Proceedings of AFLA 7

The Seventh Meeting of the Austronesian Formal Linguistics Association

Edited by
Marian Klamer

Vrije Universiteit Amsterdam
Department of Linguistics
2000
ACKNOWLEDGEMENTS

We gratefully acknowledge the funding received for the conference and the proceedings from:
The Dutch Research Foundation (NWO)
The Royal Dutch Academy of Arts and Sciences (KNAW)
The International Institute of Asian Studies, Irian Jaya Studies (ISIR), Leiden
The Centre of Non-Western Studies (CNWS) of Leiden University
The Holland Institute of Linguistics
The Faculty of Arts of the Vrije Universiteit Amsterdam
The following persons helped to organise the conference:
Anke Jongkind, Wilco van den Heuvel, Nanette Huijs, Rob Goedemans
The following persons helped to set the program:
Felix Ameka, Max Planck Institute Nijmegen/Leiden University, Geert Booij, Free University Amsterdam,
Lisa Cheng, Leiden University, Crit Cremers, Leiden University,
Mirjam Ernestus, Free University Amsterdam, Rob Goedemans, Leiden & Utrecht University,
Gertjan Postma, Leiden University, Johan Rooryck, Leiden University, Hein Steinhauer, Leiden & Nijmegen
University, Ruben Stoel, Leiden University, Rint Sybesma, Leiden University, Arie Verhagen, Leiden
University, Lourens de Vries, Free University Amsterdam, David Wilkins, Max Planck Institute Nijmegen

© The Authors

TO ORDER:
Send a cheque made out to
Vrije Universiteit, Faculteit der Letteren to:
AFLA VII Proceedings
c/o Vrije Universiteit
Secretariaat Faculteit der Letteren
Kamer 10A14
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands

Price per volume: US$ 25 (postage included)
We cannot accept creditcards
Preface

This volume consists of papers presented at the seventh meeting of AFLA (Austronesian Formal Linguistics Association), held at the Vrije Universiteit on May 11-13, 2000.

For the first time in the history of AFLA, this meeting was held outside the North-American continent, and contained contributions by speakers from eleven different countries: New Zealand, Australia, Indonesia, Brunei Darussalam, Taiwan, the USA including Hawaii, Canada, the UK, France, Germany, and The Netherlands.

Apart from the languages that are traditionally well-represented at Austronesian conferences, we were happy to see that the program also contained work on relatively small or lesser described languages, such as the minority languages of Taiwan, North-West Borneo, Eastern Indonesia, Papua and Oceania.

Special themes of this conference were Iconicity and Argument marking. The papers in this volume show that the program covered a broad range of subdisciplines -- from discourse grammar, phonology, morphology, syntax, to semantics -- and that the authors are working within various theoretical frameworks. But despite the obvious differences in expertise, interest and background, the atmosphere on the conference was typically AFLA: lively and constructive, with an average rate of attendance of about 80%. The papers in this volume deserve the same rate of attention.

This meeting has again furthered the unwritten mandate of AFLA to encourage the formal study of Austronesian languages, especially work by speaker linguists and junior scholars. Six scholars presented analyses of their native language, and more than half of the 45 participants subscribed as 'student'. This suggests that the future of Austronesian linguistics looks very bright indeed.

The eight edition of Afla will be held in the spring of 2001 at the Massachusetts Institute of Technology (MIT) in Boston, USA. The principal organiser will be Ileana Paul.

Marian Klamer, Vrije Universiteit Amsterdam

Proceedings of previous AFLA meetings:

A Selection of the papers of AFLA 2, in 1995 is published as:

The proceedings of AFLA 3 and AFLA 4 in 1996/1997 are published as:

The proceedings of AFLA 6 in 1999 are published as:
Table of Contents

Gabriele Heike Cablitz
Nominalisation of verbal clauses in Marquesan (Oceanic, French Polynesia) .................... 1

Adrian Clynes
Phonological structures and expressiveness:
The role of iconicity in 'the emergence of the marked' .................................................. 15

William D. Davies
Against long movement in Madurese ............................................................. 33

Alexandre François
Vowel shifting and cloning in Motlav: Historical explanation vs. formal description .... 49

Madelyn Kissock
Transitivity alternations in Rotuman .............................................................. 69

Thomas B. Klein and Meta Y. Harris
Fixed segmentism, markedness and faithfulness:
Nominalising reduplication in Chamorro ......................................................... 81

Anja Latrouite and Ralf Naumann
An interpretation of the voice affix /i-/ in Tagalog ......................................... 101

Diane Massam
Niuean nominalisation .................................................................................. 121

Ulrike Mosel and Jessika Reinig
Valence changing clitics and incorporated prepositions in Teop ............................. 133

Simon Musgrave
Emotion predicates and grammatical functions in Indonesian .......................... 141

Ileana Paul
Clefts vs. pseudo-clefts in Austronesian ........................................................ 155

Phil Quick
A non-linear analysis of vowel harmony and vowel harmony blocking in Pendau .... 173

Charles Randriamasimanana
Malagasy, binary branching and null subjects ........................................ 193

Der-Hwa V. Rau
Word order variation and topic continuity in Atayal ...................................... 211
Ger P. Reesink
Austronesian features in a linguistic area.................................................................231

Li-May Sung
Nominalization in Rukai and Amis...............................................................245

Adam Ussishkin
Fixed prosodic effects in Austronesian: An Optimality-Theoretic account........259

William A. Foley
Categorial Change in Oceanic Languages:
First Contact on the North New Guinea Coast....................................................271
Nominalisation of verbal clauses in Marquesan (Oceanic, French Polynesia)

Gabriele Heike Cablitz
Max Planck Institute for Psycholinguistics, Nijmegen
University of Kiel, Germany
gabcab@mpi.nl

1. Introduction

The present paper examines the nominalisation of verbal clauses (henceforth: NVC) in Northwest Marquesan 1 (henceforth: N-MQA, or just Marquesan), an Eastern Oceanic language of the Proto-Eastern Polynesian subgroup spoken in the North-Western part of the Marquesan archipelago in French Polynesia.

With respect to nominalisations, languages of the world differ as to whether nominalisations are more closely related to the syntax of noun phrases or to the syntax of verb phrases. This was suggested by Comrie & Thompson (1985). My paper also addresses the question whether nominalisations in N-MQA show more characteristics of noun phrase syntax or more characteristics of verb phrase syntax.

2. Why do I call nominalisations in N-MQA ‘Nominalisation of verbal clauses’?

The type of nominalisations I am interested in are event- and proposition-denoting nominalisations. In many languages of the world these kinds of nominalisations are not purely nominal, i.e. they do not have all the syntactic properties which non-derived noun phrases such as Engl. ‘the dog’, ‘the house’ etc. have.

In N-MQA nominalisations are not derived by morphological means such as Engl. ‘performance’ from ‘perform’. All verbal clauses in Marquesan can be nominalised without changing the form of the word functioning as the lexical head of the verbal clause.

All words which function as the lexical head of a verbal clause can also function as the lexical head of a nominalised verbal clause. Nominalisations do not derive nouns from verbs, but they transform verbal clauses into noun phrases.

At this point it is important to emphasise that nominalisations in Marquesan are syntactically derived whereas event-denoting nominalisations such as Engl. ‘performance’ are lexical nominalisations which are derived by specific morphological means.

Following Comrie & Thompson (1985) I will examine the following questions concerning NVCs in Marquesan:

1) What are the morpho-syntactic processes of deriving NVCs?
2) How many verbal and nominal properties does the lexical head of a NVC show?
3) How are arguments expressed in NVCs?
4) Do NVCs occur more with nominal or more with verbal modifiers?

---

1 The data presented here was collected on the islands of ‘Ua Pou and Nuku Hiva.
NVCs in Marquesan show characteristics of NPs as well as verbal clauses, i.e. they occupy an intermediate position between being purely nominal and purely verbal. However, the categorisation into verbs and nouns is problematic in Marquesan. Words cannot easily be classified as belonging to one particular word class such as nouns or verbs because the same form of a word can appear in several different syntactic environments such as verbal, nominal or as lexical modifiers. Therefore it is not possible to distinguish formally between the head of an NP, VP or NVC. The best way out of this problem is to define a word on the basis of its morphosyntactic properties within a particular syntactic environment. So their actual use in a particular syntactic environment classifies words as verbs, nominals or lexical modifiers.

NVCs in Marquesan are interesting to examine because they seem to be a salient feature of Marquesan grammar. About 23% of all sentential units I examined contained at least one nominalisation. Moreover, the fact that we have syntactic nominalisation in Marquesan is a reflection of the multifunctionality of words and the difficulty in categorising them into nouns and verbs. One of the aims of this paper is to demonstrate that there seems to be no clear boundaries of what nominal or verbal in Marquesan really is.

Before I turn to nominalisations in N-MQA in more detail, I would like to give some information on the basic features of N-MQA.

3. Features of N-MQA

N-MQA is an accusative case-marking language with case being marked by prepositions. Common noun phrases are marked by articles whereas verb phrases are marked for tense, aspect and mood by verbal particles (short: TAM-particles). The basic constituent order is V-S-O:

(1) 'u kanea Teiki t=a ia ha'e TAM build T. DO ART=POSS 3.sg house
Teiki has built his house.

3.1 Noun phrases: expression of arguments

The absence or presence of prepositions indicates the syntactic function of a noun phrase. Subject NPs are unmarked NPs, whereas direct object NPs are marked by the preposition i (see (1)). Oblique NPs expressing cause, instrument and recipient are also preceded by the preposition i:

(2) 'ua motu tu'u kahu 'i='ima i te ke'a (DD)
TAM cut POSS.Pr.1.sg.cothes RED-ahand/arm OBL ART stone
My sleeves were cut by/with the stone (lit.'... caused by the stone')

---

2 About 1500 sentential units of narratives were examined.
3.2 Verbal classes

Words which function as the lexical head of a verbal clause are classified on the basis of their valency. N-MQA has three major verbal classes, namely transitive, intransitive (action (e.g. he’e “go”), process (e.g. topa “fall”), state (e.g. kanahau “nice”) and neuter verbals.

**transitive verbal clause**

(3) e koti a’a Teiki i te vehie me te toki... (CH-B-005)

TAM cut Dem T. DO ART fire.wood with ART axe

Teiki is cutting the fire wood with an axe.

**intransitive verbal clause**

(4) e ‘ori te tau vehine ... (K1B-T5:19)

TAM dance ART PL woman

The women are dancing...

**Neuter verbals**

The class of neuter verbals is a small class of intransitive action verbals which take the undergoer as their subject and optionally an agent as an oblique argument (see also (2)):  

(5) 'a'o'e i hemo te kamo i= a= 'u. (DD)

be.not TAM catch/caught ART thief OBL=ART.pst=1.sg

I did not catch the thief/ The thief was not caught by me.

4. Nominalised Verbal Clauses in N-MQA

Keeping these features of Marquesan in mind I will now turn to the four leading questions asked in the beginning.

4.1 What are the (morpho-syntactic) processes of forming NVCs?

Any verbal clause can be nominalised without morphological changes of the lexical head by simply replacing the TAM-marker by the general article te:

**Verbal clause**

(6a) e popahi ia i t=a i a puke manu:... (Mak-017)

TAM command 3.sg DO ART=POSS 3.sg pile bird

He commanded his birds:...

**Nominalised verbal clause**

(6b) me te popahi o Pa'etini i t=o i a tau po'i... (Mak-026)

and ART command POSS P. DO ART=POSS 3.sg PL people

And Pa'etini's commanding of his people...

According to some researchers (Chung 1973; Clark 1981) Polynesian languages have a nominalising suffix which goes back to PPN *-Canga. In N-MQA this suffix -'ia often occurs in nominalisations, but its primary function is not to nominalise verbal clauses, but
to mark an aspectual distinction between perfective vs. imperfective aspect (see also Mosel (1992:277) for Samoan). I will discuss this below (see §4.2.1.1).

4.2 Verbal and nominal properties of NVCs

In order to decide whether NVCs in N-MQA are more closely related to the syntax of noun phrases or to the syntax of verbal clauses, it is crucial to determine the verbal and nominal properties of NVCs. This also includes the expression of arguments in NVCs. Typical nominal features in N-MQA are the marking of case, definiteness and occasionally number in NPs and their occurrence in a number of syntactic positions such as subject, object or oblique position. Further nominal characteristics of NPs are the modification by genitive attributes, by relative clauses and by a class of state verbs (e.g. kanahau “be nice”, rui “be big”, po’otu “be beautiful”). Typical verbal features are the marking of tense, aspect and mood, voice, negation and occasionally number agreement on the verb. Moreover, the expression of subjects and objects in NVCs can also tell us something about the more verbal or more nominal nature of NVCs. If the subject and object of the corresponding verbal clause are expressed in the same way as in NVCs we have evidence for the more verbal character of NVCs. However, if subject and object appear as nominal attributes we have evidence for the more nominal nature of NVCs. I will begin with the verbal properties.

4.2.1 Verbal categories

4.2.1.1 Tense, aspect and mood

NVCs in Marquesan lack tense and mood marking, but they have the linguistic means of marking perfective aspect by using the suffix -'ia (which I briefly mentioned above). According to Clark (1981) ‘the suffix is often determined by its larger syntactic context’ in which the nominalisation appears. It is true that the suffix occurs in certain syntactic constructions such as in temporal adverbial phrases expressing punctual aspect, or with numerals. It is probably no surprise that these NVCs occur with the suffix -'ia because they refer to bounded, individual and countable events. Therefore I believe that the suffix’ main function is to mark an aspectual distinction. Where the suffix -'ia occurs the situation is seen as a whole or the event is regarded as being completed as e.g. in (7) and (8):

(7) ‘u ko’ana ‘atou i te pokoki a te po’i o Taipi
    Perf find 3.pl DO ART capture POSSART people POSST.
    me te kai = i j a atu (Hak-16)
    and ART cat=Perf DIR.temp

The people of Taipi were able to capture them and later ate them up.

(8) mea 'oa te ha'apeipei = i j a a te po’i i te = i ja 'aa
    STV-P long ART CAUS=prepare=Perf POSS ART people LD ART=Dem day
    i te = i ja 'aa i te tau himene me te tau haka... (Kim-2)
    LD ART=Dem day DO ART PL song and ART PL dance

The people’s preparation of songs and dances each day was long...

NVCs which lack the suffix -'ia often have an habitual or imperfective reading as in (9):
Nominalisation of verbal clauses in Marquesan

(9) ‘ua kave mai na ‘enana ‘i te vaka,
Perf carry hither ART.PL man LD ART canoe
‘i ‘e‘i’a te ha‘a=moe ‘a i te nino .... (Hak-20)
LD ANA ART CAUS=lie EMPH DO ART body
The two men carried (him) to the canoe; there was the laying out of the body....

The following two contrasting examples with the “arrive”, illustrate that the suffix -‘ia triggers a perfective reading (see (10)), whereas the lack of the suffix -‘ia triggers an imperfective reading (see (11)):

(10) ‘i te tihe=‘ia ‘i’a ‘u pe‘au Pa‘etini... (Mak-014)
LD ART arrive=Perf there Perf say P.
When (he) arrived there, Pa‘etini said... (lit. ‘at the arrival there Pa‘etini said’)

(11) epo te tihe o te oumati, ‘omua Teatau‘apaku ‘i te tihe. (E-123)
later ART arrive POSS ART sun before T. LD ART arrive
The arriving of the sun is later, first Teatau‘apaku is arriving.

The contrast between the usage of the suffix -‘ia and the lack of the suffix -‘ia in NVCs is probably best demonstrated by an example in the same linguistic context. The following two examples are taken from a legend in which a guardian of the chief Uhutete is listing to the talk of an imprisoned boy:

(12) he mea hemopu=‘ia ia te ha‘a=hiiti a i te ikoa o
ART thing tear.apart=PASS 3.sg ART CAUS=go.up EMPH DO ART name POSS
Tahiatemata i mate te tuhine o te haka‘iki o Taipi ‘o
T. TAM dead ART sister POSS ART chief POSS T. PRES
Uhutete. (Hak-18)
U.
The thing that tore him apart was the uttering of Tahiatemata’s name who is
dead, the sister of the Taipi chief, who is called Uhutete.

In (12) the situation is viewed from within and we do not really know how often the boy utters the name of his mother, probably repeatedly. However, the guardian goes back to his chief and tells him about the boy’s uttering the name of the chief’s sister. Now, the situation is seen as a whole and thus ha‘ahiti “utter”, the lexical head of the NVC, is suffix by -‘ia:

(13) “‘ua ‘oko au i te maha‘i i te ha‘ahiti=‘ia i te ikoa
Perf hear 1.sg DO ART boy OBL ART CAUS=Perf DO ART name
ma te pe‘au o t=o ia kui ‘o Tahiatemata”. (Hak-18)
with ART say POSS ART=POSS 3.sg mother PRES T.
“I heard the boy who uttered a name by saying his mother’s name which is Tahiatemata.

The occurrence and non-occurrence of the suffix -‘ia in NVCs of N-MQA is a retention of the verbal category of perfective aspect and not simply a nominalising suffix as often claimed in the literature on Polynesian languages.

Other verbal categories are voice, negation and number agreement.

3 The report of the guardian is in direct speech.
4.2.1.2 Voice

Transitive verbs are also passivized by the suffix -'ia. The optional expression of the agent is introduced by the preposition e in passivized verbal clauses as well as in NVCs:

**Passivized verbal clause**

(14a) 'u humu='ia te puaka e Teiki (E)

Perf attach=PASSART pig AGS T.

The pig was attached by Teiki.

**Passivized NVC**

(14b) te humu'ia o te puaka e Teiki (E)

ART attach=PASSPOSSART pig AGS T.

the attachment of the pig by Teiki

The expression of the agent by an e-phrase is in fact evidence for the retention of verbal syntax.

4.2.1.3 Negation

In Marquesan verbal negation is not expressed by particles but by verbs which function as the matrix verb of a clause:

(15)'A'o 'e 'aua i kite ena me te 'enana ma 'oto. (Lav-U:049)

be not 2.dl TAM know exist with ART man Prep inside.

They do not know that there are men inside.

In nominal constituents, the nominal can be negated by the postpositioned particle ko 'e “without, cease”:

(16) 'io te henua 'enana ko 'e

Prep.loc ART country man without

in the desert (lit. ‘in the land without men’)

In NVCs the head is negated in the same way as in nominal constituents:

(17)'ua he'e Hi'i moana 'i Nuku Hiva me te vivini ko 'e o ia,

Perf go H. LD N. H. and ART understand Neg POSS 3.sg

'o te 'ite=i a paona te=nei o ia i t=a ia 'ona vehine.

PRES ART see=Perf final ART=Dem POSS 3.sg DO ART=POSS 3.sg dear wife

Hi'i moana went to Nuku Hiva without understanding that this is the last time he has seen his wife. (Hak-22)

The use of nominal negation in NVCs of Marquesan shows, in fact, the more nominal character of NVCs. However, the reason for nominal negation in NVCs might be due to the fact that verbs of negation cannot be nominalised.
4.2.1.4 Number agreement

Verbs occasionally, but not obligatorily show number agreement on the verb. If the subject has more than one referent, the verb is partially reduplicated as in (18). This is also sometimes retained in nominalisations as in (19):

(18) 'u he-he'e hua mou pakahio ma he one... (U-019)
  Perf RED-go ART PL old.woman Prep ART sand.beach
  The old women went along the beach.

(19) Ena 'io he mata'a te no-noho o te na mou maha'i. (T/H-210)
  exist Prep ART land's.end ART RED-stay POSS ART=Dem PL boy
  The boys are staying at land's end. (lit. 'there is at land's end the staying of the two boys')

4.2.2 Nominal categories

4.2.2.1 Syntactic positions in a clause

A good argument for the noun phrase status of NVCs is that they occur in a number of syntactic position in which non-derived NPs can occur. They occur in subject, object and oblique positions of a clause (see examples (20-23)):

Subject in verbal clauses
(20) 'u hakaea te piki a hua maha'i i te vehine. (T/H:349)
  Perf stop ART climb POSS same boy DO ART woman
  It stopped the making love of the boy to the woman.

Subject in nominal clauses
(21) 'ua rere hua vehine; 'io he tai te ka'o'ia atu.
  Perf flee ART woman Prep.loc ART sea ART disappear=Perf thither
  That woman flew; the disappearing was on sea.

Object
(22) 'U pe'au haka'ua hua vehine Perf say again ART.ana woman
  "Po'iti, 'ua have koe i te kutu i te upoko?" (T/H:331)
  Perf know.2.sg DO ART delouse DO ART head
  That woman said again: "Po'iti, do you know how to delouse the head?"
  (lit. "... do you know the delousing of the head")

Oblique
(23) Atahi 'a hemo ai te tuehine i te mate koea
  then TAM catch ANA ART sister DO ART illness crazy
  i te ue i t=0 ia tukane i mate nei 'o Teatau'apaku
  OBL ART cry OBL ART=POSS 3.sg brother TAM dead just Pres T.
  Then the sister became mad caused by the crying over her brother's recent death who was called Teatau'apaku. (E-122)
The examples above illustrate that NVCs behave syntactically like NPs because they are marked by case and occur in syntactic positions in which other non-derived NPs occur. Other nominal features are the marking of number and definiteness.

4.2.2.2 Number

N-MQA has a number of plural markers which precede the lexical head of an NP. NVCs cannot be marked for plurality by these plural markers. Compare (24a) and (24b):

(24a) me te popahi o Pa'etini i t=o ia tau po'i...
    and ART command POSS P. DO ART=POSS 3.sg PL people
    And Pa'etini's commanding of his people... (Mak-026)

(24b) *me te tau popahi o Pa'etini...
    and ART PL command POSS P.

However, in Marquesan NVCs can occur with numerals when the events denoted by the nominalisation signal several individual and countable events:

(25) e te' u kanea=ia i te vaka....
    Num three construct=Perf DO ART canoe
    Three constructions of a canoe (=a canoe (unspec.) was constructed three times)

When NVCs occur with numerals, the numeral replaces the article. Moreover, the head of the NVC is suffixed by '-ia because it refers to individual bounded events. As for definiteness of NVCs, I could only observe that the general article te was used in the overwhelming majority of occurrences. The article te neither marks an NP for definiteness nor for specificity. The article te is a kind of default-article which is used when all other articles are less appropriate. Therefore we can say that definiteness is weakly specified in NVCs and that the main function of the article te is to syntactically derive NVCs from verbal clauses.

4.3 The expression of arguments in NVCs

The most interesting evidence for the more verbal or more nominal character of NVCs is the examination of how arguments are expressed. According to Comrie & Thompson (1985) the expression of oblique objects and adjuncts is less interesting because there seems to be little variation across languages. Oblique objects and adjuncts do not normally undergo a transformation, i.e. they occur in the same form as in the verbal clause. This is also true for Marquesan. In Marquesan the expression of subjects and objects show nominal as well as verbal features.

4.3.1 Expression of subjects in NVCs

Subjects are expressed as genitive attributes and therefore assimilate to NP syntax. The genitive attribute mostly occurs in postnuclear position, but when subjects are pronominalised, they can occur in pre- and postnuclear position:
Nominalisation of verbal clauses in Marquesan

Possessive attribute in prenuclear position

(26) 'i t=q u ia tihe=i i Hakamou'ui.... (Mak-14)
   LD ART=POSS 3.sg arrive=Perf LD H
At his arrival/arriving in Hakamou'ui....

The genitive attribute need not follow the head of NVC directly. In (27) the direct object is closer to the nominalised lexical head than the genitive attribute:

(27) atahi 'a oko nui ai te 'i'i
   when TAM strengthen very ANA ART strength
   te ha'aha'a atu ia ia o te Hiva 'Oa.... (H-012)
   ART be.angry DIR DO 3.SG POSS ART H. O.
When they had strengthened the forces, the enraging of the Hiva 'Oa against him...

Polynesian languages often make a distinction between inalienable and alienable possessive relationships. Alienable relations are expressed by a, inalienable by g.

The o/a-distinction is partly maintained in NVCs. Subjects of transitive verbs often have an agentive role in the action expressed by the verb and therefore are often marked by a in the NVC:

(28a) ... e he'e au 'u tih mai nei 'a te vava'o a te hei
   TAM go 1.sg Perf arrive hither now EMPH ART call POSS.al ART garland
I am going because the garland's calling has reached me now.

However, not all subjects of transitive clauses are explicitly agentive and thus are marked by o:

(28b) 'ua 'oko='ia te vava'o o te tau pahu me te po'i haka. (Hak-30)
   Perf hear=PASSART call POSS ART PL drum and ART people dance
The calling of the drums and the dancers was heard.

Subjects of intransitive verbal clauses are expressed by an o-phrase, although subjects of corresponding intransitive verbal clauses can sometimes express an agentive role (like he'e "go", hiamoe "sleep"):

(29) 'ua ko'e te peke o hua vehine.... (U-126)
   Perf cease ART be.angry POSSART woman
The being angry of the woman was over.

Q-marking of subjects in NVCs is also consistent with experience verbals, neuter verbals and state verbals.

We can therefore conclude that the distinction of inalienable and alienable possessive marking in subjects of NVCs is dependent on semantic as well as syntactic factors. For

---

4 The missing of the direct object in the NVC might question whether vava'o "call" is really transitive. However, in the same legend vava'o "call" is clearly used transitively in verbal clauses:

\(\text{TAM call now ART garland DO 1.sg} \)
\(\text{The garland is calling me now.} \)
nominalised transitive verbals the choice is semantically determined, whereas for nominalised intransitive verbals the choice is syntactically motivated. All subjects in NVCs are expressed as a genitive attribute and therefore clearly show that NVCs have nominal features.

4.3.2 Expressions of objects in NVCs

As for the expression of objects in NVCs, objects are expressed in the same way as in the corresponding verbal clause and therefore have retained verbal syntax. The direct object of transitive clauses always remains in the accusative case in the corresponding NVC regardless of whether the subject is expressed or not:

(30) 'atahi 'a kite atu Anihoka ia He'ato
when TAM see thither A DO H
me te o ia mata'eina'a me te mave atu ia ia:... (H-037)
and ART=POSS 3.sg village.people with ART welcome thither DO 3.sg
When Anihoka saw He'ato and his men, (he) welcomed him:...

With respect to the expression of subjects and objects in NVCs of Marquesan we can so far conclude that NVCs shows nominal and verbal features alike. Although we have a derivation of a verbal clause into a noun phrase on the syntactic level, the assimilation to noun phrase syntax is only partial. The partial assimilation to noun phrase syntax in NVCs of Marquesan might be due to a constraint in the noun phrase syntax of Polynesian languages. According to Clark (1981:71) only one postpositioned possessive phrase can be dependent on the lexical head of a NVC. Two postpositioned possessive phrases cannot be dependent on the same lexical head:

(31) te humu'ia o te puaka a Teiki
ART attach=Perf/PASS POSS ART pig POSS T.
1) *Teiki attached the pig
2) The attachment of Teiki’s pig

Teiki is not the initiator of the action, but only the possessor of the pig over which he has control (thus the a-marking). Some ergative Polynesian languages like e.g. Samoan allow double possessive marking of agent and undergoer when one of the “possessors” is preposed in a pronominalised possessive phrase:

Samoan (Mosel 1992:275)
(32) ... l=о-na sasa o maile
ART=POSS=3.sg hit POSS dog
His hitting of the dogs

In Marquesan, double possessive marking of subject and object does not occur in NVCs which have derived from active verbal clauses. Thus constructions like ‘his hitting of the dogs’ or ‘his cooking of the fish’ are not possible to construct, unless the corresponding verbal clause was passivized as in (33):
Nominalisation of verbal clauses in Marquesan

(33) \textit{t=}\textit{o} ia humu='ia o \textit{te puaka} \\
\textit{ART=POSS 3.sg attach=PASS POSS ART pig} \\
His attachment of the pig

In passivized NVCs the undergoer is always expressed as a genitive attribute, simply because subjects are always expressed as genitive attributes in NVCs of Marquesan. However, in nominalisations of active verbal clauses we do not have the co-occurrence of subjects and objects as two genitive attributes. It is only possible to construct ‘his cooking the fish’:

(34) \textit{t=}\textit{a} ia nunu'ia i \textit{te ika} \\
\textit{ART=POSS 3.sg cook=Perf DO ART fish} \\
His having cooked the fish

Sentences like (32) and (33) are probably explained by a second constraint in nominalisations of Polynesian languages which is mentioned by Clark (1981:71): unmarked NPs cannot occur in nominalisations. Thus, in Samoan the unmarked undergoer NP has to appear as a genitive attribute. In Marquesan, on the other hand, where the undergoer NP in active verbal clauses is marked in the accusative, the language need not adopt to nominal syntax. However, if the verbal clause is passivized the syntax of nominalisations requires the unmarked undergoer NP to appear as a genitive attribute. The reasons why Marquesan only allows constructions like (34) in NVCs which derive from active verbal clauses might be due to the retention of voice distinction in NVCs.

