspect (or 'Aktionsart') but clausal aspect. Lexical aspect is expressed in Kambera by causative prefixes and applicative suffixes attached to a morphological base, not by clitics (cf. Klamer 1994).
22. Mutung is an intransitive verb which can be translated as 'burn' or 'be aflame'.
23. Di is a phonological variant of du, but du is used more frequently. The differences between du/di and ma are that du/di expresses stronger emphasis than ma and that du/di has its own stress while ma does not.

#### References

- Anderson, S.R. (1992). A-morphous morphology. Cambridge: CUP.

- Anderson, S.R. (1996). How to put your clitics in their place. *The Linguistic Review* 13, 165-192.
- Baker, M. (1988). *Incorporation: a theory of grammatical function changing*. Chicago: University of Chicago Press.
- Baker, M. (1995). *The polysynthesis parameter*. Oxford: Oxford University Press.
- Bittner, M. and Hale, K. (1996). Ergativity. *Linguistic Inquiry* 27, 531-604.
- Bloomfield, L. (1962). *The Menomini Language*. New Hampshire: Yale University Press.
- Bybee, J. (1985). *Morphology*. Amsterdam: Benjamins.
- Cutler, A.; Hawkins, J.A.; and Gilligan, G. (1985). The suffixing preference: a processing explanation. *Linguistics* 23, 723-758.
- Finney, J.C. (1997). It cometh from afar. Paper presented at the Third International Conference on Oceanic Linguistics. New Zealand, January 1997.
- Greenberg, J. (1966). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language*, Greenberg, J.H. (ed.), Cambridge: MIT Press.
- Hall, Ch. J. (1992). *Morphology and mind: a unified approach to explanation in linguistics*. London: Routledge.
- Halle, M. and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In *The view from Building 20*, K. Hale and S.J. Keyser (eds.), 111-176. Cambridge: MIT Press.
- Helsloot, K. (1995). *Metrical prosody: A template-and-constraint approach to phonological phrasing in Italian*. [Hil dissertations 16]. The Hague: Holland Academic Graphics.
- Huist, H. van der; and Klamer, M. (1996). The Uneven Trochee and the structure of Kambera roots. In *Dam Phonology*, M. Nespors and N. Smith (eds.), 39-57. The Hague: Holland Academic Graphics.
- Inkelas, S. (1993). Nimbora position class morphology. *Natural Language and Linguistic Theory* 11: 559-624.
- Klamer, M. (1994). *Kambera: a language of Eastern Indonesia*. [Hil dissertations 11]. Berlin, New York: Mouton De Gruyter. To appear.
- Klamer, M. (1996). Kambera has no passive. In *Voice in Austronesian* [NUSA Volume 39], M. Klamer (ed.), 11-30. Jakarta: Universitas Atma Jaya.
- Klamer, M. (1997a). Continuative aspect and the dative clitic in Kambera. In *Papers in Austronesian linguistics*, V. Phillips; I. Paul and L. Travis (eds.), 43-56. Dordrecht: Kluwer.
- Klamer, M. (1998). Kambera intransitive argument linking. *Studia Linguistica*. To appear.
- Klamer, M. (to appear). *A grammar of Kambera*. Berlin/New York: Mouton De Gruyter.
- Muysken, P. (1986). Approaches to affix order. *Linguistics* 24: 629-643.
- Nespors, M. and Vogel, I. (1986) *Prosodic phonology*. Dordrecht: Foris.
- Prince, A. and McCarthy, J. (1994). 'Generalized alignment.' In *Yearbook of Morphology 1993*, G. Booij and J. van Marle (eds.).
- Prince, A. and Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar*. MIT Press [to appear].
- Simpson, J. and Whitgott, M. (1986). Pronominal clitic clusters and templates. In *Syntax and Semantics 19: The Syntax of Pronominal Clitics*, H. Borer (ed.). New York: Academic Press.
- Silverstein, M. (1976). Hierarchy of features and ergativity. In *Grammatical categories in Australian languages*, R.M.W. Dixon (ed), 112-171. Canberra: Australian Institute of Aboriginal Studies.
- Spencer, A. (1992). Nominal inflection and the nature of functional categories. *Journal of Linguistics* 28, 313-342.
- Steele, S. (1995). Towards a theory of morphological information. *Language* 71: 260-309.
- Stump, G. (1992). On the theoretical status of position class restrictions on inflectional affixes. In *Yearbook of Morphology 1991*, G. Booij and J. van Marle (eds.), 211-241. Dordrecht: Kluwer.
- Zwicky, A. (1977). *On clitics*. Bloomington, Indiana: Indiana University Linguistics Club.



# PREDICATE FRONTING IN NIUEAN\*

Diane Massam  
University of Toronto

## 1. Introduction

In recent years, the issue of verb movement has been of central theoretical interest, in both SVO and VSO languages. This paper explores this issue through an examination of verb fronting in the Polynesian language Niuean, a VSO language. It is shown that the Niuean verb does not front to Comp or Tense, since another head, namely Neg, can appear between the complementizer/tense morpheme and the verb. This necessitates positing an additional functional head between Comp and VP, which we label Infl. It is posited that Niuean Infl encodes aspectual and force features, and can therefore contain a clitic complex consisting of aspectual modifiers, emphatics, a perfect marker and a question marker. We then turn to non-verbal predication in Niuean, concentrating on structures containing the morpheme *ko*. We argue that *ko* should not be considered a marker of predication, contra Seiter (1980). Instead, it is a default nominal extension, i.e. either a preposition or a case marker. These facts mean that nominal projections must be permitted to undergo predicate movement to check verbal inflection directly in Niuean by movement to Specifier of IP. It is thus seen that a single feature can be checked by more than one means.

## 2. Verb Movement

In this section we examine Niuean word order. Niuean exhibits a V/S/O/IO/Obl word order, as shown in (1) below.<sup>1</sup>

- (1) a. Ne tala aga e ia e tala ke he tagata  
Pst tell Dir3p Erg he Abs story to man  
"He told the story to the man." (L)
- b. Hifo a Lemani ki tahi mo e vaka  
go-down Abs Lemani to sea with Abs canoe  
"Lemani went down to sea with a canoe."(S52)

If we adopt the common assumption that the verb and its object originate within V', then all Niuean clauses involve verb movement. It is taken as a theoretical assumption that the verb originates in V', but there is in fact some evidence for this, in that verbs form a closer D-structure bond with their objects than with their subjects. For instance, verbs can incorporate objects, but not subjects, and verbs can form idioms in conjunction with their internal

---

\* I would like to thank members of the Syntax Project at the University of Toronto, the audience at AFLA III, UCLA, Liz Pearce, Yves Roberge, Wolfgang Sperlich of the Niuean Dictionary Project, and Lagi and Moka Sipeli for their help with this work. All errors are my own. Data sources other than fieldwork are listed in the references. This work has been supported by the Social Sciences and Humanities Research Grant #410-94-1093.

<sup>1</sup> Data is taken from a variety of sources such as C=Chapin (1974), L=Lane (1978), M=McEwen (1970), R=Rex et al (undated), S=Seiter (1980), Sp=Sperlich(to appear), W=Whittaker(1982). Page numbers are given for sources, except for Lane and Sperlich, which have no page numbers. For Sperlich, the head word is given, under which the example can be found. Some data has been verified by native speaker consultants Lingi Igasia and Lagi Sipeli. Glosses for some sentences have been added or changed for consistency. Orthography has been changed in that McEwen *ng* is written as *g*, as in standard Niuean orthography, and in that font limitations prohibit the use of macrons, so that double vowels have been substituted.

(absolutive) arguments (but not, it seems, with their external subjects).<sup>2</sup> Woolford's (1991) analysis of Niuean as having a flat VSO VP is thus not adopted.<sup>3</sup> The incorporation facts are shown below in (2), while (3) shows a verb-object idiom.

- (2) a. Kua hahala he tagata e akau.  
 Perf chop Erg man Abs tree  
 "The man is chopping the tree."(L)
- b. Kua hahala akau e tagata.  
 Perf chop tree Abs man  
 "The man is tree-chopping."(L)
- c. Faa totou he tau faiaoga e tau tohi.  
 Hab read Erg Pl teacher Abs Pl book  
 "(The) teachers often read books."(S73)
- d. \*Faa totou faiaoga e tau tohi.  
 Hab read teacher Abs Pl book  
 ("Teachers often read books.")(S74)
- (3) Loto a au ke oeli e tau matahui, ti koli.  
 like Abs I Sbj oil Abs Pl knee then dance  
 "I like to get a little drunk, then dance."(S191)

Given a verb/object constituent at the level of thematic representation, the Niuean verb undergoes movement to some c-commanding head position. Following Chomsky (1995), we assume that the features of a functional head may contain a strong feature which requires checking at S-structure, thus attracting some other element, which then undergoes a Move operation. A central problem in the study of VSO languages is to determine to which functional head the verb is attracted. The debate usually focusses on Comp and Infl (or Tense)<sup>4</sup> so we will begin our discussion with these categories.

In Niuean, the two categories of Comp and Tense are morphologically merged. The sentence begins with a particle which indicates the tense/aspect of the sentence.<sup>5</sup> These particles are given in (4)

<sup>2</sup>I do not have evidence that verb-subject idioms do not exist in Niuean, but I have not encountered any. Seiter (1980) gives (i) as a subject-verb idiom, but it arguably involves an unaccusative verb.

(i) Kua teitei pouli tei a fafo.  
 Perf nearly dark Perf Abs outside  
 "It's nearly dark." (S190)

<sup>3</sup>There are also arguments that the subject c-commands the object, arguing against a flat VSO S-structure and D-structure, again, contra Woolford (1991). In addition, there are strong reasons for assuming that at least absolutive arguments appear at S-structure in a non-thematic position (see Massam and Smallwood, to appear). This also argues against a flat structure.

<sup>4</sup>See for example, McCloskey (1993)(1994), Carnie (1993), Sproat (1985). Chung (1990) proposes a different derivation for some VSO languages.

<sup>5</sup>The semantics of the tense/aspect system is not discussed here, and I will consider the two categories to be one, labelled tense, for the purposes of this paper. Krupa (1982) considers that each particle denotes either tense or aspect, but not both. The analyses of these particles in Clark (1976), Krupa (1982), McEwen (1970), Seiter (1980), and Whittaker (1982) differ in detail. For example, Whittaker states that *kua* is present tense, while McEwen considers *kua* to indicate present, past or future. Seiter states that it is perfect, whether past, present or future. In addition, it appears that in the absence of an initial particle, the sentence might be past, present or future. I essentially follow Seiter's analysis, with one exception as follows. Seiter, Krupa and Clark give the marker, *ko e*, as a present tense/progressive marker but since the same element serves to topicalize, cleft, or



(4)

<u>PAST</u>	<u>FUTURE</u>	<u>PROGRESSIVE</u>	<u>PERFECT</u>	<u>SUBJUNCTV</u>
ne/na	to	haa ne	kua	kia

- a. Ne tagi a ia.  
Pst cry Abs she  
"She cried." (*Ko e Pusi* - a children's reader)
- b. To fano a ia  
Fut go Abs he  
"He will go." (W12)
- c. Haa ne nonofo a mutolu i hinei.  
Prog stay Abs you at this-place  
"Whilst you are staying here." (M66)
- e. Kua fano tuai a ia.  
Perf go Perf Abs he  
"He has gone." (M138)

The sentence-initial particle expresses the tense or aspect of the sentence. These particles, however, also display complementizer-like properties. Matrix clauses begin with the particles listed above, and so do sentential objects of some verbs, namely verbs of cognition, evaluation, observation, and speaking. These particles are in complementary distribution with other particles which do not have a tense function, but which are more clearly complementizers, such as the causal or factive particle *he*, seen in (5a). They are also in complementary distribution with the subjunctive particle *ke* which introduces clauses embedded under modal verbs, verbs of desire, and intention as in (5b). Finally, we find a partially distinct series of tense/aspect markers in relative clauses, as shown in (6)

- (5) a. Gagao foki nii a au he hifo a Maka ki tahi.  
sick also Emph Abs I Compl go-down Abs Maka to sea  
"I'm also sick of Maka going down to the sea." (S129)
- b. Tala age ki a ia ke hau.  
tell Dir3p to Abs him Sbj come  
"Tell him to come." (M12)

(6)

<u>PAST</u>	<u>PRESENT</u>	<u>FUTURE</u>	<u>PROGRESSIVE</u>	<u>PERFECT</u>
ne	ne	ka	ne faa e	(ne) kua

- a. Kua fakanaa e ia e tama ne tagi  
Perf pacify Erg he Abs child Nft cry  
"He pacified the child who cried." (*Sp.naa*)

form a predicate from a nominal, I do not consider it a tense marker (see Clark, 1976). It is not considered as a present tense marker in McEwen or Whittaker. Seiter notes that it behaves differently from the other tense markers in that at times the negative can precede it. This would argue that *ko e* properly belongs in the V-slot and not in the initial COMP/Tense position. See below for further discussion of *ko e*.

- b. ke he tagata ka nofo i kaina  
to person Fut stay at home  
"to the person who's going to stay at home"(S93)
- c. e tagata ne faa e onoono hake ke he mahina  
Abs man Nft Prog look up to Abs moon  
"the man who's looking up at the moon"(S93)
- d. e mena (ne) kua taute e Sione  
Abs thing Nft Perf fix Erg Sione  
"the thing Sione has fixed."(S93)

We can see that sentence-initial particles express information both as to the tense/aspect of the sentence, and to the grammatical status of the clause (matrix, subject clause, object clause, relativizing clause). In the case of object clauses, they also express information as to the selectional properties of the governing verb, i.e. whether it selects a subjunctive or a fully tensed clause.

Given the above facts, the initial particle is a portmanteau Complementizer/Tense element, CTP (i.e. Comp/Tense Phrase). It might also be the case that there is a separate CP and TP, where T raises to C, however the two analyses are identical in effect and I adopt the simpler of the two since there is no evidence for a TP specifier below C.

We can now question whether the verb fronting operation in Niuean involves fronting to CTP. In fact, this appears not to be so, at least in some cases. In order to show why this is so, we must examine the Niuean verbal complex, beginning with the preverbal elements. These are outlined below.

(7)

COMP/Tense	Negative	Auxiliaries	VERB
------------	----------	-------------	------

To nakai liu feleveia foki a taua.  
Fut not return meet also Abs we,Du,Inc  
"We will never again meet."(S16)

Note that the negative morpheme and auxiliary verb(s) intervene between the verb and the sentence initial CT particle. This means that if the verb is considered to always move to CT position, there must at some point in the derivation be an  $X^0$  of the form: [Tns-Neg-Aux-Verb]. It might be the case that the entire complex is a constituent at D-structure, in which case the entire complex moves to CT. Or, one or both of the Aux and Neg elements could appear in a head position, with the verb complex moving first to Aux, then to Neg, and so on

These views are problematic, because Neg appears to be an independent stem, itself able to host verbal clitics. There is a post-verbal perfect marker, which often co-occurs with the CT perfect marker *kua*. This element is *tuai*. It appears above in (4e). In a negative sentence, instead of appearing after the verb, it appears after the negative element.

- (8) Kua naakai tuai liu e tahi.  
Perf not Perf turn Abs sea  
"The tide has not turned." (S15)

As well as the perfect particle, there is another verbal clitic which shifts to the post-negative position in a negative sentence. This is an emphatic marker *laa* seen in (9). (Note the negative element *naakai* alternates with *ai* which is seen in (9)). It often co-occurs with *ia*. In a negative sentence *laa* appears on the Neg, while *ia* occurs on the verb.

- (9) a. Ai **laa** kitia e au e pusi.  
not yet see Erg I Abs cat  
"I have not yet seen the cat."(S15)
- b. Ne vali **laa ia** e maua e nofoa.  
Pst paint just Erg we Abs chair  
"We just now painted the chair."(S16)
- c. Naakai **laa** nofo **ia** a au he ha motu tufa a Niue.  
not yet live yet Abs I on Nsp island like Abs Niue  
"I've never before lived on an island like Niue." (S16)

A constituent question marker can also appear after the negative element.

- (10) Ai kia kitia e koe e laa kua tokoluga?  
not Q see Erg you Abs sun Perf high  
"Didn't you see the sun high up?" (S26)

It has been argued (Chung, 1970) that in other Polynesian languages the negative element is in fact a verb, since, for example, it houses verbal affixes and it takes as its complement a phrase which begins with an embedded CT element (cf. also Bauer (1993), Hohepa (1969), Pearce (1995), and Waite (1987)). In Niuean it does not appear that the negative element is a higher verb, since it does not appear with the full range of verbal clitics (outlined below) and it does not take a CTP complement. It certainly has a hint of verbal behaviour, though, in that it can attract *tuai* and *laa* (as opposed to the auxiliaries which do not do this). In this it behaves like a traditional auxiliary verb<sup>6</sup>. This independent morphological behaviour makes it unlike an affix or an element  $X^0$  adjoined to the verb.

Since none of the material surrounding Neg is phrasal, it might still be possible to maintain the claim that [Neg-*laa*-*tuai*-Aux-Verb-X-Y-Z] is a single  $X^0$  in  $V^0$ . (See below for discussion of post-verbal clitics.) But this leads to a complicated morphology since the order of morphemes within the lexical item would differ depending on whether or not it contained a negative morpheme, since in an affirmative sentence we find [Aux-Verb-X-*laa*-Y-*tuai*-Z]. It is hard to explain this change of morpheme order under a V-to-CTP view, whereas if we assume that Neg intervenes between CT and the fronted verb, we can explain why *tuai* and *laa* appear on the Neg element simply by observing that they are always on the element governed by CT (i.e. in second position).<sup>7</sup>

Finally, as well as being an independent morphological head, Neg has sentential scope rather than scope over the verb alone. It is preferable then, to consider Neg as an independent item in the syntax. We thus rule out the movement of V to CTP, at least in sentences where there is a Neg.

Auxiliaries in Niuean include desideratives, habitual, and elements meaning "look like", "nearly", "begin", among others (see Seiter, 1980). An example appears below.

- (11) Ne fia evaeva a ia ka e naakai talia he matua ke taute pihia.  
Pst want walk Abs he but not let Erg parent Sbj do so  
She wanted to go for a walk, but the parent wouldn't let her do so." (Sp *fia*)

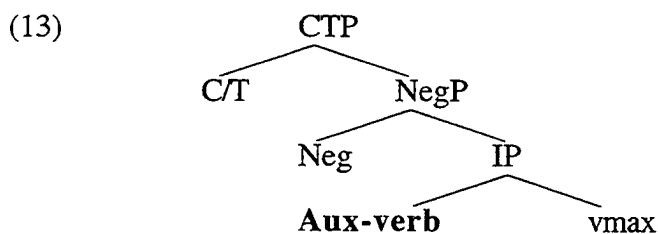
<sup>6</sup>Note that we cannot consider the Neg to be like auxiliaries in, for example, French, which, if present, usurp the inflectional position of the main verb. This is because in Niuean, even if there is a Neg, the verb still fronts. So whatever position Neg is in, it must still leave a position free for verb fronting.

<sup>7</sup>There are several emphatic elements. The first of these (*laa*) appears after Neg if it is present, but the others always appear after V.

Auxiliaries appear between Neg and V. They do not show evidence of independence from the verb.<sup>8</sup> In a sentence with a Neg, the perfect element *tuai* appears on Neg rather than the verb. In a sentence with no Neg, but with an auxiliary however, *tuai* appears on the verb, not on the auxiliary. From these facts we can generalize that *tuai* appears on the first [+V] head to the right of the perfect element in CTP. *Laa* also appears on the first [+V] element to the left of CTP. These generalizations entail that auxiliaries are either not [+V], or are not heads. Since several auxiliaries can also appear as independent verbs, I take the second option, and consider that the auxiliaries are verbal prefixes, or perhaps more correctly, that they form compounds with verbs. Under the entry for *fia* Sperlich (to appear) states that "... some speakers may want to consider such constructions either as compounds or as verbs with a prefix...". This view is supported, since the auxiliaries appear to be completely inert syntactically. This position is further (weakly) supported by Seiter's observation that one auxiliary is orthographically represented as a prefix on a verb (*fiakai* "hungry (lit. "want-eat"), *fiamohe* "sleepy" (lit. "want-sleep")). Similar support is found in the fact that several auxiliaries are redundant expressions of portions of meaning of the verb they appear with, eg. *faa* "habitual" with *mahani* "typical", or *fia* "desiderative" with *loto* "want" or with *manako* "desire". The generalizations above can now be more formally stated as below.

- (12) a. *tuai* is governed by Tense [+Perf].  
b. *laa* is governed by Tense

This requirement explains why these elements appear on the Neg element if there is one, and on the Aux-Verb complex if there is no Neg element in the sentence. It is possible that the element with *tuai* and *laa* (i.e. V or Neg) moves to the CT head in order for these morphemes to be checked. But note that, at least in sentences with Neg, there must still be some position other than CT for the verb to front to, as in the following structure, where there is a CTP head, followed optionally by a Neg head, then the Aux-V element. Following assumptions commonly adopted for other languages, we will call this position Infl, keeping in mind that the tense morphology does not actually appear in this slot. The question remains, then, as to what features are found in this Infl position. (We leave aside the question of where the subject and object should be positioned: See Massam and Smallwood (to appear)).



The functional projection to which the V moves does not appear to contain features or phonological material associated with any of the preverbal morphemes. We now turn to the postverbal morphology, arguing that Infl can contain a clitic complex, consisting of aspectual elements of various kinds, and an illocutionary force [ $\pm$  Q] operator.

The postverbal elements are shown below.

(14)

Adv <sub>s</sub> (a, b, c, g, i)	<i>aki</i> "with" (e)	<i>oti</i> "all" (f)	<i>ai</i> "pron" <sup>9</sup> (g)	Asp-Adv (h)	Emph <sub>s</sub> (c, h, i)	Perf(d, e, f, i, j)	Question (j)
----------------------------------	-----------------------	----------------------	-----------------------------------	-------------	-----------------------------	---------------------	--------------

<sup>8</sup>One auxiliary *liga* "likely", does show evidence of being an independent item, since it is followed by a tense marker, and can appear in front of Neg. I leave this auxiliary aside here.

<sup>9</sup>See Chapin (1974) and Massam and Roberge (1996) for discussion of this pronominal element.

Adverbs (directional and manner particles): *mai* "towards 1p", *atu* "towards 2p", *age* "towards 3p", *hake* "upwards", *hifo* "downwards", *lahi* "very", *fakamitaki* "well", *fakaeneene* "carefully"

Aspectual Adverbs: *tuumau* "always", *hololoa* "frequently", *agaia* "still", *agataha* "immediately"

Emphatic Particles *noa* "only" (*nii* "indeed"), *foki* "also" (*nii*), *laa* "just" (*ia*), *koa* "indeed"

- a. Kua hoge **lahi** e motu  
Perf starve greatly Abs island  
The island is greatly starving (Sp *lahi*)
- b. Ne oo **mai** a lautolu ki hinei  
Pst go(pl) **Dir1** Abs they to this=place  
"They came here." (M180)
- c. Mumui **mai nii** a lautolu he motaka ha lautolu  
follow,Pl **Dir1 just** Abs they in car of them  
"They'll just follow (us) in their car now.(S18)
- d. Kua hele **tuai** e Sione e falaoa aki e titipi haana.  
Perf cut **Perf** Erg Sione Abs bread with Abs knife his  
"Sione has cut the bread with his knife."(S243)
- e. Kua hele **aki tuai** e Sione e titipi haana e falaoa.  
Perf cut **with Perf** Erg Sione Abs knife his Abs bread  
"Sione has cut the bread with his knife"(S244)
- f. Kua iloa **oti tuai** e lautolu a au.  
Perf know **all Perf** Erg they Abs me  
"They all know me"(L)
- g. ti laga aki **hake ai** e Ataraga a Maui ki luga ..  
Then cause move up then Erg Ataraga Abs Maui to top..  
"Then Ataranga raised up Maui ..." (C98)
- h. Kata **tumau nii** a ia.  
laugh **always just** Abs he  
"He's just always laughing."(S24)
- i. Kua uku **hifo foki tuai** a au ke he toka.  
Perf dive **down also Perf** Abs I to bottom  
"I have dove down to the bottom before." (S24)
- j. Ita **tuai nakai** a patu na?  
angry **Perf Q** Abs guy that?  
"Is that guy angry?"(S25)

The first four elements all involve VP-internal relations, and accordingly, we consider them to be generated on the verb within VP. The aspectual adverbs, the emphatics, the perfect, and the Question marker (*nakai*) on the other hand, have sentential scope and we propose that they appear as a clitic complex in Infl.<sup>10</sup> This accounts for their position in the clause. *Tuai* and *laa*

<sup>10</sup>Evidence from predicate nominals, discussed below, argues against these elements being base-generated on the

are members of this clitic complex in affirmative sentences, but they form their own complex in negative sentences. The relative positions of the clitics are accounted for in the morphological component since these elements are syntactically inert i.e. they do not interact with any other syntactic constituents.

Niuean clauses contain a verb with VP-internal affixes which raises to adjoin to a clitic complex in X.<sup>11</sup> This clitic complex may contain aspectual adverbs, emphatics, a perfect marker, and a question marker. All of these elements, including the emphatics involve some form of aspect or force. The government requirements on *tuai* and *laa* will ensure that they do not appear in this clitic complex in case of a negative sentence, but instead form an independent clitic complex attached to Neg. The Niuean verb is thus part of a complex syntactic entity made up of X<sup>0</sup> adjunctions which must then be accommodated into X-bar theory. (cf. Ghomeshi (1996)). Niuean thus presents an interesting case for the study of how functional elements are distributed differently across functional projections in different languages such as discussed in Moorcroft (1995).

Niuean VSO order is thus derived by movement of V to INFL, where it adjoins to the clitic complex which is generated in INFL. We now turn to copular sentences which appear to pose a problem for this view.

### 3. Nominal Predicates

Nominal predicates in Niuean appear with the prenominal element *ko*. The *e* after *ko* is an article which does not appear before proper names and pronouns.<sup>12</sup>

- (15) a. **Ko e kamuta a au**  
 "ko" Art carpenter AbsPropArt I  
 "I am a carpenter" (S53)
- b. **Ko Pule e faiaoga**  
 "ko" Pule AbsArt teacher  
 "The teacher is Pule" (S54)
- c. **Liga ko e uga a ia koo**  
 Likely "ko" Art crab AbsArt at there  
 "That there is probably a coconut crab." (S54)
- d. **Ai ko e faiaoga a Pule**  
 not "ko" Art teacher AbsPropArt Pule  
 "Pule's not the teacher." (S54)
- e. **Ko koe nakai a ia ma Haliua?**  
 "ko" you Q AbsArt that Voc Haliua  
 "Is that you, Haliua?" (S54)

---

verb, since they also appear after a maximal projection predicate. This analysis is similar in spirit, though not in detail, with that of Stuart, (1996). She argues for movement to what she calls Mood Phrase, which houses the emphatics and the question marker. The examination of Neg, *laa*, and *tuai*, as outlined in this paper has led me to modify her analysis.

<sup>11</sup>We will not explore here the issues raised by the first four postverbal elements such as incorporation and quantifier floating.

We do not consider the *e* after *ko* to be the absolutive case marker, since it has a different distribution in that the absolutive marker, unlike this article, is not deleted before proper names or pronouns, but is realized as *a*.