4.4 Do NVCs occur more with nominal or more with verbal modifiers?

Non-derived nominals in Marquesan can be typically modified by possessive phrases, relative clauses and by a class of state verbals which comprise lexemes such as \textit{kanahau} “be nice”, \textit{rui} “be big”. However, the head of a NVC can neither be modified by relative clauses nor by state verbals:

(35) \textit{*te nunu yeve} i \textit{te ika} \\
\textit{ART cook fast DO ART fish} \\
? The fast cooking of the fish

Lexical heads of verbal clauses and NVCs are typically modified by verbal modifiers such as directionals (\textit{mai “hither”} and \textit{atu “thither”, iho “downwards, later”}) and adverbial modifiers. Non-derived nominals (e.g. \textit{te ha’e “the house”}) do not occur with directionals and other verbal modifiers. The ability of NVCs to occur with verbal modifiers is again evidence that many verbal properties have been retained in the lexical head of NVCs. Here are some examples:

\textbf{Directionals}

(36) \textit{‘u pei-pei He’ato me t=}\textit{o} ia mata’eina’a toa \\
\textit{Perf RED-ready H. with ART=POSS 3.sg village warrior} \\
\textit{no te hiti atu ‘i te toua umu huke ‘i Hiva ‘Oa} (H-036) \\
\textit{for ART go.up DIR LD ART war oven revenge LD H. O.} \\
He’ato and his village warriors were ready to go up to make a war of revenge at Hiva ‘Oa.
Adverbial modifiers

(38) 'u tu-tukina te mena va me te tahe anamai o na vaimata. (Hak-?)
Perf RED-whiz ART breath with ART flow suddenly POSS ART.PL tear
The breath was whizzing while the tears were suddenly running. (lit. ‘with the running suddenly of the tears’)

Conclusion

The lexical head of NVCs is neither purely verbal nor purely nominal because its syntactic properties, as we have seen, are verbal as well as nominal.

<table>
<thead>
<tr>
<th></th>
<th>Verbal</th>
<th>Nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>case</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>definiteness</td>
<td>+ (only te)</td>
<td></td>
</tr>
<tr>
<td>plural markers</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>numerals</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>genitive attributes</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>relative clauses</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>state verbals</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>negation</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Tense/Mood</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Aspect</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>number agreement</td>
<td>(+)</td>
<td></td>
</tr>
<tr>
<td>voice</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>verbal modifiers</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>(directionals, adverbials)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Verbal and nominal properties of NVCs

From a syntactic point of view, nominalisation of verbal clauses in N-MQA is a process by which verbal clauses become noun phrases. Although there is a shift in structural level from clause to phrase which also changes the valency of the lexical head of nominalised verbal clauses (now being the head of an NP it cannot have a subject, but only a possessive attribute), the assimilation to noun phrase syntax is only partial (the expression of subjects as genitive attributes, the loss of tense and mood marking, and the usage of nominal negation). The internal structure of nominalised verbal clauses clearly shows that many verbal properties have been retained (such as aspectual marking, verbal modification and the case-marking object NPs). Therefore we can classify the lexical head of NVCs as a verbal-nominal hybrid which is generated on a syntactic level of nominal derivation.
Nominalisation of verbal clauses in Marquesan

<table>
<thead>
<tr>
<th>Verbal clause</th>
<th>→</th>
<th>Nominalised verbal clause</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. TAM ( V_{tr} \ S - O ) (i-NP)</td>
<td>→</td>
<td>ART ( V_{tr} \ a/o-POSS ) NP (S) - i-NP (O)</td>
</tr>
<tr>
<td>2. TAM ( V_{itr} \ S )</td>
<td>→</td>
<td>ART ( V_{itr} \ a-POSS ) NP (S)</td>
</tr>
<tr>
<td>3. TAM ( V_{neut} \ S - OBL ) (i-NP)</td>
<td>→</td>
<td>ART ( V_{neut} a-POSS ) NP (S) - i-NP (OBL)</td>
</tr>
<tr>
<td>4. TAM ( V_{pass} \ S - A ) (a-NP)</td>
<td>→</td>
<td>ART ( V_{pass} a-POSS ) NP (S) -a-NP (A)</td>
</tr>
<tr>
<td></td>
<td>→</td>
<td>ART=a/o PrePoss (A) ( V_{pass} a-POSS ) NP (S)</td>
</tr>
</tbody>
</table>

Table 2: Syntactic transformation and case-marking of NVCs

Literature:

Cablitz, G. (in prep.), *Aspects of spatial reference in adult and child Marquesan*
Dordillon, R.I. (1931), *Grammaire et Dictionnaire de la Langue des Iles Marquises*, Paris
Maunsell, R. (1842), *Grammar of the New Zealand Language*, (W.C. Wilson), Auckland
Mosel, U. & E. Hovdaugen (1992), *Samoan Reference Grammar*
Zeven, F. (1987), *Introduction à la langue des îles Marquises - le parler de Nuku Hiva, Papeete (Haere Po no Tahiti)*
Phonological structures and expressiveness: the role of iconicity in ‘the emergence of the marked’

Adrian Clynes, Universiti Brunei Darussalam

'Tis not enough no harshness gives offence, the sound must seem an Echo to the sense'
Alexander Pope 1711

It is a commonplace crosslinguistically that exceptions to phonological regularities are often manifested in vocabulary broadly classifiable as ‘expressive’. Fudge’s observation echoes views dating back to at least the Prague School linguists (Uhlenbeck 1950, Trubetzkoy 1939):

Expressive words [...] have a tendency in a wide range of languages to be associated with peculiarities of phonological structure [...] including types of sounds, sound sequences and syllable-structures which can be regarded as peripheral in the language concerned (Fudge 1970:161).

In English, for example, long vowels and diphthongs never occur before [n] – except in onomatopoeic words like oink and boing. Similarly, clicks and nasal vowels do not occur in the phonology of English – except in pragmatically-charged interjectory particles, such as tut-tut! and huh? ([hâ]) respectively.

It is generally assumed that departures from phonological regularities such as these, though recurrent crosslinguistically, are insignificant in number, and so can be safely ignored in a theory of language. Fudge however provided strong statistical arguments that this is not the case, and that for English at least the correlation pervades the lexicon. In Clynes 1995 I showed that this was true too in Balinese. The correlation is in fact predicted to occur by the hypothesis of ‘Foregrounding’, that marked structure at all linguistic levels regularly regularly associates with expressive semantics (Mukarovsky 1964, Leech 1966, Fowler 1984, cf Appendix, below). Still, just why Foregrounding occurs, and how it relates to a theory of language, remains to be clarified.

Writing on related expressive morphology effects (such as expletive infixation) Zwicky & Pullum (1987:338) assert that these are ‘not within the province of the theory of grammar as ordinarily understood, though [...] certainly within the broader sphere of human linguistic abilities.’ I would argue that such phenomena in fact provide important information for a theory of grammar. Nonetheless, Zwicky & Pullum’s intuition is broadly correct: a nonlinguistic principle or principles is no doubt involved in interactions between marked structure and semantics, though clearly it interacts with linguistic principles. I argue in §2 for a physiologically-based iconic motivation for Foregrounding, which can be considered to manifest the ‘emergence of the marked’ (cf McCarthy & Prince 1994), (see §4).1

---

1 The term ‘marked’ in Praguean usage has meanings such as: ‘what departs from the neutral’ (Kean 1992:390), even ‘unusual’ or ‘irregular’ (Waugh & Lafford 1994). In Optimality Theory the term has a different, if related, understanding: any structure violating a given constraint is said to be marked with
The paper has four parts. In §1 I exemplify and give statistical evidence for the operation of Foregrounding in the phonology of Balinese. In §2 I argue for a physiologically-based iconic motivation for Foregrounding. In §3 I discuss other approaches to accounting for the correlation, and show that they all need to be augmented by principles such as those argued for in §2. In §4 I briefly sum up and draw some more general conclusions.

1. EVIDENCE FOR FOREGROUNDING IN BALINESE. In Balinese, phonological structures which are demonstrably dispreferred in the language nonetheless occur, though where they do they predictably associate with expressive meanings. For example, complex syllable onsets represent a more marked option both in Balinese and crosslinguistically (see for example Blevins 1995, Prince & Smolensky 1993, McCarthy & Prince 1994). Where they occur morpheme medially in Balinese – the most marked option crosslinguistically (Steriade 1982, 1988, Golston 1996:739) – complex onsets associate particularly strongly with ‘expressive’ meanings. A representative sample is given in (1); it includes semantic types (defined below) typically associating with morphemes of this structure.

(1)  
damprat 'yell furiously at'  BAD, SENSE (sound)  
caplak 'eat piggishly, gobble'  BAD, SENSE (movement)  
burjun 'burnt outside, raw inside'  BAD, SENSE (taste/sight)  
sinjad 'awkward, clumsy'  BAD, SENSE (movement)  
kacrit 'sound of spurring water'  SENSE, (sound)  
kables 'suddenly emerge, pop out'  SENSE, (movement)  

Similarly, exceptional ‘allophonic’ structures in Balinese predictably associate with expressive pragmatic overlay. High vowels in final closed syllables in Balinese are usually lax kucit [kucrt] ‘piglet’, kotut [kotut] ‘personal name’. The one exception to this regularity is in certain exclamatory usages:

(2)  
dut! [du(:)t]  [dut]  'exclamation of derision'  
tin! [ti(:)n]  [tn]  'sound of car horn'  

Note that the distribution of [u] and [u] is otherwise purely complementary, as is that of [i] and [i]. For further examples involving other vowels and consonants see Clynes 1995.

Claims such as those exemplified above remain anecdotal, in the absence of (i) evidence for the phonological marginality of the structures concerned (ii) statistical evidence of a correlation with specific expressive meanings. Evidence that complex syllable onsets are dispreferred in Balinese comes firstly from productivity in coining new morphemes. The canonical shape of a morpheme in Balinese, as in the closely related Javanese

---

that a ‘more-marked candidate’ output in OT terms is ‘marked’ in a Praguian sense as well, hence refer loosely to ‘marked’ or ‘more-marked’ structures as a cover term for both.

2Balinese has the following phoneme inventory: vowels /i, u, e, o, ø, a\textsuperscript{\textprime}/, consonants /p, b, m, w; t, d, n, l, r; c, j, n, s, y; k, g, q, h/.
Phonological structures and expressiveness: the role of iconicity in 'the emergence of the marked' (Uhlenbeck 1950), is a disyllabic CV(N).CVC sequence, where N is a nasal stop homorganic with the following consonant. Syllables with complex onsets, of shape CLV(C), where 'L' represents any of /l, r, w, y/, also occur. All other things being equal, we would expect four times as many morphemes of shape CLV(N).CVC, or CV(N).CLVC, to occur as do those of shape CV(N).CVC, containing only simple onsets. Instead, as Table 1 below shows, for morphemes of otherwise identical structure, those with simple onsets (row 1) consistently occur much more frequently than those with complex onsets at the left edge (row 2), while those with morpheme-medial complex onsets are even less common (row 3). In fact there are nearly 20 times as many morphemes of the preferred shape CVCVC as there are of shape CVCLVC:

<table>
<thead>
<tr>
<th>CV.CV</th>
<th>966</th>
<th>CVC.CV</th>
<th>490</th>
<th>CV.CVC</th>
<th>4410</th>
<th>CVC.CV</th>
<th>2070</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLVC.V</td>
<td>185</td>
<td>CLV.C.V</td>
<td>51</td>
<td>CLV.CVC</td>
<td>811</td>
<td>CLVC.CVC</td>
<td>398</td>
</tr>
<tr>
<td>CV.CLV</td>
<td>92</td>
<td>CV.CLV</td>
<td>23</td>
<td>CV.CLVC</td>
<td>242</td>
<td>CVC.CLV</td>
<td>179</td>
</tr>
</tbody>
</table>


A second type of evidence that these structures are dispreferred comes from reduplication: complex onsets found in the base are not copied to the reduplicant in Balinese, as the examples in (3) illustrate (cf McCarthy & Prince 1994); 3a&b illustrate straightforward derivations involving reduplication, 3c illustrates a class of ‘inherently reduplicated’ lexemes where an otherwise non-occurring monosyllabic base is reduplicated:

(3)  
   a.  slampar ‘throw’ > -se-slampar-an ‘throw (vi, DIST)’
   b.  blakas ‘chopper’ > be-blakas-an ‘(talk) trenchantly’
   c.  crukcuk /REDcruk/ ‘kind of bird’

Where morphemes consist of a reduplicated C1LVC2 sequence (3c), the expected L-segment in the medial onset is always missing: crukcuk ‘k.o. bird’ (*crukcruk, *cukcruk).

Overall, then, evidence from both (i) relative frequency in the lexicon and (ii) two distinct reduplication processes indicates that complex onsets are disfavored in Balinese, and least favoured of all where they occur morpheme-medially.

Statistical evidence of the correlation of such marked structures with specific expressive meanings was provided as follows (see Clynes 1995 for a fuller discussion). Samples of morphemes containing suspect structures, as well as a control sample, were examined for the presence of the following semantic types, often said to be associated with expressive lexis (Uhlenbeck 1950, Fudge 1970, papers in Hinton et al 1994):

(4)  
   Diagnostic expressive semantic types:
   - SENSE IMPRESSIONS: any morpheme which lexicalizes impressions of sound ("onomatopoeics"), sight, touch, taste or smell, (e.g. in English "smooth", "rough", "sweet", "sour", "bright", "dull" "dazzle")
• **MOVEMENT VERBS**: morphemes which lexicalize movements of the body, and of body parts, such as (in English) *skip*, *shuffle*, *twitch*, *limp* and so on. Also movements of other animate & inanimate entities e.g. *trickle*, *shoot* (i.e. move quickly e.g. of a comet) [actually a subset of the SENSE IMPRESSION class]

• **FEELING/EMOTION**: morphemes lexicalizing FEELINGS and emotions (e.g. *greed(y)*, *love*, *hate*, *boredom*, *boring*, *curious*, *curiosity*) [also in the SENSE IMPRESSION class]

• **BAD**: morphemes (i) with pejorative value (*cruel*, *nasty*, *showoff*), or (ii) which refer to undesirable things (e.g. *evil spirit*) or undesirable states, either mental, (e.g. *crazy*), bodily (*headache*, *smallpox*), or atmospheric (*flood*, *earthquake*).

Note that these meanings are used here (and in Clynes 1995) purely as a practical diagnostic tool: they are not claimed to represent either semantic primitives (though see Wierzbicka 1996 on *bad* and *feel* as semantic primitives), nor grammaticized subclasses within the lexicon. (Another functional class associating strongly crosslinguistically with aberrant or exceptional phonological structures is that of loanwords. I discuss reasons for grouping them with expressive lexis in Footnote 3 below: this approach is important to the discussion of Ito & Mester’s approach in §3. I stress that loanwords are not included in the statistical counts now discussed.)

A volunteer who had no knowledge of the hypothesis being tested was given a list of just the meanings (and not the associated phonological structures) of a random sample of 175 morphemes of shape *(σ)*.σ.CLVC, as well as 175 from a separate ‘control’ group of morphemes, chosen as having the most regular possible phonological structure in Balinese; in both cases Panitia 1990 was the source of data. The volunteer then classified the meanings of the two samples, either into one of the four types in (4) above, or into another class, ‘not one of the above meanings’. These results are summarised in Table 2, where column 2 gives the total number of meanings classified into one of the four diagnostic classes (here designated as ‘expressive’), for each group, and column 3 the total number meanings classified as ‘other’:

<table>
<thead>
<tr>
<th></th>
<th>‘expressive’ meanings</th>
<th>‘other’ meanings</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(σ)</em>.σ.CLVC sample</td>
<td>65%</td>
<td>35%</td>
</tr>
<tr>
<td>‘unmarked’ sample</td>
<td>37%</td>
<td>63%</td>
</tr>
</tbody>
</table>

Table 2: Classification of meanings of (i) 175 morphemes of shape *(σ)*.σ.CLVC (Row 1) and (ii) 175 morphemes of ‘unmarked’ phonological shape (Row 2). See Appendix 1 for full data & analysis. \( Z = 5.46, \ p < .0001, \) Median Test

3Crosslinguistically loanwords associate strongly with aberrant or exceptional phonological structures - see for example Chomsky and Halle 1968:373, Ito & Mester 1995a&b, Paradis 1996. It is uncontroversial that affective pragmatic elements are associated with loanword usage, at least in the initial phase of borrowing, when such words are still seen as foreign. Hock & Joseph for example argue that connotations of prestige are one of the principal motivating factors in the adoption of loanwords into a language (see Hock & Joseph 1996:271ff). Some loanwords have other affective associations - either pleasurable (*karaoke*, *tango*) or otherwise (*voodoo*, *ebola*). I conclude that a great many loanwords first enter the lexicon because of their affective pragmatic/semantic associations. To that extent they minimally share a common prelinguistic affective context with the other semantic types referred to here as ‘expressive’, and so can be plausibly grouped with them.
An appropriate statistical test (the Median Test, cf Spiegel 1972, Hatch & Lazaraton 1991:271) indicates that the probability of getting the difference in proportions between the two samples in Table 2 in a purely random way is less than one in ten thousand. We thus have good statistical support for a correlation between the otherwise dispreferred phonological structure, and the expressive meanings associating with that structure.

Very similar results are obtained when morphemes containing OCP violations are examined. Many languages disprefer the cooccurrence morpheme-internally of consonants sharing the same place of articulation; these include the related Javanese (Uhlenbeck 1950, Yip 1989), Semitic languages (Greenberg 1950, McCarthy 1994, Frisch, Broe and Pierrehumbert 1997), and English (c.f. Fudge 1970, Borowsky 1986, Davis 1990, Clynes 1995). In Balinese, as in Javanese, such sequences are a more-marked structural option. Occasionally they are eliminated from loanwords, viz. kompa 'pump' from Dutch pomp, and gumi 'world', from Sanskrit bumi, gandela (~jandela) 'window' from Portuguese janela, and klana (~clana) 'trousers' from Hindi carna.

More often, however, these structures are tolerated - particularly where they occur in morphemes with expressive meanings. This gives rise to the following striking example of the productive use of formally dispreferred phonological structures to generate expressive vocabulary.

Statistics presented in Table 3 clearly suggest an avoidance of morpheme shapes where more than one labial, apical stop, laminal (c, j, n or s), or liquid (except, as in Javanese and other related languages, when /l/ precedes /r/, cf Uhlenbeck 1950) cooccur.

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>b</th>
<th>t</th>
<th>d</th>
<th>l</th>
<th>r</th>
<th>s</th>
<th>k</th>
<th>g</th>
<th>nj</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>1</td>
<td>1</td>
<td>120</td>
<td>36</td>
<td>64</td>
<td>69</td>
<td>76</td>
<td>131</td>
<td>22</td>
<td>144</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
<td>0</td>
<td>105</td>
<td>21</td>
<td>89</td>
<td>64</td>
<td>77</td>
<td>124</td>
<td>58</td>
<td>145</td>
</tr>
<tr>
<td>t</td>
<td>21</td>
<td>16</td>
<td>17</td>
<td>19</td>
<td>59</td>
<td>41</td>
<td>110</td>
<td>138</td>
<td>37</td>
<td>125</td>
</tr>
<tr>
<td>d</td>
<td>12</td>
<td>0</td>
<td>26</td>
<td>1</td>
<td>31</td>
<td>14</td>
<td>15</td>
<td>48</td>
<td>9</td>
<td>55</td>
</tr>
<tr>
<td>l</td>
<td>15</td>
<td>15</td>
<td>78</td>
<td>47</td>
<td>0</td>
<td>100</td>
<td>65</td>
<td>80</td>
<td>48</td>
<td>116</td>
</tr>
<tr>
<td>r</td>
<td>11</td>
<td>10</td>
<td>59</td>
<td>14</td>
<td>4</td>
<td>0</td>
<td>18</td>
<td>58</td>
<td>26</td>
<td>63</td>
</tr>
<tr>
<td>c</td>
<td>9</td>
<td>7</td>
<td>65</td>
<td>10</td>
<td>34</td>
<td>38</td>
<td>17</td>
<td>135</td>
<td>22</td>
<td>122</td>
</tr>
<tr>
<td>j</td>
<td>4</td>
<td>0</td>
<td>45</td>
<td>5</td>
<td>15</td>
<td>26</td>
<td>13</td>
<td>29</td>
<td>27</td>
<td>59</td>
</tr>
<tr>
<td>n</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>s</td>
<td>51</td>
<td>28</td>
<td>187</td>
<td>52</td>
<td>63</td>
<td>118</td>
<td>7</td>
<td>97</td>
<td>52</td>
<td>176</td>
</tr>
<tr>
<td>k</td>
<td>33</td>
<td>27</td>
<td>157</td>
<td>55</td>
<td>68</td>
<td>100</td>
<td>115</td>
<td>193</td>
<td>42</td>
<td>243</td>
</tr>
<tr>
<td>g</td>
<td>22</td>
<td>3</td>
<td>101</td>
<td>6</td>
<td>64</td>
<td>89</td>
<td>66</td>
<td>93</td>
<td>80</td>
<td>188</td>
</tr>
</tbody>
</table>

Table 3. Numbers of morphemes of structure /#C...#C/ in Panitia 1990 (column=Initial C, row=Final C). The boxed areas contain homorganic combinations. /s/ is the only laminal to occur morpheme-finally.
There are in general low numbers of morphemes with such shapes, particularly so where
the initial and final segments are identical: /p......p/, /h......h/, and so on. However,
those statistics also suggest that stop consonants belonging to the remaining place
category, dorsal, can cooccur freely. In fact there are more morphemes with dorsals in
both initial and final positions, than morphemes with any other place-place
combination. It turns out, however, that morphemes with both initial and final dorsal
consonants are almost exclusively classifiable as expressives. A representative sample
of morphemes of shape /k......k/, with their meaning types, is given in (5): 4

(5) kangk 'momentarily brought to a stop' MOVEMENT
kaak,kak 'sound of someone bringing up phlegm' SOUND
kacak 'in a mess, scattered all over the place' ANIMAL*
kaewak 'cockroach' BAD
kacek 'not enough (counted wrongly, of money)' BAD
kade{klo{k 'trip up because foot caught' BAD, MOVEMENT
kagek 'suddenly come to an abrupt stop' MOVEMENT
kagok 'speechless & unable to reply' BAD
kaikik 'kind of banana' PLANT*
kakak 'laugh loudly' SOUND

Table 4 compares classifications of the meanings of a sample of 125 morphemes of
shape /k......k/, with those of the control group:

<table>
<thead>
<tr>
<th></th>
<th>‘expressive’</th>
<th>‘other’</th>
</tr>
</thead>
<tbody>
<tr>
<td>/k......k/ sample</td>
<td>62.4%</td>
<td>37.6%</td>
</tr>
<tr>
<td>‘unmarked’ sample</td>
<td>37%</td>
<td>63%</td>
</tr>
</tbody>
</table>

Table 4: Classification of meanings of (i) 125 morphemes of shape /k......k/ (Row 1) and (ii) of 175 morphemes of ‘unmarked’ phonological shape (Row 2). Z = 4.48, p < .0001.

Again morphemes with the marked (/k......k/) structure associate with a far higher
proportion of the diagnostic ‘expressive’ meanings than do those in the control sample.
The Median Test again indicates an extremely low possibility that this could be due to
random effects. The conclusion seems inescapable that this formally less-than-
wellformed structure, like a variety of other such structures in Balinese (Clynes 1995) is
preferred in the coining of expressive lexis. Further cross-linguistic evidence of the
correlation argued for here is provided for example by Uhlenbeck 1950 (for Javanese),
Fudge 1970 (English, etc), Hinton et al (North American, African languages), and
Klamer 1998 (Kambera and other Austronesian languages).

4Note that this sample includes two names of Plants and Animals, semantic types also commonly
associating with ‘marked’ phonological structures see Trubetzkoy 1939, Uhlenbeck 1950, Hinton et al
1994. These types were not included in the statistical testing. For some evidence that these behave as
expressives in linguistic patterning see the handout for this conference paper, available from me directly
at aclynes@fass.ubd.edu.bn. Some nonlinguistic evidence is discussed in Clynes 1998.
§2 ACCOUNTING FOR THE CORRELATION. In this section I will first discuss the Foregrounding hypothesis, Klamer’s (1998) account of it (though she does not use that term) in terms of iconicity, and then propose an alternative model.

The claim that unusual or marked structures correlate with expressive semantics is a very old one. It is at the heart of the Foregrounding hypothesis, as developed by Prague School writers (for example Mukarovsky 1964 [1932] and Jakobson 1965) out of earlier ideas of the Russian Formalist school of literary analysis (Shklovsky’s ‘estrangement’, cited in Van Peer 1994). Foregrounding was said to involve ‘the aesthetically intentional distortion of linguistic components’ (Mukarovsky 1964:23). Poetic license effects thus exemplify one of the principle Foregrounding strategies, ‘deviation’, or the deliberate violation of linguistic norms for pragmatic effect (Leech 1966, Van Peer 1994). However, Foregrounding is not a characteristic exclusive to literary texts. Mukarovsky recognized that it was common in journalism, and its use has since been identified in many other discourse and grammatical contexts (Jakobson 1965, Fowler 1981, 1986, Cook 1994, Van Peer 1994. See the Appendix below for exemplification from a variety of linguistic levels.). The correlations between marked phonological shape and expressive semantics demonstrated above (and for example by Fudge 1970, Clynes 1995, Klamer 1998) are thus predicted to occur by the Foregrounding hypothesis.

The Foregrounding hypothesis however is not in itself an explanation, but rather a statement of a correlation, one which invites an explanation. A partial account of why Foregrounding occurs has been offered by some of the above authors. In particular two characteristics of marked structures (in a Praguan sense) have been proposed which help explain why they are sometimes preferred ahead of more-wellformed ones. These are given in (6) and (7) (except as indicated, the wording and the terms ‘Heightened Salience’ and ‘Potential for Interpretation’ are mine):

(6) **Heightened Salience**: a more-marked structure occurring where a less-marked one is available is perceptually more salient than the less-marked alternative. [cf Mukarovsky 1932/1964, Leech 1966]

(7) **Potential for Interpretation**: A more-marked structure occurring where a less-marked one is available ‘invites[es] interpretation’ (Fowler 1986:73).