- f. **Ko e tau kamuta fakamua** a lautolu  
 "ko" Abs Pl carpenter before Abs they  
 "The were carpenters before this." (S54)
- g. [**Ko e poka-aga he tama e maka**] [ati matakutaku ai e kulii]  
 "ko" Art push-ing GenArt child AbsArt rock [reason<sub>i</sub> fear pron<sub>i</sub> AbsArt dog ]  
 "The reason the dog was afraid was the child's pushing the rock." (L)

The examples above show that the *ko e NP* is in the same morpho-syntactic slot as the verb, since it follows the auxiliary (15c) and the Neg (15d), and precedes the aspectual adverbs (15f), the emphatics (cf. (16b) below) and the question marker (15e). Note that this supports the division of the post-verbal particles into two groups, those base-generated on V and those base-generated in INFL, since the former do not seem to appear after *ko e NP*. In (15g) It is shown that the predicate nominal is a maximal projection.

These sentences are interesting because they show that Niuean does not exhibit verb fronting, but rather predicate fronting, where the predicate can be a predicate nominal. We assume that in these cases, the predicate is moving to Spec of Infl, rather than to the head position, and that it is moving to check a predicate feature in Infl, just as the verb does. The theoretical implications of this are discussed in Massam and Smallwood (to appear). A similar phenomenon in Irish is discussed in Carnie (1995).

The question arises as to the meaning of *ko*. Seiter (1980) considers it to be a predicate marker, glossing it as "Pred". This is problematic, however, when one examines further data. *Ko* is used in a variety of contexts. Various linguists have discussed various subsets of these uses, but nowhere can be found a comprehensive list of uses. Listed below are all contexts which I am aware of.

(16)

- a. **Predicate Nominals**  
 (see above)
- b. **Clefts** (=predicate nominals with sentential complement)  
**Ko e tipolo agaia nii ne inu ai a lautolu**  
 "ko" Abs lime still only Nft drink Pron Abs they  
 "It's still only lime juice that they are drinking" (S100)
- c. **Questions** (=clefts)  
**Ko hai ne fifili a koe ke vagahau?**  
 "ko" who Nft choose Abs you Sbj speak  
 "Who chose you to speak?" (S109)
- Ko hai?**  
 "ko" who  
 "Who?" (R2)
- d. **Topics** (not like clefts, see Seiter (1980))  
**Ko e matua fifine haana, mate tuai**  
 "ko" Abs parent female his die Perf  
 "As for his mother, she's dead." (S116)

- e. **Appositionals**  
 Hifo a ia ke he maaga ia ko Alofi  
 go AbsArt he to Art village that "ko" Alofi  
 "He goes down to that village Alofi." (M132)
- f. **Conjunctions**  
 Po ko Sione po ko Tavita  
 Either "ko" Sione either "ko" Tavita  
 "John or David" (M132)
- g. **Titles**  
 Ko e Pusi  
 "The Cat" (title of children's reader)
- h. **Noun Phrases in Isolation**  
 Ko koe "you"  
 Ko e tagata "a man" (W6)
- i. **Progressives/Actual Present**<sup>13</sup>  
 Ko e kumi agaia au he tama haau  
 Pres Art seek still I at child your  
 "I'm still looking for your child" (S5)

We conclude from this that *ko* is not a predicate marker, except in the negative sense of being a marker which appears on NPs which are not arguments. Clark (1976) considers it to be a preposition, and we follow this analysis here, although it might be better analyzed as a case marker. Thus, extended nominal phrases in Niuean can front to (Spec of) INFL if they constitute the predicate of the sentence, since the predicate is the locus of the aspectual nature of the sentence, whether it be verbal or nominal.

### 3. Conclusion

This paper has analyzed the nature of Niuean VSO, arguing first that this order is derived, and second, that it is derived, not by movement to Comp, nor to Tense, but rather to a lower inflectional head (Infl) which houses an aspectual/force clitic complex. Third, it was shown that verb fronting is to be more properly considered as predicate fronting, since maximal phrases which are demonstrably non-verbal also front in the language. It is hoped that with this analysis, Niuean can be placed in the typology of VSO languages and compared with other more well-studied languages of this type.

### References

- Bauer, W. 1993. *Maori*. London and New York: Routledge.  
 Carnie, A. 1993. "Nominal Predicates and Absolutive Case Marking in Irish." In C. Phillips (Ed.), *Papers on Case and Agreement II* MIT.

---

<sup>13</sup>Seiter (1980) and Krupa (1982) consider *ko* here to be a CT marker, patterning with others such as *ne*, *to*, *kua*, and the null marker. But Neg precedes "progressive" *ko e*, whereas Neg follows CT markers. So *ko e* is not a CT, but instead it is in Infl. *Ko e* only appears to be in the CT slot in affirmative sentences because it is in complementary distribution with a CT morpheme, hence it is the first element in the sentence.



- Carnie, A. 1995. *Non Verbal Predication and Head Movement*. Unpublished doctoral dissertation. MIT, Cambridge.
- Chapin, P. G. 1974. "Proto-Polynesian \*ai." *Journal of the Polynesian Society* 83:259-307.
- Chomsky, N. 1995. *The Minimalist Program*. Cambridge: MIT Press.
- Chung, S. 1970. *Negative Verbs in Polynesian*. ms. University of California, Santa Cruz.
- Chung, S. 1990. "VPs and V Movement in Chamorro." *NLLT* 8:559-619.
- Clark, R. 1976. *Aspects of Proto-Polynesian*. Auckland, New Zealand: Linguistic Society of New Zealand.
- Ghomeshi, J. 1996. *Projection and Inflection: a Study of Persian Phrase Structure*. Unpublished doctoral dissertation. University of Toronto.
- Hohepa, P. 1969. "Not in English and *kore* and *eehara* in Maori." *Te Reo* 12:1-34.
- Krupa, V. 1982. *The Polynesian Languages*. London: Routledge and Kegan Paul.
- Lane, C. 1978. *Niuean Field Notes*. unpublished ms. Victoria University of Wellington New Zealand.
- Massam, D., & Roberge, Y. 1996. "Operator Bound Clitics and Niuean *ai*." In J. Black & V. Motapanyane, eds. *Checking and Movement* Amsterdam: Benjamins.
- Massam, D., & Smallwood, C. (to appear). "Essential Features of Predication in English and Niuean." NELS26. GLSU, University of Massachusetts, Amherst.
- McCloskey, J. 1993. *Clause Structure, Ellipsis, and Proper Government*. ms. University of California at Santa Cruz.
- McCloskey, J. 1994. "Subjects and Subject Positions in Irish." *Linguistics Research Centre Working Paper* 94-05.
- McEwen, J. M. 1970. *Niue Dictionary*. Wellington, New Zealand: Dept. of Maori and Island Affairs.
- Moorcroft, R. 1995. *Clause Level Functional Categories in V2 Germanic Languages*. PhD dissertation. University of Toronto.
- Pearce, E. 1995. "Operator Binding and the Syntax of Negation in Maori." Paper presented at the Syntax and Semantics Conference. U. of Ottawa.
- Rex, T., Viviani, Aue, Vilitama, Togatule, & Talagi, eds. (undated). *Everyday Words and Phrases in English and Niuean*. USP, Niue Centre.
- Seiter, W. 1980. *Studies in Niuean Syntax*. New York: Garland Press.
- Sperlich, W. (to appear). *Niuean Dictionary*. Niue Island: Dictionary Project.
- Sproat, R. 1985. "Welsh Syntax and VSO structure." *Natural Language and Linguistic Theory* 3:173-216.
- Steuart, S. 1996. "The Structure of the Verb Complex in Niuean." *Toronto Working Papers in Linguistics* 14.2:127-158.
- Waite, J. 1987. "Negatives in Maori: a Lexical-Functional Approach." *Te Reo* 30:79-100.
- Whittaker, G. 1982. *The Niuean Language: An elementary grammar and basic vocabulary*. Alofi, Niue: USP, Niue Centre.
- Woolford, E. 1991. "VP-Internal Subjects in VSO and Nonconfigurational Languages." *Linguistic Inquiry* 22:503-540.



# The syntax and semantics of *olsem* in Bislama

Miriam Meyerhoff and Nancy Niedzielski  
University of Pennsylvania and University of California, Santa Barbara<sup>1</sup>

## 1. Introduction

In Bislama, an English-lexified creole spoken by c.150,000 people in Vanuatu, the preposition *olsem* ('like') occurs in a number of syntactic positions.<sup>2</sup> These are, at first glance, surprisingly heterogeneous. However, this paper will show that this is orderly heterogeneity at the levels of syntax and conceptual structure. We will show that

- (i) it has been possible for *olsem* to be reanalyzed in these ways because the different syntactic functions of *olsem* continue to share a core meaning, that is, the path of grammaticalization is constrained by semantic factors, and
- (ii) some reinterpretations of the meaning of *olsem* have been facilitated by the fact that *olsem* appears to have the syntax of a stem with an object clitic, that is the path of grammaticalization is constrained by syntactic factors as well.

The distribution of *olsem* will allow us to evaluate and ultimately challenge the functional claim that grammaticisation is unidirectional, i.e. that concrete meanings give way to more abstract meanings.

## 2. Origin of *olsem*

*Olsem* is derived from two English lexemes: 'all' > *ol* and 'same' > *sem*. These forms continue to exist in Bislama as separate lexical items though *ol* nows functions as the plural definite article, i.e. '(all) the'. Historical records<sup>3</sup> show *olsem* appearing very early in Bislama, and apparently during Bislama's years as a pidgin the meaning of *olsem* was fairly transparent semantically. So in (1a), the sense was something like 'you (are) all/entirely the same (as) Tiapolo'.

- (1) a. You plenty lie! You all same Tiapolo! (1867, in Paton 1895:39)  
'You lie all the time! You're like Tiapolo!'

However, even at this stage in Bislama, *ol* was beginning to bleach semantically and is found occurring with singular referents. Crowley's records show that *olsem* first occurs as a means of equating two nominals as in (1) and (2), and somewhat later appears as a means of equating two clauses. In current Bislama newspaper writing, *olsem* continues to occur most frequently to equate nominals, as shown in (3). Example (4) shows that clausal comparisons are also made using *olsem* in modern Bislama.

- (2) Me savvy fight all same white man (October 6 1912, Asterisk 1923:55)  
'I can fight like white men do'
- (3) *ol draeva oli yusum rod ya olsem wan men strim*  
'drivers use that road like a main thoroughfare'  
(*Vanuatu Weekly/Hebdomadaire* 17 June 1995)

---

<sup>1</sup> To contact the authors: mhoff@hawaii.edu; nn@STL.Research.Panasonic.COM

<sup>2</sup> We grateful to Sharon Tabi for her help as a native speaker, and to Terry Crowley, Sandra Thompson, Gillian Sankoff, Susanna Cummings and Sabine Iatridou for discussions of earlier drafts of this paper. This paper is a revision of the one presented at AFLA III at UCLA. Fieldwork by Miriam Meyerhoff was generously supported by the Wenner-Gren Foundation (grant #5742).

<sup>3</sup> We are grateful to Terry Crowley for sharing his comprehensive records of Bislama from the 19th and early 20th centuries.

- (4) *hop ... i ded olsem yu kapsaedem wota long graon* (VHW 7 Jan 1995)  
'hope is gone, like/as if you'd poured water on the ground'

In addition to these uses of *olsem*, we also find from an early stage in the development of the language, examples of *olsem* being used as in (5)

- (5) "Pourquoi n'as-tu pas fait cela?"  
"Belong me no save make all same ...." (1880s, in Imhaus 1890:39)  
"Why didn't you do that?"  
"Because I'm not allowed to do that..."

Crowley's (1990) diachronic grammar of Bislama analyzes *olsem* in (5) as an adverb, equivalent to English 'thus', and his dictionary (Crowley 1995) gives the two primary meanings of *olsem* as adverb and preposition. The analysis of the use of *olsem* in modern spoken Bislama that we present in this paper argues against calling *olsem* an adverb. We have identified a number of different functions in addition to those evinced in Crowley's (historical and 1995) data, none of which we would argue are adverbial. In section 3 of the paper, we demonstrate how in speech *olsem* is used as (i) a preposition, (ii) a complementiser, (iii) a demonstrative, and (iv) an anaphor.

We will argue that it has two discourse functions, one to focus, or shift attention and one to attenuate or hedge. In section 4 of the paper, we will consider the semantic and structural relationships between these functions and discuss what can be inferred about grammaticization paths in Bislama, and cross-linguistically. Our conclusions argue against claims made in the functional literature for unidirectional paths of change.

### 3. Functions of *olsem* in spoken Bislama

In (3) and (4), we have already noted the use of *olsem* as a preposition meaning 'like' in the national weekly newspaper. Similar uses in conversation are not uncommon, but in spoken Bislama, *olsem* is used as more than just a preposition. *Olsem* also occurs as a complementiser with a verb of speech (6a) or perception (6b).<sup>4</sup>

#### *Olsem as a complementizer*

- (6a) *Lili toktok olsem long hem, "e yu..."*  
'Lily spoke to it (thus/like this), "hey you..." (S-95-10)
- (6b) *no be mi mi luk olsem bubu ya i kakae fulap tumas*  
'no, but it seems to me (like/that) grandma here eats too much' (M-95-16)

As a preposition, *olsem* expresses identity between two nouns, events or states; in (4), it expresses identity between the speaker's state or speech activity and what they perceived or said. Note that with verbs of speaking, *olsem* is not restricted to introducing verbatim quotes. Given the (constructed) sentence (7):

- (7) *Sale i talem olsem mi sik*  
Sale said that I was sick; Sale said "I am sick"

*mi sik* can refer to either Sale or the speaker, the ambiguity being resolved in context.

<sup>4</sup> Examples of spoken Bislama are taken from recordings made with 40 speakers in the Santo-Malo region in 1994-95. Copies and transcripts are held in the Vanuatu Cultural Centre (Vila). The code following each example identifies the specific transcript in the corpus, M=Malo (village) speaker; S=Santo (urban) speaker. Further social information about the speakers can be obtained from the first author.

A cognate lexeme, *ol(o)sem*, exists in Tok Pisin. The Tok Pisin term, like Bislama *olsem*, derives from the English 'all' and 'same'. Verhaar's (1995) analysis of the structure of modern Tok Pisin shows that *olosem* has grammaticized in a similar way in Tok Pisin, and like *olsem* is now used as a complementiser. Gillian Sankoff's recordings of urban speakers of Tok Pisin from the 1970s<sup>5</sup> show that among her speakers *ol(o)sem* was overwhelmingly used to introduce reported speech, or to refer back to reported speech. We will see that the possibility of using *olsem* as a clausal anaphor has also developed in Bislama.

*Olsem* is also used as a "true" demonstrative (Kaplan 1989: 490). In (8), for instance, *olsem* is typically accompanied by an illustrative gesture.

*Olsem as a demonstrative*

- (8) *foltem han olsem ya*  
[we] fold [our] arms like this [speaker crosses arms] (M-95-11)

We also find instances of *olsem* being used in the manner of a meta-linguistic demonstrative. So for example, a typical (though, here, constructed) use of *olsem* includes the act of identifying or naming something:

- (9) *mifala i kolelem olsem: nagaviga*  
we call [them] this: 'nagaviga'

*Olsem* is found more frequently as a preposition or complementiser than it is found as a demonstrative in our corpus of modern spoken Bislama. The historical (written) record does not always provide a reliable record of all uses of a token as a demonstrative (examples like (8), or pointing demonstrations, have to be inferred from context at best), and the historical data on *olsem* is not extensive. As a consequence, we cannot state definitively when this function of *olsem* emerged. We can tentatively infer that it is a more recent development given its synchronic frequency compared with other forms that *are* attested in the historical record, but we are conscious of the fact that absence in the record – particularly when the record is so patchy – does not preclude it having been there. However, we suggest that the development of a demonstrative function for *olsem* might arise naturally from its use as a complementizer.

Verhaar (1995: 445) notes for Tok Pisin, that it is sometimes difficult to know whether in, e.g. an utterance like (9), *olsem* is a complementizer or a demonstrative (and cf. Eng. 'that' which also doubles as a complementizer and a demonstrative). It is true that intonation may resolve the ambiguity, but the ambiguity may also be tolerated. The possibility of both underlying analyses must be the factor facilitating the change of categories in the first place.

Consider, too, that there is a semantic connection which might facilitate a path between complementizer and demonstrative. As a complementizer, *olsem* asserts identity between (i) the state or action of the main verb, and (ii) the content of the subordinate clause. We might say that one function of *olsem* is to convert the subordinate clause proposition into a <e> type variable, predicated of the main verb.

This function of asserting identity is also a property of a demonstrative pronoun. Where the demonstrative is a true demonstrative, i.e. preceded or accompanied by a real world gesture, as in (8), *olsem* is the linguistic identification of the gesture.

Given that *olsem* frequently functions as the identification of paralinguistic gestures, it is perhaps not surprising to find that it is also used as a meta-linguistic

<sup>5</sup> Many thanks to Gillian Sankoff who identified a quantitative difference between Bislama and the closely related Tok Pisin, and who extracted all tokens of *olosem* from her corpus of urban Tok Pisin in order to make this comparison.

identifier. That is, we find speakers using *olsem* not just to report speech, but to stand in place of reported speech, as in (10).

- (10) *hem ya i tok bigwan mo long mi, se olsem olsem*  
'she said the most to me, she said "blah, blah, blah"' (S-95-13)

This starts to look rather similar to the use of *olsem* in (5), which Crowley suggested was an adverb. We believe that a more unified account of the grammar and a more plausible way for speakers to process *olsem* is if in (5) *olsem* is analyzed as a clausal anaphor, meaning 'like this/that', as in (11).

- (11) *mi talem man blong mi se ating bae mi stap long wok, wan samting i hapen olsem ya...*  
'I told my husband that say if I was at work and something like that [just discussed] happened...' (S-95-9)

Now, it might be possible to analyze *olsem* in (8) or (11) as a preposition whose complement has been deleted for discourse reasons. Meyerhoff (in press) shows that where the second element in an equative comparison is hearer-old information (Walker & Prince, 1996), the second element, or comparator, can be omitted, and this is not an isolated instance in the grammar. Discourse and syntactic constraints determine when complements of transitive verbs as well as prepositions may be omitted in Bislama.

A very productive aspect of verb morphology in Bislama is the transitive suffix *-em* (and its allomorphs *-im*, *-um*). The suffix occurs on transitive verbs and can change an intransitive verb into a transitive one (there is some variation among transitive verbs as to whether they take the suffix and how they mark the object theme, the constraints on this variation are discussed in Meyerhoff 1996). It is clear that *-em* was originally referential and derived from the English 3s pronoun 'him'. Gradually, this became a clitic and has now been reanalyzed as a transitive marker (Crowley 1990: 286ff). Thus, in modern Bislama an *-em* marked verb can be non-redundantly followed by a third person pronoun, e.g. (12), indicating that *-em* is not itself an incorporated argument.

- (12) *bae hem i singaotem hem*  
'she would call it' (S-95-10)

It is not entirely clear whether *-em* is an object agreement marker since its presence on transitive verbs is variable (and certainly any notion of agreement would have to be a very poor one indeed, as there is no differentiation for person or number), but *-em* is a very strong indicator that the verb selects an internal argument. For this reason, *-em* plays an important role in licensing null objects, as shown in (11).

- (13) *bae yu karem top long rop ya, yu givim e long mi*  
'you'll take the end of the vine, [and] you give [it] to me' (S-95-10)

Huang (1984) distinguishes empty objects in Chinese from true null (subject) pronominals. He argues that null objects are always the result of leftward-topicalization of a null object, a process which creates an operator-variable construction and that this is what allows the object to be interpreted. Generalizing this analysis to Bislama, we would say that the transitive marker *-em* is triggered by the presence of the null argument, but that interpretation of the object occurs independently of the verbal morphology.

Although the *-em* in *olsem* is derived from 'same' and not 'him', there is evidence that speakers have reanalyzed it as being the same morpheme as the transitive suffix. Two distinct classes of preposition exist in Bislama. One class is composed of *olsem*, *from*

'because of' and *wetem* 'with', which all look like they have the transitive suffix. This class of prepositions can be stranded. Preposition stranding is an unusual property in Oceanic languages, but the argument of these prepositions can be raised, focused or deleted, (12). The other class of prepositions consists of the locative *long* and the possessive *blong* and these prepositions can never be stranded. If any raising occurs for syntactic (13a) or pragmatic (13b) reasons, a resumptive pronoun is obligatory (a fact first commented on by Crowley 1990: 84-85).

- (12a) ...*from taem hemi smol pupu hemi mekem wan nakamal olsem* [e]  
'...because when he was a boy he made a nakamal like [this]' (VWH  
3/12/94)
- (12b) *nius ya, mi krae from* [t]  
'that's the news I cried about' [constructed example]
- (13a) *woman we mi givim mat ya long* (hem/\* $\emptyset$ )  
the woman that I gave this mat to [constructed example]
- (13b) *wan olfala transpot, hem i stap draeva long* (hem/\* $\emptyset$ )  
it's an old truck that he drives (in) [constructed example]

So it would be possible to analyze *olsem* in (8) as equating two actions, where the comparator action is omitted because the act of demonstration makes it retrievable information. But a deletion analysis is problematic for (5) or (10), where the benchmark, or first element in the comparison, would also have to be analyzed as null. In (5) and (10), *olsem* seems to be a clausal anaphor. However, if we analyze (5) and (10) differently to (8) and (11), we seem to be obscuring the similarity between the four utterances, without achieving any greater simplicity in our description of the grammar. For these reasons, we propose a unified classification of *olsem* in (5), (8-11) as a demonstrative pronoun (true demonstrative and clausal anaphor). Thus, our inventory of the syntactic functions of *olsem* in their apparent diachronic order of appearance, is: preposition, demonstrative/anaphor and complementiser.

Finally, we note two discourse functions of *olsem* one of which boosts, and the other of which hedges, the following element. In (14a), *olsem* is used to self-correct, refocusing the interlocutors' attention and clarifying common ground, and we also find *olsem* being used like Eng. 'anyway' to return to a previous topic after a distraction, again shifting or focusing the interlocutors' attention (14b).

- (14a) *jif blong olgeta... we bigwan long Bankis, olsem Mere Lava*  
'their chief... who was a big man in the Banks, or rather/anyway in Mere  
Lava' (S-94-5)
- (14b) [aside:] *hem i stap swim.*  
[to others:] *i olsem ya. Iawia i gat blong hem ating...*  
[aside:] 'she's having a bath'  
[to others:] 'Anyway. I think Iawia has one turtle...' (M-95-19)

In (15), *olsem* hedges or attenuates the force of the speaker's assertion.

- (15) *afta olsem hem i jas kambak, be afta i olsem long, i tudak, a*  
'and like he'd just come home, and it was like, it was night, eh' (M-94-2)

When *olsem* is used as a hedge, it has the same ambiguity that most hedges do; sometimes it is primarily a semantic signal indicating the speaker's lack of certainty, and sometimes it

is primarily an interpersonal device that indicates the speaker's affective orientation to other participants (Holmes 1988).

#### 4. Implications for theories and paths of change

As noted earlier, the synchronic distribution of *olsem* in Tok Pisin is heavily weighted towards its use as a complementiser with the *verba dicendi*. It seems that in Tok Pisin, the grammaticization process has been almost suppletive; the original meaning has become obsolete and the newer, derived sense has taken over. However, there is no such evidence that a similar process of replacement is occurring with *olsem* in Bislama. The frequency and distribution of all the functions of *olsem* in the speech of 24 speakers of Bislama over apparent time do not show any evidence that the older functions of *olsem* (preposition and demonstrative anaphor) are disappearing in favor of the newer ones. Nor does the data show any sociolinguistic stratification on the basis of the salient social distinctions in these communities (speaker sex, degree of urbanness, speaker's chronological age or marital status, education). There seems to be no synchronic evidence that the multiple functions we have outlined are links in a chain that is being pulled, as it were, through the speech community from a state in which there is one single function for *olsem* to a state in which there is another single function for *olsem*. The change underway is additive and has not acquired social significance. There are clear semantic connections between some of the uses of *olsem*, which we have pointed out in presenting the data, and these probably reflect the grammaticization path, but it is not clear that there is a unidirectional, continuous grammaticization path, as is assumed in some definitions of grammaticization (e.g. Heine et al. 1991:72, Hopper & Traugott 1993). The changes we have documented for *olsem* seem to need a different model of change.

The situation for *olsem* seems to resemble the radial structures proposed by Lakoff (1987) to account for semantic shift. All the functions of *olsem* in spoken Bislama continue to share the core meaning in (1), i.e. that of equating or establishing similarity between things, and they satisfy the other requirement for radial structures, i.e. that the non-central (non-preposition) meanings have different, not just more properties, than the central case, and these non-central meanings are linked to the central meaning by processes such as metonymy and metaphor (1987:84,379).

As a preposition, *olsem* claimed identity between two phrases, as a complementiser, *olsem* essentially claims identity between the state or event in the main clause and what is expressed in the subordinate clause. In extreme cases, we have seen that it even stands in for what was said, ultimate identity, if you like. As a clausal anaphor, *olsem* stands in for whole chunks of the preceding discourse, asking the hearer with one word to identify information that is hearer-old. As a means of indicating a shift in attention and as a hedge, there is also underlying a notion of likeness (either expressing more or less precise identity).

A radial structure for the semantic shift of *olsem*, as shown in Figure 1, allows us to express the links that we presume some meanings have with others, and the link all of them have with (1).

A radial structure is somewhat different to the proposals are on record to account for the grammaticization of words like *olsem*. Saxena (1991:351) proposes a functional account for the grammaticization and semantic shift of complementisers with verbs of saying and perception. Under this account, the path is a chain starting with the lexical item being used to mark verbatim quotes or as a complementiser introducing reported speech, and predicts that only in the final stages of the grammaticization process will it develop into a comparative. Clearly, Saxena's proposed route, and the principle that more abstract meanings develop out of more concrete ones (Hopper & Traugott 1993:185ff; Heine et al. 1991:65) are falsified on two counts by this data.



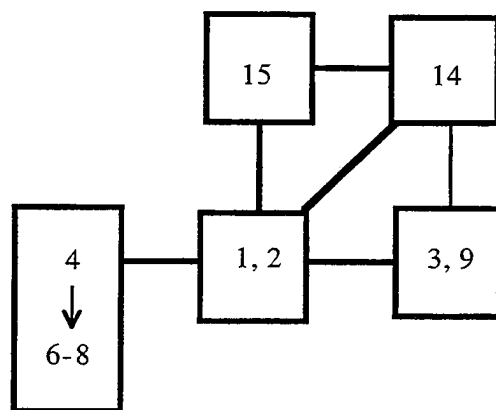


Figure 1: Proposed radial structure showing relationship of different meanings of *olsem*. [Examples in text: (1, \*) preposition; (\*) complementizer; (\*, \*-\*) clausal anaphor/demonstrative; (\*) booster/focuser; (\*) attenuator/hedge. Light line indicates less clear semantic link.]

Firstly, we have shown that the very concrete use of *olsem*, as a pure demonstrative, is probably a fairly recent development and that the demonstrative anaphor has developed from more abstract meanings, specifically, its role as a complementiser. Furthermore, we have introduced a distinction between use of an element as a complementiser introducing the complements of propositional predicates and the complements of propositional attitudes (e.g. (4b) *luk olsem* 'seem that/looks like'). We would suggest that even for languages where Saxena's proposed path holds, the distinction between the two verbs types is useful. A functional analysis would seem to predict that verbs like 'seem', which themselves entail a notion of comparison, would be quicker to co-opt a 'like' lexeme as their complementiser, viz.