We thus notice more-marked structures more readily than less-marked ones (Heightened Salience), and where they are used, we seek to make sense of that use (Potential for Interpretation). Heightened Salience appears well motivated as a manifestation of basic cognitive principles. Some have for example seen the Praguan marked-unmarked dichotomy as one manifestation of a more basic opposition, such as the Figure-Ground of Gestalt psychology. Leech (1966) draws parallels between Foregrounding processes and the Figure-Ground notions. Assuming that such connections are wellfounded, and that linguistic patterning is in part ‘driven’ by such more basic cognitive factors, (6) in itself brings no added baggage to a theory of linguistics. ‘Potential for Interpretation’ (7) too appears to be plausible and in harmony with (and ultimately motivating?) pragmatic principles such as Grice’s Maxim of Manner (1975). Nonetheless a problem remains: why should the natural interpretation of the use of more-marked structures be
that an expressive semantic/pragmatic element is intended? The same objection applies to a modified version of (7), such as (8):

(8) Potential for Interpretation (revised): A more-marked structure occurring where a less-marked one is available ‘invit[es] an expressive interpretation’.

One account which might explain (8) is offered by Klamer 1998, who suggests that the correlation between dispreferred phonological structures and expressive semantics is iconically motivated, in the sense of Pierce’s diagrammatic iconicity (see Jakobson 1965, Haiman 1994). Klamer argues that dispreferred phonological structures and expressive lexis are both perceived to be ‘peripheral’ with respect to (respectively) the rest of the phonology and the rest of the lexicon (cf Uhlenbeck 1950 and Ito & Mester 1995a&b for related views). This common feature of peripherality then motivates a natural association of the two elements: peripheral phonological structures are selected to give form to peripheral semantic structures.

Klamer’s proposal is however problematic: it is unclear why the kinds of meanings listed in (4) should be perceived as semantically ‘peripheral’, indeed it is unclear how the centre-periphery metaphor applies to semantic elements. The objection remains even if we replace ‘peripheral’ with, say, ‘exceptional’: while the phonological structures are clearly so, in what way are the diagnostic meanings ‘exceptional’ meanings? This version of diagrammatic iconicity is then insufficiently elaborated: it requires a bipartite division of meanings stored in the lexicon, but does not offer independent evidence for that division (as opposed to any other of the infinite number of possible ways of thus dividing up the lexicon).

Moreover, it is too limited in that it refers exclusively to linguistic elements, when Foregrounding effects also apply in a variety of paralinguistic and non-linguistic semiotic contexts: e.g. the use of italic and other marked fonts and graphemes to convey expressive meanings (see Appendix), and the attention-getting use of striking visual or aural elements in advertising (Jakobson 1960, Fowler 1981, 1986, Cook 1994). Indeed, the use of marked physiological signals is basic to the prelinguistic communication of emotions such as aggression, fear and sexual desire, not only in humans but in animals. A satisfactory account will have the potential to explain all such associations.

Still, Klamer’s intuition that Foregrounding should be accounted for in terms of iconicism is appealing: such a basic, prelinguistic motivation is needed to account for the recurrent correlation of marked structure and meaning at all levels (cf Zwicky &
Pullum 1987). This paper agrees then that diagrammatic iconicity does motivate Potential for Interpretation (8), but proposes that, as with the well attested cross-linguistic correlation of the high front vowel with the semantic element 'smallness', the correlation has a physiological, rather than purely linguistic grounding (Diffloth 1994).

Expressive language use conveys information about the feelings of the speaker. At the most prototypical level its use is associated with heightened affective states, such as anger, fear, joy or sexual desire. Such states naturally associate with marked physiological states (or 'structures') such as increased cardiac and respiratory rates, and with biochemical events such as production of adrenaline. Such physiological states differ then in various categorical, quantifiable ways from rest states, those associated with neutral affective states. These very natural associations are diagrammed in (10):

<table>
<thead>
<tr>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral affective states &lt;&lt; &gt;&gt; unmarked physiological states/structures</td>
</tr>
<tr>
<td>heightened affective states &lt;&lt; &gt;&gt; marked physiological states/structures</td>
</tr>
</tbody>
</table>

The natural associations in (10) are then exploited and developed in the communication of emotion: neutral affective states cooccur with and so are naturally expressed by unmarked semiotic structure, including linguistic structure. Equally, heightened affective states cooccur with and so are naturally expressed by marked semiotic structure, including linguistic structure; such structures are therefore preferred in the linguistic encoding of emotion or expressivity:

<table>
<thead>
<tr>
<th>(11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral affective states &lt;&lt; &gt;&gt; unmarked structure</td>
</tr>
<tr>
<td>heightened affective states &lt;&lt; &gt;&gt; marked structure</td>
</tr>
</tbody>
</table>

This physiologically-grounded iconicism thus explains why the use of more-marked structures where less-marked one is available should receive an expressive interpretation (cf (8)). We have then not only a primary motivation for Foregrounding effects in linguistic communication, but also for parallel effects found in the prelinguistic communication of emotions, in both humans and animals.

3. ALTERNATIVE ACCOUNTS. Here I will briefly sketch some alternatives to the account presented above, and show that minimally they all require augmentation by a principle
or principles motivated the licensing of exceptional structure such as those discussed above.

3.1 SUBGRAMMARS. One approach to accounting for departures from wellformedness patterns holding generally in the grammar, has been to assume the existence of subgrammars applying to just one subset or ‘stratum’ of the lexicon. This approach is found in some analyses in the Optimality Theory framework (Prince & Smolensky 1993), such as Itô & Mester’s accounts of various lexical phenomena in Japanese (1995a,b). They suggest that the lexicon of Japanese is subdivided into various lexical strata, each of which constitutes ‘a genuine morphological class […] which can be referred to as such in the grammar (1995b:185)’. Five such putative strata are discussed: the canonically wellformed Yamato lexis (consisting mainly of native items), plus four strata allowing a range of structures not found in the Yamato stratum: the Sino-Japanese stratum (loanwords), the large Mimetic stratum (expressives), the Foreign stratum (partly assimilated loanwords) and the Unassimilated Foreign/Alien stratum (loanwords plus some native expressives - swearwords, exclamations, and lexicalized casual speech forms). As in Balinese and English then, morphemes associating with phonological structures not tolerated in the core native lexis are generally either expressives or loanwords.

Itô & Mester assume the standard OT assumption of a single Constraint ranking across the grammar with the exception that the ranking of the class of Faithfulness constraints with respect to other constraints is not constant, but rather varies with different strata. The general principal is illustrated in (12):

(12)  
- Yamato lexis          C1 >> C2 >> C3 >> C4 >> FAITH  
- Sino-Japanese loans   C1 >> C2 >> C3 >> FAITH >> C4  
- Mimetics             C1 >> C2 >> FAITH >> C3 >> C4  
- Foreign              C1 >> FAITH >> C2 >> C3 >> C4  

In strata where Faithfulness constraints are low-ranked, such as the Yamato lexis, underlying structures violating relatively higher ranked constraints (say, C2 to C4 in (12) cannot surface. However in strata such as the Foreign stratum, a variety of structures violating those constraints surfaces, the now more-highly-ranked Faithfulness constraints imposing their emergence.

Using this approach, the occurrence in Balinese of otherwise disfavored structures such as the morpheme-medial complex onsets discussed in §1 is explained as follows. In a subgrammar ‘A’ encompassing the major part of the lexicon such structures could never surface, even if they were to occur underlingly, since they would be blocked by the constraint ranking \(*\sigma.CLV >> FAITH\) (where \(*\sigma.CLV\) is a shorthand for the complex of constraints blocking such structures; see Golston 1996 for one more detailed analysis.). Morphemes containing \(\sigma.CLV\) sequences could however surface in a Subgrammar B, where the reverse constraint ranking, FAITH >> \(*\sigma.CLV\), obtained.

\footnote{Inkelas, Orgun and Zoll (1996), as well as Itô & Mester 1995a,b provide surveys of previous work assuming subgrammars, ‘cophonologies’ (IO&Z) or ‘stratification of the lexicon’ (I&M).}
The account offered by Itô and Mester appears to be very constrained, with a single constraint hierarchy apparently prevailing throughout the lexicon, and with generally only Faithfulness constraints rerankable. It nonetheless has several major problems. Firstly, and fundamentally, it fails descriptively, in that there are many surface counterexamples to the structural types predicted to occur by Itô & Mester's model. From the examination of such exceptional data Rice (1997) concludes that the phonological evidence for the lexical stratification is not strong.

Three further major failings of the subgrammar hypothesis from the point of view of the present study are that (i) the correlation between marginal phonological structure and the expressive (Mimetic) and loanword types both in Japanese and crosslinguistically is neither predicted nor addressed (ii) more generally, the correlation of marked structure with expressive usages at other levels, linguistic and otherwise (cf Appendix), is not addressed (iii) the approach is stipulative, in that it does not explain why stratum-specific reranking of constraints should occur. The sub-grammar approach needs to be augmented with principles such as those proposed in §2, in order to eliminate these latter three problems at least.

3.2 LEXICAL PRESPECIFICATION. As an alternative to the subgrammar approach, Inkelas, Orgun & Zoll (1996 and earlier versions, 1997) and Rice (1997) argue for lexical prespecification - that morphemes which violate constraints posited to hold elsewhere in the lexicon are individually prespecified in their underlying representations for the relevant phonological structures, with highly ranked faithfulness constraints again imposing the exceptional structures. Inkelas, Orgun & Zoll 1997 argue that prespecification has advantages over a subgrammar analysis, including enabling a single constraint ranking throughout the lexicon, rather than an excessive proliferation of subgrammars. This claim is however still vulnerable to a charge levelled at the subgrammar approach, that it runs counter to the spirit of OT in allowing apparently unconstrained (if now lexically-specified) reranking of Faithfulness constraints: the ideal of a single constraint ranking throughout the lexicon is not fully realised.

6And despite the strong claim that 'in a substantial class of cases, constraint reranking can be limited to a reranking of Faithfulness constraints, within an otherwise invariant ordering of constraints' (1995b:206) at least two other constraint classes must be reranked: (i) the class of Alignment constraints must be reranked to account for unusual structural features of the extensive (and semantically expressive) Mimetic lexis, (1995b:190)), and (ii) ‘correspondence-sensitive’ constraints must be promoted to explain the behavior of unassimilated loanwords (Itô and Mester 1998). The appeal to a stratified lexicon, plus limited constraint reranking, looks then increasingly ad hoc given the significant numbers of exceptions to Itô and Mester’s generalizations, and as the number of constraint classes which must be reranked increases. Reranking of Faithfulness constraints alone would certainly not account for the Balinese data, since many ‘more-marked’ structures occur where the imposition of Faithfulness constraints cannot be plausibly invoked, such as in the coining of neologisms with marked structures (for example the productive class of Bad Nicknames, see footnote 8), where Faithfulness constraints are not yet relevant, since candidate structures for neologisms have no underlying representation which Faithfulness could impose.

7Inkelas, Orgun & Zoll 1996, 1997 discuss other general theoretical objections to this approach. Most telling perhaps is that the lexical stratification hypothesis is simply too powerful, generating too many subgrammars. They moreover show that information about exceptionality must be specified at the segmental level, rather than that of morpheme or morpheme class.
One prespecification approach which may avoid this charge is that of Golston’s Direct OT (1996). In Direct OT ‘morphemes are represented by the constraints they violate at surface structure. [...] Most roots and affixes are represented directly by No[feature] and ALIGN/ violations’ (Golston 1996:725). The highly marked Balinese pejorative nickname [ʔuʔ] (which has an otherwise non-occurring initial glottal stop, as well as an OCP violation) can for example be represented as:

\[
\begin{array}{cccccc}
\text{OCP} & \text{No}_\text{a?} & \text{NoSTOP} & \text{NoGLOT} & \text{No HI} & \text{NoRND} \\
R & R & RR & RR & R & R \\
\end{array}
\]

where R indicates a ‘distinctive constraint violation’. [ʔuʔ] is then straightforwardly a ‘marked’ form because its desiderata include distinctive violations of two very highly ranked constraints, OCP and No[ Cr].

Amongst the advantages of Direct OT are that it straightforwardly allows a single constraint hierarchy, it avoids a proliferation of diacritic marking, of subgrammars, or of FAITH types (as found for example in Fukazawa et al’s 1998 modified sub-grammar approach). Moreover, the Praugian ‘marked – unmarked’ distinction is retained in this and other prespecification approaches, whereas it is obscured by a subgrammar analysis.

However even Golston’s version of prespecification is not without its own problems. Like the alternatives surveyed above, it does not explain interactions of semantics and phonology correlations. As long as it is unmotivated, prespecification like stratification is too powerful. None of the versions of prespecification proposed above explains either why exceptional prespecification should be available at all, or why specific items come to be treated in different ways by the grammar. In the case of expressives (including loanwords, see Footnote 3) the iconic factors driving Foregrounding outlined in §2 provide such a motivation.

4. CONCLUSIONS. In the introduction it was argued that both nonlinguistic and linguistic factors must be involved in the interaction of marked phonological structure and expressive semantics, exemplified in §1. I then sketched an account of such factors in §2, where it was argued that Foregrounding effects are naturally explained by the Heightened Salience (6) and Potential for (Expressive) Interpretation (8) of marked structures; Potential for Interpretation is in turn motivated by the ‘physiologically-grounded’ iconic model proposed there. These functional principles, plus iconically-driven licensing of exceptional outputs, provide an important missing element to accounts such as Golston’s Direct OT (§3), which enables a formal description of exceptionality effects in the phonology, but does not explain them. How Foregrounding effects in other linguistic and paralinguistic contexts would be treated formally remains

---

8Virtually all Balinese are given a ‘bad name’ (adan jale) as a child; these pejorative nicknames are often ‘meaningless’ but always phonologically marked, so inelegant if only because of their marked shapes; see Clynes 1995a&b. Bad names are a source of embarrassment as an adult, and are naturally classifiable into the BAD semantic/pragmatic type.
to be described, however functional principles such as those described in §2 above are clearly needed in all such cases.

McCarthy & Prince’s characterise the process of ‘emergence of the unmarked’ as follows:

‘Even in languages where [a constraint] C is crucially dominated [e.g. by Faithfulness constraints, AC] and therefore violated, the effects of C can still be observed under conditions where the dominating constraint is not relevant. Thus, in the language as a whole, C may be roundly violated, but in a particular domain it is obeyed exactly. In that particular domain, the structure unmarked with respect to C emerges, and the structure marked with respect to C is suppressed’ (McCarthy & Prince 1994:333)

Foregrounding effects, which link with formally more-marked structure with expressive semantics, can reasonably be said to manifest the reverse effect, of ‘emergence of the marked’, given that (paraphrasing and modifying the above) that “even in a language where a constraint C is not normally violated, the effects of violation of C can be observed under conditions where violation of that constraint is licensed by other principles [those described in §2 above]. Thus, in the language as a whole, C may be obeyed exactly, but in a particular domain it is roundly violated. In that particular domain, the structure marked with respect to C emerges.” This claim is compatible with either of the formal models (prespecification, subgrammars) described above.

The data examined here are then strong evidence that functional principles not only interact with formal ones in the determination of linguistic outputs (as argued for example in the work of Boersma 1997, Steriade 1993, Flemming 1995, Jun 1995, Kirchner 1998), but that in the right context they can be seen to override concerns of purely formal wellformedness. To that extent Foregrounding effects provide evidence that functional/semantic desiderata outrank purely formal principles in the determination of linguistic outputs. Language is indeed ‘a system where everything “conspires” to convey meaning’ Wierzbicka (1988:1).

Acknowledgements

Thanks are owed to many people who have contributed directly to the present paper, including Bapak Ketut Mandi Pinatih, of Banjar Peguyangan, Singaraja, Bali, and Ketut Clynes, who provided information on Balinese. Yvonne Pittelkow of RSPAS, ANU provided invaluable help with the statistics. Emmanuel Aquino did the semantic classifications in §2; Judith Blevins, Alan Dench, Bill Foley, Sander Adelaar, Avery Andrews, Carl James, Marian Klamer, Malcolm Ross and Andy Pawley gave helpful comments on earlier versions, as did audiences in Sydney, Perth, Tucson, and Bandar Seri Begawan. Mark Aronoff pointed me to the Foregrounding literature, the writings of Paul Boersma and Chris Golston helped improve the research design. Special thanks to the organisers of Afla7 for the opportunity to present this paper there, and to the audience for valuable feedback.

References


Frisch, Stefan, Michael Broe and Janet Pierrehumbert. 1997. Similarity and phonotactics in Arabic. ms. Rutgers Optimality Archive 223.


Fukazawa, H., M. Kitahara & M. Ota 1998. Lexical stratification and ranking invariance in constraint-based grammars. CLS 34 (Rutgers Optimality Archive, 267)

Golston, Chris. 1996. Direct OT: Representation as pure markedness. Lg. 72.713-748.


Kirchner, Robert Martin. 1998. An effort-based approach to consonant lenition. Dissertation. UCLA.


Appendix

The broader context: other correlations of marked structures and expressive semantics.

Any approach that seeks to explain only correlations of phonology with expressive semantics has a further major shortcoming: those correlations are merely one expression of a phenomenon manifesting itself at all levels of linguistic structure. Expressive semantic features regularly associate with marked structures, not just at the phonological level, but also at the levels of morphology, syntax, discourse, and paralinguistics. In this appendix I illustrate a variety of such effects, which clearly require a common account.

1 EXPRESSION MORPHOLOGY. Pullum & Zwicky (1987) describe a variety of aberrant structural features, characteristic crosslinguistically of expressive morphemes (such as diminutives and, in English, ‘whimsical’ affixes of the -eria/-ette type). These include a) ‘promiscuity of input category’, b) ‘promiscuity of input base’, and c) ‘alternative outputs’.

‘Promiscuity of input category’, refers to the unusual ability of expressive morphemes to combine freely with bases of virtually any lexical category: thus ‘expletive infixes’ in
English can irrupt variously into nouns (*kanga-bloody-roo*), adjectives (*fan-friggin-tastic*), verbs (*e-friggin-vaporate*), adverbs (*abso-bleeding-lutely*) and interjections (*halle-bloody-lujah, ho-bloody-ho*), and always of course for jocular or other expressive effects.

'Promiscuity of input base' allows morphemes used expressively to combine not merely with other simple morphemes, but with more complex units such complex proper nouns (*Andy friggin' Williams, West friggin' Virginia*), other compounds (*phantom frigging army, rail bloody way*), and longer lexicalized expressions (*hopping goddam mad, fat frigging chance*). Expressive morphology also often produces 'alternative outputs': e.g. both *Kalgodammazoo* and *Kalama-goddam-zoo* are accepted by many speakers, whereas plain derivational morphology allows only a single possible output. Descriptions of comparable phenomena in other languages are found for example in Uhlenbeck (1971), Stump (1993), and Bauer (1998).

2 **EXPERIMENTAL SYNTAX.** Marked syntactic constructions also correlate with expressive semantics. In English, for example, all predicates must contain a verb phrase—except in a few highly expressive 'exclamatory' constructions, where certain nominal predicates are permitted:

(1) a. You rat!/*angel!/(*person!/??linguist!)
b. Semantics shemantics!\(^9\)
c. A semantic account, my eye!

Exclamatory particles used as single word utterances constitute no doubt the most marked syntactic 'structure', to the extent that they are generally treated as paralinguistic elements. They are almost always pragmatically charged: *Hallelujah! Shit! Hooray! Ouch!*... Often they are phonologically exceptional as well: *Shh(t)! Huh? Pssst! Phew! Brrrr! Yechh! Oh-oh!*

The mismatch between levels involved in expletive infixing—units which normally function as whole words instead functioning as modifying affixes—has parallel manifestations in syntax, when entire phrases (PP's) can be lexicalized as premodifying elements within noun phrases, but only where they are used for expressive effect (c.f. *the on-the-corner house*). The structure involves some kind of weight violation, with the heavy modifier preceding rather than following the head:

(2) a. an off-the-wall analysis
b. a totally in-your-face presentation
c. an over-the-top performance
d. that out-of-this-world feeling
e. an out-of-hand rejection

The expletive word class again predictably associates with a variety of unusual structures at the syntactic level. First, expletives again show unusual 'promiscuity' of syntactic

---

\(^9\)Examples here and below are from McMillan (1980).

\(^10\)Zwicky & Pullum (1987:338) suggest that a sequence like *Semantics shemantics!* constitutes a single NP. Yet it cannot fill a normal NP slot, and constitutes a complete utterance, with topic-comment-like propositional and intonational structure: it is therefore better understood as a kind of aberrant sentence. I assume that like other expletive affixing processes, 'shem-reduplication' does not change word class; *shemantics* is then an NP functioning as predicate: the retained plural suffix is compatible with this view.
combination, occurring for example as modifiers in phrases of a variety of types (3a-d), before interjections (3d), and as VP-level modifiers (3e):

(3)  
   a. you bloody rat (NP)  
   b. bloody fantastic (AdjP)  
   c. bloody fantastically (AdvP)  
   d. bloody hell!  
   e. why don’t you bloody do it!

And only expletives (and one or two other highly affective terms) occur in ‘imprecations’ (4a); verbs of other semantic types cannot occur in this structure (4b):

(4)  
   a. Stuff/Bugger/Frig (...) /Bless you!  
   b. *Hit you!/*Ruin you!/*Praise you!/*Congratulate you!

Thirdly, and paralleling expletive infixing, expletives alone occur in various otherwise untolerated irruptions and excrescences (to use the terminology of McMillan 1980) at the phrase (5a-b) and clause (5c-d) levels:

(5)  
   a. hurry friggin (/*right) up  
   b. Get the hell/fuck/*place out of here!  
   c. Where the hell/fuck/*place are my glasses?  
   d. Who the hell/fuck/*person/*name is he?

3 EXPRESSIVE DISCOURSE. Particular discourse genres in themselves license the use of marked structures. The best known example of this is in literary language, where “poetic license” effects exploit a wide variety of marked structures and marked semantic juxtapositions. Thus in 6a, from Gerard Manley Hopkins’ *Windhover*, the position of the adverbal NP, apart from creating processing difficulties for the reader, would be close to ungrammatical in other discourse types, as is 6b. And semantically, ‘morning’s minion’, like all metaphor, is incoherent if taken literally (Leech 1966), c.f. 6c.

(6)  
   a. I saw this morning morning’s minion  
   b. ?I read this morning the newspaper.  
   c. *I read this morning morning’s newspaper

Elsewhere Manley Hopkins systematically uses a wide range of deliberate violations, such as the (non-expletive) infixing in *wind-lilylocked-laced* (Harry Ploughman, cited in McMillan 1980). Such violations are typical of literature, though parallel uses of marked structures occur in a variety of other discourse contexts, such as jokes, advertising and journalism (Fowler 1981, 1986, Guy 1994).

4 EXPRESSIVE PARALINGUISTIC EFFECTS. In paralinguistic contexts too a variety of marked usages signal expressive functions: emotion is signaled for example by the use of fluctuations in voice pitch and quality, as well as fluctuations in volume and speed of utterance. In written texts, contextually unexpected allographs have parallel functions; *italic fonts* and *SMALL CAPITALS*, for example, are used to give salience to text, to signal ‘strange’ or unassimilated loanwords, and to flag the use of onomatopoeia (c.f. the SENSE semantic type), swear words (c.f. the BAD type) and other exclamatory usages: all functions associated with marked structures at other levels of linguistic organization. The parallels between the uses of marked allographs and those of marked phonological structures are particularly close.
0. Introduction

It has been claimed on the basis of Indonesian and Malay data such as (1) and (2) that long distance movement (at least across two clause boundaries) is possible in Javanic languages. Under that analysis, the wh-element, siapa ‘who’, binds an empty operator which moves from the most deeply embedded clause into the highest clause. Thus, in (1) and (2) there appears to be a long-distance relation between the wh-element and the empty position e in the embedded clause. ²

(1) Indonesian

Siapa yang [ Ali kira (yang) e; mem-beli mobil]]?
who think AV-buy car
‘Who did Ali think bought a car?’

(2) Malay (Cole & Hermon 1998 (27a))

Siapa [ Ali buktikan yang e; (men-)curi kereta]]?
who prove that (meng-)steal car
‘Who did Ali prove stole the car?’

Data such as (3) and (4), from Madurese and Javanese, seem to confirm this for these languages as well.

(3) Sapa, se [ Ali kera (jhaq/se) e; melle montor]]?
who REL A think COMP/REL AV.buy car
‘Who did Ali think bought a car?’

(4) Sapa, sing [ di-kira Ali (nek) e; tuku montor]]?
who REL OV-think A COMP buy car
‘Who did Ali think bought a car?’

¹I would like to thank Chris Culy, Stan Dubinsky, and Ileana Paul for helpful discussion of some of the issues considered here or comments on earlier drafts. Errors and shortcomings are regrettably mine alone. This work was supported in part by the National Science Foundation through grant SBR 98-09044 to the University of Iowa.

²Cole, Hermon, and Aman (to appear) and Paul (2000) provide evidence for the empty operator movement analysis of cleft questions for Malay and Malagasy, which extends to Madurese and, most likely, to Western Austronesian in general.
However, I will show on the basis of data from Madurese that long-distance movement may well be illusory. In fact, the Madurese data support an analysis in which the relation between the wh-element and the “extraction site” must be radically local; in particular, only a single clause node may intercede. I speculate on the strength of the evidence for Madurese that long-distance movement may be disallowed in all Javanic languages. Evidence from Javanese of the type I discuss for Madurese is included in an appendix. First, however, I show that if there is extraction from clauses, it must be from clausal subjects.

1. Extraction restricted to sentential subjects

The question in (3) appears to provide prima facie evidence that I am incorrect in my overall assertion. Saddy (1991) and Cole & Hermon (1998) have argued explicitly for long-distance movement in Indonesian and Malay, respectively. However, I will attempt to show that (3) does not provide counterevidence to my claim, but that the analysis proposed for these other languages is not tenable for Madurese.

Evidence for this comes from the interpretation of wh-questions. Note the asymmetry in the interpretation of the pair of sentences in (5).

(5) a. Bila se Siti ngera jhaq Marlena noles buku?
    when REL S AV.think COMP M AV.write book
    ‘When did Siti think that Marlena wrote the book?’

b. Bila se Siti kera jhaq Marlena noles buku?
    when REL S think COMP M AV.write book
    ‘When did Siti think that Marlena wrote the book?’

or ‘When did Siti think that Marlena wrote the book?’

(5a) is a question only about the time of the thinking event, while (5b) is ambiguous, the primary interpretation being a question about the time of the writing event. The sole difference between the two sentences is the morphology on the matrix predicate kera ‘think’—actor voice in (5a) and the bare verb form in (5b).¹

Saddy (1991) and Cole & Hermon (1998) account for the difference in interpretation through a morphological constraint that prohibits movement of an NP across a verb with actor voice morphology (the meng- prefix in Indonesian and Malay). As should be noted, in the Indonesian and Malay examples in (1) and (2), the matrix verb lacks the AV prefix; both

³Although (5b) is ambiguous, the interpretation in which the question pertains to the event in the embedded clause is strongly preferred.

⁴As explained below, actor voice (AV) occurs on transitive predicates in which the agent or actor is surface subject.
sentences would be ungrammatical were it to occur, as (6) and (7) show.

(6) *Siapa yang [ Ali mengira [ (yang) mem-beli mobil]]?
who A AV.think AV-buy car
(Who did Ali think bought a car?)

(7) *Siapa [ Ali mem-buktikan [ yang (men-)curi kereta]]?
who A AV-prove that AV-steal car
(Who did Ali prove stole the car?)

Both Saddy and Cole & Hermon claim that the matrix verbs in the grammatical sentences are active and not “passive”. However, I argue that these are indeed passive clauses, drawing on data from Madurese. Note that the Madurese question in (5b) is equivalent to the sentence in (8), which takes the object voice (or “passive” prefix) e-. That is, it is first a question about the writing event and secondarily a question about the thinking event.