(16a) It seems like/that we'll be able to join you

(16b) It is possible \*like/that we'll be able to join you

Secondly, insofar as it is clear that the core or original meaning of *olsem* is to make equative comparisons (the final point in Saxena's account of Sanskrit et al.), the claim that there is one unidirectional grammaticization path is also not true.<sup>6</sup>

However, what the Saxena's (1995) data and data on words like *olsem* in other languages suggest is that a weaker claim than the functionalist one is warranted. Cross-linguistically, it does not appear to be uncommon to find similar groupings of meanings and functions for words meaning 'like'. We have stressed the parallels between *olsem* and English 'like' in the translations of examples, but this does not appear to be purely an artefact of our translations. In Table 1, using Cerebon Javanese and Brazilian Portuguese

<sup>6</sup> Technically, the functionalist claims are not falsified as they are stated as grammaticization "principles" rather than universals. This makes grammaticization rather hard to test, particularly when the literature anticipates counter-examples and yet finesses them with comments like: "their [counter examples] relative infrequency, in fact help define our notion of what prototypical grammaticalization is" (Hopper & Traugott 1993:126). Sound historical records for any one language are rare, which makes quantifying the frequency of supporting and counter examples difficult. In many cases, paths of grammaticization and semantic shift are conjectural and based on the synchronic similarity of patterns. As this paper shows, languages may synchronically share patterns of meanings for even a single lexical item without having the same history.

as exemplars,<sup>7</sup> we show that there are similar patterns of multifunctionality for single (or closely related) words in languages other than English. In addition to Saxena's (1995) excellent data on words glossed as 'thus' in Sanskrit and several unrelated languages, this suggests that even if a cross-linguistic generalisation cannot be made about the direction of change, there is a fair amount of evidence that beneath general patterns of change there is a coherent semantic field.

Function of <i>olsem</i>	Cerebon	Br. Portuguese	English
Equative comparison	kaya	como	like, as
Complementiser w. verbs of speech or perception	ka, konon [speech] (konon, konon, 'blah blah')	assim (ó) [speech] como [perception]	that
Demonstrative	ika	assim (ó) [+gesture]	that, this
Clausal anaphor	konon	assim	that, this
Attention shifter	e	digo	anyway
Hedge	(ka ?)	cipo	like

Table 1: Realizations of the different functions of *olsem* in three other languages.

We might turn the question around at this point and ask why *olsem* does not share all the functions and grammatical roles of Eng. 'like'. An obvious contrast between the systems is that 'like' functions as an adjective, while *olsem* does not (Bislama uses *sem* 'same' as an adjective). We propose that the constraint on *olsem* is inherent in its structure. If speakers identify the *ol-* with the universal quantifier *ol* (definite plural), use of *olsem* as an adjective might be blocked. The function of adjectives is to restrict the reference of their head noun and this function is simply incompatible with the universal quantification of *ol* '(all) the'.

## 5. Conclusion

We have presented data showing that *olsem* has a rich variety of syntactic functions in spoken Bislama: preposition, complementiser, demonstrative and discourse marker. However, we have argued that since they are related to a shared core meaning, and since they sometimes rely on formal features active elsewhere in the syntax, they are not as disparate as they at first seem. We concluded that the pattern of semantic shift we have charted falsifies some functional claims about grammaticalization, but we have ventured a weaker cross-linguistic claim, namely that there is a underlying semantic field which provides a meaningful framework for these grammaticization patterns.

<sup>7</sup> Michael Ewing kindly shared his data on Cerebon with us, and Clarissa Surek-Clark her native speaker intuitions on Brazilian Portuguese.

## References

- Crowley, Terry (1990a). Serial verbs and prepositions in Bislama. In John Verhaar (ed.) *Melanesian Pidgin and Tok Pisin (Studies in Language Companion Series)*. (The Hague: Mouton). 57-89.
- Crowley, Terry (1990b). *From Beach-la-Mar to Bislama: the emergence of a national language*. (Oxford: Oxford University Press).
- Crowley, Terry (1995). *A New Bislama Dictionary*. (Suva: Institute of Pacific Studies, The University of the South Pacific).
- Heine, Bernd, Ulrike Claudi and Friederike Hunnemeyer (1991). *Grammaticalization: a conceptual framework*. (Chicago: The University of Chicago Press).
- Holmes, Janet (1988). 'Sort of' in New Zealand women's and men's speech. *Studia Linguistica*. 42, 2. 85-121.
- Hopper, Paul J. and Elizabeth Closs Traugott (1993). *Grammaticalization*. (Cambridge: Cambridge University Press).
- Huang, J.C.T. (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry*. 15, 4. 531-574.
- Kaplan, David (1989). Demonstratives. In Joseph Almog, John Perry and Howard Wettstein (eds) *Themes from Kaplan*. (New York: Oxford University Press).
- Lakoff, George 1987. *Women, Fire and Dangerous Things* (Chicago: The University of Chicago Press).
- Meyerhoff, Miriam (in press). Comparing old and new information in Bislama. In Jan Tent and France Mugler (eds) *Proceedings of the Second Oceanic Linguistics Conference, Pacific Linguistics Series A*. (Canberra: Australian National University)
- Meyerhoff, Miriam (1996). Transitive marking in Contact Englishes. *Australian Journal of Linguistics*. 16, 1. 57-80.
- Saxena, Anju (1995). Unidirectional grammaticalization: diachronic and cross-linguistic evidence. *Sprachtypologie und Universalienforschung (STUF)*. 48, 4. 350-372.
- Verhaar, John W.M. (1995). *Toward a Reference Grammar of Tok Pisin: and experiment in corpus linguistics*. (Honolulu: University of Hawai'i Press).
- Walker, Marilyn and Ellen Prince (1996). A bilateral approach to givenness: a Hearer-Status algorithm and a Centering algorithm. In Thorstein Fretheim and Jeanette Gundel (eds) *Reference and Referent Accessibility*. (Amsterdam/Philadelphia: John Benjamins).



# DP Licensing and Spec roles in Maori

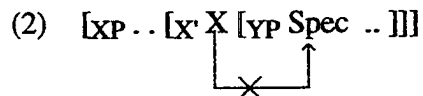
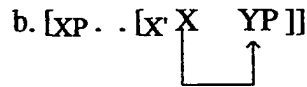
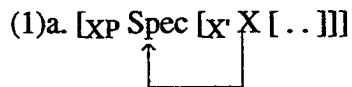
Elizabeth Pearce  
Victoria University of Wellington

In this paper I propose an account of local licensing conditions which privilege the relations between a Head and its Specifier and between a Head and the elements of the Head-chain. The analysis is based on conditions applying to PRO and to he-indefinites in Maori. The paper gives further indications of how the account that it provides can be extended to the conditions affecting the licensing of empty Ds, such as discussed in Longobardi (1994).

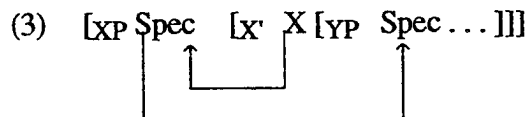
## 0. Introduction

This paper presents an analysis of syntactic conditions which distinguish between two kinds of subjects in Maori: Agentive DPs and unaccusative Theme DPs.\* The analysis is based on two constructions in which the licensing effects are distinct for DPs in the Spec and the complement position of the VP.

The particular licensing schema which emerges from the analysis privileges the head-government relations shown in (1a) and (1b), disallowing the relation shown in (2).



(1a) shows a head-Spec relation; (1b) is a head-complement relation; and (2) is a relation between a head and the Spec of its complement. Although in (2) the head X c-commands the Spec of its complement, the claim will be that it does not govern this Spec position. I will be arguing that the only way by which a head can license material in the Spec of its complement is via Spec-to-Spec transmission as shown in (3).



---

\* My thanks to Pauline Teripowai Higgins and to Timoti Karetu for sentence acceptability judgments for Maori and to Mario Saltarelli for Italian. Thanks also to the audience at the presentation of this paper at the AFLA meeting and to Jeffrey Waite.

(3) is thus in contrast with (2) in which the direct relation between the X head and the lower Spec is excluded.<sup>1</sup>

The two constructions which will provide the evidence for the licensing and anti-licensing relations shown in (1) - (3) are first, those involving PROs in non-finite clauses and second, those involving the licensing of a class of indefinite DPs. The discussion of the PRO constructions draws on material in Pearce and Waite (in preparation) and the analysis of the indefinite DPs extends on proposals in Pearce (1995).

### 1. Maori clause structure

On the evidence of the position of subjects with respect to classes of adverbial elements, I will be assuming that the subject in a tensed clause remains in the [Spec,VP] position in the surface:

- (4) *kia*    *tae<sub>i</sub>*    *rawa*    *atu*    *ia* *t<sub>i</sub>* *ki* *Aotearoa*. [Biggs 69]  
 T/A    arrive    Intens    thither    he    P    Aotearoa  
 '... that he get right to NZ'

In (4) the verb *tae* raises out of the VP to I and the post-adverb position of the subject *ia* provides the indication that the subject remains in the [Spec,VP] position.

Another piece of evidence that suggests that the VSO agent subject remains in [Spec,VP] is that, when the subject does raise, it raises to a position preceding the Tense/Aspect marker of its clause:

- (5)a. *Kātahi* *anō*    *te*    *wahine* *ka*    *whakahoki*    *mai* *i* *ngā* *pukapuka*.  
 then    again    Det    woman    T/A    return    here    DO    Det    book  
 'Then the/a woman returned the books'  
 b. \**Kātahi* *anō* (i) *ngā* *pukapuka* *ka* *whakahoki* *mai* *te* *wahine*.

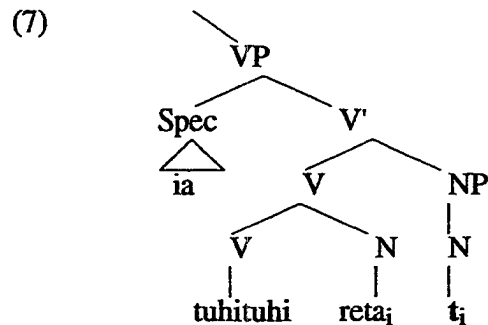
In (5a), the subject *te wahine* immediately follows the sentence initial adverbial but it precedes the T/A marker of its clause. The pre-T/A position shown in (5a) can be filled only by the subject and not, for example, by the object (5b). These facts suggest that, if the Maori clause has a Spec position which is designated for the subject above the VP, then that position precedes rather than follows the T/A clause head position. Thus, the subject *ia* in (4) is inside the VP, rather than in a subject position above the VP.

The second assumption is that the Direct Object of the verb in a transitive clause is the immediate sister of the verb. This syntactic relation is suggested in particular by the availability of Object-Noun-Incorporation (ONI) as shown in (6b):

<sup>1</sup>Rousseau (1996) argues for a similar view of the syntactic relations identified in (1) - (3). In her account, she proposes that these syntactic relations are compatible with Minimalist assumptions (Chomsky 1995) with respect to asymmetric operations.

- (6)a. E tuhituhi ana ia i ngā reta.  
 T/A write T/A 3PSg DO Det letter  
 'She is writing the letters'
- b. E tuhituhi reta ana ia. [Bauer (1798)]  
 T/A write letter T/A 3PSg  
 'She is letter writing'

A syntactic view of ONI as in (6b) would involve the raising of the N head in the structure:



The complex V so formed then raises to Infl.

A number of possible approaches can be taken as to the arrangement of other argument positions within the VP.<sup>2</sup> The discussion in this paper will however focus on the two types of nominative arguments, Agents and unaccusative Themes, which from the above, I take to be respectively in [Spec,VP] and in the complement position as sister to the V head. Any additional VP-internal arguments (whether or not there is more than one VP shell) must therefore be located above the lowest V head but below the [Spec,VP] position assigned to the Agent. This means that the bracketing indications given for a clause with a ditransitive verb in (8a) and for an unaccusative verb in (8b) are open to further interpretation on which I will not commit myself here:

- (8)a. I hoatu<sub>i</sub> [VP ia [v' t<sub>i</sub> te pukapuka] ki a Mere].  
 T/A give 3PSg Det book P Pers Mere  
 'She gave the book to Mere'  
 Waite (1994)
- b. I mahue<sub>i</sub> [VP ... [v' t<sub>i</sub> te kōtiro] i te pahi].  
 T/A left Det girl P Det bus  
 'The girl missed the bus'

In (8a) the nominative subject ia is in [Spec,VP]; but in (8b) the nominative subject te kōtiro is the Theme complement of the unaccusative verb. In both (5a) and (5b) the preposition-governed argument will be located somewhere in the structure above the lowest V' node.

<sup>2</sup>For a range of alternative proposals, see Larson (1988), Den Dikken (1995), Belletti and Shlonsky (1995).

## 2. Subjects in tensed versus non-tensed clauses

In some respects clausal complements in Maori which have an irrealis interpretation show a familiar subjunctive versus non-finite alternation such as is found in Italian in examples (9a,b):

- (9)a. Maria vuole [ aiutare la famiglia ].  
 Maria wants help-Inf the family.  
 'Maria wants to help the family'
- b. Maria vuole [che ti aiuti la famiglia].  
 Maria wants that you help-3Sg/Subjunct the family  
 'Maria wants that the family help you'

When the subject of the embedded clause is coreferential to the main subject as in (9a), the verb of the embedded clause must be an infinitive. When the two subjects are non-coreferential as in (9b), the embedded clause has an inflected subjunctive verb. Corresponding to the Italian examples in (9) are the forms for Maori shown in (10):

- (10)a. Kei te pīrangi a Mere [ ki te/\*kia āwhina i tōna whānau ].  
 T/A want Pers Mere help DO Poss family  
 'She wants to help her family'
- b. Kei te pīrangi a Mere [ kia/\*ki te āwhina tōna whānau i  
 T/A want Pers Mere help Poss family DO  
 a koe].  
 Pers 2PSg  
 'She wants that her family help you'

The examples in (10) appear to parallel those for Italian in (9) in that the form of the T/A marker preceding the verb depends on whether or not the two subjects are coreferential. In Maori when the subject of the embedded clause is non-coreferential to the main clause subject the verb of the embedded irrealis clause is preceded by kia. When the two subjects are coreferential and the embedded subject is non-overt, the verb is preceded by ki te.

Leaving aside the question of possible syntactic tests to distinguish kia and ki te from simple Tense/Aspect markers,<sup>3</sup> I will treat both of these forms as Infl elements, kia thus being like a subjunctive marker, and ki te corresponding to non-finite inflection.

It has long been noticed that, although Maori shows comparability with languages like Italian for (10a,b), as with other Polynesian languages (Samoan: Chung (1978); Tahitian, Tokelauan, Tikopian: Hooper (1982)), it also has an unexpected restriction on the kinds of predicate that can occur after ki te as in (10b). In essence the restriction is

<sup>3</sup>The form kia also occurs as an optative as in:

- (i) Kia toru ngā ika! [Biggs 35]  
 kia three Det/Pl fish  
 'Let there be three fish'
- (ii) Kia tae mai pea ia. [Bauer (2053)]  
 kia arrive hither perhaps 3Sg  
 '(I hope) she comes'



that unaccusative verbs may not occur in an embedded clause after *ki te*.<sup>4</sup> Some examples are given in (11):

- (11)a. Passive  
 \*Ka pīrangi ia [ki te āwhinatia e tōna whānau ].  
 T/A want 3PSg help-Pass by Poss family  
 'She wishes to be helped by her family'
- b. Neuter/Stative  
 \*E pīrangi ana a Moana [ki te mahue i tōna tane ].  
 T/A want T/A Pers Moana left P Poss male  
 'Moana wants to be abandoned by her husband'
- c. Experiencer  
 \*Ka pīrangi ia [ki te mōhio ki te kōtiro rā ] [Bauer (1983), (45)]  
 T/A want he know P Det girl there  
 'He wanted to know that girl'

The restriction exhibited in Maori in (11) is absent in Italian, a language which has clearly identifiable unaccusative verbs:

- (12)a. Passive  
 Voglio [essere aiutato dalla famiglia ].  
 I-want to-be helped by-the family  
 'I want to be helped by the family'
- b. Unaccusative  
 Voglio [arrivare alle otto].  
 I-want to-arrive at-the eight  
 'I want to arrive at eight'
- c. Experiencer  
 Voglio [piacere a Carlo].  
 I-want to-please to Carlo  
 'I want to be pleasing to Carlo'/'I want Carlo to like me'

The assumption is that a PRO subject is available in non-tensed unaccusative clauses in Italian in the same way as it is available in non-unaccusative non-tensed clauses, as in (13a):

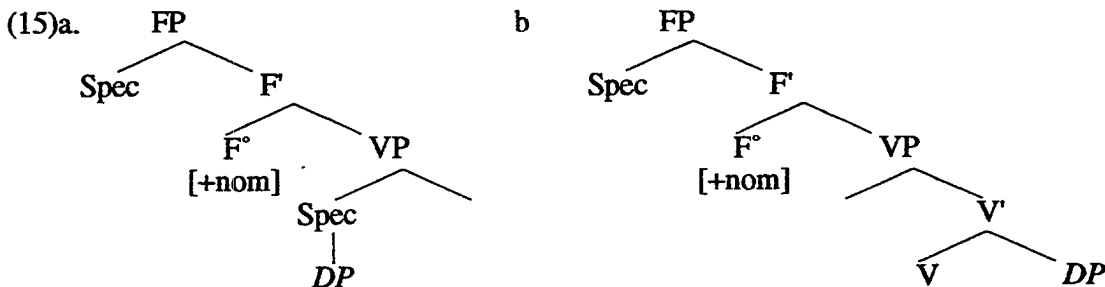
- (13)a. ...[<sub>IP</sub> PRO<sub>i</sub> aiutare<sub>j</sub> [<sub>VP</sub> t<sub>i</sub> t<sub>j</sub> la famiglia]] TRANSITIVE  
           PRO to help                   the family
- b. ...[<sub>IP</sub> PRO<sub>i</sub> arrivare<sub>j</sub> [<sub>VP</sub> t<sub>j</sub> t<sub>i</sub> alle otto]] UNACCUSATIVE  
           PRO to arrive                   at eight

That is, the subject PRO of the embedded clause will raise to [<sub>Spec,IP</sub>] in parallel with the use of the [<sub>Spec,IP</sub>] position for subjects in tensed clauses in Italian. In the non-finite CP clause [<sub>Spec,IP</sub>] is standardly an ungoverned, un-Case-marked position.

<sup>4</sup>These constructions are examined in greater detail in Pearce and Waite (in preparation). See also Chung (1978), Reedy (1979), Hooper (1982) and Bauer (1993: §1.1.2.2).



This analysis of non-finite PROs now requires that we examine the reverse situation. That is, I ask the question: how are the overt subject DPs licensed in the tensed clause? The corresponding structures with overt DPs are as shown in (15):



Although in the surface forms, the DPs remain inside the VP, both DPs must have access to the F head for nominative Case checking. Suppose that a strong [+nom] F projects a Specifier position, whereas a weak [-nom] F fails to project a Specifier. The DPs in (15) can now enter into a covert relation with Spec,FP for Case checking. The lexical/non-lexical contrast for the [Spec,VP] subject reduces to the presence versus the absence of the Spec,FP. This proposal is in accordance with the suggestion put forward in Chomsky (1995) that the strength of D features in a functional head may be the determining factor in whether or not a Spec is projected. A projected Specifier has a checking function which must be implemented, overtly or covertly.

This analysis also has some plausibility in view of both syntactic and morphological characteristics which can be highlighted in a comparison, say with Italian. That is, Italian has strong D or other F features which lead to the projection of the relevant Specifier even when the F is [-nom]. The relevant contrasts are identified in (16):

(16)		<u>Italian</u>	<u>Maori</u>
	(i)	Spec,IP subject	Yes: SVO
	(ii)	Subject-Verb agreement	No
-->	(iii)a.	D-features	weak
	b.	[-nom]	- Spec

The relative weakness of D-features in Maori is suggested by the failure of overt raising of the subject in the simple tensed clause (16(i)) as well as by the absence of overt subject-verb agreement (16(ii)). However, whereas the Minimalist approach aims to derive the contrasting surface configurations from the checking properties, my analysis of the failure of the PRO to occur in (14b) seems, as it stands, to require the use of a mechanism of government, such as indicated by the conditions (1) - (3).

As an alternative, however, we can consider that the relation between a verb and its sister DP is to be regarded as another kind of checking relation. That is, all of the legitimate relations in (1) and (3) count as licensing structures in terms of Minimalist checking. The mechanism of government is thus reinterpreted as a checking relation which can be satisfied in the head-Spec structure of (1a) or in the head-complement relation of (1b). The structure in (3) is a covert variant of (1a) in that the Spec of YP has access to the Spec of XP position.

### 3. The licensing of indefinite he-NPs

The second case involving a contrast in the licensing relations affecting [Spec,VP] and the verb-complement position is found in the conditions affecting the licensing of indefinite he-NPs. Once again, following Chung, Mason & Milroy (1995), the contrast is between unaccusative and non-unaccusative constructions. The analysis to be presented in this section extends on that put forward in Pearce (1995).

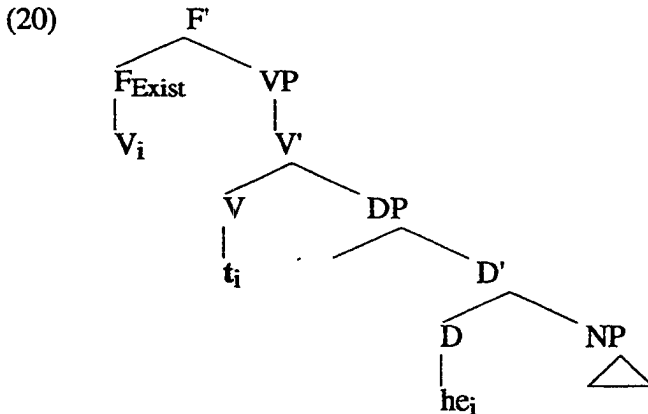
An indefinite he-NP can occur only as the nominative DP of the clause. As shown schematically in (17), in simple clauses these indefinites cannot be the nominative DP in [Spec,VP]: (17a), but they can be the nominative DP of an unaccusative: (17b) :

- (17)a. \*[ T/A Verb<sub>i</sub> [vp he-NP t<sub>i</sub> ... ]]  
 b. [ T/A Verb<sub>i</sub> [vp t<sub>i</sub> he-NP ... ]]

Thus the examples given in (18), corresponding to (17a), are ungrammatical and those given in (19) instantiate the grammatical schema of (17b):

- (18)a. \*Kei te patu he tamariki i te kau. [Bauer (1983), (72)]  
 T/A beat *he* children DO Det cow  
 'Some children are beating the cow'  
 b. \*I whiu he wahine i tāna mōkai ki te moana [CMM (19b)]  
 T/A throw *he* woman DO her pet P Det sea  
 'A woman threw her youngest child into the ocean'
- (19)a. Passive  
 Kua mahia e Pani he kapu tū mā rātou. [Bauer (1086)]  
 T/A make-Pass by Pani *he* cup tea P them  
 'Pani has made them a cup of tea'
- b. Neuter/Stative  
 I mahue he kōtiro i te pahi.  
 T/A left *he* girl P Det bus  
 'A girl missed the bus'
- c. Experiencer  
 I pīrangī he tāne ki tēnei wahine. [CMM (27b)]  
 T/A desire *he* male P this woman  
 'A man desired this woman'

If we take the indefinite he as a kind of positive polarity item, we can say that it needs to be licensed by an existential head or operator. If the existential projection is housed within the functional structure of the clause, it enters into a direct licensing relation with the sister of the verb through the chain created by the movement of the verb, such as shown in (20):

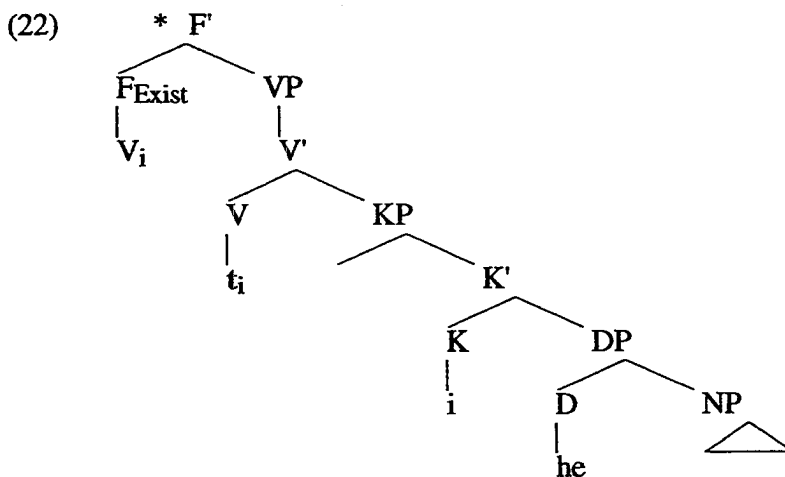


In (20) the lowest V head of the chain directly head-governs the D head of its DP complement.

There is an overt analogue of the head linking relationship shown in (20). Following Waite (1994), the predicational constructions illustrated in (21) involve the raising of a head to a he which here serves as the T/A nexus of the clause.

- (21)a. He [N māhita]<sub>i</sub> [NP ia t<sub>i</sub>].  
 T/A teacher he  
 'He is a teacher'
- b. He [A whero]<sub>i</sub> [AP tēnei t<sub>i</sub>].  
 T/A red this  
 'This is red'
- [Waite (1994)]

The nominative DP of (20) is in contrast to the accusative DP in (22) in which the chain relation between the V and the D head is blocked by the overt Case morphology:



Thus, the licensing of the unaccusative he subject takes place in a head-government structure, in which the he is the D head of the complement of the verb. Such is not the case for the he-NP subjects in [Spec,VP] in (18). In parallel with the explanation for the

inaccessibility to government of the PRO in [Spec,VP], I assume that there is no licit chain relation for the [Spec,VP] indefinite *he*-NP subject: the chain relation shown in (20) does not work for the [Spec,VP] position (17a) because the non-overt existential F fails to project a Specifier in which to house an operator that could enter into a binding relation with the DP in the [Spec,VP] position.

There is, however, another kind of construction in which an indefinite *he*-NP originating from [Spec,VP] is licensed. These are constructions which include a higher operator, such as a quantifying expression or a negative, as in (24); corresponding to (23.II) in the terms of Chung, Mason & Milroy (1995):

- (23) Description of *he*-indefinites following Chung, Mason & Milroy (1995)<sup>6</sup>  
 I: The existential *he*-indefinite can only be the nominative argument of a passive or unaccusative verb.  
 II: The polarity *he*-indefinite is an operator-bound nominative argument.
- (24) Operator licensing of *he*  
 a. Kāore he tamaiti i kai i ngā tuna rā.  
     Neg *he* child T/A eat DO Det eel there  
     'A child did not eat those eels'

<sup>6</sup>An additional factor is the distinction in the behaviour of stage versus individual-level predicates, as shown in (i) and (ii).