(8) Bila se e-kera Siti jhaq Marl e na noles buku?
when REL ov-think S COMP M AV.write book
‘When did Siti think that Marlena wrote the book?’
or ‘When did Siti think that Marlena wrote the book?’

I wish to claim that in (5b) and (8) the matrix is “passive”, in the same way that the clauses in (9b) and (9c) are passive.

dog that AV.bite A
‘The dog bit Ali.’
A dog that bite
‘The dog bit Ali.’
A ov-bite dog that
‘The dog bit Ali.’

The data in (9) illustrate the basic voice system in Madurese. Actor voice (AV) morphology occurs on the verb when the agent or actor of a transitive sentence (and some intransitives)
occurs as the subject, as in (9a). AV is marked by a nasal prefix or the prefix a-. (9b,c) are what have sometimes been referred to as passives (especially in Indonesian (Chung 1976, Kana 1986, Sneddon 1996)). In both, a non-actor is the subject and the actor occurs either immediately preverbally in the bare verb form (9b) or postverbally in the object voice (OV) form (9c), where object voice is invariably marked by the prefix e-. 6

If the matrix clauses in (5b) and (8) are passives, they actually demonstrate extraction from a sentential subject and not a complement clause. The reason that extraction appears to be from a complement is that the sentential subject has been extraposed. Thus, I argue that (5b) has the analysis in (10).

(10) [ bila e' je Siti kera ei ] [ jhaq Marlena noles buku ei ]

In (10), the clausal constituent jhaq Marlena noles buku e, originates as the complement of kera ‘think’, moves to subject position e’, and is extraposed to postverbal position. 7 When in subject position, bila ‘when’ (or more accurately the operator that it binds) is fronted. Although the sentence is clearly open to alternative analyses at this point, in what follows I will motivate the appropriateness of elements of (10). There are three types of evidence favoring subjecthood of the clause containing the extraction site.

1.1 Clausal complements as subjects of passives

First of all, the entire complement clause may occur as the subject of the clause, provided that the verb is in the bare form (11a) or the object voice form (11b), as demonstrated here with declarative sentences.

(11) a. Jhaq Marlena noles buku Siti kera.  
    COMP M AV.write book S think  
    ‘Siti thinks that Marlena wrote the book.’

b. Jhaq Marlena noles buku e-kera Siti.  
    COMP M AV.write book OV-think S  
    ‘Siti thinks that Marlena wrote the book.’

1.2 Verbal suffixes

Second, there are predicates that require additional verbal morphology when the sentential complement occurs in subject position. The verb bale ‘say’ takes a sentential complement, as

6 Additional details regarding the Madurese voice system are available in Davies 1999a.

7 I take no position here regarding whether the extraposed clause is higher or lower in the tree structure than the wh-element.
Against long movement in Madurese

in (12).

(12) Siti a-bala [jhaq Hasan entar daq Jakarta].
    S AV-say COMP H go to J
    'Siti said that Hasan went to Jakarta.'

When the sentential complement occurs in subject position, bala obligatorily takes the suffix -aghi, (13a); when the suffix is not present, ungrammaticality results, (13b).

(13) a. [Jhaq Hasan entar daq Jakarta] Siti bala-aghi.
    COMP H go to J S say-AGHI
    'Siti said that Hasan went to Jakarta.'

b. *[Jhaq Hasan entar daq Jakarta] Siti bala.
    COMP H go to J S say
    (Siti said that Hasan went to Jakarta.)

The data in (14) show a similar situation with the verb kasta 'regret'. When the suffix -e occurs on the predicate, the sentential complement can be the subject (14b); but when the suffix is not present, the sentence is ungrammatical, (14c).

(14) a. Ita kasta [jhaq Bambang mangkat].
    I regret COMP B leave
    'Ita regrets that Bambang left.'

b. [Jhaq Bambang mangkat] Ita kasta-e.
    COMP B leave I regret-E
    'Ita regrets that Bambang left.'

c. *[Jhaq Bambang mangkat] Ita kasta.
    COMP B leave I regret
    (Ita regrets that Bambang left.)

The relevance of this is that it is only possible to form a question with a sentence-initial wh-phrase relevant to the embedded clause when the requisite "clausal subject" morphology occurs on the matrix predicate.

(15) a. *Daqrema se Siti bala jhaq Hasan entar daq Jakarta?
    how REL S say COMP H go to J
    (How did Siti say that Hasan went to Jakarta?)
William Davies

b. Daqrama se Siti bala-agli jhaq Hasan entar daq Jakarta?8
   how REL S say-AGHI COMP H go to J
   ‘How did Siti say that Hasan went to Jakarta?’

Only (15b), in which the verb takes the suffix -aghi, is grammatical.9 Parallel data are available with the verb kasta ‘regret’.

(16) a. Bila se Ita kasta jhaq Bambang mangkat?
    when REL I regret COMP B leave
    ‘When did Ita regret that Bambang left?’

b. Bila se Ita kasta-e jhaq Bambang mangkat?
    when REL I regret-E COMP B leave
    ‘When did Ita regret that Bambang left?’

(16a), which lacks the suffix, can only be interpreted as a question about the time of Ita’s regret; that is, the matrix event is questioned. The complement event can only be questioned if the suffix occurs on the matrix verb, as in (16b). The obligatoriness of the additional verbal morphology in (15b) and (16b) requires an account. Given the fact that this morphology is also obligatory when the complements occur as subjects, (13a) and (14b), a reasonable hypothesis is that the complements in (15b) and (16b) are also subjects, albeit extraposed subjects.

1.3 “Control” structures

A third argument for the subject status of the complements is available based on a widespread construction involving matrix verbs. Most matrix verbs in Madurese allow an additional argument to occur provided that the argument is coreferent to an embedded pronominal element. This is illustrated for the verbs kera ‘think’ in (17) and bala ‘say’ in (18). Note that once again, bala must take the suffix -aghi.

---

8This sentence is ambiguous between the interpretation given in (15b) in which it is a question about the event described in the complement clause and an interpretation in which it is a question about the event described in the matrix clause. As these types of questions always carry this systematic ambiguity, henceforth I will ignore the interpretation regarding the matrix event unless it is of particular relevance.

9To form a grammatical question in which the matrix event is questioned without the -aghi suffix, AV morphology is necessary on the matrix verb, as in (i).

(i) Daqrama se Siti a-bala jhaq Hasan entar daq Jakarta?
   how REL S AV-say COMP H go to J
   ‘How did Siti say that Hasan went to Jakarta?’
(17) Siti nggera Marlena jhaq (abaqang) noles buku.
   S AV.think M COMP she AV.write book
   'Siti thinks about Marlena that she wrote the book.'

(18) Siti a-bala-aghi Hasan jhaq (abaqang) entar daq Jakarta.
   S AV.say-AGHI H COMP he go to J
   'Siti said about Hasan that he went to Jakarta.'

In Davies (2000), I argue that there is no raising in Madurese but that Marlena and Hasan are dependents of the matrix clause and control a pronoun (usually nonovert) in the complement.10 These additional objects can be promoted to subject, as in (19) and (20), the bare verb counterparts of (17) and (18), respectively.

(19) Marlena Siti kera jhaq (abaqang) noles buku.
   M S think COMP she AV.write book
   'Siti thinks about Marlena that she wrote the book.'

(20) Hasan Siti bala-aghi jhaq (abaqang) entar daq Jakarta.
   H S say-AGHI COMP he go to J
   'Siti said about Hasan that he went to Jakarta.'

What is significant for present purposes is the interpretation given to questions when this additional argument occurs as the subject of the matrix clause.

(21) Bila se Marlena Siti kera jhaq (abaqang) noles buku?
   when REL M S think COMP she AV.write book
   'When did Siti think about Marlena that she wrote the book?'

(22) Daqrema se Hasan Siti bala-aghi jhaq (abaqang) entar daq Jakarta?
   how REL H S say-AGHI COMP he go to J
   'How did Siti say about Hasan that he went to Jakarta?'

Importantly, (21) and (22) are questions only about the matrix predicate. Were it the case that long-distance movement was possible here, these questions should be as ambiguous as (5b) and (15b); that is, it should be possible to interpret these as questions about either the matrix event or the embedded event, with a potential preference for the embedded interpretation. We can account for the lack of ambiguity in these questions if extraction is limited to sentential subjects. In (21) and (22), the sentential complements can never be subjects because that position is the province of the NPs Marlena and Hasan. Since the complements are not sentential subjects, the interpretation of the wh-phrase is limited to the matrix predicate.

10 Aspects of this construction are also discussed in Davies 199a,b.
1.4 Questioning from in situ sentential subjects

A final piece of evidence for the restriction of extraction to sentential subjects comes from the fact that these questions can be formed when the sentential element still resides in subject position, as in (23-25). All of these sentences are ambiguous regarding whether it is the matrix or embedded predicate that is the focus of the question; however, as with the previous ambiguous examples, the preferred interpretation is the embedded predicate.\(^{11}\)

(23) Bila se Marlena noles buku Siti kera?
when REL M AV.write book S think
‘When did Siti think that Marlena wrote the book?’

(24) Daqrøma se Hasan entar daq Jakarta Siti bala-ahi?
how REL H go to J S say-AGHI
‘How did Siti say that Hasan went to Jakarta?’

(25) Bila se Bambang mangkat Ita kasta-e?
when REL B leave I regret-E
‘When did Ita regret that Bambang left?’

Additionally, base-generated sentential subjects show the same properties that have been established for sentential complements above. This is illustrated in the following data:

(26) Jhaq Ali ng-rosaq komputer nyøngøng-ahi Siti.
COMP A AV.ruin computer AV.surprised-AGHI S
‘That Ali ruined the computer surprised Siti.’

(27) Daqrøma se Ali ng-rosaq komputer nyøngøng-ahi Siti?
how REL A AV.ruin computer AV.surprised-AGHI S
‘How did that Ali ruined the computer surprise Siti?’

(28) Daqrøma se nyøngøng-ahi Siti jhaq Ali ng-rosaq komputer?
how REL AV.surprised-AGHI S COMP A AV.ruin computer
‘How did it surprise Siti that Ali ruined the computer?’

In (27), there is a clear case of questioning an element in a sentential subject. In (28), the sentential subject has been extraposed and the interpretation is the same as (27), where there is no extraposition.

\(^{11}\)These sentences can frequently be disambiguated with appropriate intonation. However, as written with no intonational cues, they are ambiguous.
Against long movement in Madurese

While it might seem relatively strange to claim (at this juncture) that extraction from clauses is only possible from sentential subjects, this actually brings Madurese (and other Javanic languages) into synchronicity with many languages in the Western Austronesian subgroup. Keenan (1972) has claimed for Malagasy and Kroeger (1993) for Tagalog that extraction from clauses is only possible from sentential subjects. Thus, while the analysis thus far provides no explanation for this peculiarity, it does put these languages on an equal footing.

2. Partial movement and extraposition

However, as stated above, I do not wish to leave the analysis here. Having established the relevance of subjecthood above, I now provide evidence for the claim that there is no real movement out of the embedded clause. That is, there is no long-distance movement at all in Madurese. Rather what obtains in the case of (5b) and (15b) and many other examples is so-called “partial” movement followed by extraposition of the sentential subject. I attempt to motivate this analysis below.

2.1 Partial movement

As is well-known, Western Austronesian languages have what has been characterized as “partial wh-movement” in which a wh-phrase occurs in initial position of an embedded clause yet has wide scope. As (29) shows, Madurese is no exception to this.

(29) Siti ngera bila se Marlena noles buku?
    S AV.think when REL M AV.write book
    ‘When does Siti think that Marlena wrote the book?’

Despite the fact that bila ‘when’ occurs in the embedded clause, (29) is a wide scope question, and a perfectly natural construction. (29) is synonymous with the preferred interpretation of (23), where the entire complement clause occurs as the subject rather than the object.

(23) Bila se Marlena noles buku Siti kera?
    when REL M AV.write book S think
    ‘When did Siti think that Marlena wrote the book?’

Thus, (23) can also be analyzed as an instance of partial movement. That is, the appropriate analysis of (23) is (30), where the wh-phrase is still contained in the sentential subject, rather than (31), where it occurs in the matrix clause.

(30) = (23) [[ bila, [Op, se Marlena noles buku e_i], Siti kera e_j ]]
(31) bila, [ Op, se [ [Marlena noles buku e_i], Siti kera e_j ]]
Following this line of thought, the appropriate analysis of (5b) is (32), where partial movement and extraposition cooccur.\footnote{Here I again leave unspecified precisely where in the structure the extraposed clause occurs inasmuch as this does not bear on present concerns.}

\begin{equation}
(32) = (5b) \quad \left[ [ \text{bila}_i \text{ e}_k \right]_i \text{ se } \text{Siti kera } \text{e}_j \left] \right. \left. [ \text{Op}_i (\text{jhaq}) \text{Marlena noles buku } \text{e}_k \right]_k
\end{equation}

The structure in (32) contains the elements of (30) and adds to it the extraposition of the constituent \((k)\) which includes the null operator and the remainder of the complement clause.\footnote{One might object that the usual state of affairs is that the relative particle \text{se} is used in the complementizer position of questions but in (5b,32) \text{jhaq} shows up as the complementizer of the extraposed sentential subject and the \text{se} particle is claimed to be part of the matrix clause. First off, \text{se} is possible in the position occupied by \text{jhaq} here. Second, precisely this set of complementizers (\text{jhaq} in the embedded clause and \text{se} in the matrix) occur in (i), in which the sentential subject has not been extraposed.}

(32) is to be contrasted with (33), in which the \textit{wh}-element commands the entire sentence in syntactic structure.

\begin{equation}
(33) \quad \text{bila}_i \left[ \text{se } [ [ \text{e}_k ]_i \text{ Siti kera } \text{e}_j \left] \right. \left. [ \text{Op}_i (\text{jhaq}) \text{Marlena noles buku } \text{e}_k \right]_k \right]
\end{equation}

Evidence for the partial movement/extraposition analysis is available from (i) extraposition of sentential subjects, (ii) extraposition of relative clauses, and (iii) extraposition of sentential complements with partial \textit{wh}-movement.

\subsection*{2.2 Extraposition of sentential subjects}

First, sentential subjects can routinely be extraposed, as illustrated in (34b) and (35b).

\begin{equation}
(34) \quad \text{a. } \left[ \text{Jhaq Marlena noles buku } \right]_i \text{ e-kerja Siti.} \\
\quad \text{COMP M AV.write book ov-think S} \\
\quad \text{‘Siti thinks that Marlena wrote the book.’} \\
\quad \text{b. E-kerja Siti } \left[ \text{Jhaq Marlena noles buku } \right]. \\
\quad \text{OV-think S COMP M AV.write book} \\
\quad \text{‘Siti thinks that Marlena wrote the book.’}
\end{equation}
Against long movement in Madurese


  COMP H not pass test-DEF lie

  'That Hasan did not pass the test is a lie.'

  b. Lecek [jhaq Hasan loq lolos tes-a].

  lie COMP H not pass test-DEF

  'It's a lie that Hasan did not pass the test.'

In fact, as is the case in English, extraposed sentential subjects are largely preferred, most likely for reasons of processing.

2.3 Extraposition of relative clauses

Second, relative clauses in Javanic languages are formed using the same strategy as focussed wh-questions, and relative clauses can be extraposed. A relative clause in subject position is illustrated in (36).


  package that REL OV-send B OV-open I

  'Ita opened that package that Bambang sent.'

In (36) paket jhuwa se ekerem Bambang ‘that package that Bambang sent’ is the subject of the verb in object voice, e-bukaq ‘open’. The parallel between this construction and the questions in which the sentential complement is in subject position is unmistakable. As (37) illustrates, the relative clause can be extraposed to clause-final position, leaving the head in clause-initial position.

(37) Paket jhuwa e-bukaq Ita [se e-kerem Bambang].

  package that OV-open I REL OV-send B

  'That package was opened by Ita that Bambang sent.'

It is also possible to extrapose relative clause in object position. In (38b), the modifying clause occurs sentence-finally, separated from its head, buku, by the PP daq Ali ‘to Ali’.

(38)  a. Marlena a-barriq buku [se Siti toles] daq Ali.

  M AV-give book REL S write to A

  'Marlena gave the book that Siti wrote to Ali.'

  b. Marlena a-barriq buku daq Ali [se Siti toles].

  M AV-give book to A REL S write

  'Marlena gave the book to Ali that Siti wrote.'
2.4 Extrapolation of sentential complements

The final piece of evidence for the partial movement/extrapolation analysis comes from sentential complements. As illustrated above in (29), partial movement with wide scope interpretation is possible with sentential complements. In the same way that relative clauses in object position can be extrapolated and thus separated from their heads, sentential complements can be separated from a wh-phrase, as in (40).

(39) ?Biq telpon Siti a-bala-agni daqrama [se Marlena ng-rosaq komputer daq Ali ]?
     with telephone S AV-say-AGHI how REL M AV-ruin computer to A
     'How did Siti tell Ali by telephone that Marlena ruined the computer?'

(40) Biq telpon Siti a-bala-agni daqrama daq Ali [ jhaq/se Marlena ng-rosaq komputer ]?
     with telephone S AV-say-AGHI how to A COMP/REL M AV-ruin computer
     'How did Siti tell Ali by telephone that Marlena ruined the computer?'

The sentence in (39) is only marginally acceptable due to the heavy nature of the complement clause in relation to the following PP. However, it serves as the basis for comparison with the completely acceptable (40). In (40), the wh-element daqrama 'how' precedes the PP daq Ali 'to Ali' and the sentential complement follows. In both (39) and (40) it is unambiguously the event in the embedded clause which is the focus of the question, as the matrix clause already includes a manner adverbial, the instrumental PP biq telpon 'by telephone'.

If the analysis is correct, this predicts that it should be possible to extrapolate a sentential subject that includes a partially moved wh-phrase in its initial position. As (41) shows, this is correct.

(41) Nyangang-agni Siti daqrama se Ali ng-rosaq komputer?
     AV.surprise-AGHI S how REL A AV-ruin computer
     'How did it surprise Siti that Ali ruined the computer?'

---

1 Preferable to (39) is (i), in which the heavy constituent is shifted to clause-final position.

(i) Biq telphon Siti a-bala-agni daq Ali daqrama se Marlena ng-rosaq komputer?
     with telephone S AV-say-AGHI to A how REL M AV-ruin computer
     'How did Siti tell Ali by telephone that Marlena ruined the computer?'
Against long movement in Madurese

3. Conclusion

The foregoing data lend plausibility to an analysis of Madurese cleft questions which rejects all long distance movement in favor of a radically local relation. The natural question to ask is whether this is a desirable move. Is anything gained under the proposed analysis? I would claim that there is. If we leave aside the proposal that all wh constructions are radically local, we are left with the fairly convincingly established fact that all extraction from clauses must be from sentential subjects. While this accords with observations about other Western Austronesian languages, this condition in and of itself is relatively unusual—certainly not what would be expected given what we know of the world’s languages in general. Attempts to explain the restriction of movement to sentential subjects in various Austronesian languages frequently wind up relatively complex and largely mere descriptions of the facts. Adoption of the present proposal eliminates the need for any arcane machinations and perhaps makes the affected languages seem a little less unusual.

The “partial” movement aspect of the proposed analysis is, of course, needed in any event, so its inclusion adds nothing to the grammar of Madurese. Therefore, in some respects restriction to radical locality results in a simpler grammar. What goes begging, obviously, is any explanation for why these A-bar relations should be restricted in this way. I have no good explanations for this at this juncture. My speculations run along the following lines: if it turns out that the partial movement/extraposition story would work somehow for all these ‘sentential subject’ languages, then one might look to the voice system for an explanation since all these languages have rich voice systems. On the other hand, if it is just the Javanic languages that are open to this kind of analysis, we might take the results that we find in the “control” construction to be indicative of some larger constraint in the language against inter-clausal movement. The absence of long-distance movement then follows naturally. What remains is to find what aspect of Madurese grammar drives this restriction.

References


In Davies 2000, I provide evidence on the basis of Cebuano that the same type of restriction against inter-clausal movement may actually obtain in the Philippine languages as well. If this speculation/result holds up, it might be that all (or most) of the languages with the restriction on extraction from clauses to subject clauses are open to the same explanation as Madurese.
Appendix: Javanese

Javanese cleft wh-questions show the same attributes as the Madurese. The data included below parallel the Madurese data discussed in the main text, and the same arguments can be made as were made on the basis of Madurese. Following the number of the Javanese example is the Madurese example that it is related to.

(J1) = (5a) Kapan sing Siti ngira nek Marlena nulis buku?
    when REL S AV.think COMP M AV.write book
    ‘When did Siti think that Marlena wrote the book?’

(J2) = (8) Kapan sing di-kira Siti nek Marlena nulis buku?
    when REL OV-think S COMP M AV.write book
    ‘When did Siti think that Marlena wrote the book?’
    or ‘When did Siti think that Marlena wrote the book?’

(J3) = (11b) Nek Marlena nulis buku sing di-kira Siti.
    COMP M AV.write book REL OV-think S
    ‘That Marlena wrote the book was thought by Siti.’

(J4) = (12) Siti kandha nek Ali ng-antem asu.
    S say COMP A AV-hit dog
    ‘Siti said that Ali hit the dog.’
Against long movement in Madurese

\[(J5) = (13a)\] Nek Ali ng-antem asu di-kandha-kna Siti.
\[\text{COMP A AV-hit dog OV-say-KNA S}\]
‘Siti said that Ali hit the dog.’

NB: In these data (from Surabayan Javanese), \(-k(n)\)a is equivalent to Madurese \(-ah\)gi. In Central Javanese this is realized as \(-k(\text{a})\)ke, while still other speakers use \(-k(\text{e})\)ne.

\[(J6) = (13b)\] *Nek Ali ng-antem asu di-kandha Siti.
\[\text{COMP A AV-hit dog OV-say S}\]
(Siti said that Ali hit the dog.)

\[(J7)\] Piye sing Siti kandha nek Ali ng-antem asu?
how REL S say COMP A AV-hit dog
‘How did Siti say that Ali hit the dog?’

NB: This is grammatical in Javanese since kandha usually occurs without AV morphology. However, this is a question about the saying event not the hitting event.

\[(J8) = (15b)\] Piye sing di-kandha-kna Siti nek Ali ng-antem asu?
how REL OV-say-KNA S COMP A AV-hit dog
‘How did Siti say that Ali hit the dog?’

\[(J9) = (18)\] Siti ngandha-kna Ali nek (dheweke) ng-antem asu.
\[S \text{ AV.say-KNA A COMP he AV-hit dog}\]
‘Siti said about Ali that he hit the dog.’

\[(J10) = (20)\] Ali di-kandha-kna Siti nek (dheweke) ng-antem asu.
\[A \text{ OV-say-KNA S COMP he AV-hit dog}\]
‘Siti said about Ali that he hit the dog.’

\[(J11) = (22)\] Piye sing Ali di-kandha-kna Siti nek (dheweke) ng-antem asu?
how REL A OV-say-KNA S COMP he AV-hit dog
‘How did Siti say about Ali that he hit the dog?’

\[(J12) = (23)\] Kapan sing Marlena nulis buku di-kira Siti?
when REL M AV.write book OV-think S
‘When did Siti think that Marlena wrote the book?’

\[\text{COMP B leave AV.sad-KNA I}\]
‘That Bambang left made Ita sad.’
(J14) = (27) Kapan sing Bambang mangkat nyosah-na Ita?
   when REL B leave AV.sad-KNA I
   ‘When did that Bambang left make Ita sad?’

(J15) = (28) Kapan sing nyosah-na Ita nek Bambang mangkat?
   when REL AV.sad-KNA I COMP B leave
   ‘When did it make Ita sad that Bambang left?’

(J16) = (34b) Di-kira Siti nek Marlena nulis buku.
   OV-think S COMP M AV.write book
   ‘Siti thinks that Marlena wrote the book.’

(J17) = (36) Buku iku [ sing di-tulis Marlena ] di-waca Siti.
   book that REL OV-write M OV-read S
   ‘That book that Marlena wrote was read by Siti.’

(J18) = (37) Buku iku di-waca Siti [ sing di-tulis Marlena ].
   book that OV-read S REL OV-write M
   ‘That book was read by Siti that Marlena wrote.’

(J19) = (39) Karo telpun Siti ngandha-kna piye sing Ita lunga nang Jakarta nang Ali?
   with telephone S AV.say-KNA how REL I go to J to A
   ‘How by telephone did Siti say to Ali that Ita went to Jakarta?’

(J20) = (40) Karo telpun Siti ngandha-kna piye nang Ali {nek/sing}
   Ita lunga nang Jakarta?
   with telephone S AV.say-KNA how to A COMP/REL I go to J
   ‘How by telephone did Siti say to Ali that Ita went to Jakarta?’
Vowel shifting and cloning in Motlav:
historical explanation vs. formal description

Alexandre FRANÇOIS
LACITO-CNRS, Paris
Université Paris-Sorbonne

Abstract

In his first account of Motlav grammar, Codrington (1885), followed by Kasachérou (1962), described the phonological rule of ‘vowel shifting’: a number of prefixes change their vowel into that of the next lexeme, e.g. na- + vôy > nû-vôy (‘volcano’). We first show that this rule only accounts for half of the lexicon, namely CV- roots, whereas for CCV-roots no change occurs, leaving an unvarying vowel instead.

We then discuss a diachronic hypothesis in order to account for these two distinct sets of lexemes: stress rules in former stages of the language first brought about a morphological alternation between two kinds of roots (CV- vs. CCV-). Eventually, this alternation was reinterpreted as reflecting a phonological constraint just emerging from inherited forms; as a consequence, the whole system is currently undergoing some standardization.

Finally, we provide a synchronic reanalysis of these recently emerged rules, thanks to newly defined theoretical tools. For instance, our choice of a multi-linear approach allows us to take into consideration the notion of ‘floating phoneme’, in order to account for the behavior of these prefixes; and the same notion appears to be efficient in analyzing ‘vowel transfer’ (e.g. na- + hinag > ni-hnag), another phonological rule of the same language.

Motlav is an Oceanic language spoken in northern Vanuatu (Banks Is), on the island of Motalava, by approximately 1800 speakers. It is geographically and historically close to Mota, a more conservative language which was first described in detail by Reverend Codrington (1885; 1896); this author also gave a first account of the grammar of Motlav (1885: 311-323).

Contrary to Mota, the morphology of Motlav is quite difficult, and seems to obey complex segmental and phonotactic rules, leaving bundles of lexical exceptions. Our aim is to show that this apparent complexity may in fact be reduced to a small set of rules and constraints, which mostly affect the quality and position of vowels within the word.

Yet, such a simplification of the analysis requires some strengthening or redefinition of our theoretical tools, which form the framework of our analysis; we will then come across such notions as syllabic template, floating vowels, and hierarchy of cognitive operations.

1 An earlier but more detailed version of this article was published (François 1999), relating the morphology of vowels in Motlav to dictionary-making issues.
1. Phonological outline of Motlav

1.1 Consonants

Motlav contains 16 distinct consonants, which may appear in any position of the linguistic chain – i.e. syllable-initial or final. Their phonological value is given below, together with their spelling when it is not obvious.

<table>
<thead>
<tr>
<th>voiceless stops</th>
<th>prenas. voiced stops</th>
<th>fricatives</th>
<th>nasals</th>
<th>lateral</th>
<th>glides</th>
</tr>
</thead>
<tbody>
<tr>
<td>labiovelar</td>
<td>bilabial</td>
<td>alveol.</td>
<td>velar</td>
<td>glottal</td>
<td></td>
</tr>
<tr>
<td><em>p</em></td>
<td><em>t</em></td>
<td><em>k</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>b</em></td>
<td><em>d</em></td>
<td><em>g</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>m</em></td>
<td><em>n</em></td>
<td><em>ŋ</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>w</em></td>
<td><em>l</em></td>
<td><em>y</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>v</em></td>
<td><em>s</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>l</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This information is being given here for the reader’s convenience; but it may be useful to underline the fact that, synchronically speaking, there is no morphological rule in Motlav involving the quality or position of consonants, such as assimilation.

Let us only underline the fact that the two glides /w/ and /y/ strictly behave as consonants in this language: they never form diphthongs, and follow the same rules as other Cs when filling in the syllable template. Note also that the phoneme /v/ surfaces as [p] at the end of a syllable, and this phonetic allophone is spelled p in the orthography: e.g. na-pno ‘country’ corresponds phonologically to /na-vnø/.

1.2 Vowels

However complex the rules involving vowels may be, the vowel system in itself is quite simple, since it contains no diphthongs, nor long or nasalized vowels, but instead seven oral, short vowels, represented below.

```
i u
I 1  á
I 1 1  á
```

Codrington (1885: 311) only mentioned six vowels, while other analyses (Crowley forthcoming; Kasarhérou 1962) proposed a seven-element system, but with different vowel qualities. In another article (François 1999: 443), we have claimed that the two mid-high vowels should be better described as being [-ATR], both on phonetic and phonemic grounds: the [ATR] feature is involved in the morphological process of vowel harmonization, which we will not deal with here.

---

2 In his grammar sketch of Motlav (“Mwotlap”), Terry Crowley (forthc.) considers the phoneme /p/, although it shows no phonological contrast with /v/.
1.3 Phonotactics

Motlav only accepts one type of syllabic pattern, i.e. \( l(C)V(C) \). Since both consonants are optional here, the phonology of Motlav actually contains four patterns: \( V, CV, VC, CVC \); but the essential point here is to underline that there is no allowance for consonant clusters within a syllable, and that \(-C_1C_2-\) sequences may only show at the syllable boundary.

As a consequence, a CCV- morpheme will only surface as such if its first consonant can attach to a preceding open-syllable (CV-) prefix. For example, the radical \( mtiy \) 'sleep' will stay unchanged in \( ni-mtiy \) '(s/he) sleeps', because it may be segmented into two CVC syllables \( lnimtiyl \); whereas an unprefixed form, as the one for imperative 'Sleep!', will have to undergo a process of vowel epenthesis, with the vowel of the radical cloning between the first consonant cluster: \(*mtiy \rightarrow Mtiy!\), segmented \( lnmtiy \). We won't detail this rule here\(^3\), but understanding it gives a first notion of the kind of phonotactic constraints defining Motlav morpho-phonology.

We can characterize the notion of phonological word in Motlav, as a segmental unit including one or several morphemes, which are together bound to a strict \([CVCl... lCVC]\) pattern. Incidentally, this definition makes it possible to distinguish prefixes from clitics on phonological grounds: a prefix is integrated to the phonological word, and as such takes its position inside the syllabic pattern thus defined – e.g. \( le-\) is a prefix in \( lle-plno\) 'in the country', without epenthesis. Whereas a clitic only integrates into the stress-defined word\(^4\), but never into the phonotactic word: e.g. \( ne\) is a clitic in \( lle-nelnol\) 'of the country', because it remains without the word boundary defined on phonotactic grounds.

2. Distribution of vowels within the word

The four main phonological rules involving vowels in Motlav are designated below:

- **Vowel Harmonization:** \( iplu-k \) 'my fellow' \( \rightarrow \delta pl\-n \) 'his fellow';
- **Vowel Epenthesis:** \( mtiy \rightarrow mtiy \) 'coconut';
- **Vowel Shifting:** \( na-+wolv \rightarrow n\-wolv \) 'moon';
- **Vowel Transfer:** \( na-+hnag \rightarrow ni-\hnag \) 'yam'.

The first two rules have already been alluded to, and won't be detailed here; this paper will deal mostly with vowel shifting, and will eventually present the operation of vowel transfer (§5).

2.1 Evidence for vowel shifting

The phenomenon of vowel shifting was first acknowledged by Codrington (1885: 311), and focused upon by Jacqueline Kasarhérou (de la Fontinelle), in her short presentation 'Les changements vocaliques de trois préfixes en motlav' (1962). Both authors presented this rule as involving a series of prefixes consisting of one single consonant (C-): e.g. the noun article \( n-\), the verbal prefix \( n-\) (considered by Kasarhérou to be the same as the article), the perfect marker \( m-\), would either surface as such (C-) before a root beginning with a \( V\), or else would borrow their vowel from the root, through a cloning / shifting operation:


4 Motlav words receive stress on their last syllable: see fn.9.
Table 1 - Rules for vowel shifting, with noun article n-

<table>
<thead>
<tr>
<th>radical</th>
<th>n- + radic.</th>
<th>radical</th>
<th>article+rad.</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>V_r-</td>
<td>n-V_r-</td>
<td>ulsi</td>
<td>n-ulsi</td>
<td>summit</td>
</tr>
<tr>
<td>CV_r-</td>
<td>nV_r-ICV_r-</td>
<td>bé</td>
<td>né-bé</td>
<td>water</td>
</tr>
<tr>
<td>CCV_r-</td>
<td>nV_r-CCV_r-</td>
<td>qti</td>
<td>ni-qti</td>
<td>head</td>
</tr>
</tbody>
</table>

Further evidence of this phenomenon include the following items, among thousands. The first rows involve the noun article (supposedly n-), the last one uses a numeralizer of the form v-:

Table 2 - Sample evidence for vowel shifting

<table>
<thead>
<tr>
<th>ni-git</th>
<th>house</th>
<th>né-sém</th>
<th>money</th>
<th>ne-get</th>
<th>taro</th>
</tr>
</thead>
<tbody>
<tr>
<td>no-gom</td>
<td>disease</td>
<td>nö-vöy</td>
<td>volcano</td>
<td>nu-kumay</td>
<td>sweet potato</td>
</tr>
<tr>
<td>na-lař̩</td>
<td>(a) fly</td>
<td>ni-qi-t-</td>
<td>my head</td>
<td>né-qte-n</td>
<td>his head</td>
</tr>
<tr>
<td>vœ-vœ</td>
<td>two</td>
<td>ve-vet</td>
<td>four</td>
<td>vé-vœh</td>
<td>how many?</td>
</tr>
</tbody>
</table>

This insertion of a shifted vowel can be analyzed as an epenthesis rule to avoid consonant clusters at the beginning of a word. This is exactly what happens with another of the phonological rules in Motlav, which we have labeled ‘vowel epenthesis’: a CCV- root like mtig ‘coconut’ needs to copy its vowel into the consonant cluster, to avoid it at the beginning of a word (mtig → mhtub). This recalls also what happens in Kalam, a non-Austronesian language from Papua New Guinea (Pawley 1993: 91):

In the context C_CVC, the release vowel may be a very short, unstressed near copy of V or a short, unstressed central or high central vowel, e.g., mleñ ‘dry’ is [mleñ] or [mleñ].

In Motlav, no central vowel is heard, and the epenthetic vowel is always clearly a clone of the following V. These prefixes "seem not to have a proper vowel, but instead to borrow it from the radical" (Kasarhërou 1962). This evidence should suffice for positing suffixes of the form C-, i.e. with a single consonant, and no specific vowel.

2.2 Exceptions to vowel shifting

Despite the obviousness of the latter analysis, we have to face a quantity of irregular forms, whose prefix vowel is different from that of the radical. A few of them are presented in Table 3, where irregular forms are underlined:

Table 3 - Some exceptions to vowel shifting

<table>
<thead>
<tr>
<th>+ article</th>
<th>meaning</th>
<th>+ locative</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>na-naw</td>
<td>sea</td>
<td>le-naw</td>
<td>in the sea</td>
</tr>
<tr>
<td>na-sñal</td>
<td>sea</td>
<td>le-sñal</td>
<td>in the sea</td>
</tr>
<tr>
<td>na-he-k</td>
<td>my name</td>
<td>le-he-k</td>
<td>my name</td>
</tr>
<tr>
<td>na-gmel</td>
<td>men’s house</td>
<td>le-gmel</td>
<td>in the men’s house</td>
</tr>
<tr>
<td>na-lo</td>
<td>sun</td>
<td>le-lo</td>
<td>in the sun</td>
</tr>
<tr>
<td>na-pññë</td>
<td>country, island</td>
<td>le-pññë</td>
<td>in the country</td>
</tr>
</tbody>
</table>

We have chosen only a few radicals among those which show at least one irregular form; each radical is presented here with two different prefixes, the noun article n- and the loca-
tive preposition -1-. A regular pair for both prefixes would be nō-vōy ‘volcano’ - lō-vōy ‘on the volcano’. Table 3 does show regular forms too, like na-naw or le-gmel; but the rest is not, and falls among the ‘exceptions’ pointed at by both Codrington and Kasarhérou.

The trouble is, such ‘exceptions’ are not few, but represent up to 50% of the data. Such observation obviously needs to be accounted for, unless the rule for vowel shifting appears to be just a weak random trend in Motlav phonology. In fact, another sort of regularity appears in Table 3, provided it is not read in rows, but in columns. For a given prefix, all the irregular forms will show the same vowel; e.g. every noun preceded by the article n-, will have as a first vowel either a copy of the radical vowel (e.g. nō-vōy), or the vowel lal (e.g. na-gmel, na-lo); as for the locative -1-, the ‘default vowel’, as it may be called, is an lal (e.g. le-naw, le-lo).

2.3 Shifting prefixes vs. unvarying prefixes

Evidently, Motlav does not have just one neutral vowel which could be assigned at the system-level, like Indonesian Pendau does with /o/ (Quick 2000); instead, each shifting prefix is provided with its own ‘default vowel’. The quality of this vowel is assigned to the prefix in the lexicon, and cannot be predicted through any phonological rule: it is thus necessary to represent it in the citation form of each morpheme. From now on, we will choose to speak about the noun article na- or the locative preposition le-, instead of n- and -1- respectively, since this is the only way to know what their default vowel is. This is a first difference with former accounts which were given of Motlav phonology, since we now consider shifting prefixes to be of the form CV-, with both C and V being specific.

Nevertheless, one improvement has to be made in our representation of these CV-prefixes. Talking about an article na- would make it necessary to posit an extra rule of deletion / assimilation for this /a/ before certain radicals, e.g. na- + qōh ~ nō-qōh ‘night’. This rule would have a form like

$$ < C_0 V_o + C_1 (C_2) V_1 C_3 ... \rightarrow C_0 [V_1] C_1 (C_2) V_1 C_3 ... > $$

However, an argument against such a generalization is given by other prefixes of the form (C)V-, which never change their vowel:

- verb prefix ni- (3rd singular + present): e.g. ni-in ‘drinks’, ni-et ‘sees’, ni-van ‘goes’, ni-gen ‘eats’, ni-qōh ‘(it) becomes night’;

As a consequence, vowel shifting in Motlav cannot be described by the general rule stated above. Some CV- prefixes are unvarying, while others do show a ‘weak’ vowel, which is likely to assimilate to the vowel of the radical. This is why the best representation is to oppose these two kinds of CV- prefixes, typographically, by contrasting:

- prefixes of the form /CV-/ which never shift their vowel, e.g. ni- (3SG:PST); they will be called ‘unvarying prefixes’;
- prefixes of the form /CVC-/ which are sensitive to vowel shifting, e.g. nA- (noun article); they are the ‘shifting prefixes’.
Alexandre François

The uppercase letter codes for the special status of this ‘default vowel’, i.e. a vowel which sometimes surfaces and sometimes does not; the exact theoretical status of this vowel will be the main issue of §4. Motlav has eight shifting prefixes, which are listed below:

Table 4 - The eight vowel-shifting prefixes of Motlav

<table>
<thead>
<tr>
<th>FORM</th>
<th>MEANING</th>
<th>PREFIXED TO</th>
</tr>
</thead>
<tbody>
<tr>
<td>nA-</td>
<td>Article (‘a / the’)</td>
<td>nouns</td>
</tr>
<tr>
<td>bE-</td>
<td>Purposive (‘for’)</td>
<td></td>
</tr>
<tr>
<td>lE-</td>
<td>Locative (‘in’)</td>
<td>verbs,</td>
</tr>
<tr>
<td>mE-</td>
<td>Perfect</td>
<td>adjectives,</td>
</tr>
<tr>
<td>nE-</td>
<td>Stative</td>
<td>(nouns)</td>
</tr>
<tr>
<td>tE-</td>
<td>Future</td>
<td>locatives</td>
</tr>
<tr>
<td>vE-</td>
<td>Ablative</td>
<td>numbers</td>
</tr>
</tbody>
</table>

Incidentally, notice that some of these prefixes can combine. This is mainly the case for tE- (‘Ablative’), which is prefixed to locative words, be they directly locative —e.g. to-Motlap ‘from Motalava’— or derived by means of the preposition lE- (‘Locative’), thus bringing about a sequence tE-lE-. In this case, the preposition which is adjacent to the root (lE-) is sensitive to vowel shifting, and its vowel is always reproduced on the first prefix (lE-):

\[
tE- + lE- \rightarrow tE-lE- \rightarrow \text{‘from the house’}
\]

\[
tE- + lE- + voy \rightarrow tE-lE-voy \rightarrow \text{‘from the volcano’}
\]

\[
tE- + lE- + naw \rightarrow tE-lE-naw \rightarrow \text{‘from the sea’}
\]

At this stage of the presentation, it is impossible to know whether the vowel /e/ on the first prefix (te-) is an instance of its own ‘default vowel’ —which would mean that no vowel shifting occurred at all— or if it is a copy of the following vowel —which means that vowel shifting only affects the second prefix (lE-). Further demonstration will show the second assumption to be correct.

2.4 Penneable vs. blocking lexemes

One of the conditions for vowel shifting, obviously, is that the prefix must belong to the limited list of shifting prefixes (Table 4). But what about the radical? Is there any restriction on the set of radicals that can shift their vowel, and is that restriction systematically predictable?

Table 3 above had already provided a set of exceptions to the vowel-shifting process. However, only the underlined forms (e.g. le-naw) were supposed to be irregular, while other forms, including for the same radical (e.g. na-naw), were considered to follow the rules. But now that we know the underlying form of each prefix, it becomes ambiguous whether the first /a/ in na-naw is really a clone of the vowel in the radical (naw), or if it is the default vowel of the noun article (nA-). Now, there is strong evidence pointing at the latter solution: whenever a given lexical root blocks the vowel-shifting process with one prefix, it does so with all other prefixes. Conversely, whenever a lexeme shifts its vowel to one prefix, it will do so with any other (shifting) prefix. Table 5 gives further evidence of this claim:
Vowel shifting and cloning in Motlav

Table 5 - Permeable vs. blocking lexemes

<table>
<thead>
<tr>
<th>permeable lexemes</th>
<th>trans.</th>
<th>bare root</th>
<th>article nA-</th>
<th>Pp bE-</th>
<th>Stative nE-</th>
<th>Perfect mE-</th>
</tr>
</thead>
<tbody>
<tr>
<td>cold</td>
<td>momiyy</td>
<td>no-momyy</td>
<td>bo-momyy</td>
<td>no-momyy</td>
<td>mo-momyy</td>
<td></td>
</tr>
<tr>
<td>night</td>
<td>qôh</td>
<td>né-qôh</td>
<td>bô-qôh</td>
<td>nô-qôh</td>
<td>mô-qôh</td>
<td></td>
</tr>
<tr>
<td>think</td>
<td>dêmdêm</td>
<td>né-dêmdêm</td>
<td>bê-dêmdêm</td>
<td>nê-dêmdêm</td>
<td>mê-dêmdêm</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>blocking lexemes</th>
<th>trans.</th>
<th>bare root</th>
<th>article nA-</th>
<th>Pp bE-</th>
<th>Stative nE-</th>
<th>Perfect mE-</th>
</tr>
</thead>
<tbody>
<tr>
<td>work</td>
<td>mgungu</td>
<td>na-mgungu</td>
<td>be-mgungu</td>
<td>ne-mgungu</td>
<td>me-mgungu</td>
<td></td>
</tr>
<tr>
<td>sleep</td>
<td>mntntiy</td>
<td>na-mntntiy</td>
<td>be-mntntiy</td>
<td>ne-mntntiy</td>
<td>me-mntntiy</td>
<td></td>
</tr>
<tr>
<td>want</td>
<td>myôs</td>
<td>na-myôs</td>
<td>be-myôs</td>
<td>ne-myôs</td>
<td>me-myôs</td>
<td></td>
</tr>
</tbody>
</table>

This means that all lexical items of Motlav, without exception, fall into either of two morphological categories, defined in relation with the vowel-shifting process. It is a 'permeable' lexeme, if it allows its first vowel to be copied onto the preceding shifting prefix: e.g. qôh → nô-qôh, bô-qôh, mô-qôh, but never *na-qôh. Conversely, a lexeme will be labeled 'blocking' if it is incompatible with vowel shifting; in this particular case, the prefix can only take its own default vowel: e.g. myôs → na-myôs, be-myôs, me-myôs, but never *nô-myôs. The feature [permeable] vs. [blocking] is carried by each lexeme in the lexicon - unless it can be predicted by the theory; the latter issue will be discussed in 2.5.

At this point of the discussion, two remarks may be made. First, this rule definitely helps answer the question about the nature of the vowels in a word like na-naw 'sea' (Table 3). Since we are no longer dealing with irregular 'forms', but with irregular lexemes, then a word like le-naw 'in the sea', which has blocked the vowel-shifting process, proves the root naw belongs to the set of 'blocking lexemes'; as a consequence, the a of the article on na-naw cannot be due to vowel assimilation, and is necessarily an instance of the default vowel of the prefix. The same logic applies to another ambiguous form like le-gmel 'in the men's house': since the test with the article nA- gives na-gmel instead of *ne-gmel, then it becomes obvious that the root gmel is blocking, and that the first e in le-gmel comes from the prefix itself. Thanks to this test, all the ambiguous forms of Motlav can easily be solved.

Second point: only blocking lexemes can show the underlying form of each prefix; and this turns out to be essential even to morphosyntactic analysis, since they help distinguish between the noun article nA- and the stative prefix nE-. As is clear in Table 5, these two prefixes always have the same surface forms when they are combined to permeable roots. This point led J. Kasarhérout to a confusion, when she presented a prefix n- as an "actualizing prefix" ('préfixe actualisant') which is "compatible with all full words of this language", regardless of their syntactic category. In fact, deeper morphological evidence reveal that there are two distinct morphemes, one of which is compatible exclusively with nouns (nA-), whereas the other mainly goes with adjectives and stative verbs (nE-). Far from challenging the noun-verb opposition, this pair of prefixes confirms that it is quite strong in Motlav.

2.5 Is the [blocking] feature predictable?

In summary, we have demonstrated that the process for vowel shifting only occurs in one case, i.e. the combination of a permeable lexeme + a shifting prefix:

5 Table 4 suggests that nE- is also compatible with nouns, though this is only true of a dozen items. On the other hand, the fact that all nouns be predicative (if they bear the article nA-) is not a sufficient argument to say that verbs and nouns are not separated in Motlav; and the existence of a few 'precategorial roots' like the ones presented in Table 5 (e.g. qôh, myôs...) is not a good argument either, as long as the major part of the syntax continues to contrast nouns and verbs.
Table 6 - Combination <permeable lexeme + shifting prefix>

<table>
<thead>
<tr>
<th>Unvarying prefix (ni-)</th>
<th>Permeable lexeme (mi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- (ni-mi)</td>
<td>- (ni-qo)</td>
</tr>
<tr>
<td>Shifting prefix (me-)</td>
<td>- (me-mi)</td>
</tr>
<tr>
<td></td>
<td>+ (me-qo)</td>
</tr>
</tbody>
</table>

Now, the next question would be to find out whether this feature of the lexeme [blocking] vs. [permeable] is predictable from the form of the root, or not. This would make it clearer whether vowel shifting belongs to the domain of phonology—if it is predictable—or of morphology—if it is not. A first answer to our question is ‘No’, i.e. the form of a root is not sufficient to predict with certainty if it will shift its vowel or not. The evidence for this claim is the following (morphological) minimal pair:

- nê-lêt ‘pudding’: permeable lexeme lêtl
- na-lêt ‘firewood’: blocking lexeme rêtl.

Two homophonous roots lêtl show two divergent behaviors in relation with vowel shifting: one is permeable, while the other belongs to the set of blocking lexemes. Although there are few minimal pairs like this one, it should be enough evidence to draw the following conclusion: It is not 100% possible to predict the feature [± blocking] from the root itself; its value is assigned to each root within the lexicon. In our presentation of Motlav lexemes, a blocking root will be preceded with a small symbol (°).

However, though there may not be a systematic rule for this prediction, we can find at least a very strong tendency in Motlav lexicon. It appears that permeable lexemes are generally roots beginning with only one consonant, whereas blocking lexemes normally begin with two. Table 7 illustrates this contrast between ‘CV roots’ and ‘CCV roots’:

Table 7 - Regular correlation between root structure and vowel shifting

<table>
<thead>
<tr>
<th>Permeable lexemes: CV-</th>
<th>Blocking lexemes: CCV-</th>
</tr>
</thead>
<tbody>
<tr>
<td>vis</td>
<td>*dyê</td>
</tr>
<tr>
<td>lêt</td>
<td>*tôyig</td>
</tr>
<tr>
<td>siseg</td>
<td>*myôs</td>
</tr>
<tr>
<td>véyttit</td>
<td>*môti</td>
</tr>
<tr>
<td>yô</td>
<td>*vnô</td>
</tr>
<tr>
<td>vap</td>
<td>*hôyo</td>
</tr>
<tr>
<td>Motlap</td>
<td>*blekat</td>
</tr>
</tbody>
</table>

This table means that all CV- roots (first column) are regularly permeable, i.e. allow their first vowel to copy onto the prefix: e.g. ni-vis ‘owl’, vé-yô ‘two’, to-Motlap ‘from Motlapava’; conversely, CCV- lexemes normally block the process of vowel shifting, thus being always associated with the default vowel of the prefix: e.g. nE- + *tôyig → ne-tôyig ‘is easy’; nA- + *vnô → na-pnô ‘country, island’.

2.6 A new sort of exceptions

This list could go on with thousands of lexemes: the tendency we have just mentioned is true for 100% of the verbs, adjectives and numerals, and about 95% of the nouns. In this
case, it makes sense to speak about a phonological rule, for which there are a certain number of exceptions (<5% of the nouns). But it must be clear that we are not following the same logic as previous authors, such as Codrington and Kasarhérou, who considered as ‘exceptions’ all roots which did not shift their vowel. These roots, which represent half of the lexicon, are now integrated in a newly-defined rule, under the name of ‘blocking roots’.

Now, what we consider to be exceptions, is a much smaller set of nouns, which do not correspond to the correlation between phonotactic structure of the root (CV- vs. CCV-) and compatibility with vowel shifting. These exceptions can be either permeable roots, which have the unexpected form CCV- (a dozen nouns); or blocking roots, which start with only one consonant (up to forty nouns). The most common of these irregular lexemes are presented in Table 8:

Table 8 - A few exceptional nouns: CCV- permeable & CV- blocking roots

<table>
<thead>
<tr>
<th>permeable lexemes: CCV-</th>
<th>blocking lexemes: CV-</th>
</tr>
</thead>
<tbody>
<tr>
<td>gen-</td>
<td>'he-</td>
</tr>
<tr>
<td>vni-</td>
<td>'lo-</td>
</tr>
<tr>
<td>qafi-</td>
<td>'let</td>
</tr>
<tr>
<td>tge-</td>
<td>'to-</td>
</tr>
<tr>
<td>hyayyu-</td>
<td>'veh</td>
</tr>
<tr>
<td>blet-</td>
<td>'hōnū</td>
</tr>
<tr>
<td>skul-</td>
<td>'lo-</td>
</tr>
</tbody>
</table>

For instance, the word for ‘garden’ tge, with the article nA-, does not give the expected *na-tge, but an unpredictable form ne-tge; in the other direction, the word for ‘mountain’ is not *no-tō, as would be normal for a CV- root, but na-tō. Incidentally, we notice that among the morphological pair ne-lēt / na-lēt mentioned above, the first root is regular, whereas the second one belongs to the set of exceptional blocking lexemes (see Table 8).

The origin of most of these exceptions will be discussed in §3.3. What we would like to develop right now, is a historical hypothesis which would account for the major morphological split between permeable and blocking roots, and for its regular correlation with phonotactic structure. After this diachronic approach, we will focus on a synchronic representation of vowel shifting in Motlav.

3. A diachronic account for vowel shifting in Motlav

3.1 Historical stress and syllable loss in Motlav

The question is: what historical processes can account for vowel shifting in Motlav; and especially, why do CCV- roots systematically block this rule, while only CV- roots allow it? To answer this question, it is useful to remember that former linguistic stages of Motlav followed Proto Oceanic in having mainly open syllables of the form (C)V (Ross 1998: 17). The closed (CVC) syllables of Motlav are obviously an innovation, and were evidently formed through the loss of unstressed vowels. Basic stress rules of pre-Motlav included a primary stress on the penultimate syllable of the phonological word, and secondary stresses on every second syllable towards the left of the word, e.g. *Mōtalāva.

In more recent times, all post-tonic syllables were deleted, causing closed (CVC) syllables to appear. Here are a few place names:
These examples show how words with four open (CV) syllables were eventually reduced to two closed (CVC) syllables, or from six to four. Other words shorten from two to one syllable (*CVCV > CVC); other cases will be discussed below. In all these examples, only stressed syllables were retained in Motlav, while un stressed post-tonic vowels were definitel y lost.

3.2 Historical stress and vowel shifting

Now, as far as prefixes are concerned, there were two possibilities. In a given prefixed word, either the stress would strike this prefix, or it would strike the following syllable – in which case the prefix would be in a pre-tonic position. The first case can be illustrated by a phrase in Pre-Motlav meaning ‘(the) country’, *nda vania (< POc *panua); in Motlav we have:

*nda vania > na-pnô ‘country, island, village’

This example shows that when the prefix was accented, it maintained the quality of its own vowel, without any copy occurring: in modern Motlav, this corresponds to what we have been calling the ‘default vowel’ of the prefix. On the other hand, since the first vowel of the radical was in a post-tonic position, it regularly lost its vowel, thus defining a new root of the form CCV-. This hypothesis is the most powerful to explain why CCV- roots systematically block the process of vowel shifting: the genesis of such roots implied necessarily a stressed prefix, and hence no vowel assimilation at all.

The second case, i.e. the unaccented prefix, occurred whenever the phonological word (including the prefix) had an odd number of syllables, e.g.

*na vânâ-ña > nê-véna-n ‘his country’

In this situation, the stress on the first syllable of the root (*-va-) caused it to survive in the modern form (-vê-), and this evolution is the origin of all CV- roots in Motlav. On the other hand, the unaccented (pre-tonic) vowel of the prefix, after maybe becoming a schwa, proved its articulatory weakness in assimilating totally to the next syllable. Notice that the total deletion of the prefix vowel (*nvênan) was impossible, because of the phonotactic

---

6 Notice that these two words are considered by speakers to be totally distinct from one another; the translation of ‘his country’ uses now the alienable pattern na-pnô no-no-n, and the word nê-véna-n is a poetic word for ‘his fatherland’. We have discovered up to eight etymological ‘doublets’ of this kind in Motlav (François 1999: 456), all of which are opaque to the speakers.