Stage versus Individual level predicates

- (i) STAGE  
 a. I    whero    he kanohi    i te    makeretanga o te    tarau.  
     T/A red    *he* face    P Det drop-Nmlz of Det trouser  
     'She/they blushed because she lost her trousers' [CMM (36e)]  
 b. I    makariri he kōtiro i tā    koutou haringa i ngā    paraikete.  
     T/A cold    *he* girl P Det-Gen you-Pl take-Nmlz DO Det-Pl blanket  
     'A girl got cold because you took away the blankets' [CMM (36c)]
- (ii) INDIVIDUAL  
 a. ?\*I   whero    he waka.  
     T/A red    *he* car  
     'A car was red' [CMM (39e)]  
 b. \*I    makariri he kōhatu.  
     T/A cold    *he* stone  
     'A stone was cold' [CMM (39c)]

The contrast between (i) and (ii) requires further investigation. Whilst the examples in (i) have the form of the neuter/stative type of unaccusative, with an *i*-marked Cause argument, our first assumption would be that the forms in (ii) are also unaccusatives. It would, for example, seem reasonable to suppose that the syntactic and thematic relations between a colour and the referent which has the colour is the same in all cases. On the other hand, Levin and Rappaport (1994) argue that a characteristic common to unaccusative verbs is that they encode either internal or external causation. As they point out, de-adjectival unaccusative verbs are formed from stage-level rather than from individual level adjectives. This suggests that, in line with the opposing stage/individual interpretations for the adjectives in (i) and (ii), those in (i) count as unaccusative adjectives, but those in (ii) do not have an unaccusative structure. The argument structure of these adjectives is a topic for further research. Also requiring further investigation is the possible role of Event structures within or external to the VP, such as proposed in Kratzer (1989) and in much recent work.

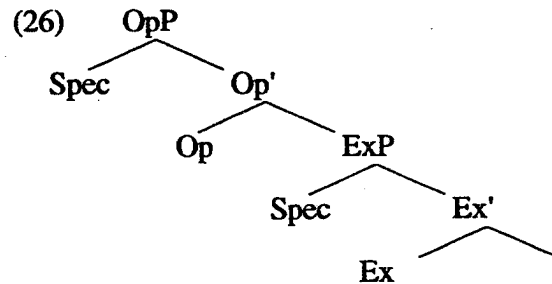
- b. la tau, ia tau, i tito waiata hou he wahine.  
 that year that year T/A compose song new he woman  
 'Every year a woman composed new songs' [CMM (30c)]
- c. Ki te karanga he reo, kei puta iho koe.  
 T/A call he voice don't come out down you  
 'If a voice calls, don't you come down' [CMM (31a)]

Chung, Mason and Milroy propose to unify their two-way characterization as in (25):

- (25) The variable introduced by a he-indefinite must be unselectively bound or quantificationally closed by a sentence level operator. [CMM, fn.19 (a)]

What I have been trying to do here is to identify the precise syntactic conditions for these indefinites. The account that I have given of the licensing requirement for the (23.I) type involves the absence of a Specifier for the existential F (as well as the absence of a head-to-head relation with the D (= he)). With the constructions represented in (23.II), such as those in (24), we might suppose that these constructions contain operators which have the capacity to link up with the Existential head, empowering it to project a Specifier position. The he-indefinite in these cases has access to the projected Spec and thus to a relation with the existential head. The availability of such a Specifier position is supported by the typical position of the subject in these constructions, as in (24a).<sup>7</sup>

A partial representation of the suggested schema is as follows:



The idea is that, whereas in the case of the (23.I) construction the  $F_{EX}$  projection is 'made visible' through the head-chain relation with the he of the unaccusative below it, the (23.II) type higher operator constructions have the capacity to identify the  $F_{EX}$  projection from above. In both cases the  $F_{EX}$  projection must be identified by some overt material in an appropriate licensing configuration.

<sup>7</sup>Note that only some of the triggers for subject preposing have the capacity to license he-indefinites:

- (i) Kātahi anō te/\*he wahine ka whakahoki mai i ngā pukapuka.  
 then again Det woman T/A return here DO Det book  
 'Then the/a woman returned the books'

Kātahi (anō) is an adverbial licenser for subject preposing. The licensing operator for he-indefinites must be an operator with quantificational force.

Returning again to the unaccusative constructions, recall that these do not need an independent mechanism to trigger the projection of the Spec,ExistP. This is because the *he* of the indefinite enters into a head-to-head relation with the Verb-Infl-Exist chain. The relationship in this instance is covert because the *he* Determiner head does not raise out of the DP, nor can the whole DP raise to a Spec position of an ungoverned XP.

#### 4. Some further implications

The kinds of binding mechanisms just described find parallels with phenomena that have been analysed for other languages. Thus, Longobardi (1994) invokes a licensing role for chains involving head-to-head relations with respect to properties of determiners in Italian. Consider in particular the examples in (27) and (28):

##### Longobardi (1994)

- (27)a. \*Acqua viene giù dalle colline. [L (14a)]  
           water comes down from the hills
- b. Viene giù acqua dalle colline. [L (14b)]  
           comes down water from the hills
- c. Ho preso acqua dalla sorgente. [L (14c)]  
           I took water from the spring
- (28)a. \*Consideravo studenti intelligenti. Belletti (69a) L fn11.(i)  
           I considered students intelligent
- b. Consideravo gli studenti intelligenti. Belletti (71a)  
           I considered the students intelligent

In both (27b) and (27c) *acqua* is the direct complement of the verb, the unaccusative subject in (27b) and direct object in (27c). The null determiner is licensed here in the same syntactic configuration as applies to the licensing of unaccusative indefinite *he* in Maori. In (27a) the null determiner is not lexically governed. In (28) *studenti* occupies a small clause Specifier and is accessible for accusative Case checking, as shown by (28b), but the null determiner in (28a) cannot be licensed in a head-to-head relation. For Longobardi, there is a requirement for lexical head government of the empty category D, but, as he notes, for cases like (28a) "the required relationship with a head seems stronger than many usual definitions, since it does not allow an empty D to be licensed by a verb across small clause boundaries" (fn. 11). Whereas constructions like (27a) can be ruled out on the basis of a failure of lexical head-government under c-command, (28a) is like the ungrammatical constructions in Maori in which the indefinite DP is located in [Spec,VP]. In both (28a) and the schema (17a), the relevant Spec position is in a c-command relation with a potential lexical governor. The 'stronger' relation which Longobardi points to has been spelt out in the proposals which I have been developing here.<sup>8</sup>

<sup>8</sup>A further account which is comparable to the analysis developed in the present paper is the analysis put forward in Rizzi (1991) in relation to the mechanisms applying to head-government of a [Spec,IP] trace. Under Rizzi's account, a crucial condition of the capacity of the C to govern the trace is that the C must acquire the Agr feature from the IP below it. In the resulting configurations, the Spec-head relation has the pre-eminent role over any alternative mechanism based simply on c-command from C to the [Spec,IP] position.



Longobardi's analysis focuses on the relationship between the D and the N of the NP that it heads. When the D is empty, its interpretation can be satisfied, either through the raising of N to D (generic interpretation) or through an anywhere rule which assigns an existential interpretation to the empty D (Longobardi 1994 (102)). In the latter case lexical government is required for the empty D. In the analysis that I have presented in this paper the conditions and the interpretation for *he*-indefinites match up with those applying to the Italian existential empty D. In Longobardi's analysis of Italian the lexical government requirement is forced because the D is empty. The Maori *he* occurring in the same syntactic position is lexical. The head-government requirement (or the chain) from V to D (= *he*) is suggested by the failure of *he*-NPs to occur in the presence of overt accusative Case-marking (cf. (21)). We may assume however that the LF representation of such quantified expressions requires at least a coindexation (if not also raising) to an extra-VP position. In the portrayal of the conditions affecting the existential *he* in Maori I have supposed that the representation must include a linking mechanism to that extra-VP position. One particular piece of evidence in support of this proposal rests on the presence of overt *he* in the T/A position in examples like those in (21a,b).

In the interests of a universal characterization of mechanisms affecting both the Maori *he* and the Italian existential empty D, I thus wish to propose that the two phenomena are most appropriately unified in terms of the present analysis of the Maori construction. It seems to me that this interpretation has the further advantage that it provides a better representation of the quantificational characteristics of the semantics of the constructions involved, including both those with overt and with non-overt quantifiers. A similar focus on the unification of operator-bound and bare indefinite DPs is proposed also in Déprez (1996), based on material from yet more languages.

In summary, I have argued that the syntactic conditions applying to PRO DPs and to *he*-indefinites in Maori can be accounted for in terms of relations which allow for government between a head and its Spec (1a) and between a head and its complement (1b), but which do not allow for an unmediated government relation between a head and the Spec of its complement (2). These conclusions have been drawn from the consideration of VSO structures in Maori in we have been able to view the properties of DPs located in [Spec,VP] as distinct from in other positions within the VP or external to the VP. In the preceding paragraphs I have suggested that the mechanisms applying to Maori are generalizable to other (SVO) languages in which similar effects are manifested in different, but comparable, construction types.

\* \* \* \* \*

### References

- Bauer, Winifred (1983), 'Experience verbs in Maori,' *Te Reo* 26, 3-28.  
 Bauer, Winifred (1993) (with William Parker and Te Kareongawai Evans) *Maori*.  
 Routledge, London & New York. [= Bauer]  
 Belletti, Adriana (1988), 'The Case of unaccusatives,' *Linguistic Inquiry* 19, 1-34.  
 Belletti, Adriana, & Ur Shlonsky (1995), 'The order of verbal complements: a comparative study,' *Natural Language and Linguistic Theory* 13, 489-526.  
 Biggs, Bruce (1969/1973), *Let's Learn Maori*, Reed Books. [= Biggs]

- Chomsky, Noam (1995), 'Categories and transformations,' in N. Chomsky, The Minimalist Program, MIT Press, Cambridge, Mass., pp. 219-314.
- Chung, Sandra (1978), Case marking and Grammatical Relations in Polynesian, University of Texas Press, Austin.
- Chung, Sandra, Te Haumihiata Mason, & J.W. Milroy (1995), 'On Maori he and the uses of indefinites,' The Journal of the Polynesian Society 104, 429-459. [= CMM]
- Déprez, Viviane (1996), 'Subject/Object asymmetries in indefinite licensing,' GLOW Newsletter 36, 28-29.
- Dikken, Marcel den (1995), Particles, Oxford University Press, New York & Oxford.
- Hooper, Robin (1982), Neuter Verbs and Stative Aspect in Polynesian, MA thesis, Auckland University.
- Kratzer, Angelika (1989), 'Stage-level and Individual-level predicates,' ms., University of Massachusetts, Amherst.
- Larson, Richard K. (1988), 'On the Double Object construction,' Linguistic Inquiry 19, 335-391.
- Levin, Beth, & Malka Rappaport Hovav (1994), Unaccusativity: At the Syntax-Lexical Semantics Interface, MIT Press, Cambridge, Mass.
- Longobardi, Giuseppe (1994), 'Reference and Proper Names,' Linguistic Inquiry 25, 609-665.
- Pearce, Elizabeth (1995), 'Operator binding and the syntax of negation in Māori,' paper presented at 'Negation: Syntax and Semantics', University of Ottawa.
- Pearce, Elizabeth, & Jeffrey Waite (in preparation), 'Kia and ki te complementation in Maori', ms, Victoria Univ. of Wellington.
- Reedy, Tamati M. (1979), Complex Sentence Formation in Maori, PhD dissertation, University of Hawaii.
- Rizzi, Luigi (1991), 'Residual Verb Second and the Wh-criterion,' Technical Reports in Formal and Computational Linguistics 2, University of Geneva. Also in Adriana Belletti & Luigi Rizzi (eds), Parameters and Functional Heads: Essays in Comparative Syntax, Oxford University Press, New York & Oxford, 1996, pp. 63-90.
- Rousseau, Anna (1996), 'Specifiers, head-dependencies and asymmetric relations,' GLOW Newsletter 36, 56-57.
- Waite, Jeffrey (1994), 'Determiner phrases in Maori,' Te Reo 37, 55-70.

Department of Linguistics  
 Victoria University of Wellington  
 PO Box 600  
 Wellington  
 New Zealand

Elizabeth.Pearce@vuw.ac.nz

## Syntax vs. Semantics in Tagalog wh-extraction

Norvin Richards, MIT

Tagalog exhibits a phenomenon which I will refer to as “topicalization”, which is illustrated in 1<sup>1</sup>:

1. a. Bumili **ang lalaki** ng tela  
AT-bought T man G cloth  
‘The man bought cloth’
- b. Binili ng lalaki **ang tela**  
GT-bought A man T cloth  
‘A man bought the cloth’

1.a and 1.b are both well-formed Tagalog sentences; they differ in that *lalaki* ‘man’ has been topicalized in 1.a, while *tela* ‘cloth’ is the topic in 1.b. Topicalization involves attaching the marker *ang* to the topicalized nominal (or *si*, for a proper name). Also, verbal morphology roughly indicates the thematic role of the topicalized nominal; in 1.a, the infix *-um-* shows that the logical subject is the topic, while in 1.b the infix *-in-* marks the direct object as topic.

Topicalization interacts in interesting ways with wh-extraction, as shown in 2:

2. a. Sino ang bumili ng tela?  
who T AT-bought G cloth  
‘Who bought cloth?’
- b. \*Sino ang binili **ang tela**?  
who T GT-bought T cloth
- c. Ano ang binili ng lalaki?  
what T GT-bought A man  
‘What did the man buy?’
- d. \*Ano ang bumili **ang lalaki**?  
what T AT-bought T man

As 2 shows, extraction of certain arguments forces the use of the form of the verb which topicalizes the extracted argument. Subject extraction, for example, is only possible with *bumili*, the form of the verb which topicalizes the subject; topicalization of the object, as in 2.b, makes subject extraction impossible. In 3, we see that similar restrictions hold on the formation of relative clauses in Tagalog:

3. a. ang lalaki-ng bumili ng tela  
T man-that AT-bought G cloth  
‘the man that bought the cloth’
- b. \*ang lalaki-ng binili **ang tela**  
T man-that GT-bought T cloth
- c. ang tela -ng binili ng lalaki  
T cloth-that GT-bought A man  
‘the cloth that the man bought’
- d. \*ang tela -ng bumili **ang lalaki**  
T cloth-that AT-bought T man

---

Many thanks to my Tagalog informants, Marlon Abayan, Alex Diaz, Luis Sarmenta, and Nicole Lazo. I would also like to thank Irene Heim, Martha McGinnis, David Pesetsky, and the audience at AFLA III for helpful discussion. Responsibility for any remaining errors is mine alone. This work was supported in part by a graduate fellowship from the NSF.

<sup>1</sup> I will use Schachter’s (1976) terminology for describing the Tagalog topicalization system; thus verbs will be marked AT for Actor-Topic, GT for Goal-Topic, DT for Direction-Topic, etc.; similarly, T stands for Topic, A for Actor (roughly, the logical subject), G for Goal (roughly, the direct object), and so forth.

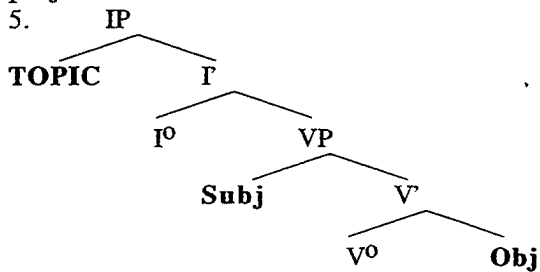
In cases where extraction is of an element which cannot be topicalized, extraction has no effects on topicalization. For instance, there is no way of topicalizing times in Tagalog, and the wh-word *kailan* 'when' is compatible with topicalization of either the subject or the object:

4. a. *Kailan bumili ang lalaki ng tela?*  
 when AT-bought T man G cloth  
 'When did the man buy cloth?'  
 b. *Kailan binili ng lalaki ang tela?*  
 when GT-bought A man T cloth  
 'When did a man buy the cloth?'

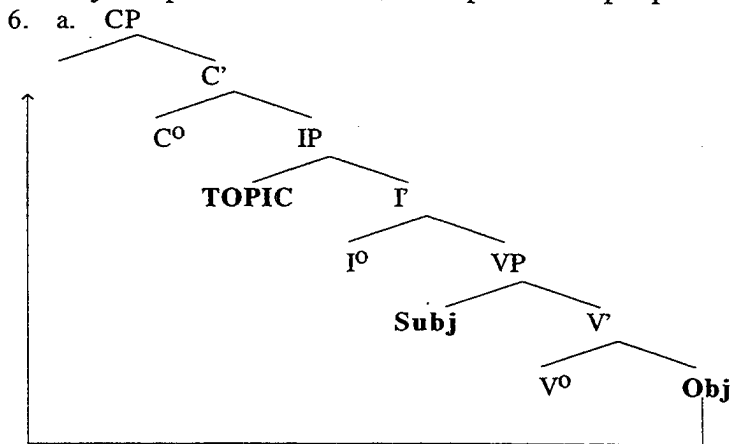
In the next section I will outline an account of these facts developed by Nakamura (1994, 1995, 1996). I will then go on to explore some problems with this account and attempt to provide solutions to them.

## 2. Nakamura

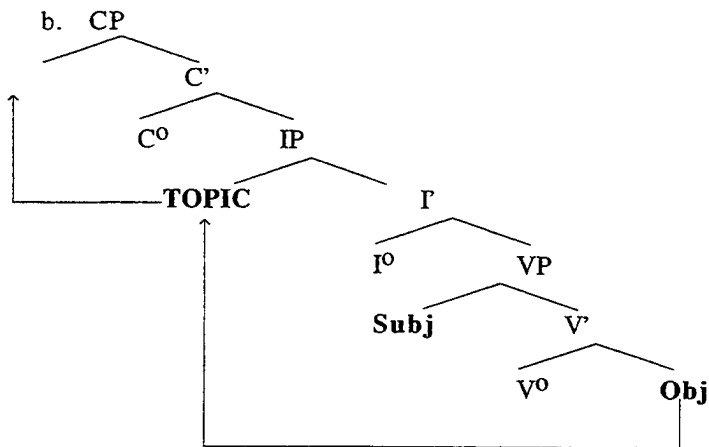
Nakamura offers an Economy-based account of the facts presented above, which I will use as a starting point. He assumes, following Guilfoyle, Hung, and Travis (1992), that topicalization involves movement to the specifier of some IP projection, more or less as in 5<sup>2</sup>:



When wh-extracting a nominal which can land in topic position, either of the derivations in 6 ought to be available in principle. The wh-word could proceed directly to Spec CP, as in 6.a, or stop in the Topic position on its way up, as in 6.b:



<sup>2</sup> Throughout, I will diagram Tagalog sentences as though Topicalization is an overt movement. This is simply for ease of exposition; nothing crucial hinges on the question.



Nakamura's claim is that the availability of the derivation in 6.b blocks the derivation in 6.a, for reasons having to do with Shortest Move; a derivation consisting of two comparatively short moves, like 6.b, is preferable, all other things being equal, to a derivation consisting of a single long move, like 6.a. In other words, a wh-word which can be topicalized must be topicalized in order to be extracted. The similar facts concerning relativization follow in a similar way. In cases in which a wh-word cannot be topicalized (for instance, in extraction of *kailan* 'when'), a derivation like that in 6.b is impossible. 6.a is therefore available, as the next best option; as always in Economy-based frameworks, if the optimal derivation is unavailable, a sub-optimal derivation can be used instead. *Kailan*, then, can be directly extracted without stopping in the topic position, since this is the best the grammar can do, and topicalization of any nominal in the sentence can proceed freely.

This approach to the Tagalog extraction facts has much to be said for it; it is extremely simple and elegant, deriving the relevant facts from general principles in a very straightforward way. In the next section I will discuss a syntactic distinction drawn in Tagalog between two classes of questions which does not follow from Nakamura's account.

### 3. Clefts and nonclefts

Nakamura's account, as we have seen, correctly predicts the grammaticality of both 7.a and 7.b:

7. a. Sino ang bumili ng tela?  
 who T AT-bought G cloth  
 'Who bought the cloth?'  
 b. Kailan bumili ang lalaki ng tela?  
 when AT-bought T man G cloth  
 'When did the man buy cloth?'

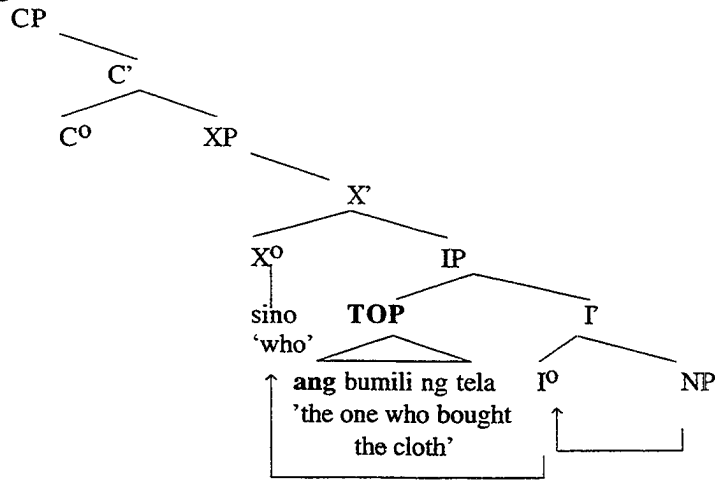
In 7.a, according to Nakamura, the subject has been topicalized and then wh-extracted, triggering the use of the Actor-Topic form of the verb. In 7.b, on the other hand, the wh-word cannot be topicalized and therefore need not be.

There is a structural distinction, however, between 7.a and 7.b which is not predicted by Nakamura's theory. The structure of 7.b appears to be similar in the relevant respects to that of its English counterpart. 7.a, on the other hand, appears to be a cleft, similar in structure to 8:

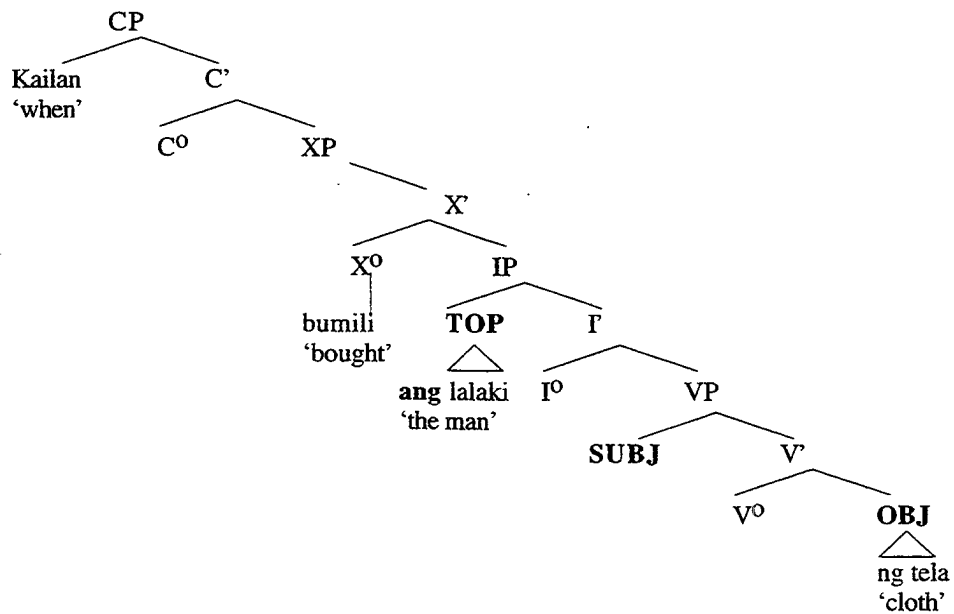
8. Lalaki ang bumili ng tela  
 man T AT-bought G cloth  
 'The one who bought the cloth was a man'

Both 7.a and 8 consist of a headless relative clause *ang bumili ng tela* 'the one who bought the cloth' which serves as the subject of a nominal predicate: *lalaki* 'man' in 8, and *sino* 'who' in 7.a. For discussion of the cleft nature of many Tagalog questions, see Schachter and Otones (1972), Seiter (1975), Kroeger (1991), Richards (1991b), Nakamura (1996). 7.a and 7.b, then, have structures something like those given in 9:

9. a.



b.



In the next sections I will briefly discuss two arguments for giving 7.a and 7.b these structures.

### 3.1 Clitic Placement

Tagalog possesses a number of second-position clitics, including the pronouns *siya* 'T-he/she' and *ka* 'T-you' (10.c from Sityar 1989, 18):

10. a. Uuwi *siya* bukas  
AT-will-go-home T-he/she tomorrow  
'He/she will go home tomorrow'
- b. Hindi *siya* uuwi bukas  
not T-he/she AT-will-go-home tomorrow  
'He/she will not go home tomorrow'
- c. Bukas ng alas otso ng gabi *ka* aalis  
tomorrow at eight night T-you AT-will-leave  
'You will leave tomorrow at eight o'clock at night'

Such cliticization is typically sensitive to clause boundaries, as can be seen in 11:

11. a. Sinabi ni Juan na matalino *siya*  
GT-said A Juan that intelligent T-he/she  
'Juan said that he/she was intelligent'
- b. \*Sinabi *siya* ni Juan na matalino  
GT-said T-he/she A Juan that intelligent

Clitics apparently attach, then, to the first constituent in the clause, which allows us to use them as tests for clause boundaries. Their behavior seems to confirm the claim made above; wh-extraction of the kind which interacts crucially with topicalization involves a cleft, while the other type of wh-extraction does not:

12. a. Sino ang nakita *niya*?  
who T GT-saw A-he/she  
'Who did he/she see?'
- b. \*Sino *niya* ang nakita?  
who A-he/she T GT-saw
13. a. Kailan *niya* nakita ang kalabaw  
when A-he/she GT-saw T water-buffalo  
'When did he/she see the water buffalo?'
- b. \*Kailan nakita *niya* ang kalabaw  
when GT-saw A-he/she T water-buffalo

Here we see that in questions of the type claimed here to involve clefting (12), cliticization is to the verb rather than to the wh-word; we may conclude that a clause boundary separates the verb and the wh-word, which is what the structure in 9.a shows. In questions which do not involve clefts (13), no such clause boundary intervenes between the verb and the wh-word, and cliticization is therefore to the wh-word.

### 3.2 *ang*

The structures above also allow us to develop an account of the distribution of *ang*, which, as mentioned above, usually marks nominals in topic position:

14. a. Bumili **ang lalaki** ng tela  
AT-bought T man G cloth  
'The man bought cloth'
- b. Binili ng lalaki **ang tela**  
GT-bought A man T cloth  
'A man bought the cloth'

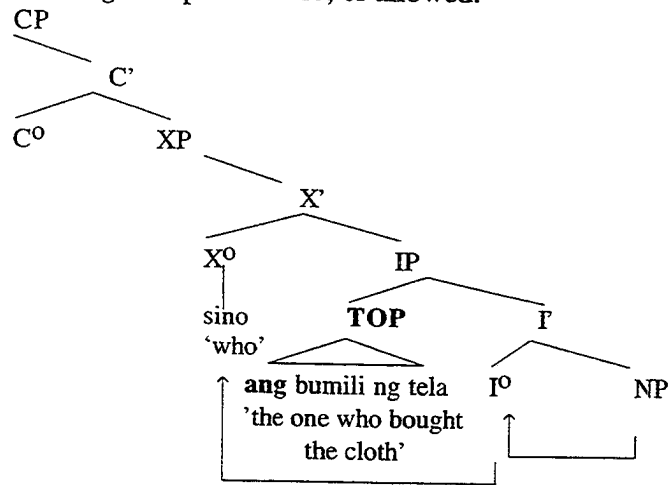
Wh-clefts obligatorily involve *ang* placed immediately before the verb, while wh-extraction without clefting disallows *ang* in this position:

15. a. Sino **(\*ang)** bumili ng tela?  
who T AT-bought G cloth  
'Who bought cloth?'
- b. Kailan **(\*ang)** bumili ang lalaki ng tela?  
when T AT-bought T man G cloth  
'When did the man buy cloth?'

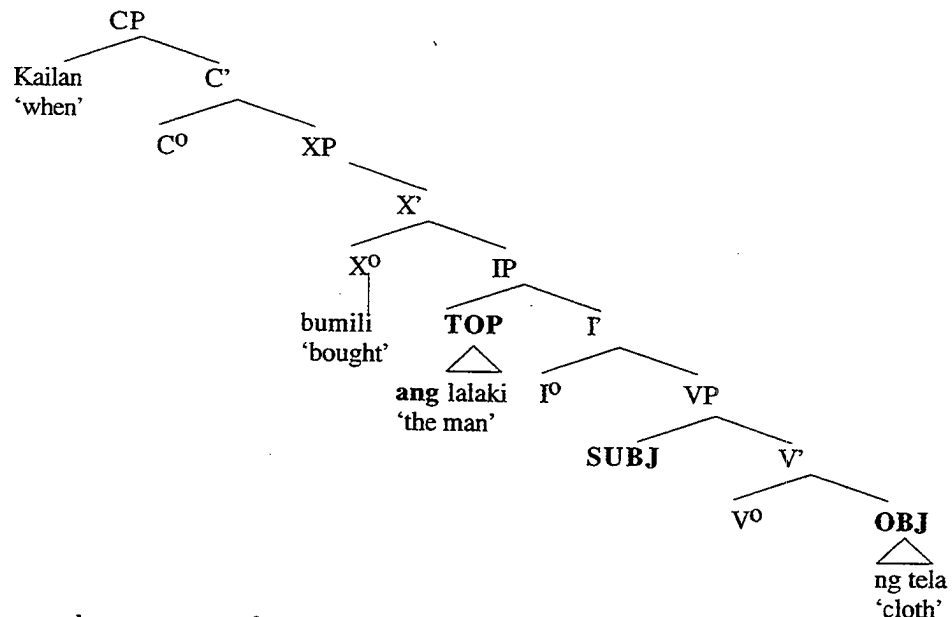
The structures given above in 9 (repeated below as 16) explain the distribution of *ang* straightforwardly. In a cleft, the string of words following the wh-word is a

topicalized headless relative NP, and like all topicalized NPs is marked with *ang*. In a non-cleft, the string of words following the *wh*-word begins with the verb of the main clause; no *ang* is expected here, or allowed:

16. a.



b.



We can use the presence of *ang* before the verb, then, as a diagnostic for a cleft. It is interesting to note that in relative clauses, as well, *ang* is not used, suggesting that relative clause formation does not obligatorily involve clefting:

17. ang lalaki-ng (\*ang) bumili ng tela  
 T man that T AT-bought G cloth  
 'the man who bought the cloth'

Surveying the range of cases discussed so far, then, we arrive at the following conclusions: *wh*-extraction which affects topicalization involves a cleft, while relativization (which affects topicalization) and *wh*-extraction which does not affect topicalization does not. What accounts for this distribution of clefts?

#### 4. Difficulties faced by Tagalog *wh*-words

In what follows I will claim that Tagalog makes extensive use of clefting in forming *wh*-questions because *wh*-words are very difficult to extract in Tagalog, in



a sense to be made clear shortly. As a result, the monoclausal structure which is typically used for wh-extraction is generally unavailable.

The topic position will be taken to be a barrier to A'-extraction of NPs; this may be an Economy effect, as in Nakamura's account (another possibility would be an account based on Relativized Minimality). I will further claim that semantic considerations prevent topicalization of wh-words, but not of relative operators. As a result, a wh-word can neither be moved past the topic (for syntactic reasons) nor through it (for semantic reasons). Wh-extraction of NPs from a position below the topic is therefore completely ruled out, and a cleft is employed to resolve the difficulty. Extraction of a relative operator, on the other hand, is permitted by the semantics to proceed via the topic position, which explains the lack of clefts in relative clauses. Wh-extraction of non-NPs (e.g., of the adverb *kailan* 'when') is permitted by the syntax to skip the topic, and the conflict between the syntax and the semantics thus fails again to arise.

We have seen that syntactic theory provides two closely related possible mechanisms for forcing wh-extraction of NPs not to skip the topic position (namely, Economy and Relativized Minimality). In the next section, I will develop some arguments which suggest that topicalization of a wh-word might be impossible for semantic reasons.

### 5. syntax vs. semantics: specificity?

Tagalog topics obligatorily receive a specific reading:

18. Binili ng lalaki ang kalabaw  
 GT-bought A man T water-buffalo  
 'A/the man bought the/a certain/\*a water buffalo'

Here the topic *ang kalabaw* 'T water-buffalo' cannot be a non-specific indefinite. Without trying to investigate too deeply the semantic nature of specificity, it seems fairly clear that a semantic relation to a referent familiar to the speaker is somehow involved (cf. Enç 1991, Ludlow and Neale 1991, Vangsnes 1994, among many others). Given that the purpose of a wh-word is presumably to signal the speaker's ignorance of a particular referent, it seems not implausible to assume that the semantics of a wh-word might be incompatible with specificity. In fact, a number of tests for specificity seem to point in this direction.

#### 5.1 Existentials

One well-known case in which specific NPs are preferred over non-specifics is in existential contexts:

19. a. There is a man in the garden  
 b. \*There is the man in the garden

Interestingly, as noted by Ross (1967) and Stockwell, Schachter, and Partee (1968), wh-words and relative operators also contrast in this environment; wh-words, like non-specific NPs, may appear in existential contexts, while relative operators, like specific NPs, may not:

20. a. Who is there in the garden?  
 b. \* I don't know the man who there is in the garden

This would seem to support the conclusion that wh-words are non-specific while relative operators are specific.

#### 5.2 Double-object constructions

Double-object constructions in which the Theme is specific seem to be another context which is sensitive to specificity; for many speakers, 21.a contrasts unfavorably with 21.b (Frazier 1993, Beckman 1994):

21. a. ??The nurse brought a doctor the patient  
 b. The nurse brought the doctor the patient

This is effectively the reverse of the existential context, apparently preferring specific nominals over non-specific ones. Interestingly, *wh*-words and relative operators seem to contrast again in this context:

22. a. ?? Who<sub>i</sub> did the nurse bring t<sub>i</sub> the patient?  
 b. ? This is the doctor OP<sub>i</sub> that the nurse brought t<sub>i</sub> the patient

Here, again, *wh*-words appear to act non-specific, while relative operators behave like specific nominals.

### 5.3 Partitives

In partitive constructions, as in the double object constructions discussed in the previous section, specific NPs are preferred over non-specifics:

23. a. \*I ate four of some pies  
 b. I ate four of the pies

As before, *wh*-words appear to act like non-specific nominals in this regard, and relative operators appear to be specific:

24. a. ?? What<sub>i</sub> did you eat four of t<sub>i</sub> yesterday?  
 b. ? These are the pies that I ate four of t<sub>i</sub> yesterday.

### 5.4 Tagalog *wh*-in-situ

Finally, some evidence internal to Tagalog suggests that *wh*-words must be impossible to topicalize. Tagalog allows *wh*-in-situ in certain restricted cases<sup>3</sup>. Interestingly, *wh*-words cannot be topicalized<sup>4</sup>:

25. a. Bumili **ang lalaki** ng ano sa tindahan?  
 AT-bought T man G what D store  
 'What did the man buy at the store?'  
 b. \*Binili ng lalaki **ang ano** sa tindahan?  
 GT-bought A man T what D store

This suggests, again, that Tagalog *wh*-words are incompatible with topicalization.

### 5.5 conclusion

In the above section I have tried to show that *wh*-words might be incompatible with the semantic properties of Tagalog topicalization, while relative operators might well be compatible with those semantic properties. The above discussion has, for reasons of space, been somewhat hurried, and a number of interesting issues have unfortunately been glossed over, but the basic trend appears fairly clear.

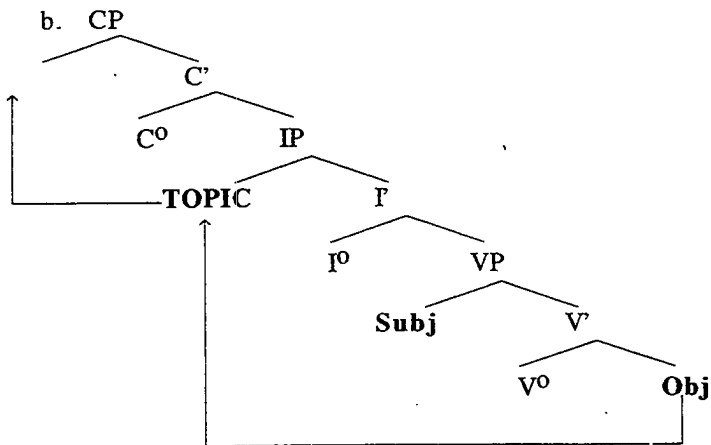
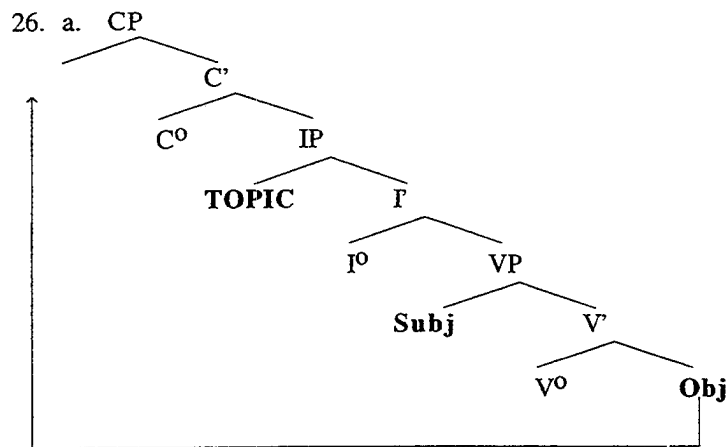
We can now return, then, to the two possible derivations involving a *wh*-word which has been base-generated in direct object position (6, repeated as 26):

<sup>3</sup> In particular, *wh*-in-situ is apparently only possible in main clauses; that is, i. contrasts with 25.a, for my informants, in only allowing an echo-question reading:

i. Pinaniniwalaan ni Juan na bumili ang lalaki ng ano sa tindahan?  
 DT-believes A Juan that AT-bought T man G what D store  
 'Juan believes that the man bought *what* at the store?'

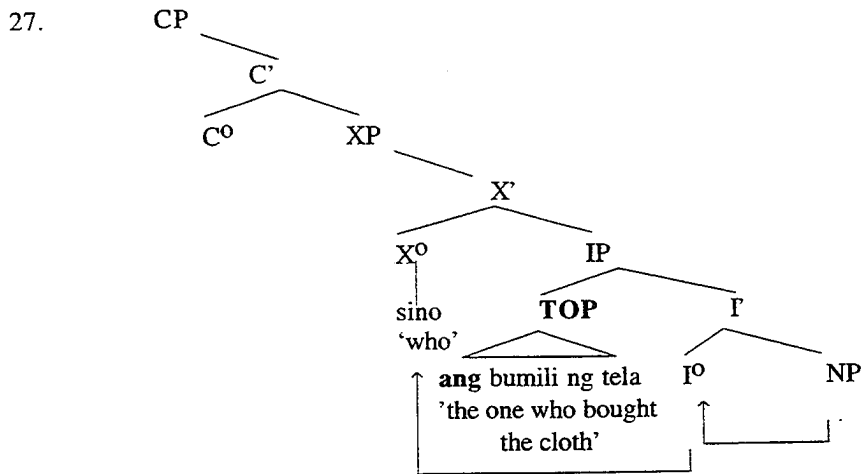
In this regard Tagalog appears to be not unlike French (cf. Bosković 1995).

<sup>4</sup> What is also interesting, of course, is that they *need* not be topicalized; that is, LF-moved *wh*-words are apparently subject to less strict constraints than overtly moved ones. I will not attempt to explain this here.



26.a, as we said before, can be ruled out for syntactic reasons (perhaps Economy-based reasons). We have now seen some arguments which suggest that 26.b might be ruled out as well for *wh*-words, on semantic grounds<sup>5</sup>. We have thus reasoned our way to the conclusion that a *wh*-word which is base-generated in direct object position cannot possibly be overtly extracted to Spec CP; both of the available derivations are blocked. Tagalog is thus unable to make use of structures in which a *wh*-word is base-generated in direct object position, and must instead use a cleft, as in 27:

<sup>5</sup> If we do use Economy to rule out 26.a, 26.b (if it is to be the preferred derivation, as in Nakamura's account) will have to be well-formed in the respects which are relevant for the comparison; in Chomsky's (1996) terminology, it will have to "converge as gibberish", blocking other derivations for syntactic reasons but being in fact uninterpretable by the interface.



Here the *wh*-word is base-generated as a predicate, and the element which is extracted from a direct object position is a relative operator.

Clefting, then, is used when *wh*-extraction from a given site is impossible; in the case we have just seen, this is because the syntax and the semantics conspire to rule out both of the possible extraction paths for a *wh*-word base-generated in direct object position. We saw before that relativization and certain *wh*-extractions do not involve clefting:

28. a. *ang lalaki-ng (\*ang) bumili ng tela*  
       T man that T AT-bought G cloth  
       ‘the man who bought the cloth’  
    b. *Kailan binili ng lalaki ang tela?*  
       when GT-bought A man T cloth

In both of these cases, the conspiracy between the syntax and the semantics breaks down. In 28.a, a relative operator is being extracted, and the semantics fails to rule out 26.b; relative operators are compatible with specificity and can therefore be topicalized. In 28.b, the syntax apparently allows the equivalent of 26.a; extraction may skip the topic position (possibly because topicalization of *kailan* is impossible, as in Nakamura’s Economy-based account).

We have seen that clefting may be understood as a “last resort” operation used to evade strict constraints on *wh*-extraction. In the next section I will explore this notion further, discussing some cases which would seem to be problematic.

### 6. No syntax-semantics clash?

The clefting cases above were ones in which the syntax and the semantics made conflicting demands on the extracted item; syntax demanded extraction via the topic position, while semantics ruled out topicalization of *wh*-words.

As Nakamura points out in his work, there are several cases in which topicalization is restricted in certain ways. Two such cases are given in 22-23. 22 is a “recent perfective” form, and 23 involves a non-verbal predicate<sup>6</sup>:

22. *Kabibili lang ng lalaki ng tela*  
       RP-bought just A man G cloth  
       ‘The man just bought the cloth’

<sup>6</sup> The claim that *pasan* ‘carrying on the shoulders’ is not a verb is primarily based on morphology; *pasan* lacks the aspectual and topicalization morphology which is typically required on verbs.

23. a. *Pasan*                    *ng lalaki ang anak*  
 carry-on-shoulders A man T child  
 'The man is carrying the child on his shoulders'  
 b. \**Pasan*                    *ang lalaki ng anak*  
 carry-on-shoulders T man G child

In 22, neither of the NPs is topicalized; this shows that verbs in the recent perfective form, unlike most Tagalog verbs, need not have a topic at all. In 23, the non-verbal predicate *pasan* 'carrying on the shoulders' must take the Goal NP *anak* 'child' as the topic; topicalization of the actor, as in 23.b, is impossible<sup>7</sup>.

Nakamura shows that, as we expect in an Economy-based framework, extraction without topicalization is possible precisely in these cases, where topicalization is impossible:

24. a. *Sino ang kabibili lang ng tela?*  
 who T RP-bought just G cloth  
 'Who just bought the cloth?'  
 b. *Ano ang kabibili lang ng lalaki?*  
 what T RP-bought just A man  
 'What did the man just buy?'
25. a. *Ano ang pasan ng lalaki?*  
 what T carry-on-shoulders A man  
 'What is the man carrying on his shoulders?'  
 b. *Sino ang pasan ang anak?*  
 who T carry-on-shoulders T child  
 'Who is carrying the child on his shoulders?'

In 24.b and 25.b, we are apparently extracting non-topicalized direct objects, a move which is typically ruled out, possibly by Economy. On the assumption that Economy simply requires that the shortest available moves be made, the well-formedness of 24.b and 25.b is unsurprising. Topicalization is not an available move, so it is not required by Economy.

What is surprising, on the account developed here, is the fact that the questions in 24-25 are clefts; note that they contain the topic-marker *ang* after the *wh*-word, which we have used as a diagnostic for clefting. I have claimed here that clefting is a product of a clash between the syntax and the semantics; the syntax requires topicalization, while the semantics forbids it, and a cleft is used to resolve the conflict. These appear to be cases in which the syntax does not require topicalization. Why, then, do they involve clefts?

Considerations of space prevent me from considering both of these cases in any detail. For the recent perfective case, then, I will simply note along with Cena (1995) that topicalization is in fact possible in these constructions, though not required:

26. a. *Kabibili lang ng lalaki ng tela*  
 RP-bought just A man G cloth  
 'A man just bought cloth'  
 b. *Kabibili lang ng lalaki ang tela*  
 RP-bought just A man T cloth  
 'A man just bought the cloth'

This raises the possibility that the topicalization position is available as a landing site, and that use of this position as an escape hatch by the extracted element is in fact required by the syntax, as in all the cases we have examined thus far.

<sup>7</sup> Irrelevantly for our purposes, 23.b can of course have the meaning 'the child is carrying the man on his shoulders'.

### 6.1 non-verbal predicates

Let us turn to the task of dealing with the facts of extraction with non-verbal predicates. To review the facts again; as first noted by Cena (1979), Tagalog has a class of non-verbal predicates in which only one of the two NP arguments can be topicalized, but either can be extracted:

27. a. *Pasan ng lalaki ang anak*  
 carry-on-shoulders A man T child  
 'The man is carrying the child on his shoulders'
- b. \**Pasan ang lalaki ng anak*  
 carry-on-shoulders T man A child  
 (ok on reading: 'The child is carrying the man on his shoulders')
- c. *Ano ang pasan ng lalaki?*  
 what T carry-on-shoulders A man  
 'What is the man carrying on his shoulders?'
- d. *Sino ang pasan ang anak?*  
 who T carry-on-shoulders T child  
 'Who is carrying the child on his shoulders?'

The intuition pursued in earlier accounts of this fact, including Cena (1979) and Nakamura (1994, 1995, 1996), is that extraction need only proceed via the Topic position if it possibly can; in this case, topicalization of the Actor is impossible, and thus unnecessary for extraction. Here I will sketch a possible alternative account, according to which the peculiar behavior of non-verbal predicates has to do with their syntactic structure.

Facts about Tagalog clitic placement provide some evidence that verbal and non-verbal predicates may have different syntactic structures. Recall that Tagalog has a class of second-position clitics, which necessarily follow the first constituent in the sentence. This constituent may be a head, as in 28.a-b, or a maximal projection, as in 28.c.

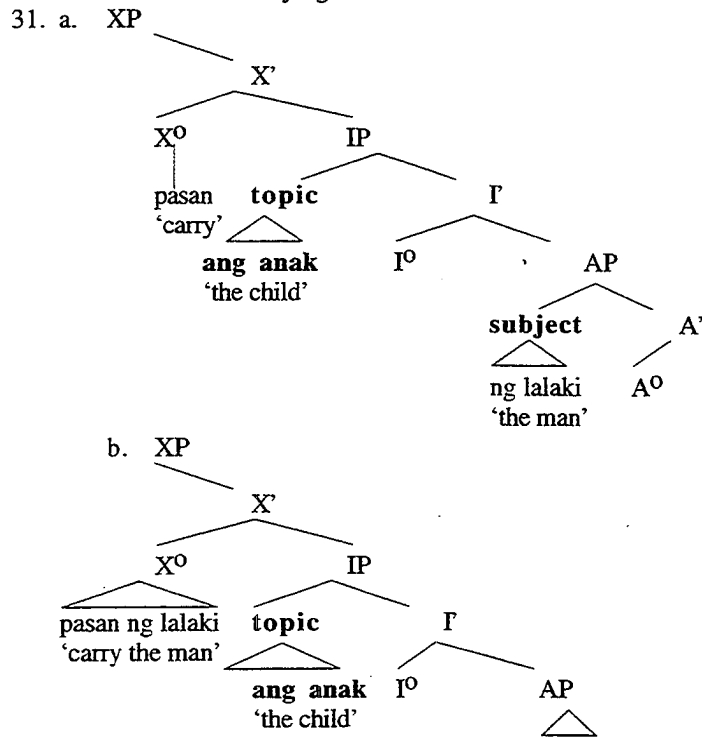
28. a. *Bumili siya ng tela*  
 AT-bought T-he/she G cloth  
 'He/she bought cloth'
- b. *Hindi siya bumili ng tela*  
 not T-he/she AT-bought G cloth  
 'He/she didn't buy cloth'
- c. [*Bukas ng alas otso ng gabi*] *siya aalis*  
 tomorrow at eight night T-he/she AT-will-leave  
 'He/she will leave tomorrow at 8 p.m.'

Cliticization, then, may be used as a test for constituency. It is interesting to note, then, along with Kroeger (1993), Sityar (1989), and Schachter and Otnes (1972), that non-verbal predicates differ from verbal ones in that cliticization may optionally treat the entire fronted predicate as a constituent:

29. a. *Pumunta siya sa Maynila*  
 AT-went T-he/she D Manila  
 'He/she went to Manila'
- b. *Takot siya sa kulog*  
 afraid T-he/she D thunder  
 'He/she is afraid of thunder'
- c. \*[*Pumunta sa Maynila*] *siya*  
 AT-went D Manila T-he/she
- d. [*Takot sa kulog*] *siya*  
 afraid D thunder T-he/she  
 'He/she is afraid of thunder'

Developing a theory of this fact is beyond the scope of this paper. Here I will simply adopt the proposal of Carnie (1995), who suggests that these predicates may be analyzed as complex heads; in 29.d, on this proposal, the phrase *takot sa kulog* 'afraid of thunder' has been analyzed as a head. A sentence like 30, then, may have either of the structures in 31. 31.a shows the more orthodox of the possibilities, in which the head *pasan* has undergone raising into the head of some IP-level category, just as a verb would. In 31.b, the entire predicate *pasan ng lalaki* 'carried on the shoulders by the man' has been analyzed as a complex head, and has undergone head-movement just like *pasan* in 31.a.

30. *Pasan ng lalaki ang anak*  
 carry-on-shoulders A man T child  
 'The man is carrying the child on his shoulders'

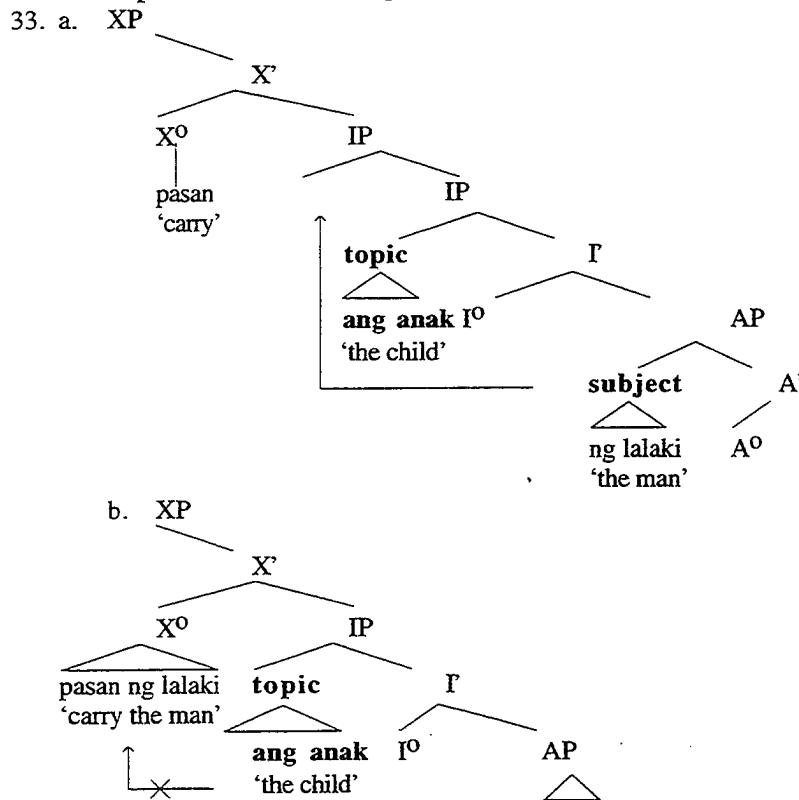


*pasan ng lalaki* 'carried on the shoulders by the man', on this theory, has a structure available to it which is unavailable to a verbal predicate (namely, that in 31.b), in which the topic does not c-command the other argument. Perhaps it is this syntactic ambiguity which renders the non-topic extractable, rather than the restricted topicalization possibilities. We might hypothesize that extraction of non-topics in non-verbal predicates always involves the structure in 31.b, in which the non-topic has been moved past the topic by being incorporated into a complex head. In fact, there is some evidence for this hypothesis.

Tagalog word order is fairly free; the sentences in 32, for instance, are well-formed and synonymous:

32. a. *Pasan ng lalaki ang anak*  
 carry-on-shoulders A man T child  
 'The man is carrying the child on his shoulders'
- b. *Pasan ang anak ng lalaki*  
 carry-on-shoulders T child A man

I assume, as is standard, that these variations in word order are a result of movement, perhaps adjunction to an IP projection. Assuming that such movement must be to a position c-commanding the trace, we expect it to be possible only when the non-verbal predicate head has raised by itself, as in 33.a, and not when the entire predicate has undergone head-movement, as in 33.b:



Thus, the word order in 32.b can only be a result of the structure in 33.a, not that in 33.b. On the hypothesis that the structure in 33.b is necessary for extraction of the non-topic NP, we reach the conclusion that extraction can only take place from the position immediately following the predicate head; the topic cannot intervene between *pasan* and the extraction site. How do we test this prediction?

Tagalog allows a form of “possessor raising” in which the possessor of a topic NP can be extracted. This is shown in 34:

34. Iyon ang doktor na bago **ang kotse**  
 that T doctor that new T car  
 ‘That is the doctor whose car is new’

These constructions can be used as a diagnostic for the location of the extraction site, assuming that extraction takes place from inside the possessed NP.

Interestingly, extraction of the possessor of the non-topic argument of a non-verbal predicate is only possible if the non-topic is adjacent to the predicate head, just as the account sketched above predicts:

35. a. Iyon ang doktor na [pasan ng asawa] **ang bigas**  
 that T doctor that carry-on-shoulders A spouse T rice  
 ‘That is the doctor whose spouse is carrying the rice on his/her shoulders’  
 b. \*Iyon ang doktor na pasan **ang bigas** ng asawa  
 that T doctor that carry-on-shoulders T rice A spouse



This suggests that the above account is on the right track; the peculiar extraction properties of non-verbal predicates arise from the availability of a structure in which the predicate raises as a unit. This process moves non-topic NPs to a higher position than that attainable by non-topics in a sentence with a verbal predicate. We have seen, then, that the syntax does indeed constrain extraction of non-topic NPs even in sentences with non-verbal predicates, in which they could not in principle be topicalized<sup>8</sup>.