7 The syllable itself remains, but the quality of the vowel may change, according to the one in next syllable. Among the many possible combinations, our examples involve *dù u > e (word-internally), *dù u > e (word-finally); *di i > e; *dii a > ə; *ti e > ə; *tii i > ɨ; *tii a > ɘ; *tii o > ə.
Vowel shifting and cloning in Motlav

rules applying in Motlav (see §1.3). In a way, vowel shifting could be considered as a special case of vowel epenthesis involving prefixes: an underlying form like *n-vēna-n surfaces as nē-vēna-n, following *mitty → mitty 'sleep'.

Further evidence of both evolutions are presented in Table 9.8

Table 9 - Former stress accounts for vowel shifting

<table>
<thead>
<tr>
<th>POc etymon</th>
<th>Pre-Motlav</th>
<th>Motlav</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>blocking lexemes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*kasupe</td>
<td>*nā gasūwe</td>
<td>na-ghōw</td>
<td>rat</td>
</tr>
<tr>
<td>*kurita</td>
<td>*nā wurīta</td>
<td>na-wyēt</td>
<td>octopus</td>
</tr>
<tr>
<td><em>tob</em>a</td>
<td>*nā toqa-ku</td>
<td>na-tge-k</td>
<td>(my) bell</td>
</tr>
<tr>
<td>*daraq</td>
<td>*nā darā-ku</td>
<td>na-nye-k</td>
<td>(my) blood</td>
</tr>
<tr>
<td>permeable lexemes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*bawa</td>
<td>*na qōe</td>
<td>no-qo</td>
<td>pig</td>
</tr>
<tr>
<td>*piraq</td>
<td>*na vīa</td>
<td>nē-vē</td>
<td>k.o. taro</td>
</tr>
<tr>
<td>*suri</td>
<td>*na sūri</td>
<td>nī-hī</td>
<td>bone</td>
</tr>
<tr>
<td>*kawil</td>
<td>*na gāu</td>
<td>ne-ge</td>
<td>fish-hook</td>
</tr>
</tbody>
</table>

To sum up, we have demonstrated that the process of vowel shifting is historically a consequence of former stress: a radical would become 'permeable' whenever the word stress would strike its first syllable, leaving its prefix unstressed. This historical explanation accounts for the correlation, on the one hand, between permeable roots and CV- structure, and on the other hand, between blocking roots and CCV- structure in modern Motlav.10

3.3 Accounting for exceptions

Our hypothesis proves its explanatory power in accounting not only for regular processes, but also for many exceptions. Several (modern) CV- roots which unexpectedly block vowel shifting, actually used to bear stress on their very prefix, but have recently lost an extra mora. This happened when both consonants of the root were identical, causing a geminate cluster to be shortened (*C,CV > C,V):

*na rerēša > *na-reṛa > *na-reṛ > na-yeṛ "turmeric"
*na lolō-na > *na-rollo-n > na-lo-n "the inside of it"
*na sasā-ku > *na-se-sek > *na-se-k > na-he-k "my name"

This happened, too, when either the first or the second syllable of the root happened to have no consonant (*V instead of *CV), thus bringing about a 'pseudo' CV root which contained an 'invisible' consonant. Examples of [C]CV roots include:

8 Notice that we spell Pre-Motlav and Motlav according to the conventions shown in §1: thus q = [p]; i = [i]; a = [a]; u = [u]; e = [e]; o = [o]. In Mota, a language using the same conventions, the lexical items of Table 9 are gasuwe, wērīta, toqa-k, qoq, via, sur, gau, with no article (Codrington 1896).
9 Modern Motlav words are systematically accented on the last syllable, not on the penultimate, contrary to what has been stated in other articles (Crowley forthcoming). The evolution of stress from penultimate to final syllable is easy to understand from the deletion rules presented here; the same path was taken from Latin to French, e.g. L civitate(m) [kerti'tarte] > F cité [sit'et] 'city'.
10 More examples and reflections have been proposed in a former paper (François 1999).
Alexandre François

\*nci alítio > na-lēt ‘firewood’
\*nci alóa > na-lo ‘sun’
\*nci aňári > na-ňey ‘Canarium almond’
\*nci awía > na-ď ‘turtle’

while examples of C[C]V roots would be as follows:

\*nci tańwe > na-tô ‘mountain’
\*nci sańña > na-hôń ‘Wrasse fish’
\*ncia Roua > te-Yô ‘from Roua island’

3.4 Emergence of a phonological rule in synchrony

As would be expected, the same process happened to verb roots, bringing about blocking CV lexemes:

\*mé maúři > *me-mir ‘(it) lived, grew’ (POc *maqurip)

But remarkably, all resulting lexemes were felt to be exceptions to the regular (just emerging) correlation between phonotactics and vowel shifting: as a consequence, all verbs and adjectives underwent a morphological standardization, which created non-etymological permeable CV- lexemes:

\*me-mir > *mi-mir > mi-miy ‘(it) was in bud’

This process of standardization has been affecting all roots, except the most common nouns, which is not so surprising. However, the pressure of the phonological correlation here under discussion is still so strong, that even these exceptional noun roots are now beginning to conform to the norm, showing it to be still lively and productive. For instance, the ‘correct’ form te-Yô (see above) is sometimes heard to-Yô, despite the etymology.

Parallel to this ongoing standardization, it is worth noticing that loanwords are often—though not always—forced into the same correlation between phonotactic structure of the root and compatibility with vowel shifting. For example, a CV- words like doctor, when combining with nA- article, will have its vowel copied onto the prefix: no-dokta; conversely, the word policeman gets its first vowel deleted to form a CCV- root, which will block the process of vowel shifting: na-qlismen.

The latter evidence suggests that the historical explanation, however powerful it may be to account for present data, must always be completed by a synchronic presentation. Present-day speakers of any language do not just make a passive use of inherited paradigm, which would only find their logic in diachrony. Instead, each generation endeavors to figure out formal similarities and constant correlations out of the amount of forms they are supposed to be using. This is how productive rules emerge, either in phonology, morphology or semantics, in such a way that the diachronic path leading to the present situation is totally ignored. Because of functional constraints in memory and ease of processing, a new consistent system is built up, slowly but surely.

Most of the time, however, this kind of standardization process is still ongoing in the language, in such a way that only a part of the lexicon—normally the less commonly used—has already been conformed to the emergent rule, whereas several items are used frequently enough to resist for a longer period. This is what happens with about fifty nouns in Motlav, which have not yet undergone the same overall revision that verb or adjective roots have.
4. Synchronic account of vowel shifting: a multilinear approach

The aim of next paragraph is to give a synchronic description of the most productive rule for vowel shifting in modern Motlav.

4.1 Distinguishing tiers

In order to achieve a formal representation of vowel shifting in Motlav, it makes sense to distinguish between at least two layers, one for consonants and one for vowels. Indeed, it has already been shown that Motlav phonology allows only vowels, not consonants, to copy, assimilate or migrate from one place to another within the phonological sequence.

This idea recalls the presentation that Nick Clements made about Kolami, a Dravidian language in which such words as *kinik, suulup, melep, ayak*, are supposed to follow a "rule for propagation of vowel nodes" (Clements 1991). For this purpose, Clements makes use of what a famous article by McCarthy (1989) called planar V/C segregation, to account for data in some Semitic and American Indian languages. Here is Clements' presentation of McCarthy's conceptions:

> When [the template] is introduced derivationally, consonants link to it on one family of planes and vowels link to it on another (non-intersecting) one. At this point, consonants and vowels are entirely segregated in phonological representations, and are brought together only by the later process of tier conflation which 'folds' the consonant and vowel planes together.

Vowels and consonants organize into two different tiers, where they first follow their own specific rules; then both tiers eventually 'conflate' to conform to the *syllabic template* of the language, if there is one. Motlav template has already been presented in §1.3: the basic syllabic pattern of this language is ICVCI, with both Cs being optional.

Now, as far as vowel shifting is concerned, let us sum up the rule which affects more than 95 per cent of the lexemes:

- roots beginning with one C ('CV roots') are permeable, i.e. allow their vowel to migrate to a shifting CV- prefix;
- roots beginning with two Cs ('CCV roots') are blocking roots, i.e. prevent their vowel from migrating to the prefix.

An elegant way to represent this phonological process, would be to say that a "vowel node" can propagate to the left of the word boundary, provided it has only a single C to cross over; conversely, a cluster of two consonants behaves as a "blocking node" (Clements 1991), which hinders this vowel propagation.

Let us contrast the behavior of two nouns regarding vowel shifting: one is permeable *nA-+vøy → nō-vøy* 'volcano'; the other one is a blocking lexeme *nA-+mtig → na-ntig* 'coconut'. In the first case, the vowel of the radical surfaces not only in its own position, but copies also to the left; in the second case, a cluster of two Cs makes the cloning process impossible:
In the remainder of this paper, our main point of interest will be the theoretical nature of the ‘default vowel’ of the prefix. What kind of phoneme is this, which sometimes surfaces and sometimes doesn’t? The point is to escape from the feeling of having a deus ex machina phoneme which only exists when needed.

4.2 French ‘liaison’ and the notion of floating phoneme

Some help may be found thanks to a typological insight on similar processes in other languages. French ‘liaison’ is a rule which governs surfacing vs. non-surfacing of certain word-final consonants, according to the phonotactic structure of the following word. For example, the feminine form of adjective ‘small’ is spelt petite, phonologically /patit/ with a final /t/ which must always be pronounced—we shall call it a ‘true’ phoneme; two synonymous phrases meaning ‘my girlfriend’ are

- ma petite copine /mapotitanp/ 
- ma petite amie /mapotiam/.

Conversely, the masculine form petit shows a special /t/, which will surface only if the second word begins with a vowel: e.g. two phrases for ‘my boyfriend’ are

- mon petit copain /maptitokopian/ without a /t/, 
- mon petit ami /maptitami/ with a /t/.

The rule for liaison, which concerns the masculine form in the last example, has been described, in a multilinear framework (Encrevé 1988), as involving two fundamental notions:

- first, the notion of syllabic template, which consists of a (more or less constraining) string of Cs and Vs;
- second, the notion of floating phoneme, whose main property is to surface only on the condition that a slot has been left empty in the template, after other phonemes have taken their place.

In the case of our word petit, we can consider there is a floating consonant T at the end of the underlying masculine form /patit/ — vs. feminine /patit/. This means that this T will only surface if a Consonant slot has been left empty by other ‘true’ phonemes, within the syllabic template. The latter situation occurs when the next word begins with a vowel, thus leaving an empty C-slot, for the final T to slide into:
Vowel shifting and cloning in Motlav

4.3 Floating vowels in Motlav

Our aim is not here to discuss French liaison, but to show how relevant Pierre Encrevé's assumptions are for Motlav data. The behavior of this 'floating consonant' in French, which sometimes surfaces and sometimes does not, makes it indeed very similar to the 'default vowel' of shifting prefixes in Motlav. The vowel /a/ of noun article, and /e/ of other prefixes (see Table 4 in §2.3), can be described here as floating vowels. This means that these vowels will surface only if a V-slot is left empty in the syllabic template, once all other phonemes have been realized.

Three situations are possible:

(a) The V-slot has already been taken by a full vowel

This is typically the case when the radical begins with a V, e.g. ulsi 'summit':

```
CV.CV (C)
nulsi
```

All full-right phonemes start taking their position into the CVC\|\(CVC\) pattern; then no room is left for the floating vowel. This example helps underline an essential point regarding theory: a floating phoneme may take empty slots in a given syllabic template, but it cannot create one; thus *na-ulsi is excluded. This is precisely what opposes them to 'true' phonemes, which necessarily have a slot of their own. The following examples will help us further build on this theory.

(b) The floating vowel of the prefix is superseded by a vowel shifted from the root

Another situation is when the radical is 'permeable', i.e. normally starts with one consonant. In this case, the first V of the radical migrates to the prefix(es), and supersedes the floating vowel.

```
CV (C) | CV.C
n v y
```

What has happened here is best represented in terms of a chronology of cognitive operations:
1. consonants (here n-v-y) take their own slot into the syllabic template CVCICVC, leaving V-slots empty.
2. the (first) radical vowel (here ə), which is a true phoneme, automatically comes into its own slot, between v and y; only one slot is now left empty, the prefix V.
3. the lexeme vowel has priority over any other V, to fill in the empty V-slot on the prefix; yet, this happens only if it is allowed by phonotactic structures, i.e. if it is not hindered by a cluster of two Cs ('blocking node').
4. if the leftward propagation of the radical vowel was blocked during stage nb. 3, then the floating vowel eventually fills the empty V-slot (see below).

(c) The slot is free
If operation number 3 above has failed, then the vowel-slot on the prefix is still empty when the floating vowel comes in. This is the only way the ‘default vowel’ of each shifting prefix may surface:

<table>
<thead>
<tr>
<th>C</th>
<th>V</th>
<th>C</th>
<th>I</th>
<th>C</th>
<th>V</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>m</td>
<td>t</td>
<td>g</td>
<td>A</td>
<td>i</td>
<td></td>
</tr>
</tbody>
</table>

The notion of floating phoneme is therefore very useful in order to account for a vowel which belongs to the lexicon, but shows intermittently. The last section of this paper will confirm how useful this notion is to describe another phonological rule of Motlav, namely vowel transfer.

5. Vowel transfer: floating vowels on lexemes

Without going into too much detail, it is worthwhile looking at another phonological rule of Motlav, which we have labeled ‘vowel transfer’ (see first lines of §2). This rule involves the same eight prefixes which were listed in Table 4 (§2.3), but different lexemes, which are not concerned with vowel shifting. These lexemes can be either nouns, adjectives or verbs, but are not more than twenty in all: the process of vowel transfer is much more limited than what we have been studying so far.

5.1 A mobile and intermittent vowel
The principle of vowel transfer is that when a (shifting) prefix is added to one of these lexemes, the first vowel of the radical does not only copy to the prefix –which corresponds to ordinary vowel shifting–, but also has a rule for deletion:

\[
\text{nA- + hinag} \rightarrow *\text{ni-hinag} \rightarrow \text{ni-hnag} \quad \text{‘yam'}
\]

A simpler way of presenting the same process, would be to consider that the radical vowel directly migrates to the prefix, with no need for an extra deletion rule: nA- + hinag →
ni-hnag. This presentation implies that the V of the radical has the particular property of being mobile, which is new in our system. In order to show its special status, we will write it with an uppercase letter, e.g. hNag 'yam'.

A short list of the lexemes involved by the latter rule, which we may call 'transferable roots', include the following:

| Table 10 - Some lexemes involved in the rule for vowel transfer |
|-------------------|----------------|----------------|----------|
| llwo              | big             | mēnay          | clever   |
| hNag              | yam             | vēhōg          | flesh    |
| sslecthêe         | centipede       | lēwo-          | tooth    |
| tlwag             | one             | dēlo-          | neck     |

Table 10 shows that the 'mobile vowel' (MV) involved is always /i/ or /u/, i.e. high and mid-high front vowels. Consequently, former accounts of Motlav morphology were erroneous, when they considered a vowel like /a/ on the article (e.g. na-gmel 'men's house') to come from the deletion of this vowel on the radical (*gamel < POC *kamaliK); actually this never happens for /a/ or /e/, which are never more than the default vowel of the prefix itself.

'Transferable roots' have other uncommon properties with regard to their mobile vowel. When the prefix is unvarying, e.g. ni- '3rd singular + present', the MV just doesn't appear:

\[
\text{ni-} + \text{mēnay} \rightarrow \text{ni-mnay} \quad \text{'(he) becomes intelligent'}
\]

Conversely, this MV will surface on the right of the first consonant, when the lexeme is unpreixed, or when its prefix has the form CVC-. Table 11 gives a summary of these rules, with the root mēnay 'intelligent, clever'. The last column shows whether the 'mobile vowel' (MV) appears before or after the first consonant of the radical (C₁).

| Table 11 - Rules involving the 'mobile vowel' of transferable roots |
|-------------------|----------------|----------------|----------|
| prefix | prefixed word | meaning | MV/C₁ |        |
| nA-    | nēmtnay       | cleverness | before  |        |
| nE-    |               | is clever  | before  |        |
| mal-   | mal-mēnay     | is already clever | after |        |
| Ø      | mēnay         | clever     | after   |        |
| ni-    | nīmtnay       | gets clever| no      |        |

5.2 Floating vowels in lexemes

This property, for a phoneme, to surface or not according to its phonological environment, reminds us again of the floating vowels carried by our shifting prefixes (§4.3). The 'mobile vowels' of transferable roots could well be analyzed the same way, since they can be shown to share exactly the same properties. One of these, is the impossibility for floating vowels to create a slot of their own in the syllabic template: all they can do is take V-slots

11 See Codrington (1885: 311) with words like na-bte (Mota patai) 'breadfruit'; and Crowley (forthc.) with words like na-than 'man'.
that have already been created by the phonological structures of the language, and which are left empty after preliminary operations have taken place.

Once again, three cases are possible:

(a) \textit{All slots have been taken by true phonemes}

If the prefix is not a shifting prefix, but has the form CV- (e.g. \textit{ni-} ‘Third singular present’), then both C and V, which are ‘true’ phonemes, take the first two slots of the ICVCICVC sequence. Afterwards, the first two consonants of the radical (\textit{m} and \textit{n} in our example) take their own slot, in such a way that there is no room left for the floating vowel of the lexeme to surface:

\[
\begin{array}{c}
\text{CVCICVC} \\
ni\text{-mnay} \\
\hat{E}
\end{array}
\]

Floating vowels, by definition, cannot create their own position in the template; if they did, we would have a form like \textit{*ni-mênay} instead of \textit{ni-mnay}, and there would be no difference between this \(\hat{E}\) and a ‘true’ \(\hat{e}\).

(b) \textit{A V-slot is left empty on the right of C₁}

If the first ICVC syllable has already been filled in by a CVC- prefix (e.g. \textit{mal} ‘Complete aspect’), or if there is no prefix at all, then the first consonant of the radical (\(C₁=\textit{m}\)) has to begin a new ICVC syllable. Now, we know that phonotactic rules in Motlav exclude consonant clusters within the syllable; as a consequence, \(C₂\) (here \textit{n}) takes the next C-slot, leaving a V-slot empty in the middle. This gives the floating vowel an opportunity to surface on the right of \(C₁\):

\[
\begin{array}{c}
\text{CVCICVC} \\
mal\text{-mnay} \\
\hat{E}
\end{array}
\]

(c) \textit{A V-slot is left empty to the right of C₁: competition between two floating Vs}

The third case is when the lexeme takes a C- prefix (which does not exist in Motlav) or a shifting CV- prefix. In this case, the first three ‘true’ phonemes involved are three consonants (here \(n, m, n\)), which automatically fit into the syllabic template \textbf{ICVCICVCICV}. The V-slot on the prefix is left empty again, allowing for a floating vowel to surface on the left of \(C₁\):

\[
\begin{array}{c}
\text{CVCICVC} \\
nm\text{-nay} \\
\hat{A} \hat{E}
\end{array}
\]

\textit{nA-} + m\textit{ênay} \(\rightarrow\) \textit{nêmnaay} ‘cleverness’

66
Vowel shifting and cloning in Motlav

The latter example provides us with some new information. First, floating vowels in Motlav have the property of being mobile, which was not visible in the previous discussion; a good representation would be to say that this MV (e.g. Ė) is ‘attached’ to the first consonant of the lexeme (e.g. m).

Secondly, in a situation where two floating vowels are competing for the same empty V-slot, the vowel of the radical has priority over that of the prefix, excluding a form like *na-mmay. This kind of hierarchy LEXEME > PREFIX is instructive about the way morphology is cognitively perceived.

Conclusion

As far as the vocabulary is concerned, Motlav can be said to be a conservative language, since it shares many items with neighbouring Mota or with Proto Oceanic. However, historical effects of former stress have largely modified not only the shape of the words, but also the very mechanisms of the whole phonological system. Inherited structures have recently undergone several processes of standardization and reanalysis, bringing about novel phonotactic constraints—a CVC syllabic pattern— and morpho-phonemic rules, including vowel shifting. Moreover, we have demonstrated that a special sort of phoneme was created through history, namely our ‘floating vowels’; these are present in less than thirty items of the language, but their frequency makes them crucial to understand the whole morphology and grammar of the language.

Finally, we would like to put forward the idea that formal analysis is not there to confirm or invalidate a theory defined a priori, but must help build this theory out of the very data. Linguistic structures are already present in the way people talk, and it is neither necessary, nor scientifically satisfying, to hypothesize them out of the blue, ‘more geometrico’. Theory and formalism in linguistics must serve the empirical observation, and not the reverse.

References

KASARHEROU, Jacqueline 1962. Les changements vocaliques de trois préfixes en motlav. Te reo 5, 32-34.
Alexandre François


Transitivity Alternations in Rotuman*

Madelyn Kissock
Oakland University
kissock@oakland.edu

1.0 Introduction

The relationship between verbal affixation and structural case assignment has occupied a central position in discussions of Western Austronesian 'focus' systems (e.g., for a recent discussion of these issues in Malagasy, see Paul 2000). Considerably less attention has been paid to related processes in Oceanic languages. This paper is a preliminary attempt to explore the effects of verbal affixation on transitivity alternations in Rotuman. It turns out that several other phenomena, such as negation and voice neutrality, have interesting roles to play with regard to transitivity. In addition, the discussion of transitive suffixes will shed new light on certain aspects of incorporation. Evidence for the claims made in the paper will be based on data from Rotuman folktales and examples from Churchward (1940).

Section 2 gives a brief general background on Rotuman. Section 3 describes various triggers for valence changing. Section 4 presents a number of interesting aspects of incorporation in Rotuman. In Section 5 I point out the likely need for revision to some of the restrictions on affixation asserted by Churchward (1940). Section 6 sketches out briefly my views on the diachronic development of the -'aki causative. Finally, Section 7 presents some concluding remarks.

2.0 Background

Rotuman is a member of the Central Pacific group within Oceanic. It is traditionally treated as forming a subgroup with Fijian. Rotuman adheres fairly strictly to a surface SVO order. Nominals and pronominals show no overt case marking but both show alternation in form due to 'phase' distinctions.1 In addition, some pronominals have distinct genitive and suffixed forms. Unsuffixed verbs are generally voice-neutral. For example, Churchward (1940:123) points out that iris 'a could mean either 'they ate' or 'they were eaten.' Illustrations are given in (1) and (2) below.

(1) iris rou le' ta 'e Losa
they leave child the P Losa
'they left the child at Losa'

* The author would like to thank the audience at AFLA 7, especially Mark Hale, for valuable comments on this paper. Needless to say, errors which persist in this version are the responsibility of the author.

1Any differences in form as a result of phase are abstracted away in the glosses here as they are not relevant to the discussion. For some detailed discussions of Rotuman phase, see McCarthy (1995 and 1999), Hale and Kissock (1998), and Hale (2000).
Churchward (1940) classifies verbs into three categories. The first two categories distinguish unsuffixed verbs which are typically used intransitively from unsuffixed verbs which are typically used transitively. He states of the former type, "Verbs that are seldom or never used transitively are classed as intransitive." Examples (3-5) are illustrations.

(3) *gou fesia' se irisa*
   I hate to them
   'I hate them.'

(4) *nōnō ma 'oria ō' rua al (Tiafot 16.8)*
   after awhile their(2) parent two die
   'After awhile, their parents died...'

(5) *Tinrau, fā 'on sau, seksek se ut ne 'on hanue ta (Tiafot 16.11)*
   Tinrau son belong king walk-along to end of his town the
   'Tinrau, the king's son, walked along to the outskirts of his town.'

At the same time, he gives an example of a 'usually intransitive' verb used transitively: *matā'u – ia matā' se irisa* 'he looks after them' but, *ia matā' iris* 'he looks after them.'

Of the latter (unsuffixed verbs typically used transitively), he says "Verbs which are used transitively though they have no trans.suffix – and there are many such verbs in Rotuman – are said to be "informally" trans. Exs. are *hili, hil*, to choose, and *ho'a, hoa', to take or carry." A textual occurrence of this last is given in (6).

(6) *ne 'aawar hoa' ia se tei (Āetos 176)*
   ind.q. you-two carry it to where
   '...where you two are carrying it to.'

Once again, Churchward qualifies strict categorization by adding, "Some verbs, again, which are usually trans. (inf.t.) are occ. used intransitively: e.g., *tuki 'to stop (a person from doing something)' is generally trans., as in *gou la tūk iris*, 'I will stop them'; but one may also say (with a slightly different force) *gou la tūk se irisa*, 'I will make an effort to stop them...'."

It is apparent from his examples and discussions that Churchward takes the presence of a prepositional phrase (along with absence of a bare object) to be clear evidence of intransitivity.

3.0 Changes in Valency

As will be illustrated below, the number of arguments of a verb may be either increased or reduced in Rotuman by means of suffixation or by the presence of a counterfactual, such as negation.

---

2Churchward typically cites both complete and incomplete forms, as here.
3.1 Transitive Suffixes

The third of Churchward’s verb categories (referred to in Section 2.0 above) comprises all verbs which are transitive due to their form. He explains this as follows (Churchward 1940:22): 3

The principal suffixes that serve to mark verbs as transitive are -a, -na, and -kia, some verbs taking one, some another. When -a is used, it is attached to the com. phase; -na to the com. phase (its e-form if it has one); -kia to the inc. phase.

Moreover, “Verbs having these suffixes are said to be ‘formally’ trans., their transitiveness being shown by their form.” As such, examples (7) and (8), both of which have transitive suffixes, may be contrasted with their unsuffixed counterparts in (3) and (6) above.

(7) gou fesi’en iris
    I hate them
    ‘I hate them’

(8) ‘amia la hoa’kia ãe (Åetos 216)
    We-two will carry you-sg
    ‘We will carry you’

Unlike in the case of unsuffixed verb forms described in Section 2.0, Churchward offers no qualifying remarks which allow for some exceptions. Instead, he states quite deliberately: “Formally trans. verbs, however, cannot be used intransitively: they are immutably transitive” (Churchward 1940:116). 4 We will see below that the ‘formal’ transitive suffix has a somewhat wider and more interesting distribution than is initially implied in this section.

3.2 The -‘aki Suffix: Causatives and Applicatives

Rotuman has a single form, -‘aki, which functions either as causative marker or an applicative marker, increasing or reducing, respectively, the valency of the verb. Churchward (1940:105-6) notes the following functions for this suffix:

(a) “Causative. This usage is very common, though by no means universal, with intr. verbs: see I.14.1 (a). In a few instances it occurs with inf.t. verbs (cp. I.11.3): e.g., from tole ‘to carry on the shoulder,’ we get tol’aki ‘to cause to be carried on the shoulder,’…”

(b) “Relational or Prepositional: pani ‘to paint’, pân’aki ‘to paint with,’ fäega ‘to speak,’ fäeag’aki ‘to speak about,’ mua ‘to go in front,’ mua’aki ‘to go in front of, to lead’.”

3 It is likely that there are, at most, two different forms of this transitive suffix, the remaining form being a predictable allomorph. This matter requires further exploration.

4 He does, however, offer an interesting comment on the transitive suffix and voice neutrality. He states “It appears, however, that the pass. sense is more frequently expressed by an inf.t. verb...than by a f.t. one…” (Churchward 1940:22) As he gives no examples, it is difficult to determine the characteristics of a passive, formal transitive verb.
He goes on to note that:

Many words terminating in -'aki may be used in two or more of the above senses. E.g., 'ātē'aki, from ā'tē, to eat things, may come under either (a), (b), or (c), meaning (a) to cause to eat, to feed, (b) to eat with, to use (cutlery, plates, etc.) in eating, (c) to eat for (so long).