We are still left with the question, however, with which this section began; why is clefting necessary in these structures? Why is the *wh*-word not simply extracted from the non-verbal predicate? A partial answer may be found in the parallel behavior of non-verbal predicates in Irish, to which we now turn.

### 6.2.2 Non-verbal predicates revisited: Irish

Carnie (1995) notes that Irish non-verbal predicates, like their Tagalog counterparts, may behave as a syntactic unit in a way which is impossible for verbal predicates. Irish word order is typically VSO (Carnie (1995), p. 110):

35. *Leanann an t-ainmní an briathar i nGaeilge*  
 follow-PRES the subject the verb in Irish  
 'The subject follows the verb in Irish'

However, when the predicate is non-verbal the entire predicate must precede the subject (Carnie (1995), p. 183):

36. *Is [dochtúir capall] Seán*  
 C doctor horses-GEN Sean  
 'Sean is a doctor of horses'

Carnie's theory, both of this and of the Tagalog facts discussed above, is that non-verbal predicates in these languages may be analyzed as a complex heads and undergo head-movement to a clause-initial position. As support for this claim he notes that non-verbal predicates are opaque to extraction in Irish, as shown in 37.b. The ill-formedness of 37.b is not merely a Subjacency effect; as 37.c shows, extraction from a relative clause is perfectly well-formed in Irish, as long as the relative clause is not part of the predicate (Carnie (1995), p. 194):

37. a. *Is [amhrán a<sup>L</sup> bhuaifidh an píobaire] (é) "Yellow Submarine"*  
 C song that play-FUT the piper AGR  
 "'Yellow Submarine' is a song which the bagpiper is going to play'  
 b. *\*Cén Píobaire arb [amhrán a<sup>L</sup> bhuaifidh sé] "Yellow Submarine"*  
 which piper rel song that play-FUT he  
 'Which bagpiper is "Yellow Submarine" a song which he is going to play?'  
 c. *Cén Píobaire arb é "Yellow Submarine" [an t-amhrán a<sup>L</sup> bhuaifeadh sé]*  
 which piper rel AGR the song that play-FUT he  
 'Which bagpiper is "Yellow Submarine" the song which he is going to play?'

Returning to Tagalog, then, we are now in a position to explain why extraction in 38 must involve clefting:

38. *Sino ang pasan ang anak?*  
 who T carry-on-shoulders T child  
 'Who is carrying the child on his shoulders?'

We saw evidence in the previous section that sentences like 38 involve a structure analogous to that in 36, in which the non-verbal predicate is fronted as a unit. I suggested above that this might be a way of "smuggling" the non-topic argument past the topic, thus rendering it free to extract. We have now seen, however, that such structures are impervious to *wh*-extraction (in Irish, at least, and by

<sup>8</sup> This approach thus somewhat undermines one of the major arguments for the Economy-based approach.

assumption in Tagalog as well). Extraction of the non-topic argument of a non-verbal predicate is thus faced with a paradox; the non-topic must be moved along with the rest of the predicate in order to get past the topic, but this apparently renders direct wh-extraction impossible. Clefting apparently resolves the paradox.

A remaining question, of course, is why clefting is any easier than wh-extraction in this context. Considerations of space prevent me from addressing this question at any length. One possibility, following Martohardjono (1988) and Richards (1991a, b), would be to say that the Tagalog relative operator used in clefts and relative clauses differs from its English counterpart in that it does not undergo wh-movement. This would explain why clefting avoids the island effects created by non-verbal predicate fronting; the operator involved remains in situ, and need not attempt to escape the island.

## 7. Conclusions

In this paper I have investigated certain properties of Tagalog wh-extraction which are problematic for existing theories. We have seen that Tagalog wh-movement sometimes involves a cleft. I have suggested that clefting may be a solution to cases in which conflicting demands on the extracted element make direct extraction impossible. We have seen two such cases. One is a clash between the syntactic and semantic components of the grammar; the syntax rules out extraction past the Topic position, while the semantic properties of topicalization rule out topicalization of wh-words and thus prevent wh-extraction from using the Topic position as an escape hatch. Another case had to do with extraction of non-topics in sentences with non-verbal predicates. I noted certain clitic placement facts which suggested that non-verbal predicates in Tagalog may undergo raising to sentence-initial position as a unit. I suggested that this property was responsible for the peculiar extraction properties of non-verbal predicates; non-topic arguments may raise with the rest of the predicate to a position above that of the topic. Direct extraction from this position is impossible, however; as in Irish, raising a predicate XP as a unit makes the predicate opaque for extraction.

## References

- Beckman, Jill. 1994. Double objects, definiteness, and extraction: a processing perspective. *UMOP* 19.
- Bošković, Željko. 1995. Superiority effects with multiple wh-fronting in Serbo-Croatian. ms., U.Conn.
- Carnie, Andrew. 1995. Non-verbal predication and head-movement. Ph.D. dissertation, MIT.
- Cena, Resty. 1979. Tagalog counterexamples to the accessibility hierarchy. *Studies in Philippine Linguistics* 31:119-124.
- Cena, Resty. 1995. Surviving without relations. ms.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Enç, Mürvet. 1991. The semantics of specificity. *Linguistic Inquiry*. 22:1-26.
- Frazier, Lyn. 1993. Processing Dutch sentence structures. *Journal of Psycholinguistic Research* 22:85-108.
- 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.
- Kroeger, Paul. 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford, CA: Center for the Study of Language and Information.

- Ludlow, Peter, and Stephen Neale. 1991. Indefinite descriptions: in defense of Russell. *Linguistics and Philosophy* 14. 171-202.
- Martohardjono, Gita. 1988. untitled ms., Cornell University.
- Nakamura, Masanori. 1994. An economy account of wh-extraction in Tagalog. In *Proceedings of the twelfth West Coast Conference on Formal Linguistics*, 405-420. Stanford, CA: Center for the Study of Language and Information.
- Nakamura, Masanori. 1995. Reference set, Minimal Link Condition, and parameterization. Paper presented at the Workshop on Optimality in Syntax, MIT, Cambridge, MA.
- Nakamura, Masanori. 1996. Economy of chain formation. Ph.D. dissertation, McGill University.
- Richards, Norvin. 1990. Tagalog subjects. ms., Cornell University, Ithaca, NY.
- Richards, Norvin. 1991a. Tagalog relative clauses. ms., Cornell University, Ithaca, NY.
- Richards, Norvin. 1991b. Wh-extraction in Tagalog. ms., Cornell University, Ithaca, NY.
- Ross, John. 1967. Constraints on variables in syntax. Ph.D. dissertation, MIT.
- Schachter, Paul. 1976. The subject in Philippine languages: Topic, actor, actor-topic, or none of the above. In *Subject and topic*, ed. by Charles Li, 491-518. New York: Academic Press.
- Schachter, Paul, and Fe Otnes. 1972. *Tagalog reference grammar*. Berkeley and Los Angeles: University of California Press.
- Seiter, William. 1975. Information questions in Philippine languages. ms., University of California, San Diego.
- Sityar, Emily. 1989. Pronominal clitics in Tagalog. Master's thesis, University of California, Los Angeles.
- Soh, Hooi-Ling. 1996. Certain restrictions on A-bar movement in Malay. This volume.
- Stockwell, Robert, Paul Schachter, and Barbara Partee. 1968. *The Major Syntactic Structures of English*. New York: Holt, Rinehart, and Winston.
- Vangsnes, Øystein. 1994. The syntactic source and semantic reflexes of the definiteness effect. Thesis, University of Bergen.



## 0. INTRODUCTION

This paper proposes that nominative case in Cebuano is restricted to a certain class of pronouns. Wolff 1966 refers to these pronouns as *preposed genitives* although his term covers three distinct uses: the *possessive predicate* pronoun which functions as a predicate, the *preposed genitive* pronoun which functions as an argument in a nominal phrase, and the *nominative* pronoun which functions as an argument in a clause. I reserve the term *preposed genitive* for the second meaning and refer to both preposed genitives and nominative pronouns as the class of *preposed pronouns*.

The claim that a preposed pronoun can bear nominative case is based on its dependence upon finite verb inflection. A nominative pronoun is licensed if and only if it is in Spec, Tense Phrase (TP) and a finite verb is in Tense (T<sup>0</sup>). Syntactic restrictions on nominative pronouns are helpful in articulating the clause structure of Cebuano and revealing the extent of verb movement. In addition, the restriction that nominative pronouns can only occur as agents of passive verbs leads me to propose that voice relations in Cebuano are licensed in an A-bar position which is lower than Spec, TP.

## 1. THE INTERNAL STRUCTURE OF PREPOSED PRONOUNS

Cebuano has several classes of personal pronouns, more or less derivable from a set of disyllabic stems.<sup>2</sup> The various pronoun classes correspond to the various argument cases, except for the possessive predicate pronouns which are used as possessive/locative predicates. As shown in the following chart, the preposed pronouns differ from other pronoun classes in the way that they indicate morphological case.

(1)

Pronoun Class	Morphological form	Example: 3 SG stem: íya
Preposed:	stem + linker nga/-ng	íyang
Poss Pred:	(sa) + stem	(sa) íya
n-Genitive:	/n/ + stem	níya
Obj/Obl:	kan- + stem	kaniya,
or	sa + stem	sa íya
Topic: (some irregular stems)		
1st and 2nd SG	stem + final stress	
1st and 2nd PL	/k/ + stem + final stress	
3rd	/s/ + stem + final stress	siyá

The first column in the chart names the pronominal class: (nominative/genitive) preposed pronouns, possessive predicate pronouns, n-genitive pronouns, object or oblique pronouns, and topic pronouns. The second column broadly describes the morphological form of a pronoun class, including stem changes and accompanying particles; and the third column illustrates the third person singular member of each pronoun class.

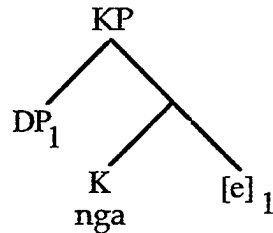
<sup>1</sup>I gratefully acknowledge the help of Hilda Koopman, Peter Sells and Anna Szabolsci. Thanks to John Wolff and his consultants Ida Wolff and Marra Pepito for answering my e-mail queries on data and to my Cebuano consultants in Florida, Cora Pelayo, Thelma Johnson and Victoria Timonera. Finally, thanks to my mom for finding these Cebuano speakers. All errors and inaccuracies are my own.

<sup>2</sup>Short forms of first and second person singular genitive and topic pronouns drop the initial syllable of a pronoun stem.

In general, morphological case in Cebuano is marked at the left edge of the nominal phrase. For example, some pronouns indicate case by adding an initial consonant or prefix to the pronominal stem: the initial /n/ of n-genitive pronouns (hence the name “n-genitive”), the prefix *kan-* of object/oblique pronouns, the initial /k/ of first and second person plural topic pronouns, and the initial /s/ of third person topic pronouns.<sup>3</sup> Other nominal arguments have a preceding case marker. As shown in (1), a possessive predicate pronoun is optionally marked by *sa*, while an object/oblique pronoun which is not prefixed with *kan-* requires the *sa* case marker. A nonpronominal argument indicates case on its preceding case marker or on a preceding demonstrative which encodes case. Even a personal name is preceded by a case marker. In short, all nominal arguments in Cebuano project a Kase Phrase (KP).

Unlike other pronouns, the preposed pronoun never occurs with a preceding case marker nor adds prefixes or initial consonants to its stem. In other words,  $K^0$  is not overtly realized at its left edge. The preposed pronoun consists of a bare stem followed by the linker *nga*, which usually appears in its clitic form as a velar nasal. My analysis of the structure of a preposed pronoun is given in (2).

(2) Structure of a preposed pronoun



A preposed pronoun is generated as a KP like all other arguments in Cebuano. The linker *nga* is a case marker in  $K^0$  which takes a DP complement, i.e. the bare pronoun stem. The DP moves to Spec, KP, so that it is to the left of the linker in  $K^0$ . This movement to Spec, KP is a bit mysterious. Why must DP move to Spec, KP? Perhaps the  $K^0$  *nga* requires its specifier to be filled in order to satisfy its clitic features. Why is it that only pronouns occur with preposed genitive or nominative case? Perhaps DP movement is restricted to pronouns and nonarguments; nonpronominal arguments can only move as whole KPs or PPs.

## 2. THE SYNTACTIC DISTRIBUTION OF NOMINATIVE PRONOUNS

A preposed pronoun is often interchangeable with an n-genitive pronoun in nominals or clauses. However, comparing the distribution of nominative pronouns and n-genitive pronouns suggests that they are licensed in different ways.

### 3.1 Surface position

One difference between nominative pronouns and n-genitives is their surface positions. As shown in (3), nominative pronouns can occur sentence initially but n-genitives cannot:

- (3) a. Akong                    gipalit    ang tila.  
 NOM.1.SG=LK R.P-buy TOP cloth  
 'I bought the cloth.'  
 b. \*Nako gipalit ang tila.  
 ('I bought the cloth.')

<sup>3</sup>Topic pronouns may actually indicate case by final stress, with the initial consonant changes marking person or number features. Thus, topic pronouns may also move to Spec, KP, as argued for preposed pronouns (cf. the structure in (2)).

- c. \*Sa babaye/Ni Maria gipalit ang tila.  
(‘The woman/Maria bought the cloth’)

The nominative pronoun *akong* ‘I’ appears sentence initially in (3a). N-genitive pronouns are second position clitics, so they do not appear sentence initially, as shown by *nako* in (3b). However, even a nonpronominal n-genitive is disallowed in sentence initial position, as shown in (3c).<sup>4</sup>

N-genitives can occur postverbally but nominative pronouns cannot, as shown in (4):

- (4) a. Gipalit nako ang tila.  
R.P-buy GEN.1.SG TOP cloth  
‘I bought the cloth.’  
b. \*Gipalit akong ang tila.

In (4a), the n-genitive pronoun *nako* ‘I’ can follow the verb, but the nominative pronoun cannot occur in this position, as shown by the ungrammaticality of (4b).

When it occurs sentence initially, a nominative pronoun can host second position clitics, as shown in (5a).

- (5) a. Ako na lang siyang tawgon.  
NOM.1.SG CL CL TOP.3.SG=LK call-IR.P  
‘I’ll just call him/her.’ (Wolff 1966:251)  
b. Tawgon na lang siya nako.  
call-IR.P CL CL TOP.3.SG GEN.1.SG  
‘I’ll just call him/her.’ (Wolff 1966:251)

In (5a), three second position clitics *na lang siya* intervene between the nominative pronoun stem *ako* ‘I’ and its  $K^0$  linker *ng*. The n-genitive pronoun *nako* ‘I’, itself a second position clitic, occurs as the last member of the clitic group, as shown in (5b).

(6) shows that either an n-genitive or a nominative pronoun can occur in a position which is neither postverbal nor sentence initial.

- (6) a. Gahapon akong gipalit ang tila.  
yesterday NOM.1.SG=LK R.P-buy TOP cloth  
‘Yesterday I bought the cloth.’  
b. Gahapon nako gipalit ang tila.  
‘Yesterday I bought the cloth.’

As shown in (6a) and (6b), either a nominative pronoun or an n-genitive can occur after a sentence initial temporal adjunct such as *gahapon* ‘yesterday’.

### 3.2 Prepositives and dependent mood inflection

Wolff 1966 I:234 lists some contexts in which the n-genitive must occur instead of the nominative pronoun. These contexts fall into two categories: sentences containing certain “prepositives,” so called because they precede the verb, and exclamatory sentences. Both contexts can be shown to involve verbal inflection that is too deficient to license a nominative pronoun.

<sup>4</sup>This suggests that the n-genitive  $K^0$  must be also be governed, like the  $K^0$  of a preposed pronoun. The preposing  $K^0$  satisfies its clitic government requirement by attracting its own DP to its specifier whereas the n-genitive satisfies its clitic government requirement by an m-commanding host.

The prepositives listed in Wolff as being unable to occur with a nominative pronoun include negative particles and modals, both of which behave like auxiliary verbs containing mood features. For example, the negative particle *dili* signifies irrealis mood and selects for irrealis inflection on the verb while the negative particle *wala* signifies realis mood and selects for subjunctive inflection on the verb (realis inflection for some speakers). Modals such as *kinahanglan* 'must/should', *mahimo* 'can', and *gusto* 'want' select for irrealis mood inflection on the verb.

N-genitives can occur with negation or modals, as shown by the (a) examples below. But nominative pronouns generally do not occur with negation or modals, as shown by the (b) and (c) examples.

- (7) a. *Dili nako paliton ang tila.*  
 IR.NEG GEN.1.SG buy-IR.P TOP cloth  
 'I won't buy the cloth.'  
 b. \**Dili akong paliton ang tila.*  
 c. \**Akong dili paliton ang tila.*
- (8) a. *Wala ko siya pangutana.*  
 R. NEG GEN.1.SG TOP.3.SG ask-SU.P  
 'I didn't ask her/him.' (Wolff 1966: 234)  
 b. \**Wala ako siyang pangutana.*  
 c. \**Ako siyang wala pangutana.*
- (9) a. *Kinahanglan nakong paliton ang tila.*  
 must GEN.1.SG=LK buy-IR.P TOP cloth  
 'I must buy the cloth.'  
 b. \**Kinahanglang akong paliton ang tila.*  
 c. \**Akong kinahanglang paliton ang tila.*
- (10) a. *Mahimo natong pangutan'on ang tiguwang.*  
 possible GEN.1.PL.INC=LK ask-IR.P TOP old  
 'We can ask the old one.' (Wolff I 256)  
 b. \**Mahimong atong pangutan'on ang tiguwang.*  
 c. \**Atong mahimong pangutan'on ang tiguwang.*
- (11) a. *Gusto nakong paliton ang tila.*  
 want GEN.1.SG=LK buy-IR.P TOP cloth  
 'I want to buy the cloth.'  
 b. \**Gusto akong paliton ang tila.*  
 c. *Akong gustong paliton ang tila.*  
 'I want to buy the cloth.'

Thus, the nominative pronoun is generally not allowed in sentences which contain negation or a modal. (11c) is an exception, since it shows that a nominative pronoun can precede *gusto*.<sup>5</sup>

<sup>5</sup>A more puzzling exception is exemplified by the following text from Wolff 1967: 365, paragraph 18.

*Dili mahimo kana=ng akong buhatan.*  
 IR.NEG possible TOP.that=LK NOM.1SG=LK do-IR.LOC  
 'I cannot do that.'

In this sentence, a nominative pronoun *akong* occurs with a negated modal. Note that the topic *kana* 'that' precedes the nominative pronoun. According to Wolff (p.c.), speakers are inconsistent about accepting this



Not all prepositives are incompatible with a nominative pronoun. A nominative pronoun can occur after a sentence initial adjunct, as shown in (6a) repeated below.

- (6) a. *Gahapon akong gipalit ang tila.*  
 yesterday NOM.1.SG=LK R.P-buy TOP cloth  
 'Yesterday I bought the cloth.'

Whether or not a prepositive can cooccur with a nominative pronoun depends on whether or not mood inflection on the verb marks the true mood of the clause or is restricted by the prepositive. Although *gahapon* 'yesterday' has inherent temporal features, mood in (6a) is independently expressed by the realis verbal affix *gi-*.

In preverbal position, a temporal adjunct like *gahapon* 'yesterday' can mark the tense of the clause while the verb appears in subjunctive mood (cf. Wolff 1966:527).<sup>6</sup> From what we have seen with negation and modals, we can predict that a nominative pronoun should not be permitted in this context. This prediction is borne out by the data in (12).

- (12) a. *Gahapon nako palita ang tila.*  
 PAST-yesterday GEN.1.SG buy-SU.P TOP cloth  
 'Yesterday I bought the cloth.'  
 b. \**Gahapon akong palita ang tila.*

In (12), the verb *palita* 'buy' is inflected for subjunctive mood. A subjunctive verb can occur with an n-genitive pronoun, as shown in (12a), but it cannot occur with a nominative pronoun, as shown in (12b).

Like temporal adjuncts, some deictic pronouns can express tense when they occur in preverbal position. A preverbal deictic can be followed by either a subjunctive verb or a finite verb which expresses the true mood of the clause. As expected, a finite verb can occur with a nominative pronoun, but a subjunctive verb cannot, as shown in (13).

- (13) a. *Didto iyang gipalit ang tila.*  
 PAST.there NOM.3.SG=LK R.P-buy TOP cloth  
 'S/he bought the cloth there.'  
 b. \**Didto iyang palita ang tila.*  
 PAST.there NOM.3.SG=LK buy-SU.P TOP cloth  
 'S/he bought the cloth there.'

- (14) a. *Diin imong gipalit ang beer Ting?*  
 PAST.where NOM.2.SG=LK R.P-buy TOP beer Ting  
 'Where did you buy beer, Ting?'  
 b. \**Diin imong palita ang beer Ting?*  
 PAST.where NOM.2.SG=LK buy-SU.P TOP beer Ting  
 'Where did you buy beer, Ting?'

In (13)-(14), the deictic pronouns *didtu* 'there' and *diin* 'where' occur preverbally, so they can indicate tense. The (a) examples contain the finite verb *gipalit* 'bought' which is compatible with a nominative pronoun. The (b) examples contain the subjunctive verb *palita* 'buy' which cannot occur with a nominative pronoun.

---

sentence, but when they do accept it, they assign contrastive focus to the topic: 'No it's impossible for me to do THAT (as opposed to something else)'.

<sup>6</sup>Subjunctive mood is realized differently for active and passive verbs. Passive verbs use a separate set of subjunctive mood affixes while active verbs use the same set of affixes used in irrealis mood.

In summary, the appearance of a nominative pronoun seems to depend upon verbal inflection. When verbal inflection is restricted by a prepositive, the preposed pronoun is disallowed, but when a verb independently carries the inflectional features of the clause, the nominative pronoun is licensed. This dependency upon verbal inflection supports the analysis of preposed pronouns as bearing nominative case since nominative case is generally seen as the structural case licensed by finite inflection.

### 3.3 Inflection lacking voice and mood features

The second context given by Wolff as not allowing nominative pronouns involves exclamatory sentences, as shown in (15):

- (15)a. Kanindut        nimog [nimo ug] balay!  
          nice    GEN.2.SG=INDEF    house  
          ‘What a nice house you have!’ (Wolff 1966:234)  
       b. \*Imong kanindut ug balay!

An n-genitive pronoun can appear in an exclamatory sentence such as (15a), but a nominative pronoun cannot, as shown by the ungrammatical (15b). An exclamatory sentence is characterized by its lack of a topic and by the verbal affix *ka-* which lacks voice and mood features.

The recent perfective construction in Cebuano is similar to an exclamatory sentence in its lack of a topic and lack of voice or mood features on the verb. Not surprisingly, recent perfective sentences are also not allowed to contain nominative pronouns.

- (16) a. Kapital lang    nako    ug    tila.  
          PERF-buy CL GEN.1.SG INDEF cloth  
          ‘I just bought some cloth.’  
       b. \*Akong kapital ug tila.  
          ‘I just bought some cloth.’

The n-genitive NP *nako* can occur in a recent perfective sentence, as shown in (16a), but the nominative pronoun is not possible, as shown in (16b).

Is the nonoccurrence of the nominative pronoun due to the “deficiency” in verbal inflection or the lack of a topic? These factors may be related, as shown by comparing (15) and (16) with (17), in which a relative clause lacks an overt topic but contains a fully inflected verb and also a nominative pronoun.

- (17)    Naa    bay        imong        nadala?  
          PRES.there Q=PART NOM.2.SG=LK R.P-bring  
          ‘Did you bring anything?’  
          (Lit. ‘Is there anything brought by you?’) (Wolff 1966:523)

(17) consists of the deictic pronoun *naa* ‘there, present tense’, the question marker clitic *ba*, and the headless relative clause *y imong nadala* ‘(thing) brought by you’.<sup>7</sup> The topic of the headless relative clause may be represented as a WH gap corresponding to ‘anything’ in Wolff’s translation. The verb is fully inflected not only for realis mood but also for patient voice. Voice features of the verb specify the thematic properties of the topic gap, so that the relative clause does not truly lack a topic whereas exclamations and recent perfective sentences do. Thus, the licensing of a nominative pronoun is dependent upon a verb having

<sup>7</sup>The case marker *y* is phonologically and orthographically attached to *ba*.. See Sityar (to appear) for an analysis of *y* indefinites.

inflectional features that specify the mood of the clause and the thematic properties of the topic.

### 3. AGENTS OF PASSIVE VERBS

Cebuano allows verbal as well as nonverbal (i.e. nominal, adjectival, and prepositional) predicates, but only verbal predicates are overtly inflected for mood, voice, aspect, etc. It appears as though the nominative pronoun cannot occur as an argument of a nonverbal predicate. The fact that the nominative pronoun occurs only in verbal predicate sentences suggests that nominative case must be licensed by strong inflectional features, thus supporting the analysis that the licensing of a nominative pronoun depends upon fully specified verbal inflection.

However, nominative case is not available in all verbal predicate sentences. As noted by Wolff, a nominative pronoun must occur as the agent of a passive verb.<sup>8</sup> It cannot occur with an active verb.

- (18) a. Nagpalit akó ug tila.  
 R.A-buy TOP.1.SG INDEF cloth  
 'I bought some cloth.'
- b. \*ákong nagpalit ug tila.  
 NOM.1.SG=LK R.A-buy INDEF cloth  
 ('I bought some cloth.')
- c. Miabot na siya.  
 R.A-arrive CL TOP.3.SG  
 'She/He has already arrived.'
- d. \*Iyang miabot.  
 NOM.3.SG=LK R.A-arrive  
 (He/She arrived.)
- (3) a. Akong gipalit ang tila.  
 NOM.1.SG=LK R.P-buy TOP cloth  
 'I bought the cloth.'

In (18), the transitive verb *nagpalit* 'bought' and the intransitive verb *miabot* 'arrived' are inflected for actor voice. (18a) and (18c) are acceptable with the actors *ako* 'I' and *siya* 'she/he' in topic case. But (18b) and (18d) are unacceptable because the actors *akong* 'I' and *iyang* 'she/he' appear in nominative case. Thus, nominative case is not available to either transitive or intransitive verbs in actor voice. In (3a), the verb *gipalit* is inflected for patient voice. It marks the patient *ang tila* 'the cloth' with topic case and the actor pronoun *akong* 'I' with nominative case.

Unlike nominative case, preposed genitive case is not restricted to the agent of a transitive predicate. In nominals, a preposed genitive pronoun can be the external argument of a transitive noun or the argument of an intransitive noun, as shown in (19).

- (19) a. ang imong pagpatay sa iya.  
 TOP GEN.2.SG=LK GER-kill O 3.SG  
 'your killing of him/her'  
 not 'his/her killing of you'

<sup>8</sup>There are basically three sets of passive voice classes, patient (P) voice which selects an argument that undergoes a change in state, theme (T) voice (a.k.a Wolff's instrumental passive) which selects an argument that undergoes a change in location, and locative (L) voice which selects an argument which is the locus of an action.

- b. ang iyang pagabot  
 TOP GEN.3.SG=LK GER-arrive  
 'his/her arrival'

In (19a), the preposed genitive pronoun *imong* 'your' must be interpreted as the agent of *pagpatay* 'killing', not as the patient. Yet in (19b), the preposed genitive pronoun *iyang* 'his/her' occurs with the unaccusative noun *pagabot* 'arrival' and is presumably an internal argument. This contrasts with (18d) in which the nominative pronoun is disallowed with the unaccusative verb *miabot* 'arrived'. Thus, the preposed genitive case is restricted to actors but not restricted to transitive contexts or external arguments.

#### 4. PREPOSED PRONOUNS IN STRUCTURAL SUBJECT POSITIONS

A nominative pronoun can appear in embedded contexts such as a relative clause or a complement clause, as shown in (20) a and b.

(20) a. relative clause

Nangita ko sa batang akong gipalitag [gipalitan ug]  
 R.A-see TOP.1.SG O child=LK NOM.1.SG=LK R-buy-L=INDEF  
 kalamay.  
 rice pudding

'I am looking for the boy I bought some kalamay from.' (lit. '... the boy who was bought some kalamay from by me.') (Wolff 1966:306)

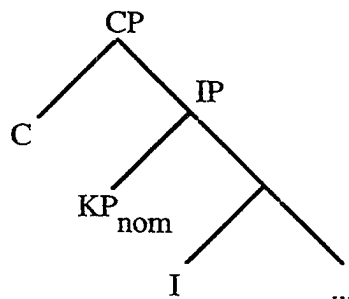
b. complement clause

Nagingon si Mariang iyang paliton ang tila.  
 R.A-said TOP MARIA=LK NOM.3.SG=LK buy-IR.P TOP cloth  
 'Maria said she/he will buy the cloth.'

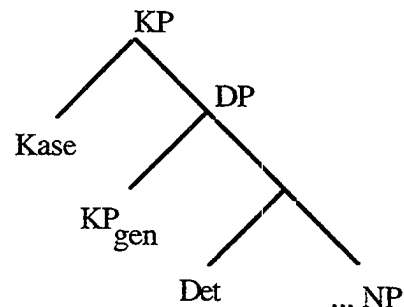
In (20a), the nominative pronoun *akong* occurs in the relative clause modifying *sa bata* 'the child'. In (20b), the nominative pronoun *iyang* occurs in the clausal complement of the verb *nagingon* 'said'. Both the relative clause in (20a) and the complement clause in (20b) are headed by the complementizer linker *nga* which occupies  $C^0$ . These examples suggest that a nominative pronoun occurs lower than  $C^0$ .

In particular, the nominative pronoun occupies Spec, IP, the canonical position for nominative case assignment. By analogy, preposed genitive pronouns occur after case markers, suggesting that they are in Spec, DP, where genitive case is assigned. A comparison of the tree structures in (21)a and b highlights the similarity of subject case positions in clauses and nominal phrases, as previously observed by Szabolsci 1983 and Abney 1987, among others.

(21) a.



b.



The outer projections, CP and KP, both serve to close off an extended projection and relate it to external elements. Note that the linker *nga* appears as either a complementizer C<sup>0</sup> or a case marker K<sup>0</sup>. The structural subject position of a clause or a nominal phrase is analyzed as the highest specifier position following these outer projections.

(22) illustrates the Spec, IP position of a nominative pronoun in a headed relative clause and a complement clause. (23) illustrates the Spec, DP position of a preposed genitive pronoun in a nominal phrase.

(22)

a. Headed relative clause (cf. (20a))

Nangita ko sa bata [CP [C ng ] [IP akong [I gipalita]g kalamay ]]

'I am looking for the boy I bought some kalamay from.' (lit. '... the boy who was bought some kalamay from by me.')

b. complement clause (cf. (20b))

Nagingon si Maria [CP [C ng ] [IP iyong [I paliton] ang tila]]

'Maria said she/he will buy the cloth.'

(23) nominal (cf. (19a))

[KP [K ang ] [DP imong [D ] pagpatay sa iya ]]

'Your killing of him/her'

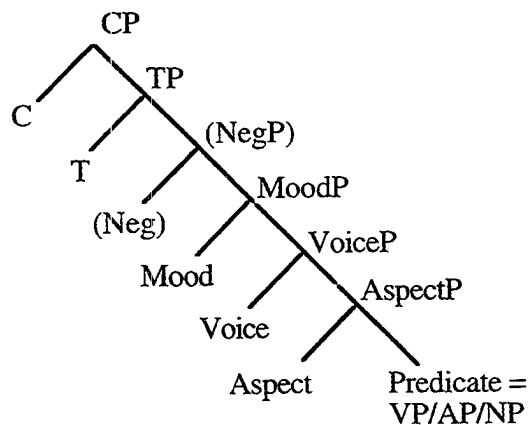
In (22) a and b, a nominative pronoun in Spec, IP occurs after a complementizer linker in C<sup>0</sup>. In (23), the preposed genitive pronoun *imong* 'your' is analyzed as occupying Spec, DP, following the case marker *ang* in K<sup>0</sup>. Thus, a nominative pronoun occurs in the structural subject position of a clause while a preposed genitive pronoun occurs in the structural subject position of a nominal phrase.

## 5. PREPOSED PRONOUNS AND HEAD MOVEMENT

### 5.1 INFL in Cebuano

Let us now examine INFL in Cebuano in more detail, adopting more articulated IP projections (cf. Pollock (1989)). The tree in (24) expands the IP drawn in (21a).

(24)



Aside from the standard projections, Tense Phrase (TP) and Negation Phrase (NegP), (24) uses three other IP projections, MoodP, VoiceP, and AspectP. Verbs encode mood but not tense features. One reason to posit MoodP as a separate projection from TP is that mood

and tense features can both be indicated in a clause, and when this happens, tense features appear on elements which precede a mood-inflected verb. Furthermore, prepositives such as modals, negation, deictic pronouns, etc. appear to select for certain mood features on the verb.

VoiceP represents the interface between the lexical semantic features of the predicate and the discourse relevant features of a particular clause. The role of VoiceP is to check voice agreement, defined as follows:

(25) Voice agreement

- a. The nominal features of VoiceP must agree with the thematic features of an A-bar bound argument in Spec, VoiceP.
- b. The verbal features of VoiceP must agree with the lexical semantic features of a verb.

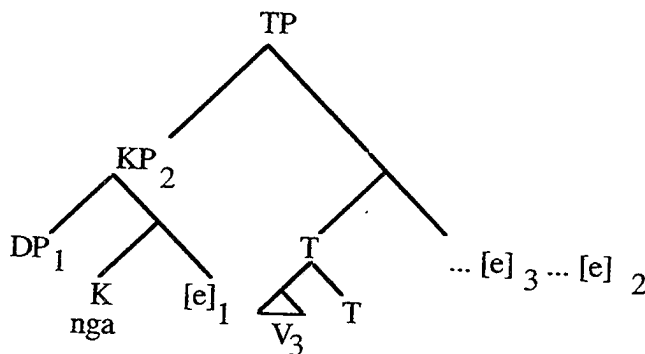
(25a) is usually satisfied when the thematic features of VoiceP agrees with the thematic features of a topic in Spec, VoiceP. The topic is A-bar bound to Spec, TopicP, from which it receives its presupposed interpretation. As discussed in Sityar (to appear), voice agreement is not restricted to topics but can also apply to *y* indefinites which are A-bar bound by an operator in FocusP. These thematic features of VoiceP are realized in verbal voice inflection.

Verbal predicates may also inflect for aspect, hence the projection, AspectP. A recent perfective sentence such as (16) is possibly a nominalization which takes AspectP but no other IP projections, thus accounting for the verb's lack of voice and mood features.

### 5.2 Verb movement and nominative case

A verbal predicate in Cebuano adjoins to the various inflectional heads in IP, usually landing in T<sup>0</sup>. Nominative case is licensed when a nominative pronoun is in Spec, TP and the verb is in T<sup>0</sup>.

(26) Licensing configuration for nominative case assignment



T<sup>0</sup> satisfies its verbal features by having the verb + I<sup>0</sup> complex adjoin to it, and Spec, TP satisfies its nominal features by having the nominative KP move there. The internal structure of a nominative KP is as given in (2) in which the bare pronominal stem (DP) has moved to Spec, KP.

The representation in (26) is supported by the fact that the nominative KP and the verb must be adjacent. This follows from (26) if there is no adjunction to T-bar. For example, neither negation nor modals can intervene between the nominative pronoun and the verb

(cf. the (c) examples in (7-10)). Other examples of this adjacency restriction is illustrated in (27) and (28).

- (27) a. Hinay nga iyang gihatag ang pan ngadto sa babaye.  
 slow LK NOM.3.SG.=LK R.T-give TOP bread to O woman  
 'She/he slowly gave the bread to the woman.'  
 b. Iyang gihatag nga hinay ang pan ngadto sa babaye.  
 c. \*Iyang hinay nga gihatag ang pan ngadto sa babaye.

(28) \*Akong gahapon gipalit ang tila.

(27)a and b respectively show that a manner adverb can precede a nominative pronoun or follow the verb. I analyze (27a) as being biclausal, with the sentence initial adverb acting as a higher predicate.<sup>9</sup> In (27b), the verb is higher than the VP-adjoined adverb because it has moved to T<sup>0</sup>. As expected from (26), (27c) shows that a manner adverb cannot occur between the verb and the nominative pronoun. (28) shows that a temporal adjunct cannot intervene between the nominative pronoun and the verb (cf. (6a)).

### 5.3 Head movement in nominal phrases

Unlike nominative pronouns, preposed genitive pronouns do allow elements to intervene between them and the head noun, as shown in (29).

- (29) a. intervening adjective  
 ang iyang dakong paghigugma  
 TOP GEN.3.SG=LK big=LK GER-love  
 'his/her great love'  
 b. plural marker:  
 Di na maihap ang akong mga apo.  
 NEG CL IR.T-count TOP GEN.1.SG=LK PL grandchild  
 'My grandchildren can't be counted.'

In (29a), the adjective *dako* 'big' and its linker occur between the genitive pronoun *iyang* 'his' and the head noun *paghigugma* 'love'. In (29b), the plural marker *mga* occurs between the genitive pronoun *akong* 'my' and the head noun *apo* 'grandchild'.

To explain this difference in word order, I assume that a verb can move higher in the clause than a noun does in the nominal phrase. The head noun does not move higher than NumP, headed by the plural marker *mga* in plural nouns, and it does not move past an intervening adjective. There may be D<sup>0</sup> to K<sup>0</sup> movement, however, since case markers and demonstratives encode both case and determiner features.

## 6. SECOND POSITION CLITICS

The strict adjacency between a nominative pronoun and verb appears to be violated by second position clitics (cf. (5a)).

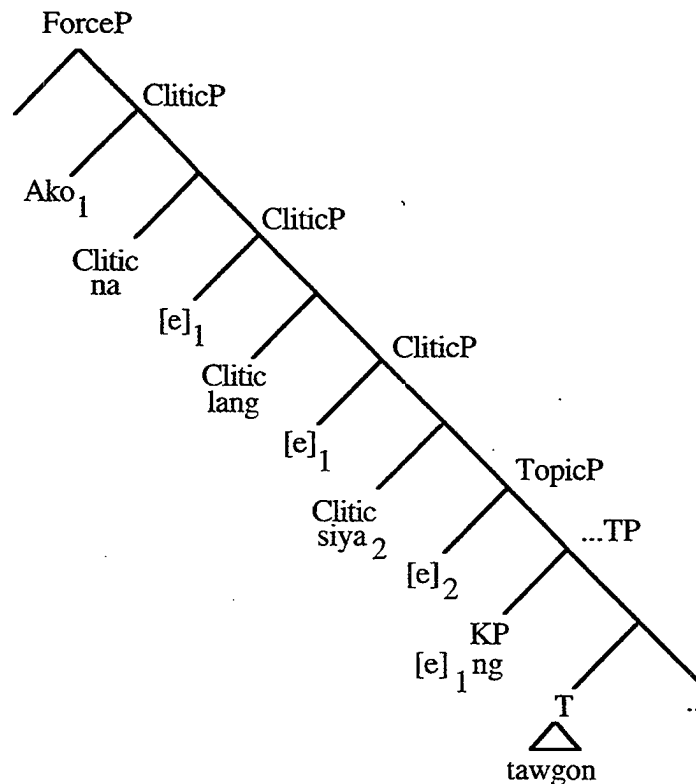
- (5) a. Ako na lang siyang tawgon.  
 NOM.1.SG CL CL TOP.3.SG=LK call-IR.P  
 'I'll just call him/her.' (Wolff 1966:251)

<sup>9</sup> This construction is related to others in which the manner adverb is inflected as a verb, but a full account of this construction is beyond the scope of this paper.

Note however, that clitics occur between the DP pronoun and the nominative linker in  $K^0$ , not between the nominative  $K^0$  and the verb. I interpret this ordering to mean that the DP pronoun moves out of KP to a higher position which c-commands the clitics, stranding its nominative  $K^0$  in Spec, TP.<sup>10</sup>

More specifically, I adopt the following account. Rizzi (1995) proposes that the CP domain can expand into several projections if the proper features are present. I assume that in Cebuano, the presence of clause-level clitics activates (at least one) clitic projection CliticP within CP. CliticP attracts clause-level clitics to Clitic<sup>0</sup> and the closest available phonological host to Spec, CliticP.<sup>11</sup> Thus, I represent the sentence in (5a) as (30):

(30)



ForceP, the outermost CP projection, specifies the clause type or force (cf. Rizzi (1995)). Elements in ForceP such as the interrogative complementizer do not count as possible hosts for clitics in Cebuano. Thus I represent the multiple CliticPs in (30) as appearing below ForceP. The order of these CliticPs is determined by morphological rules. In (30), the DP pronoun *ako* moves to each Spec, CliticP because it is the closest phonological host for the next higher clitic.

<sup>10</sup>It is not clear why the nominative linker must be stranded. But in general, the linker *nga* appears as the last member of a clitic cluster, regardless of its syntactic category.

<sup>11</sup>It does not seem to matter whether the clitic or host is a head or a phrase, as long as prosodic conditions are met.



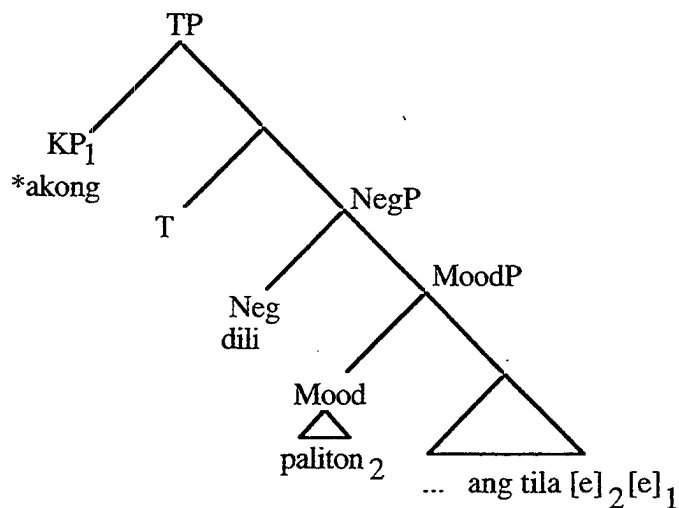
## 7. ACCOUNTING FOR RESTRICTIONS ON NOMINATIVE PRONOUNS

### 7.1 Failure of verb movement to T<sup>0</sup>

The failure of the verb + I<sup>0</sup> complex to move to T<sup>0</sup> can account for many of the contexts in which nominative case is not licensed in Cebuano. For example, recall that nominative case is restricted to verbal predicate sentences which are fully inflected for voice and mood. Nonverbal predicates apparently do not undergo head movement since they do not inflect for voice or mood. Meanwhile, the verb in a recent perfective sentence only moves as high as Aspect<sup>0</sup>. Thus, nominative case cannot be licensed in these contexts.

Preposed pronouns are also disallowed in negative clauses. Neg<sup>0</sup> is between T<sup>0</sup> and Mood<sup>0</sup> since negation places selectional restrictions on mood inflection. A negative particle in Neg<sup>0</sup> blocks verb movement to T<sup>0</sup>.

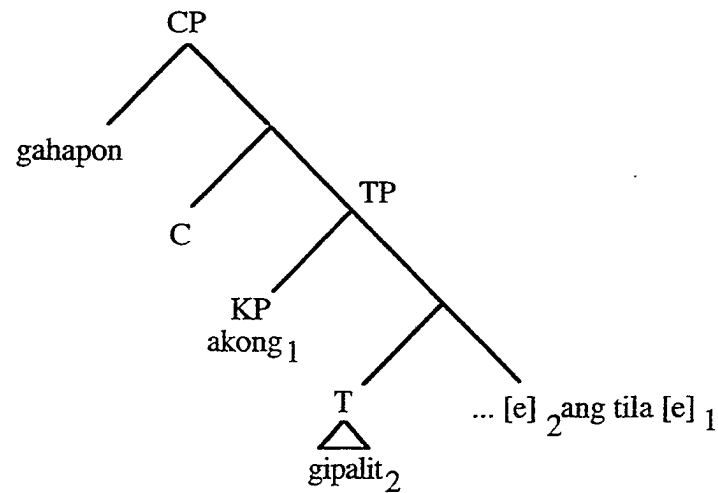
(31) (cf. (7c))



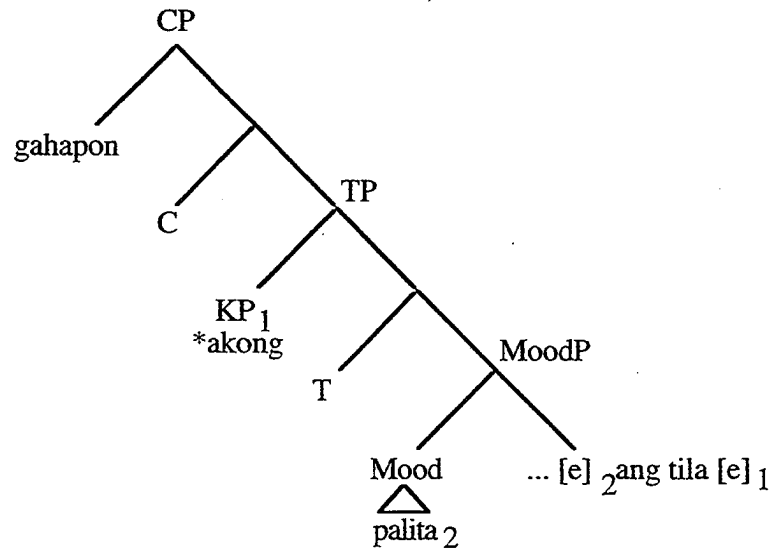
The verb *paliton* ‘buy’ is adjoined to Mood<sup>0</sup>, but it cannot move to T<sup>0</sup> to license the preposed pronoun *akong* ‘I’ because it is blocked by *dili* in Neg<sup>0</sup>.

Recall that preverbal temporal adjuncts or deictics seem to express tense features and can occur with either a finite or subjunctive verb. The tree in (32a) shows the temporal adjunct preceding a finite verb; the tree in (32b) shows the temporal adjunct preceding a subjunctive verb.

(32)a. Temporal phrase with a finite verb (cf. (6a))



b. Temporal phrase with a subjunctive verb (cf. (12b))



The basic story is this: in (32a), the finite verb *gipalit* has moved to T<sup>0</sup>, and thus it can license the nominative pronoun. But in (32b), T<sup>0</sup> has no independent tense features, so it cannot attract the verb. Because the verb remains in Mood<sup>0</sup>, it must inflect for subjunctive mood, and nominative case cannot be licensed.

But what is the role of the temporal adjunct/deictic pronoun in these contrasting configurations? A preverbal temporal adjunct or deictic pronoun occurs in Spec, CP and c-commands TP whereas a postverbal temporal adjunct or deictic pronoun occurs lower than TP.<sup>12</sup> Stowell (1992) claims that tense is a two place predicate that relates two temporal arguments, his *Zeit Phrases* (ZPs). The subject ZP is *reference time* while the object ZP is the *event time*. Tense specifies that reference time is before, after or at the same time as the

<sup>12</sup>In Rizzi's (1995) account of the expanded CP domain, Topic and Focus project separate CPs. It is unclear to me whether a sentence initial temporal adjunct in Cebuano is being topicalized, focused or either of these options, so the trees in (32) are shown with a simple CP.

event time. With Stowell's account of tense in mind, I analyze a temporal adjunct in postverbal position as modifying event time, i.e. identifying or restricting the time in which the event occurs. In postverbal position, a deictic pronoun places a similar role, but it can only modify event location, not event time. However, when a temporal adjunct or deictic pronoun occurs in Spec, CP, it behaves as modifier or predicate of tense itself. This is because a temporal adjunct like *gahapon* 'yesterday' (i.e. 'day before reference time') and a deictic pronoun like *didto* 'there, far from speaker and hearer, (and preverbally, PAST)' can also relate a reference time/location to an event time/location. The presence of a finite verb indicates that  $T^0$  is specified for tense features whereas the presence of a subjunctive verb indicates that  $T^0$  has no tense features. In the first case, the temporal adjunct or deictic pronoun in Spec, CP must express a temporal relation which is compatible with the temporal relation expressed in  $T^0$ . But when  $T^0$  occurs without tense features, the temporal adjunct/deictic pronoun in Spec, CP supplies the tense of the clause. This may be accomplished by  $T^0$  movement to  $C^0$  and specifier-head agreement in CP.

The noncooccurrence of preposed pronouns with modals is more complicated to explain. Ordinarily, a modal cannot cooccur with a preposed pronoun either preceding or following it, as shown in (9b-c), (10b-c), and (11b). However, in (11c), the modal *gusto* can license a preceding nominative pronoun.

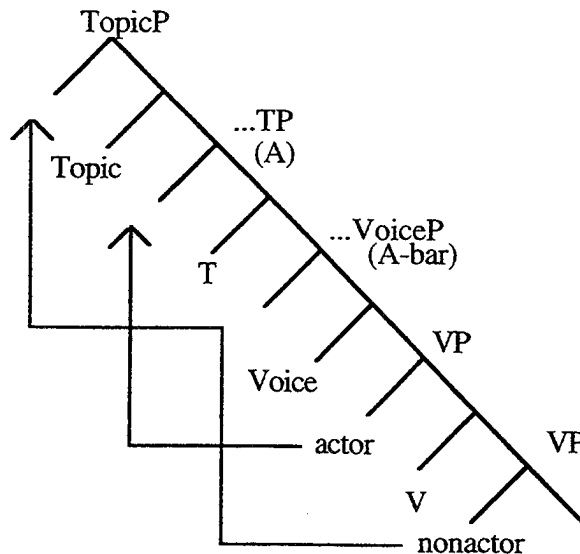
- (11) c. *Akong gustong paliton ang tila.*  
'I want to buy the cloth.'

In (11c), *gusto* seems to be a raising verb since it can license the nominative pronoun in its Spec, TP. But why then can't the other modal verbs behave like *gusto*? Even in languages such as German in which a modal can inflect for tense overtly, tense inflection on modals appears to mark politeness or evidentiality and not just simple tense relations. So I suppose that a Cebuano modal like *mahimo* or *kinahanglan* lacks the proper tense features to license nominative case in its own clause. Furthermore, modals generally appear with nonfinite clausal complements, suggesting that they block finite tense features in the lower clause. However, a fully developed account of the interaction of modality and tense is beyond the scope of this paper. Suffice it to say that finite tense features are either deficient or lacking in the complement clause of a modal, so that the lower verb cannot move to  $T^0$  to license nominative case.

## 7.2 Failure of nominal movement to Spec, TP

As noted before, nominative case is restricted to agents of passive verbs. This means that an actor is marked for nominative case and a nonactor is marked for topic case. The tree in (33) represents the path of movement for these two arguments.

(33)



The nominative actor moves to Spec, TP while the topic nonactor moves first to Spec, VoiceP and then to Spec, TopicP.

The representation in (33) assumes that actors c-command nonactors inside VP, an assumption which is clearly supported by binding facts among nontopic arguments in Cebuano. Chomsky's (1995) Minimal Link Condition states that only the shortest moves are possible, so that a target position must attract the closest element bearing the appropriate feature:

(34) *K attracts F* if *F* is the closest feature that can enter into a checking relation with a sublabel of *K*. (Chomsky 1995:297)

Since an actor c-commands a nonactor, the actor is always closer to Spec, TP than the nonactor is. Thus, only an actor can move to Spec, TP to check the nominal feature of TP. A nonactor is barred from making the longer move to Spec, TP, so it cannot bear nominative case in Cebuano.

I assume that Spec, VoiceP is an A-bar position in Cebuano. There is no operator relation in VoiceP itself, but only A-bar bound elements such as topics,  $y$  indefinites bound by an operator in FocusP, or WH variables can occupy Spec, VoiceP. Voice agreement is obligatory, as defined in the following Voice Criterion:

(35) Voice Criterion

- a. Spec, VoiceP must contain a KP bearing the thematic features specified by  $\text{Voice}^0$ .
- b. A KP bearing the thematic features specified by  $\text{Voice}^0$  must be in Spec, VoiceP.

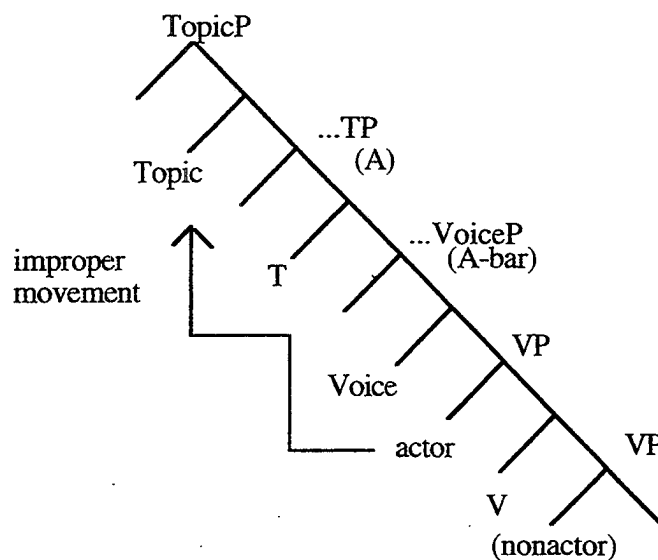
A nonactor which bears the thematic features specified by  $\text{Voice}^0$  must move to Spec, VoiceP.

Now consider the cases in which the actor of an active verb is not allowed to bear nominative case (cf. (18)b and d).

- (18)  
 b. \**ákong*            *nagpalit ug tila.*  
 NOM.1.SG=LK R.A-buy INDEF cloth  
 ('I bought some cloth.')
- d. \**Iyang*            *miabot.*  
 NOM.3.SG=LK R.A-arrive  
 (He/She arrived.)

As shown in (36), these sentences lead to an improper movement violation.

(36)



Assuming the Voice Criterion in (35), the actor of an active verb must move to Spec, VoiceP. However, subsequent movement to Spec, TP is improper because it would entail movement from an A-bar position in Spec, VoiceP to an A position in Spec, TP. Thus, the only way for an actor to bear nominative case is if another KP can satisfy the Voice Criterion, i.e. the nominative actor can only occur in a passive sentence.

The fact that only to agents of passive verbs can bear nominative case can thus be attributed to the interaction of three factors: the Minimal Link Condition, the Voice Criterion, and the a ban against improper movement.