The causative function of -'aki does not appear to be its most common function. It competes, in this function, with the Rotuman reflexes (one ‘direct’ and one ‘indirect’) of the well-reconstructed POC causative prefix *paka-: a’a- and faka-. Churchward classifies causative verbs in -'aki as ‘informally transitive’ and notes that they have ‘formal transitive’ counterparts. Thus we get ala ‘to die’, al'aki ‘to kill’ (‘informal transitive’), and al'akia ‘to kill’ (‘formal transitive’ with transitive suffix -a).

The voice neutrality of unsuffixed verbs in Rotuman interacts in interesting ways with the -'aki causatives. For example, in many cases it appears that it is the passive reading of the bare stem which serves as the input to causative formation, as in the example cited by Churchward in the discussion of causatives above: tole ‘to carry on the shoulder’ — and therefore, by voice neutrality, also ‘to be carried on the shoulder’ — causative tol'aki ‘to cause to be carried on the shoulder,’ built only to the passive reading. By contrast, the verb 'imo ‘drink’ (but also, by voice neutrality, ‘to be drunk’ in the non-alcoholic sense) gives rise to an ambiguous -'aki form: 'iom'aki means both (a) ‘to cause (obj. a person) to drink’ and (b) ‘to cause (obj. water) to be drunk.’ Such alternations may tell us something important about the core argument structure of the stems involved.

The ‘relational’ or ‘prepositional’ use is, however, by far the most common use of -'aki in Rotuman. The types of relationships it may express appear to be centered around ‘instrumental’ and ‘dative’ (broadly construed, i.e., including both goals and beneficiaries, for example). It appears to promote an originally oblique argument with this type of theta-role to direct object position. Some additional examples include:

afa ‘to make a mark/impression’ (‘on’ expressed with a PP): af'aki ‘to make a mark with (obj. pencil, rubber stamp, etc.)’

atu ‘to grasp or embrace tightly with both arms’ : at'aki ‘to clasp (obj. arms) tightly around (se) something’

lemi ‘to lick’ : lem'aki ‘to lick with’ (obj. the tongue)

mamē ‘to mourn’ : mamē'aki ‘to mourn for’

momono ‘to plug up, block’ : momon'aki ‘to use for plugging up or blocking’
Transitivity Alternations in Rotuman

The 'durational' function mentioned by Churchward may be able to be collapsed with the prepositional uses mentioned above. If we assume that the prepositions used to express these durational properties are dative or instrumental markers, the generalization would be simply that anything marked by these prepositions may be elevated to object position by -'aki.

It is of considerable interest that not any prepositional relationship can be captured by -'aki suffixation. Simple spatial relationships (under, behind, by, from, around) and non-dative directionals (e.g., from, away from, through) do not appear to be expressible by -'aki and a direct object. If we accept, on the evidence of Western Austronesian 'focus' systems, that theta-roles like 'goal' and 'instrument' must be structurally represented (cf. Paul 2000), it is likely that it is precisely these structural positions which -'aki deals with.

Not surprisingly, there is no indication, either in Churchward or in the texts, that it is possible to attach a transitive suffix to the applicative -'aki forms (as one may to causative -'aki).

3.3 Counterfactuals – Negation and Optative

Rotuman has bipartite negation, as in French ne...pas. The Rotuman forms are kat...ra, or, in the future, kal...ra), with (minimally) the verb placed between the two elements. With informally transitive (i.e., unsuffixed) verbs, negation may detransitivize the verb, forcing its object into a PP as in example (10) below. This detransitivization is also seen with the optative pa 'to wish', as in (11).

(9) ia rē garue ta
    he do work the
    'He does the work'

(10) ia kat rē ra 'e garue ta
    he NEG do NEG P work the
    'He did not do the work.'

(11) ia pa rē 'e garue ta
    he OPT do P work the
    'He wishes to do the work.'

The exception to this is if the object is indefinite, such as the object in (12), which is marked by preposed indefinite ta.

(12) ia kat al'qk ra ta le'
    he NEG kill NEG a person
    'He did not kill anyone.'

5The surface position of the verb in negated strings (between the two 'parts' of the negation) suggests that Rotuman has V-raising at least to T, if not higher.
Textual examples of this contrast can be seen in (13) and (14).

(13) pō 'e hān ta kat 'ēs vāvōne-t ra (Haoag 8)
    since woman the NEG have husband-a NEG
    'since the woman did not have a husband'

(14) ka gou kal 'ēs hoi‘ak ra 'ē ‘āea (Pure 40)
    and I NEG-FUT have any more NEG P you-sg
    'and I will not have you any more'

Formally transitive verbs (those marked with the transitive suffix) do not detransitivize under these same conditions, as shown in (15).

(15) ia pa kat al‘akia ra iris
    he wish NEG kill-tr NEG them
    'He does not wish to kill them.'

Churchward notes a semantic differences between the type found in (10) (an unsuffixed, normally transitive verb which is detransitivized under negation) and that found in (15) (a transitive suffixed verb which remains transitive under negation). The relevant semantic contrast can be seen from the translation of (16) and (17) below (where upper case is used to indicate emphasis).

(16) ia kat al‘ak ra 'ē irisa
    he NEG kill NEG Prep them
    '(The man said he would kill them but) he did NOT kill them.'

(17) ia kat al‘akia ra iris
    he NEG kill-tr NEG them
    '(The man wounded them but) he did not KILL them.'

The phenomenon of detransitivization by counterfactuals is found in many other languages (e.g., Russian and Lithuanian), however, I am not familiar with it in any other Oceanic languages. If it can be demonstrated that the inability of the negated verb to assign accusative case to indefinites arises for some well-defined structural reason, the fact that Rotuman (unlike its Oceanic neighbors) shows such a phenomenon could be telling as to fundamental aspects of its structure.

4.0 Incorporation and Affixation

4.1 Types of Incorporated Elements

Negation serves as a useful diagnostic for incorporated elements (as noted, for incorporated nouns, by Churchward 1940:122). Unsurprisingly (in an Oceanic context), bare indefinite nouns regularly incorporate into the informally transitive (unsuffixed) verb form. The relevant contrast can be seen in (18) vs. (19) and (20) vs. (21) below.
Transitivity Alternations in Rotuman

(18) ia kat 'inea ra ta te
   he NEG know NEG some things
   'He doesn’t know things.'

(19) ia kat 'inea te ra
   he NEG know things NEG
   'He doesn’t know things.'

(20) ka le' mo rua 'i kat pōpō rog ra e... (Āeatos 143)
    but child ptcl 2 these NEG get report NEG about-it...
    'but the children did not get any report about it'

(21) Ka fa ta kal lua te ra (Haoag 66)
    and boy the NEG-fut spit-out thing NEG
    'and the boy wouldn't spit anything out'

It is quite surprising, however, that more complex indefinite objects can also incorporate, including DP's, both relatively short additions to the bare N as in (22), see also (13) above, and significantly longer ones (23):

(22) Ma raksa 'ia 'e 'itar kat 'es rahi-t ra (Le Maafi 109)
    but unfortunate that we-2 NEG have fire-any NEG
    'but unfortunately we haven’t any fire'

(23) ma kat 'es sui tore-t ra (I’or 59)
    but NEG have bone leftover-a NEG
    'but he didn’t have a leftover bone'

Explicit morphological plurals, also not generally subject to incorporation in Polynesian and Micronesian languages, may be incorporated in Rotuman, as can be seen in the following example in which lele‘a is the lexical plural of le‘e ‘child.’

(24) Ka sāghān uan ‘on Rah kat 'es lele‘a' ra (Haoag 11)
    and sister middle his Raho NEG have children NEG
    'and Raho’s middle sister had no children'

In addition, while many Oceanic languages attest to the incorporation of directionals (a type of adverb) into the verb, Rotuman allows a much broader range of adverbs to incorporate, as can be seen in example (15) above, in which hoi’aki ‘any more; again’ is incorporated, as well as from:
It appears from these facts (and assuming that position relative to negation is a reliable diagnostic for incorporation) that Rotuman has generalized the common Oceanic process involved in incorporating bare non-specific (i.e., indefinite) N’s to all indefinite objects, including phrasal ones. In addition, Rotuman appears to have generalized with widespread Oceanic process underlying the incorporation of directional adverbials to a variety of temporal and manner adverbials as well.

4.2 Affixation to Incorporated Elements

Churchward asserts (1940:107) that a number of suffixes, including directionals, the ‘ingressive’ in -'ia, the ‘pronominal suffixes’ which express the ingressive for certain types of verbs (these are in fact reflexive pronouns), and the transitive suffixes used form ‘formally transitive’ verbs follow the adverb, should the verb to which they properly should be attached be accompanied by one. He gives the following examples:

(26) iris sur miji-m (*iris surum mij)
    they come promptly-DIR
    ‘they enter promptly’

(27) fā ta fek fakapau-'ia
    man the angry very-INGR
    ‘the man became very angry’

(28) gou jōn vāve-atou
    I flee quickly-I
    ‘I fled quickly’

(29) fā ta fut hoi’aki-a lā ta (*fā ta futia hoi’āk lā ta)
    man the pull again-TR rope the
    ‘the man pulled the rope again’

Such sentences are also attested in the legends:

(30) ka ia kal pō hoi’āk-'ia ra ia (Sau 158)
    and he NEG-fut get again-INGR NEG her
    ‘and he would not be able to get her again’

(31) ka fā rua kat rē a’lelei-'ia ra hān rua (Sianpualetalf45)
    and man 2 NEG treat badly-INGR NEG wife 2
    ‘and two men began to ill treat their two wives’
Transitivity Alternations in Rotuman

(32) ka iria kat foar pau sio ra se hàn rua (‘Åeatos 27)
and they NEG tell fully DIR NEG P woman 2
‘and they didn’t explain fully to the two women’

Suffixation of this type strongly supports the idea that these adverbs are, in fact, incorporated—note particularly that the transitive suffix will follow the adverb if present.

The incorporation of adverbs is found in the more archaic Indo-European languages as well as in Rotuman. In the Indo-European case, this incorporation goes hand-in-hand with the incorporation of prepositions. This makes an analysis under which the ‘applicative’ -‘aki constructions (involving the elevation of dative and instrumental arguments to direct object status) are taken to arise via the mechanism of ‘preposition incorporation’ (see Baker 1988) plausible.

As expected, incorporated nouns also form the basis for the attachment of these affixes (Churchward 1940:122):

(33) ia fā’ puku-m se goua
he write letter-DIR to me
‘he wrote to me’

(34) fā’ ta ‘inea tē-ia
man the know thing-INGR
‘the man comes to know things’

(35) gou fā’ puku-etou
I write letter-I
‘I start letter-writing’

Note, particularly, the use of the transitive suffix attached after an incorporated noun in (34) which Churchward says is practically synonymous with (35).

(36) ia rak’ak tē-an iris
he teach thing-TR them
‘he teaches them things’

(37) ia rak’ak tē se irisa
he teach thing P them
‘he teaches things to them’

Oddly, Churchward (1940:107) notes that the suffix ‘aki ‘is not moved along in this way, but remains where it was.’ Note that his own example, cited in the discussion of the causative suffix in Section 2.3 above (‘åeåti’), involves the affixation of ‘aki to a verb+incorporated noun pair, for which his dictionary provides additional examples (e.g., hō’sol’aki ‘to heap or pile up’
Madelyn Kissock

< ho 'i ‘accumulate, fetch, heap’ + solo ‘mound’ + -‘aki). Since on this page he is explicitly discussing ‘shifting’ of these affixes to a position following adverbs (rather than incorporated nouns), it is possible that -‘aki will only ‘shift’ around incorporated nouns (not adverbs). No examples involving adverbs, either way, are known to me at this time.

The shifting of verbal affixes to a position following incorporated adverbs and nouns is clearly of great significance in revealing the precise structural status of these incorporated elements. It may also bear on the long-debated issue of whether incorporation is a lexical or a syntactic process. While the ability of ‘V+incorporated Noun/Adverb’ to take a transitive suffix may be taken by some to be clear evidence of a derivation in the lexicon (i.e., compounding), the apparently obligatory (at least in the case of adverbs) and fully productive nature of the incorporation clearly points in the direction of syntactic derivation. These constructions may thus force us to rethink what is and what is not a test for this critical distinction.

5.0 Restrictions on Affixation

Churchward (1940:117) notes that “F.t. verbs cannot take any of the tense-forming suffixes.” The set of suffixes involved in the ‘tense’ marking cases mentioned by Churchward include -‘ia, -tia, -a, and the ‘reflexive’ object pronouns. The first three look like versions of the POC transitive suffixes. While a detailed investigation of the use of these forms in Rotuman is still outstanding, I am disinclined at this time to accept Churchward’s characterization of them as marking ‘tense’ at all. In some cases I strongly suspect they represent the POC *-Cia suffixes, whose precise function differs from language to language even within Polynesian (e.g.), and has proven remarkably difficult to characterize definitively. The suffix was, in any event, incompatible with any transitive marking on the verb to which it attached. These forms need to be investigated in Rotuman — without such an investigation, their significance for the problems under discussion cannot be determined.

6.0 Origin of Causative -‘aki

It is difficult to see how we might connect the causative function of the -‘aki suffix with its applicative functions synchronically. However, since the suffix has no plausible POC etymology (note that the POC ‘remote transitive’ suffix *-aki(ni) is not an ideal candidate, given the Rotuman glottal stop), it must have originated within Rotuman. It is hard to see how two unrelated suffixes of this shape could have independently developed in this brief time. So I offer the following diachronic speculation on the origin of causative -‘aki.

1 -‘aki arises as an incorporated instrumental preposition. (This is its most widely attested synchronic function, and probably the most difficult to get from any other diachronic source.)

2 -‘aki comes to mark the promotion of other oblique structural cases, such as the dative.

We have seen this type of syntactic extension several times already in Rotuman: originally, there are no instances of kat VERB ra ADVERB in the extensive corpus of Rotuman legends, indicating the adverbs obligatorily incorporate.
Transitivity Alternations in Rotuman

from comparative evidence, it is necessary to assume that only directional adverbs were incorporated — now all adverbs are; again based on comparative evidence it seems that originally only (indefinite) bare-nouns were incorporated, now any indefinite DP may be.

3 Many instrumental applicative -'gki verbs were subjected to a reanalysis based on the following systematic ambiguity (using the examples cited on pg. 3 above):
(a) He makes an impression on the paper with a pencil = (b) The pencil makes an impression on the paper.
(a) He embraces the child tightly with his arms = His arms clasp tightly around the child.
(a) He licks at the candy with his tongue. = (b) His tongue licks at the candy.
(a) He plugs up the hole with a cork = (b) A cork plugs up the hole.

The -'gki forms were originally instrumental applicatives to the (a) sentences, but were reanalyzed as causatives to the (b) sentences. (For a similar reanalysis of instrumentals, see Garrett 1990.) That is, they are reinterpreted as meaning 'he makes the pencil make an impression on the paper,' 'he makes his arms clasp tightly around the child,' 'he makes his tongue lick at the candy,' and 'he makes a cork plug up the hole.'

4 This new 'causative' -'gki comes to be used in cases in which no instrumentality was ever involved, such that ala 'to die' can form al'gki 'to kill.'

7.0 Conclusion
I have tried to show in this paper that Rotuman offers a wealth of interesting data, much of it Oceanic-like but with a unique Rotuman twist, which bears on fundamental questions surrounding structural case and argument structure. This primarily descriptive account is the necessary first step in developing the formal syntactic analysis which I hope to be able to provide in a subsequent paper.

References
Madelyn Kissock


Fixed segmentism, markedness and faithfulness: Nominalising reduplication in Chamorro

Thomas B. Klein and Meta Y. Harris, The University of Manchester

Fixed segmentism in reduplication occurs when invariant segments appear where copying is expected otherwise. Fixed segmentism has been important for a number of issues in theoretical phonology including prespecification (Marantz 1982, among others) and underspecification (Pulleyblank 1988, among others) in rule-based phonology and the approach to markedness in Optimality Theory (OT) (Alderete et al. 1999). Given the theoretical significance of fixed segmentism, any newly uncovered pattern of this type of phonological structure is of inherent interest to phonological theory. A major result of the authors’ fieldwork on Chamorro reported in this paper is that Chamorro nominalising reduplication exhibits two fixed vowels, the unmarked high front vowel [i] and the marked low front vowel [a]. Given that this pair of fixed segments includes a cross-linguistically marked vowel, the pattern from Chamorro presents a challenge to the idea formalised in the OT framework by Alderete et al. (1999) that fixed segmentism decreases phonological markedness. On the other hand, given that the fixed vowels of Chamorro nominalising reduplication include the cross-linguistically unmarked high front vowel, the OT approach to fixed segmentism appears as a viable framework for the formal analysis of this pattern. In addition to exploring the connection between phonological markedness and fixed segmentism, Alderete et al. (1999) have claimed that any phonological restriction resulting in fixed segmentism in reduplication must be matched as a possible restriction on default patterns or segmental inventories elsewhere. As is demonstrated below, the patterns of default vowel phonology found in Chamorro do not match the pattern of fixed segmentism observed with nominalising reduplication. Thus, Alderete et al.’s approach predicts that the set of fixed vowels found in nominalising reduplication should constitute the vowel inventory of some other language or languages. As is shown in this paper, this prediction is met through vowel inventories which are restricted to non-schwa vowels of maximally contrasting height.

The model of reduplication invoked in Alderete et al. (1999) is couched in the framework of correspondence theory presented in McCarthy & Prince (1995). However, McCarthy & Prince’s (1995) model of correspondence has been shown to be problematic for cases in which the reduplicant morpheme appears to be more faithful to the input than to the surface reduplicative base (Rainy & Idsardi 1997, Spaelti 1997). In McCarthy & Prince’s (1995) full model of correspondence, the reduplicant may be subject to input-reduplicant faithfulness to force faithfulness to the input root. As pointed out in Rainy & Idsardi (1997), this notion is problematic because it evaluates faithfulness by comparing the input of one morpheme to the output of another, thus weakening the idea of faithfulness itself by postulating cross-morphemic comparison (see Spaelti 1997 for further criticism of the full model). The data presented in this paper show that prefixal reduplicants in Chamorro take certain cues for their segmental from the input rather than the surface representation. Thus, low front [a] appears in the nominalising reduplicant with low vowel roots even though the surface base vowel may be reduced to schwa (cf. [katsa?] ‘to husk a coconut’ versus [‘ka+katsa?] ‘coconut husker’). Furthermore, mid vowels in the reduplicant of continuous aspect reduplication in Chamorro match the input root, but need not match the surface base.

* This research has been made possible financially through grants to the first author from the British Arts and Humanities Research Board (AHRB) (Research Award MRG-AN5855/APN9404) and the Research and Graduate Support Unit (RGSU) of the University of Manchester. The support of these institutions is gratefully acknowledged. Our gratitude to the native speakers and scholars of Chamorro whose data is represented here: John Cepeda, Maria Fejera, Larry Gamboa, Rosa S. Palomo, John Perez and Lumin Perez. Special thanks to Anthony Grant and the audience at AFLA7, in particular Adam Usishtikin, for discussion of this material. None of these institutions and individuals is responsible for any of the shortcomings of this research.
as evidenced in ['konni?] 'to catch, to take' versus ['ko+konni?] 'catching, taking'. Thus, the data from Chamorro are evidence for the necessity to examine fixed segmentism in light of recent approaches to apparent input-reduplicant faithfulness. As a consequence, we present a model of reduplication in which word faithfulness (Struijke 2000) is indispensable to the understanding of fixed segmentism.

Chamorro is an Austronesian language spoken on the Mariana Islands of Guam, Saipan, Rota and Tinian and in expatriate communities in the United States. The available descriptions of contemporary Chamorro (Topping 1973, 1980, Topping et al. 1975) provide only a fragmentary description of fixed segmentism in nominalising reduplication. Furthermore, notwithstanding the great insight displayed in these sources, these works also employ certain descriptive generalisations that are not correct, as is shown below. Thus, there is a need for a more complete and more accurate description of fixed segmentism in Chamorro. The present paper addresses this need using data from our fieldwork on Guam in the Summer of 1999. The results of this fieldwork not only show that the vowel inventory of the nominalising reduplicant is restricted to [i] and [a], but also show that the nominalising reduplicant induces vowel fronting or umlaut in the base and that the base is subject to Vowel Lowering. Given that umlaut and Vowel Lowering have received much attention in recent theoretical works (Chung 1983, Halle & Vergnaud 1987, Crosswhite 1998, Klein 2000), the data from nominalising reduplication presented here have implications beyond the focus of this paper.

This paper is structured as follows. §1 lays out preliminaries that are essential for the understanding of the vocalism of nominalising reduplication. This is followed in §2 by a brief review of previous descriptions of this pattern. §3 presents and describes the data from our fieldwork and lays out the generalisations concerning the vocalism of nominalising reduplication. In §4 we present our analysis of fixed segmentism in nominalising reduplication paying particular attention to issues of markedness and faithfulness. §5 presents some conclusions and suggests directions for future work.

1. Preliminaries

In this section we describe the inventory of Chamorro monophthongs under primary and secondary stress and in unstressed position and comment on the distribution of tense and lax vowels and the phonemic status of mid vowels. Furthermore, we lay out some of the basic stress facts of Chamorro and comment on the behaviour of umlaut and Vowel Lowering. In particular, we present data showing that Vowel Lowering in the reduplicative base may result in a mismatch between the reduplicant and the base vowel in Chamorro reduplication. Finally, we present the feature system used to specify the vowels of Chamorro.

Given that Chamorro is a stress language, it is expected that the vowels of Chamorro covary depending on the degree of stress, as is common in other stress languages. Under primary stress Chamorro has a symmetrical six-vowel system of monophthongal underliers in lexical contrast (Topping 1973, Chung 1983). This system may be graphically represented as in figure 1. below.

![Fig. 1. Chamorro symmetrical system of six vowel underliers under primary stress](image)

Note that the system displayed in fig. 1. emphasises the phonological symmetry of the series of three front vowels versus three back vowels distributed evenly along the height dimension. This system abstracts away from phonetic reality in that nonlow vowels show...
tense and lax allophones which are distributed allophonically. The data in Chung (1983) and Topping (1973) show that nonlow vowels in Chamorro are lax in closed syllables and tense in open syllables. There is variation in the data presented in these works with respect to the tense versus lax distinction in open unstressed syllables. Whereas Topping (1973) maintains that nonlow vowels are lax in open unstressed syllables, the transcriptions in Chung (1983) show tense nonlow vowels in unstressed open syllables. The data in our corpus show evidence for the sum of these descriptions. Thus, nonlow vowels are lax in closed syllables regardless of the level of stress. Furthermore, nonlow vowels in open syllables are generally tense in stressed syllables, whereas they may be tense or lax in unstressed syllables. Note that for the purposes of the analysis presented in this article, we abstract away from the tense-lax distinction.

Chamorro shows stress reduction in that the inventory of low vowels is reduced under secondary stress and in unstressed position. Thus, the data in our corpus generally confirm Chung’s (1983) observations that low vowels under secondary stress merge to [a] and that the low vowels merge to [a] in unstressed syllables. Furthermore, we also found data in evidence of Topping’s (1973) observation that the low vowels may appear as [a] in unstressed position (see below). Note that schwa may not occur under stress. In contrast to the low vowels, all nonlow vowels of Chamorro may occur under secondary stress and in unstressed position.

Some previous authors have claimed that mid vowels are not phonemic underliers in present-day Chamorro, at least not in the native vocabulary (Witucki 1973, 1974, Chung 1983). However, other works disagree with this view and maintain that mid vowels are possible underliers in Chamorro including the native vocabulary (Topping 1973, Halle & Vergnaud 1987). The primary evidence for the scepticism towards mid vowel underliers stems from the fact that their distribution is restricted in that they are primarily observed in stressed closed syllables whereas high vowels generally occur in the place of mid vowels elsewhere, such as in open syllables and in unstressed position. However, a careful inspection of the available evidence shows that mid vowels may occur in stressed open syllables in items that are not borrowed from Spanish or English. Thus, lexical items such as bélu ‘tottering, balanced (upright)’ and [bo?an] ‘froth, bubbles’ show that mid vowels may occur in stressed open syllables in roots. Furthermore, stress-attracting prefixes such as ké- ‘about to, try’ and é- ‘look for, hunt’ (Topping 1973) contain mid vowels in open syllables as in [ké+konné?] ‘try to catch’ (cf. [konné?] ‘to catch’). In addition, the stress-attracting reduplicant in what Topping (1973) labels continuative aspect reduplication shows mid vowels in open stressed syllables. Consider the data in (1).

(1) Mid vowels in continuative aspect reduplication

(a) [?entalu?] ‘to interfere’
(b) [pe+piska] ‘to fish, to hunt’
(c) [ko+konné?] ‘to catch, to take’
(d) [so+sudda?] ‘to find’

As shown in (1), mid vowels regularly occur in the auto-stressing open syllable reduplicant in continuative aspect reduplication. Given this evidence for the occurrence of mid vowels in open syllables, we conclude that mid vowels are possible underliers in Chamorro even though their distribution is somewhat restricted, thus essentially following the view in Topping (1973) and Halle & Vergnaud (1987).

The default location of primary stress in Chamorro words is on the penultimate syllable ((2) (a)). However, certain roots have antepenultimate or final stress, as shown in (2) (b) and (2) (c), respectively.

(2) Chamorro primary stress (Data reproduced from Chung 1983: 38)

(a) pulónnum ‘triggerfish’
(b) dánkulu ‘big’
(c) peskadót ‘fisherman’
Certain Chamorro prefixes attract primary stress onto themselves, thus disrupting regular penultimate stress. Note that the reduplicant morphemes of continuative aspect reduplication (see (1)) and nominalising reduplication belong to the class of stress-attracting prefixes.

Chamorro has two well-known alternations in its vowel system, vowel fronting or umlaut (Topping 1973, Chung 1983, Klein 2000, among others) and Vowel Lowering (Chung 1983, Crosswhite 1998, among others). Umlaut in Chamorro is the fronting and, for rounded vowel underliers, the concomitant derounding of input back vowels induced by certain morphological units including the determiner i ‘the’. Consider the data in (3).

(3) Chamorro umlaut

(a) [ˈgumaʔ] ‘house’ [iˈgimaʔ] ‘the house’
(b) [ˈpotta] ‘door’ [iˈpetta] ‘the door’
(c) [ˈatsuʔ] ‘rock’ [iˈatsuʔ] ‘the rock’

The data in (3) show that Chamorro umlaut fronts the back vowel underliers to result in the corresponding front vowels.

The basic generalisation concerning Vowel Lowering is that nonlow vowels surface as mid in stressed closed syllables and as high elsewhere (Chung 1983). Vowel Lowering applies optionally in syllables that are unstressed in morphologically complex forms, but are under main stress in corresponding forms of lesser morphological complexity. Consider the data in (4).

(4) Chamorro Vowel Lowering (Chung 1983: 49)

(a) [kʷɛntus] ‘to speak’ [kʷɛntusɪ] or [ɛnˈtusɪ] ‘to speak to’
(b) [lɔkˈluk] ‘to boil’ [lɔkˈlukŋa] or [lʊkˈlukŋa] ‘its boiling’

The data in (4) (a) and (4) (b) show that Vowel Lowering may cause an alternation between mid versus high front vowels and mid versus high back vowels, respectively.

Umlaut and Vowel Lowering interact with nonprimary stress in a quite complex manner. Given that the primary focus of the present paper is on the fixed segmentism of the stress-attracting nominalising reduplicant, the details of the interaction with nonprimary stress are for the most part abstracted away from in the remainder of this paper. However, the occurrence of umlaut and Vowel Lowering under nonprimary stress is important for the understanding of the vocalism of the base in nominalising and continuative aspect reduplication. In particular, note that Vowel Lowering optionally occurs in the base of continuative aspect reduplication (see also (1) above). Consider the data in (5).