## CONCLUSION

In this paper, I have argued that a nominative pronoun is licensed for nominative case by a finite verb in  $T^0$ . Contexts which fail to license a nominative pronoun could be seen as failure of the verb to move to  $T^0$  or the failure of a KP to move to Spec, TP. Aside from proposing a structural account of nominative case, this paper also outlines (some of) the functional projections in Cebuano clauses and nominal phrases and incorporates an account of Cebuano voice relations.

Abbreviations: TOP - topic, GEN - genitive, NOM - nominative, INDEF - indefinite, O-object/oblique, PL - plural, SG - singular, INC - inclusive, PRES - present, R - realis, IR

- irrealis, SU - subjunctive, IMP - imperative, A - active voice, P - patient voice, T - theme voice, L - locative voice, GER- gerund, LK - linker, CL - clitic.

References:

- Abney, S.P. 1987. *The English Noun Phrase in its Sentential Aspect*. Ph.D. dissertation, MIT.
- Bell, S. J. 1983. Advancements and ascensions in Cebuano. In *Studies in Relational Grammar I*. ed. by D. M. Perlmutter. Chicago: University of Chicago Press.
- Chomsky, N. 1995. Categories and Transformations. In *The Minimalist Program*. Cambridge, Massachusetts: The MIT Press.
- Kiss, K. E. 1995. "Introduction," in *Discourse Configurational Languages*. ed. by K. Kiss. Oxford: Oxford University Press.
- Kroeger, P. 1993. *Phrase Structure and Grammatical Relations in Tagalog*. Stanford, Dissertations in Linguistics, CSLI Publications.
- Kuroda, S.-Y. 1992. Judgement forms and sentence forms. In *Japanese Syntax and Semantics*. Dordrecht: Kluwer Academic Publishers.
- 1972. The categorial and thethetic judgement. *Foundations of Language* 9:153-185.
- Pollock, J.-Y. 1989. Verb movement, universal grammar, and the structure of IP. in *Linguistic Inquiry* 20, 3: 365-424.
- Rizzi, L. 1995. "The fine structure of the left periphery." ms.
- Sells, P. to appear. Raising and the order of clausal constituents in the Philippine languages. To appear in *Proceedings of AFLA 2*, Dordrecht: Kluwer Academic Publishers.
- Sityar, E. to appear. Topics and y indefinites in Cebuano. To appear in *Proceedings of AFLA 2*, Dordrecht: Kluwer Academic Publishers.
- Stowell, T. 1992. Aspects of Tense Theory, presented at GLOW 15.
- Szabolsci, A. 1983. "The possessor that ran away from home" in *The Linguistic Review* 3:89-102.
- Wolff, J. 1966. *Beginning Cebuano*. Part 1. Yale University Press, New Haven.
- 1967. *Beginning Cebuano*. Part 2. Yale University Press, New Haven.

# Certain restrictions on A-bar movement in Malay\*

Soh, Hooi Ling  
Massachusetts Institute of Technology  
hlsoh@mit.edu

## 0. Introduction

This paper deals with certain restrictions on A-bar movement in Malay. Specifically, it examines why the prefix meN- blocks A-bar extractions of certain NPs but not PPs. Two types of analyses have been proposed to account for this blocking effect. One type relies on the idea that A-bar extractions must pass through [Spec,IP] (Keenan and Comrie 1977, Chung 1975, Hung 1987, Guilfoyle et al. 1992, Nakamura 1995). The other type argues that the restriction is due to meN- absorbing the case assigned by the verb (Voskuil 1993). In this paper, I argue that contrary to previous proposals, extractions need not pass through [Spec,IP] in Malay. I also show that meN- does not absorb Case. Rather, I propose that the blocking effects are due to Relativized Minimality (Rizzi 1990) and meN- blocks A-bar extractions over it because it occupies an A-bar position.<sup>1</sup> The implication of this analysis is that a potential blocker is relativized not only in terms of the position (A or A-bar), but also the category (NP or PP).

## 1.0 Extraction Restrictions<sup>2</sup>

I examine two types of sentences in Malay with regards to the extraction facts. These sentence types are called stem sentences and meN- sentences.

### 1.1 Stem sentences

Examples of stem sentences are shown in (1) and (2). The verb appears in its bare form.<sup>3</sup> The subject of a stem sentence can be either definite or indefinite. The structure of (1) is represented in (3) assuming the VP-Internal Subject Hypothesis.

---

\* I am grateful to my syntax generals committee Noam Chomsky, Michael Kenstowicz and Alec Marantz for criticisms and discussions throughout the preparation of this paper. Thanks are also due to Shigeru Miyagawa, David Pesetsky, Norvin Richards and the participants of the syntax generals workshop for very helpful comments and discussions. I am also grateful to the participants of the AFLA III conference for very helpful feedback and comments. I would like to thank my informants Noor Hayati Yaacob, Nori Nadzri and Jeffry Mohd. Zain. All errors are my own.

<sup>1</sup> See Honegger (1996) for an alternative analysis.

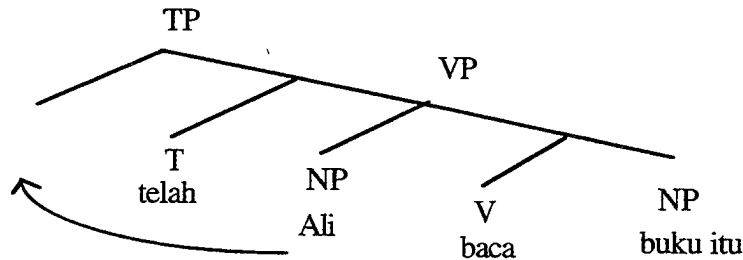
<sup>2</sup> Abbreviations:  
PERF perfective                      AT    agent topic            Q    question morpheme  
CLAS classifier                      TT    theme topic

<sup>3</sup> I assume following Chung (1978) that the stem sentence is not a variant of the object preposed construction with subject shift even though the verb in the object preposed construction also appears in its bare form. The object preposed construction is analyzed as a type of passive in Chung (1976, Guilfoyle et al. 1992) while the stem sentence is analyzed as a type of active in Chung (1978). An example of an object preposed construction is shown below.

(i) Buku itu Ali baca.  
book the Ali read  
'The book was read by Ali.'

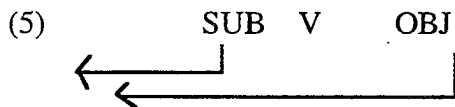
The verb in the object preposed construction appears in its bare form (with a null passive prefix) and the object moves to a sentence initial position.

- (1) Ali telah  **baca**  buku itu.  
Ali PERF read book the  
'Ali has read the book.'
- (2) Seorang anggota polis telah  **tangkap**  pencuri itu.  
a-CLAS member police PERF catch thief the  
'A policeman has caught the thief.'
- (3)



In stem sentences, both the subject and the object can be extracted freely, as shown in (4). (4a) shows that the subject has been extracted and (4b) shows that the object has been extracted. The possibility of subject and object extractions in stem sentences are presented in (5).

- (4) a. Siapa;kah yang  $t_i$  telah  **baca**  buku itu?  
who -Q that PERF read book the  
'Who has read the book?'
- b. Apakah yang Ali telah  **baca**   $t_i$ ?  
what-Q that Ali PERF read  
'What has Ali read?'



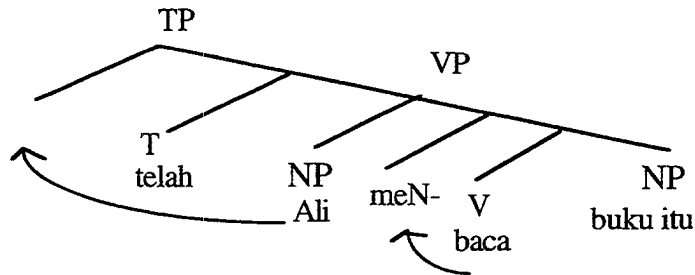
### 1.2 meN- sentences

meN- sentences differ from stem sentences in that the verb bears the prefix meN-. There does not seem to be any meaning difference between meN-sentences and stem sentences. Stem sentences appear to be preferred in colloquial speech while meN-sentences are preferred for formal communication. As shown in (6) and (7), both definite and indefinite subjects may appear in meN-sentences. The structure of (6) is shown in (8), where meN- occupies a position above the verb and below the internal subject position.

- (6) Ali telah  **membaca**  buku itu  
Ali PERF meN-read book the  
'Ali has read the book.'
- (7) Seorang anggota polis telah  **menangkap**  pencuri itu.  
a-CLAS member police PERF meN-catch thief the  
'A policeman has caught the thief.'



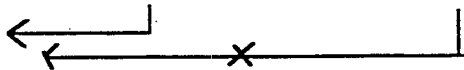
(8)



In simple meN- sentences, the subject can be extracted freely as shown in (9a), but the object cannot be extracted over meN- as shown in (9b). To extract the object, the verb must not bear the prefix meN- as shown in (9c). The restriction is presented in (10).<sup>4</sup>

- (9) a. Siapa<sub>i</sub>kah yang t<sub>i</sub> telah **membaca** buku itu?  
 who-Q that PERF meN-read book the  
 'Who has read the book?'
- b. \*Apa<sub>i</sub>kah yang Ali telah **membaca** t<sub>i</sub> ?  
 what-Q that Ali PERF meN-read  
 'What has Ali read?'
- c. Apa<sub>i</sub>kah yang Ali telah **baca** t<sub>i</sub> ?  
 what-Q that Ali PERF read  
 'What has Ali read?'

(10) SUB meN-V OBJ (Saddy 1991, Hung 1987)



Saddy (1991) observes that in a structure where meN-V is followed by a CP complement, extractions from within the CP complement are not possible when the matrix verb bears the prefix meN-. (11a) shows that the embedded subject cannot be extracted if the matrix verb bears the prefix meN-. (11b) shows that the embedded object cannot be extracted when the matrix verb bears the prefix meN-. To extract the embedded object, both the matrix verb and the embedded verb must not bear the prefix meN- as shown in (11c). To extract the embedded subject, only the matrix verb must not bear the prefix meN- as shown in (11d). The restrictions are presented in (12). These restrictions confirm the generalization that the extraction paths must not cross meN-.

- (11) a. \*Siapa<sub>i</sub>kah yang Ali **menganggap** t<sub>i</sub> menyukai Karim.  
 who-Q that Ali meN-believe meN-like Karim  
 'Who does Ali believe like Karim?'

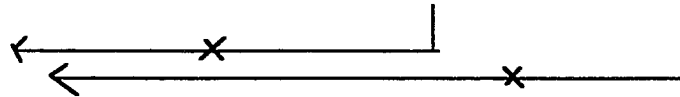
<sup>4</sup> It is worth noting that meN- does not block LF A-bar movement as shown by the presence of *wh* in-situ with meN-.

(i) Ali telah membaca apa?  
 Ali PERF meN-read what  
 'What has Ali read?'

It is possible that meN- which does not appear to have any semantic content is absent at LF.

- b. \*Siapa<sub>i</sub>kah yang Ali **menganggap** Minah suka t<sub>i</sub>?  
 who-Q that Ali meN-believe Minah like  
 'Who does Ali believe Minah like?'
- c. Siapa<sub>i</sub>kah yang Ali **anggap** Minah suka t<sub>i</sub>?  
 who-Q that Ali believe Minah like  
 'Who does Ali believe Minah like?'
- d. Siapa<sub>i</sub>kah yang Ali **anggap** t<sub>i</sub> menyukai Karim?  
 who-Q that Ali believe meN-like Karim  
 'Who does Ali believe likes Karim?'

(12) SUB meN-V [SUBJ (meN-) V OBJ] (Saddy 1991)



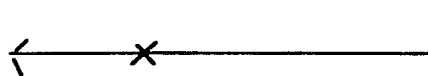
The same restriction holds true in meN- sentences with more embedding. NP A-bar extractions are blocked when any higher verb bears the prefix meN-. This is shown in (13a, b and c). The extraction is fine when none of the higher verbs bears meN- as shown in (13d). The restrictions are presented in (14).

- (13) a. \*Siapa<sub>i</sub> kah yang Ali **menganggap** John sangka t<sub>i</sub> mencintai Fred?  
 who-Q that Ali meN-believe John think love Fred  
 'Who does Ali believe John thinks loves Fred.'
- b. \*Siapa<sub>i</sub> kah yang Ali **anggap** John **menyangka** t<sub>i</sub> mencintai Karim?  
 who-Q that Ali believe John meN- think love Karim  
 'Who does Ali believe John thinks loves Karim.'
- c. \*Siapa<sub>i</sub> kah yang Ali **menganggap** John **menyangka** t<sub>i</sub> mencintai Karim?  
 who-Q that Ali meN-believe John meN- think love Karim  
 'Who does Ali believe John thinks loves Karim.'
- d. Siapa<sub>i</sub> kah yang Ali **anggap** John sangka t<sub>i</sub> mencintai Karim?  
 who-Q that Ali believe John think love Karim  
 'Who does Ali believe John thinks loves Karim.'

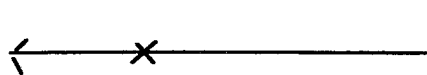
(14) a. SUB meN-V [SUBJ V [SUBJ (meN-)V OBJ] (Saddy 1991)



b. SUB V [SUBJ meN-V [SUBJ (meN-)V OBJ] (Saddy 1991)



c. SUB meN-V [SUBJ meN-V [SUBJ (meN-)V OBJ] (Saddy 1991)



## 2. Previous Analyses

In the following sections, I examine two previous analyses on the extraction restrictions in Malay. One is an economy account by Nakamura and the other is a variable binding account by Voskuil.<sup>5</sup>

### 2.1 Nakamura (1993, 1995): an Economy Account

Nakamura's economy analysis relies on the idea that A-bar extractions are only possible from the topic position ([Spec,IP]) (cf. Guilfoyle et al. 1992). Direct extractions of the object over the subject are excluded because one can minimize the length of chain links by first passivizing and then moving the object to the subject position before extracting the object. Within Nakamura's analysis, *meN-* is an agent topic marker; *di-* and a null prefix are theme topic markers.<sup>6</sup> The presence of *meN-* indicates that [Spec, IP] is occupied by the agent; the presence of *di-* and  $\emptyset$  indicates that [Spec, IP] is occupied by the theme (Guilfoyle et al. 1992). The theme cannot be A-bar extracted when the verb bears the agent topic marker (*meN-*) and the agent cannot be A-bar extracted when the verb bears the theme topic marker (*di-* or  $\emptyset$ ). To extract the theme, the verb must bear a theme topic morpheme (*di-* or  $\emptyset$ ). Within Nakamura's analysis, the reason why extraction paths must not cross *meN-* is because passivization is the most economical way to extract the agent in the most deeply embedded clause. As a result, the upper verbs bear the null theme topic marker. The structure of a sentence with extractions from within the embedded clause is shown in (15). In (15), [Spec, IP] is occupied by a *pro* which is coindexed with a CP argument. *pro* is replaced at LF by the movement of the entire CP to the [Spec, IP] position.

- (15) S-structure  
 Siapa<sub>i</sub>kah yang [IP *pro*<sub>i</sub> Ali  $\emptyset$ -anggap [CP<sub>i</sub> t<sub>j</sub> mencintai Karim]].  
 who-Q that Ali  $\emptyset$ -believe meN-love Karim  
 'Who does Ali believe loves Karim?'

There are two problems with this analysis. There is evidence from stem sentences that extractions need not pass through [Spec, IP] in Malay. The evidence comes from the placement of aspectual marker, weak cross over, and *that*-trace asymmetry. Also, Nakamura's explanation for why the upper verbs must not bear *meN-* when an embedded NP is extracted predicts certain extraction restrictions that are not attested.

#### 2.1.1 Object extractions need not pass through [Spec,IP]

##### 2.1.1.1 Positioning of Aspect markers

The first piece of evidence that object extractions need not pass through [Spec, IP] comes from the positioning of aspect markers in stem sentences. In stem sentences the agent NP comes before the aspectual marker as shown in (16), indicating that it is not in [Spec,VP], but in [Spec, IP].

- (16) Ali telah **baca** buku itu.  
 Ali PERF read book the  
 'Ali has read the book.'

<sup>5</sup> Nakamura (1995) mainly addresses the problem of extractions in Tagalog with some discussions on extractions in Malay with the assumption that Malay extractions behave the same as that of Tagalog.

<sup>6</sup> This null prefix is the one appearing in the object preposed construction (see footnote 3).



2.1.1.3 That-trace asymmetry

If all A-bar movements pass through [Spec,IP], one expects no that-trace asymmetry. This is because the movement of the object will share with the movement of the subject the trace in [Spec,IP] which is neither properly governed nor antecedent governed. However, a that-trace asymmetry is found (Md.Salleh 1987), indicating that A-bar movements do not need to pass through [Spec,IP].

- (22) a. \*Siapakah awak kata bahawa akan datang?  
 who-Q you say that FUT come  
 'Who did you say that will come?' (Md.Salleh 1987:85)
- b. ?Siapakah awak kata bahawa Ali jumpa?  
 who-Q you say that Ali meet  
 'Who did you say that Ali met?' (Md.Salleh 1987:85)

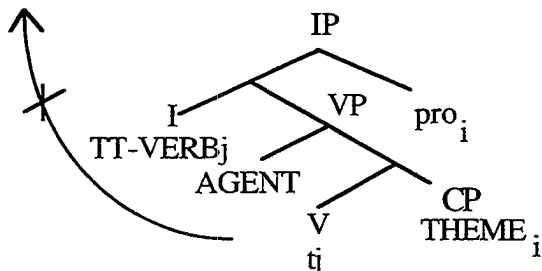
2.1.2 *Long distance extraction*

Nakamura's explanation for why the prefix meN- must be absent when an embedded NP is extracted predicts certain extraction restrictions that are not attested. Nakamura's analysis of long distance extractions involving meN- parallels his analysis of the contrast in (23) for Tagalog. In Tagalog, the agent of the matrix clause cannot be extracted when the matrix verb bears the TT morpheme as shown in (23b).

- (23) Tagalog
- a. Sino ang nagsabi na bumili ng kotse si Linda?  
 who Comp AT-said Comp AT-bought Acc-car Top-Linda  
 'Who said that Linda bought a car?'
- b. \*Sino ang sinabi na bumili ng kotse si Linda?  
 who Comp TT-said Comp AT-bought Acc-car Top-Linda  
 'Who said that Linda bought a car?' (Nakamura 1993, 1995)

Nakamura (1993, 1995) argues that this restriction is due to [Spec, IP] being occupied by a *pro* which is coindexed with the theme topic in (23b). This *pro* in [Spec, IP] blocks the extraction of the matrix agent. The structure of (23b) is represented as in (24).

(24) Tagalog (Nakamura 1993, 1995)



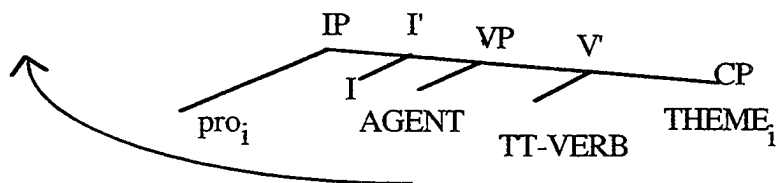
Given that *pro* blocks matrix agent extractions in Tagalog, we expect *pro* to also block the extraction of the matrix agent in Malay in *meN-* sentences involving a CP complement as in (25).

- (25)(=15) S-structure  
 Siapa;kah yang [IP  $pro_i$  Ali  $\emptyset$ -anggap [CP<sub>i</sub>  $t_j$  mencintai Karim]].  
 who-Q that Ali  $\emptyset$ -believe meN-love Karim  
 'Who does Ali believe loves Karim?'

However, contrary to expectation, *pro* does not block the extraction of the matrix agent in Malay as shown in (26) and presented in (27).

- (26) Siapa;kah yang [IP  $pro_i$   $t_j$   $\emptyset$ -anggap [CP<sub>i</sub> Minah mencintai Karim]]?  
 who-Q that  $\emptyset$ -believe Minah meN-love Karim  
 'Who believes that Minah loves Karim?'

(27) Malay



This suggests that the matrix verb does not bear a null theme topic marker and why the prefix *meN-* must be absent in the upper verb when an argument from a lower clause is extracted needs a different explanation.

## 2.2 Voskuil (1993): a Variable Binding Account

Voskuil proposes that *meN-* is assigned verbal Case if affixed to a transitive verb. The object must form a chain with *meN-* in order to receive a Case index.

- (28) a. Yanto membaca [NP buku itu]  
 Yanto meN-read book that  
 b. Yanto ... [meN<sub>i</sub> [V NP<sub>i</sub>]]

The chain <*meN*, NP> receives a theta role in the object position and is assigned Case in *meN-*. Given the assumption that syntactic variables must be case marked, the non-Case marked object cannot function as a syntactic variable and thus A-bar movement of the object is not possible.

One problem with this analysis is that heavy NP shift which involves A-bar movement is allowed as shown in (29).

- (29) a. Saya menyepak bola yang dia campak kepada saya dengan kuat.  
 I meN-kick ball that s/he throw to me with strength  
 'I kicked the ball that s/he threw to me very hard.'  
 b. Saya menyepak  $t_i$  dengan kuat [bola yang dia campak kepada saya]<sub>i</sub>.  
 I meN-kick with strength ball that he throw to me  
 'I kicked very hard the ball that s/he threw to me.' (Md.Salleh 1987:137)

Also, the analysis does not explain why extraction paths cannot cross meN- (cf. (9), (11), (13)).

### 3. meN- as an A-bar Blocker

I propose that the reason why meN- blocks A-bar extractions over it is because it occupies an A-bar position within the theory of Relativized Minimality (Rizzi 1990). I assume that a trace has to be both head governed and antecedent governed in Malay. The definitions of antecedent government and Relativized Minimality are presented in (30) and (31).

- Relativized Minimality (Rizzi 1990:6-7)

(30) Antecedent Government: X antecedent-governs Y iff

- (i) X and Y are coindexed
- (ii) X c-commands Y
- (iii) no barrier intervenes
- (iv) Relativized Minimality is respected

(31) Relativized Minimality: X  $\alpha$ -governs Y only if there is no Z such that

- (i) Z is a typical potential  $\alpha$ -governor for Y,
- (ii) Z c-commands Y and does not c-commands X.

The intuitive idea behind Relativized Minimality is that a particular kind of government is blocked by the element which typically has the potential for government of that kind. So, A-antecedent government is blocked by something in an A-position and A-bar antecedent government is blocked by something in an A-bar position. The reason why meN- blocks A-bar extractions over it is because it occupies an A-bar position.

However, there is an NP-PP asymmetry in terms of the blocking by meN-. meN- blocks NP extractions but not PP extractions. (32a) shows that the extraction of an object NP is blocked by meN-. (32b) and (32c) show that both argument and adjunct PP extractions are fine across meN-.

- (32) a. \*Apakah yang Ali telah membaca  $t_i$  ?  
 what-Q that Ali PERF meN-read  
 'What has John read?' (=9b)
- b. Kepada siapakah Minah memberi kucing kesayangannya?  
 to who-Q Minah meN-give cat beloved-her  
 'To whom did Minah give her beloved cat?'
- c. Bagaimanakah Ali menjawab soalan itu?  
 how-Q Ali meN-answer question the  
 'How did Ali answer the question?'





- b. Patung itu<sub>i</sub> dibeli Minah t<sub>i</sub>.  
doll the di-buy Minah  
'The doll is bought (by) Minah.'

As shown in (35) and (36), *di* blocks A-bar NP extractions but not PP extractions.

- (35) A-bar movement (NP)  
\*Siapa<sub>i</sub>kah yang patung itu dibeli t<sub>i</sub>?  
who-Q that doll the di-buy  
'Who was the doll bought (by)?'

- (36) A-bar movement (PP)  
Oleh siapa<sub>i</sub>kah patung itu dibeli t<sub>i</sub>?  
by whom-Q doll the di-buy  
'By whom was the doll bought?'

#### 4.1.2 The Intransitive Marker ber-

Another prefix which behaves like meN- is the intransitive marker ber-. Ber- attaches to intransitive verbs and some verbs which take a CP complement. When ber- attaches to verbs, it does not appear to add any meaning to the sentence. As shown in (37) and (38), ber- blocks A-bar NP extractions but not PP extractions.

- (37) A-bar movement (NP)
- Mereka berharap bahwa semua cadangan mereka itu akan diterima.  
they ber-hope that all suggestion their the will be-accepted  
'They hope that all their suggestions will be accepted.'
  - \*Apa<sub>i</sub> kah yang mereka berharap t<sub>i</sub> akan diterima?  
what-Q that they ber-hope will be-accepted  
'what do they hope will be accepted?'
  - Apa<sub>i</sub> kah yang mereka harap t<sub>i</sub> akan diterima?  
what-Q that they hope will be-accepted  
'what do they hope will be accepted?'
- (38) A-bar movement (PP)
- Ali berlari ke pasar.  
Ali ber-run to market  
'Ali ran to the market.'
  - Ke mana<sub>i</sub> kah Ali berlari t<sub>i</sub> ?  
to where-Q Ali ber-run  
'Where did Ali run to?'

#### 4.1.3 The Comparative Prefix se- (as..as)

The comparative prefix se- also behaves like meN- in blocking NP extractions and not PP extractions as shown in (39) and (40). In (39c), the extraction of NP across se- is prohibited and in (40) which involves a PP, the extraction across se- is fine.

- (39) A-bar movement (NP)
- a. Ali setinggi Minah.  
Ali se-tall Minah  
'Ali is as tall as Minah.'
- b. Siapa<sub>i</sub>kah yang t<sub>i</sub> setinggi Minah?  
who-Q that se-tall Minah  
'Who is as tall as Minah?'
- c. \*Siapa<sub>i</sub>kah yang Ali setinggi t<sub>i</sub>?  
who-Q that Ali se-tall  
'Who is Ali as tall as?'
- (40) A-bar movement (PP)
- Sejak bila<sub>i</sub>kah Ali setinggi Minah t<sub>i</sub>?  
since when-Q Ali se-tall Minah  
'Since when was Ali as tall as Minah?'

#### 4.2 *The Existence of Prefixes which block A-bar movement but not A-movement*

If the blocking is due to certain prefixes occupying an A-bar position, one expects these prefixes not to block A-movement when A-movement is licensed. There is some support from the passive morpheme di- and the intransitive marker ber-. Both prefixes block A-bar NP movement across them but not A-movement of NPs.

##### 4.2.1 The Passive Morpheme di-

The passive morpheme di- does not block A-movement of NPs but blocks A-bar movement of NPs as shown in (41) and (42).

- (41) (=34a) A-movement (NP)
- Patung itu<sub>i</sub> dibeli t<sub>i</sub> oleh Minah.  
doll the di-buy by Minah  
'The doll is bought by Minah.'
- (42) (=35) A-bar movement (NP)
- \*Siapa<sub>i</sub>kah yang patung itu dibeli t<sub>i</sub>?  
who-Q that doll the di-buy  
'Who was the doll bought (by)?'

##### 4.2.2 The Intransitive Marker ber-

The intransitive marker ber- also does not block A-movement of NPs but blocks A-bar movement of NPs as shown in (43) and (44).

- (43) A-movement (NP)
- Laporan itu<sub>i</sub> belum bertulis t<sub>i</sub> lagi.  
report the not (yet) ber-write yet  
'The report hasn't been written yet.'