(5) Optional Vowel Lowering in continuative aspect reduplication

(a) [ˈpeskə] ‘to fish, to hunt’ (i) [ˈpe+piska] ‘fishing, hunting’
(b) [ˈkɔnniʔ] ‘to catch, to take’ (i) [ˈko+kɔnniʔ] ‘catching, taking’,

As shown in (5) (a), [i] and [e] may occur in the reduplicative base of roots such as [ˈpeskə], whereas (5) (b) shows that [u] and [o] may occur in the reduplicative base of roots such as [ˈkɔnniʔ], resulting in mismatches between the reduplicant vowel and the reduplicative base vowel as in (5) (a) (i) and (5) (b) (i). Importantly, these data show that the reduplicant vowel in Chamorro reduplication may be more faithful to the input root vowel than to the reduplicative base vowel. This type of faithfulness has received much attention in the recent literature on reduplication (Rainey & Idsardi 1997, Spaelti 1997, Struijke 1998, 2000, Yip 2000) and plays a significant role for the analysis of fixed segmentism to be presented below.
Given that the theoretical analysis of the fixed segmentism observed in Chamorro nominalising reduplication to be presented below is based to a significant extent on proposals in Alderete et al. (1999), it is useful for the purposes of maximum comparability between our analysis and Alderete et al.’s to make analogous assumptions concerning phonological feature specifications. However, given that the feature system used by Alderete et al. is not worked out in great depth, a few necessary details must be filled in to get the analysis off the ground.

Alderete et al. (1999: 335f.) follow the system of stricture-based features presented in Clements & Hume (1995), but invoke the featural organisation of the Particle Phonology framework (Schane 1984; see also element theory (Harris & Lindsey 1995, among others) and Dependency Phonology (Ewen 1995, among others)) in maintaining that mid vowels combine the features of high and low vowels. Given this system of phonological features, the feature specifications for the Chamorro vowel underliers from fig. 1 are presented in (6). Headedness relations among elements must be added to the system outlined in Alderete et al. to represent the contrast between /e/ and /a/ in Chamorro. Following conventions in Dependency Phonology, these headedness relations are represented by an underline.

(6)

<table>
<thead>
<tr>
<th></th>
<th>u</th>
<th>e</th>
<th>o</th>
<th>a</th>
<th>o</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[LAB]</td>
<td>[LAB]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[DORS]</td>
<td>[DORS]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[COR]</td>
<td>[COR]</td>
<td>[COR]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[PHAR]</td>
<td>[PHAR]</td>
<td>[PHAR]</td>
<td>[PHAR]</td>
<td>[PHAR]</td>
</tr>
</tbody>
</table>

The table in (6) presents the feature specifications of the Chamorro vowels arising from the stricture-based features [COR], [LAB], [DORS], and [PHAR] assumed in Alderete et al. (1999). Given that Alderete et al. invoke the featural organisation of Particle Phonology, the mid vowels /e/ and /o/ are both specified for [PHAR]. In addition, note that /e/ and /a/ are both specified through the features [COR] and [PHAR]. Thus, to encode the contrast between /e/ and /a/, we invoke the notion of headedness. Thus, [COR] is the head in the mid front vowel, whereas [PHAR] is the head in the low front vowel. Furthermore, note that both low vowels are headed by [PHAR].

This section has laid out essential preliminaries to the understanding of the Chamorro vowel and stress system. The next section presents a summary of the previous descriptions of the vocalism of nominalising reduplication and identifies issues and open questions that arise from these earlier descriptions.

2. Previous descriptions of nominalising reduplication

Fragments of Chamorro nominalising reduplication have been described in Topping (1973, 1980). Additional data relevant to nominalising reduplication may be found in Topping et al. (1975). The purpose of this section is to present the descriptive picture concerning nominalising reduplication according to these sources and to identify issues and open questions arising from this presentation. Note that the representation of the data in this section follows the transcription conventions of the sources.

Topping (1973: 181) characterises nominalising reduplication as follows.

The rule for reduplication for nominalization is to repeat the stressed vowel and preceding consonant (if one is present). The primary stress falls on the first vowel of the reduplicated form. If the stem vowel is back, then it gets fronted.
Topping exemplifies this description of nominalising reduplication with the following data. Note that, in accordance with the default location of primary stress in Chamorro, all of the roots in the data in (7) have primary stress on the penultimate syllable.

(7) Nominalising reduplication: Data from Topping (1973: 181 f.)

(a) gupu ‘to fly’
(b) adda’ ‘mimic’
(c) kanno ‘eat’
(d) tuge’ ‘write’

\[ g_{\text{fgipu}} ‘flyer’ \]
\[ a’d\text{’adda’} ‘mimicker’ \]
\[ k\text{’akanno’} ‘eater’ \]
\[ t\text{’ftuge’} ‘writer’ \]

Note that the apostrophe in Topping’s works stands for the glottal stop. Furthermore, note that Topping generally does not transcribe word-initial glottal stop in cases of glottal stop insertion to avoid vowel-initial words in Chamorro. Thus, items such as adda’ ‘mimic’ and a’dadda’ ‘mimicker’ in (7) (b) have a word-initial glottal stop phonetically (cf. Topping 1973: 36).

The data in (7) support three important characteristics of nominalising reduplication. First, these data show that nominalising reduplication employs a CV reduplicant. Secondly, the data in (7) show that the nominalising reduplicant is prefixing. Finally, the data in (7) show that the nominalising reduplicant attracts primary stress onto itself. Given that normally all vowel underliers of Chamorro occur under primary stress, it is expected \textit{ceteris paribus} that all six contrastive vowels of Chamorro are found in the stress-attracting reduplicant of nominalising reduplication. However, the data in (7) (a) and (d) show that this is not the case. Instead, the base vowel /u/ is replaced by [i] in the reduplicant under nominalising reduplication. Note that information regarding the quality of the low vowel in the reduplicant of nominalising reduplication may not be gleaned from the data in (7). This is because the representation of the Chamorro data in Topping (1973) employs for the most part orthographic conventions which do not distinguish between [a] and [æ]. Furthermore, note that no items with mid vowels subject to reduplication are presented in the data in (7).

Topping (1980: 328) expands on the description in Topping (1973) by stating that “if the first vowel of the stem is o or u, the first vowel of the reduplicated form will be i”, but illustrates the behaviour of stem o only through one example ((8) (a) below). In our search of Topping et al. (1975), we have found one more example for the behaviour of stem o (see (8) (b)).

(8) Nominalising reduplication with /o/: Data from Topping et al. (1975), Topping (1980)

(a) godde ‘to tie’
(b) logru ‘give odds (gambling)’

\[ i\text{’gigedde} ‘the tier’ \]
\[ l\text{’logru ‘odds maker’} \]

The data in (8) show that for bases with /o/ the reduplicant of nominalising reduplication surfaces with the vowel [i]. Thus, the data in (7) and (8) are a good indication for the behaviour of stem u and o; but do not conclusively illustrate the behaviour of the other vowels of Chamorro under nominalising reduplication. However, some additional data may be gleaned from Topping et al. (1975).

Topping et al. (1975) distinguish between low front vowels and low back vowels under primary stress through the use of æ and a, respectively. Thus, data from this work may be used to throw light on the behaviour of low vowels in the reduplicant of nominalising reduplication (cf. (9) (a)).
Nominalising reduplication: Data from Topping et al. (1975)

(a) (i) [æ]dda’ 'mimic' [æ]’adda’ 'mimicker'
(ii) fahan 'buy' [fæ]’fahan 'buyer'
(b) (i) li’of 'dive' li’i’of 'diver'
(ii) penta 'to paint' pipenta 'painter'
(c) (i) peska 'to fish, to hunt' pifeska 'fisherman, hunter'
(ii) peska 'to fish, to hunt' pifeska 'fisherman, hunter'
(iii) bende 'to sell, to vend' bfbende 'shopkeeper, seller'
(iv) de’on 'pinch, tweak' dédé’on 'pincher'

The data in (9) (a) show that low vowels appear as the front low vowel in the reduplicant of nominalising reduplication regardless of the location of the base low vowels along the front-back dimension. Note that there are a sizeable number of additional examples with low stem vowels behaving like the data in (9) (a) to be found in Topping et al. (1975). The datum in (9) (b) indicates that for bases with /i/ the reduplicant of nominalising reduplication appears as [i]. Note that there is one more example of this kind to be found in Topping (1980). The data in (9) (c) show variation with respect to the behaviour of base /e/ in the reduplicant of nominalising reduplication. The data in (9) (c), (i-iii) show that base /e/ may licence [i] in the reduplicant. On the other hand, the data in (9) (iv) and the variant pifeska ‘fisherman, hunter’ in (9) (c) (ii) suggest that base /e/ may appear as [e] in the reduplicant as well. Note that whereas the data with base /e/ shown in (9) (c) seem to exhibit variation under nominalising reduplication, there are too little data available from the previous literature to establish this conclusion firmly. In addition, it must be noted that the behaviour of base /e/ has not been explicitly described in any of Topping’s works.

For all of the data in (7)-(9), it must be noted that the vowel quality of the base vowels in the reduplicant is represented through orthographic convention rather than phonetic transcription. This leaves the phonetic quality of the base vowels in the reduplicated forms undetermined in two respects. First, it is not clear from the orthography which allophone of the low vowels appears initially in bases like <adda’> ‘to mimic’ to reduplicated forms such as <[æ]’adda’> ‘mimicker’. Secondly, given that mid vowels alternate with high vowels under nonprimary stress, the orthography employed in Topping’s works glosses over the phonetic height of the vowels in bases such as <penta> ‘to paint’ in reduplicated forms such as <pipenta> ‘painter’.

Topping (1973) has advanced the following explanation for the restricted inventory of vowels observed in the reduplicant of nominalising reduplication.

[T]he stressed vowel of the reduplicated form is fronted because the nominalized form is usually preceded by the article i. It will be recalled that i causes vowel fronting, as in guma’ ‘house’, i gima’ ‘the house’. It does not seem feasible to try to account for the vowel fronting as the result of reduplication. (Topping 1973: 182)

Topping’s statement is problematic from three perspectives. First, the data presented above suggest that the vowel inventory of the reduplicant of nominalising reduplication is limited to [i] and [a] and is, thus, not accounted for as a simple vowel fronting. Descriptively speaking, vowel raising seems to be involved as well, given that /o/ and /e/ appear to surface as [i]. Secondly, the observed fronting and raising does not seem to depend on the presence of the determiner i ‘the’, given that the vast majority of the reduplication data in Topping (1973) and Topping et al. (1975) appear without this determiner. Finally, vowel fronting as observed in the nonreduplicative phonology of Chamorro results in an /o/-[e] alternation for mid vowels, whereas /o/ in Topping’s nominalising reduplication data appears to alternate with [i].

In sum, the data from the previous sources suggest that the vowels that may appear in the prefixal stress-attracting CV reduplicant of nominalising reduplication are limited to i and æ where i occurs with nonlow base vowels and æ occurs with low base vowels. However, this generalisation is not firmly established because there are few data available for
nonlow base vowels and because some data points appear to show reduplicant e with the base vowel e. In addition, the orthographic conventions employed in Topping's works prevent any firm conclusions as to the quality of the vowels in the bases of the reduplicated forms. Furthermore, Topping's claim that the restriction observed in the reduplicant vowels is due to umlaut or vowel fronting induced in the reduplicant by the determiner i is questionable in light of the fact that most of the data from nominalising reduplication presented in Topping's works appear without the determiner and that [e] appears at best inconsistently in the reduplicant. The data from our field research to be presented in the next section have been collected to shed light on these issues.

3. Vocalism in nominalising reduplication

The previous section has worked out three underinvestigated issues in the vocalism of Chamorro nominalising reduplication. First, the set of vowels that may appear in the nominalising reduplicant needs to be established with a greater degree of certainty than is possible from the existing sources. Secondly, the quality of the initial base vowels in reduplicated forms needs to be scrutinised. Finally, it is questionable that the restrictions concerning the vowels in the nominalising reduplicant are caused by the umlaut-inducing determiner i. The data from our fieldwork presented in this section contribute to the resolution of these issues.

It is shown in this section that nominalising reduplication is a case of fixed segmentism in that only [i] and [a] may occur as vowels in the nominalising reduplicant. Furthermore, it is shown that reduplicant [i] occurs only with nonlow root vowels and that [a] occurs only with low root vowels. This pattern is interesting for two reasons. First, given that usually the full inventory of Chamorro vowels is displayed under primary stress, it is surprising to find that the stress-attracting nominalising reduplicant is severely restricted as to the quality of vowels that may occur in it. Secondly, the dependence of the occurrence of the two reduplicant vowels on the height of the root vowel is of interest because this type of reduplicative pattern does not seem to have been found elsewhere to this point. Thus, the inventory and the distributional restrictions on the vowels in the nominalising reduplicant require a novel explanatory account. As an empirical prerequisite to establishing the case for fixed segmentism, one must compare the vowels found in the reduplicant with the vowels in the unreduplicated roots and in the bases of reduplication. Given the data in this section, there is consistent evidence for a systematic mismatch between the vowels in the reduplicant and the vowels in the corresponding roots and reduplicative bases.

The data for the present study are extracted from face-to-face interviews with six native speakers of Chamorro conducted on location on Guam in the Summer of 1999. All speakers are natives of the island of Guam and the group of speakers is composed of three women and three men with ages ranging from late twenties to mid sixties. The interviews were conducted by the authors of the present study using primarily translation-elicitation techniques to elicit Chamorro target items. All interviews took place in a quiet room and were recorded using a high-quality tie-clip microphone and a high-quality portable digital audio tape (DAT) recorder. The data from the DAT tapes were copied to analogue tape to provide a working copy of the material. The data from the analogue tapes were then digitised and analysed using standard software packages for speech analysis. The data presented below were obtained through auditory analysis of the digital sound files. This analysis was performed independently by the two authors of this study. The transcriptions of our data presented below follow IPA conventions. Consider now the data in (10) and (11).

\[(10) \text{ Root vowel } /i/ \]

\[\begin{align*}
(a) & \quad ['li?i?] \quad \text{‘to see’} & \quad ['li+li?i?] \quad \text{‘seer’} \\
(b) & \quad ['?i?i?] \quad \text{‘to slice’} & \quad ['?i+i?i?] \quad \text{‘slicer’} \\
(c) & \quad ['?istudja] \quad \text{‘to study’} & \quad ['?i+i?studja] \quad \text{‘studyer’}
\end{align*}\]
Fixed segmentism, markedness and faithfulness: Nominalising reduplication in Chamorro

(11) Root vowel /u/

(a) ['gupu] 'to fly' (i) [gi+gipu] 'flyer'
(b) ['tug!?] 'to write' (ii) [ti+tig!?] 'writer'

The data in (10) and (11) show that with high root vowels the reduplicant vowel is [i] regardless of whether the root vowel is front /i/ or back rounded /u/. Note that the base vowel in the initial syllable of the reduplicated form is high and front whenever the root vowel in the same position is high and front. However, we observe an optional alternation in the quality of the first base vowel if the corresponding root vowel is /u/. In such cases, /u/ may alternate with [i], as the comparison between (11) (a) (i) versus (11) (a) (ii) exemplifies. We suggest that this alternation in the base is analogous to the vowel fronting or umlaut alternation observed in the nonreduplicative phonology of Chamorro. Consider now the data in (12) and (13).

(12) Root vowel /e/

(a) ['?egga?] 'to watch' [?i+?egga?] 'watcher'
(b) ['peska?] 'to fish, to hunt' (i) [pi+piska?] 'fisherman'

(13) Root vowel /o/

(a) ['kørn?] 'to catch' (i) [ki+kørn?] 'catcher'
(ii) [ki+kønn?] 'catcher'
(iii) [ki+kønn?] 'catcher'
(iv) [ki+kønn?] 'catcher'
(b) ['kø?ko?] 'to harvest' (i) [ki+kø?ko?] 'harvester'
(ii) [ki+kø?ko?] 'harvester'

The data in (12) and (13) show that mid root vowels appear with the reduplicant vowel [i] in nominalising reduplication. In other words, mid vowels appear to be banned from the reduplicant and [i] is observed instead. Thus, the data in (12) and (13) show that roots with initial high and mid vowels tolerate only [i] in the reduplicant of nominalising reduplication. Furthermore, analogous to (11), the data in (13) also show optional vowel fronting of the back vowel in the reduplicative base. In particular, compare the front vowels found in the reduplicative bases in (13) (a) (i) and (ii) and (13) (b) (ii) with the corresponding back vowels in (13) (a) and (b). Moreover, the data in (12) and (13) also show systematic variation in the vowel height in the closed initial syllables of the reduplicative bases. Thus, the initial high vowels in the bases in (13) (a) (i) and (iii) alternate with the corresponding mid vowels in (13) (a) (ii) and (iv). We suggest that this alternation is analogous to the Vowel Lowering alternation observed in the nonreduplicative phonology of Chamorro. Note that the combined effects of optional Vowel Lowering and vowel fronting may produce four different variants of root /o/, as seen in (13) (a).

The upshot of the data in (10) through (13) is that the vowel in the prefixing nominalising reduplicant is [i] regardless of whether the initial root or base vowel is [i]/[i], [u]/[u], [e] or [ø]. Consider now the data on low vowels in (14) and (15).

(14) Root vowel /a/

(a) ['?adzik] 'to choose' ['?a+?adzik] 'chooser'
(b) ['?adda?] 'mimic' ['?a+?adda?] 'mimicker'
The data in (14) and (15) show that low root vowels appear with the reduplicant vowel [a] under nominalising reduplication. In other words, the front low vowel [a] reflects the low height of the root vowel in the nominalising reduplicant regardless of whether the root low vowel is front or back. Furthermore, note that the initial base vowel may appear as [a] or [œ]. There are two crucial points to be taken away from the data in (14) and (15). First, the lowness of the reduplicant vowel corresponds to the lowness of the initial root vowel, but not always to the vowel in the replicative base. Secondly, the reduplicant is low front [a] regardless of whether there is /a/ or /œ/ in the initial syllable of the root. This is in sharp contrast to the occurrence of [i] in the reduplicant with nonlow root or base vowels observed earlier.¹

The data in (10)-(15) show without a doubt that the vowels in the stress-attracting reduplicant of Chamorro nominalising reduplication are limited to [i] and [a]. In addition, the data in (10)-(13) show that [i] appears as the reduplicant vowel with nonlow root vowels, that is, /i/, /u/, /e/ and /o/. In contrast, the data in (14) and (15) show that the reduplicant vowel is [a] if the root vowel is a low vowel, that is, /a/ or /œ/. Furthermore, data such as (15) (d) show that the lowness of the reduplicant vowel is not always copied from the replicative base, but instead may be carried over from the input root. Thus, the six vowel system under primary stress in the nonreduplicative phonology of Chamorro is reduced to a two vowel system under primary stress in the nominalising reduplicant. Note in particular that reduplicant [i] occurs consistently with mid vowel roots, as the data in (12) and (13) show.² The table in (16) graphically displays these results.

(16) Vocalism of Chamorro nominalising reduplication

<table>
<thead>
<tr>
<th>Root</th>
<th>Reduplicant</th>
<th>Base</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>[i]/[i], [e], [u]/[u], [œ]</td>
<td></td>
</tr>
<tr>
<td>/a/, /œ/</td>
<td>[a], [œ]</td>
<td></td>
</tr>
</tbody>
</table>

The data in (10)-(15) have all been produced by the speakers of Chamorro without the determiner i. Thus, Topping’s idea that the segmental restrictions in the nominalising reduplicant are observed because vowel fronting or umlaut must be induced by the determiner i is shown to be unfounded. Furthermore, given the limitation of the vowels in the nominalising reduplicant to [i] and [a], the set of vowels produced by umlaut is at odds with the set of vowels appearing under fixed segmentism in nominalising reduplication. Thus, umlaut results in high, low, and mid front vowels whereas only high and low front vowels may occur in the reduplicant of nominalising reduplication. In sum, the fixed segmentism observed is an integral part of the morphophonological process of nominalising reduplication and not induced by umlaut. Interestingly, though, the nominalising reduplicant itself appears to be an umlaut-inducing element in that unumlauted vowels may appear in the initial syllable of the reduplicative base. Thus, the nominalising reduplicant needs to be

¹ Note that the variable appearance of [a] versus [œ] in the initial vowel in the replicative base corresponding to an initial low vowel in the root observed in (15) is probably due to variable secondary stress versus no stress, respectively, in adjacency to the main-stressed reduplicant. The details of the issue of the level of stress in replicative stems are left for future work.

² There is one exception to this generalisation in the corpus under consideration. Thus, [’tentalu?] ‘to interfere’ was nominalised as [’e+’tentalu?] ‘interferer’ in one instance.
Fixed segmentism, markedness and faithfulness: Nominalising reduplication in Chamorro

added to the list of umlaut-inducing elements compiled in the earlier literature on Chamorro umlaut (Topping 1968, Chung 1983, Klein 2000).

In sum, the data presented in this section show that the nonlow root vowels are optionally affected by umlaut or Vowel Lowering in the reduplicative base. In addition, low root vowels appear as [a] or may be reduced to schwa in the reduplicative base. Furthermore, we have seen that the marked reduplicant vowel [a] occurs in correspondence to low vowels in the initial root or base syllable. Finally, the unmarked reduplicant vowel [i] occurs elsewhere, that is with initial high or mid root vowels. Thus, markedness in the reduplicant vowel is decreased unless overruled by faithfulness to the input or the reduplicative base in terms of vowel lowness. These findings are the empirical backdrop for the OT analysis of fixed segmentism in Chamorro nominalising reduplication presented in the next section.

4. Fixed segmentism, markedness and faithfulness

The preceding description of the fixed vocalism in the reduplicant of Chamorro nominalising reduplication has shown that the vowel in the nominalising reduplicant is [a] if the first root or base vowel is a low vowel, but appears as [i] elsewhere. This pattern clearly incorporates a reduction in phonological markedness in that /u/ plus the cross-linguistically marked vowels /e/, /o/, and /a/ are banned from appearing in the nominalising reduplicant even though this reduplicant is under primary stress. To capture this pattern we propose that the appearance of [a] is due to two factors. First, the lowness of the vowel in the root or the base is matched in the reduplicant through word or base-reduplicant faithfulness, respectively. Secondly, /a/ is excluded through markedness so that only the front low vowel of the two low vowels is available to appear in the reduplicant. In all other environments, unmarked [i] is the default vowel in the reduplicant of nominalising reduplication.

This section is organised as follows. §4.1. explores the connection between the fixed segmentism observed in Chamorro nominalising reduplication, vowel inventories in other languages and the question of default vowel phonology in Chamorro. §4.2. is an initial demonstration of the effects of the markedness constraint hierarchy of Alderete et al. (1999) in relation to the markedness reduction observed in the vocalism of the nominalising reduplicant. In §4.3. we show the necessity of word faithfulness in the sense of Struijke (2000) and the usefulness of base-reduplicant faithfulness to account for the phonology of low vowels in nominalising reduplication. §4.4. rounds out the analysis by showing how unmarked [i] emerges as the reduplicant vowel in the correct environments.

4.1. Fixed segmentism, default phonology and vowel inventories

Alderete et al. (1999) base their framework on three generalisations concerning the relation between reduplicative morphemes and segmental inventories. These generalisations are reproduced in (17)-(19) (cf. Alderete et al. 1999: 332f.).

(17) Reduplicant/Inventory relation I

Except when copying the base, the reduplicant’s inventory is a proper subset of the whole language’s.

(18) Reduplicant/Inventory relation II

Any phonological restriction on the whole of one language is a possible restriction on the reduplicant of another language.

(19) Reduplicant/Inventory relation III

Any phonological restriction on the reduplicant of one language is a possible restriction on the whole of another language.

The case of fixed segmentism in Chamorro nominalising reduplication falls under the generalisation in (17) in that the vowel inventory of the nominalising reduplicant is a subset of the vowel inventory of Chamorro as a whole. Given the phonological nature of fixed
segmentism in Chamorro nominalising reduplication, the fact that only part of the Chamorro vowel inventory is observed in the nominalising reduplicant must be due to some kind of phonological restriction. Consequently, in accordance with the generalisations in (18) and (19), Alderete et al.’s (1999) framework predicts that the restriction of the Chamorro nominalising reduplicant to [i] and [a] must be matched as some aspect of the phonology of Chamorro itself or as a restriction on the non-reduplicative phonology of another language.

Alderete et al. (1999) discuss two general areas in which segmental restrictions on reduplicants may be met in the nonreduplicative phonology of the same language or some other language, namely, default segmentism and segmental inventories. Consequently, we investigate if the restriction to [i] and [a] is matched as some kind of default segmentism in Chamorro or if the restriction is mirrored as a vowel inventory in other languages.

The most obvious candidates for default vowel phonology are vowel insertion and vowel reduction under nonprimary stress. Chamorro is not attested to have any process of vowel insertion so that the vowel inventory of nominalising reduplication may not be compared to vowel insertion. As far as vowel reduction is concerned, we have seen in §1 that Chamorro displays stress reduction only in the inventory of low vowels that may occur under nonprimary stress. Thus, one option is for /a/ and /a/ to merge to [a] under secondary stress and to [a] in unstressed position. The other option for the low vowels is the reduction to [a] in unstressed position. Importantly, all high and mid vowels may occur under secondary stress and in unstressed position in the nonreduplicative phonology of Chamorro. Thus, the pattern of vowel reduction displayed in the nonreduplicative phonology of Chamorro does not match the vowel inventory of the nominalising reduplicant.

As far as the comparison between the vowel inventory under nominalising reduplication and the vowel inventories of other languages is concerned, there appear to be languages whose nonschwa vowel inventory matches the vowel inventory of the Chamorro nominalising reduplicant. Thus, the literature reports on languages in which the inventory of contrastive nonschwa vowels consists of just /i/ and /a/. Such vowel inventories have been reported for Margi (Ladefoged & Maddieson 1996: 286) and Iatmul and Abelam (Foley 1986: 49). In what follows, languages with this type of vowel inventory are referred to as MIA languages. It appears then, that the restricted vowel inventory of Chamorro nominalising reduplication is matched by the vowel inventory of MIA languages in that in both sets of inventories round or back vowels are not allowed and in that nonschwa vowels are maximally dispersed along the height dimension, resulting in the absence of noncentral mid vowels in both sets of inventories. Thus, it may be concluded that Alderete et al.’s (1999) generalisations (18) and (19) are confirmed in that the restrictions on the vowel inventory of Chamorro nominalising reduplication are mirrored as the vowel inventories of MIA languages and vice versa.

4.2. Markedness decrease in Chamorro fixed segmentism

As laid out above, the fixed segmentism in Chamorro nominalising reduplication is a markedness decrease in the sense that the unmarked vowel [i] occurs in the reduplicant unless this is overruled by faithfulness to a low vowel in the root or the base. In OT, segmental markedness is expressed through universal hierarchies of markedness constraints. The universal markedness hierarchy concerning place invoked by Alderete et al. (1999) is given in (20). Recall that Alderete et al. assume, following Clements & Hume (1995), that vowels bear the same place features as consonants.

(20) Place-markedness hierarchy (Alderete et al. 1999: 335)

\[*\text{Pi/LAB}, *\text{Pi/DORS} >> *\text{Pi/COR} >> *\text{Pi/PHAR}\]

According to the hierarchy in (20), round and back vowels are more marked than high front vowels which in turn are more marked than low vowels. Importantly, recall Alderete et al. follow the idea that mid vowels are specified by the features [Cor] and [Phar]. Thus, according to Alderete et al., the hierarchy in (20) always induces worse marks for any mid
vowel than its peripheral counterparts because mid vowels violate the constraints *PL/Cor and *PL/PHAR. Furthermore, Alderete et al. (1999: fn. 9) allude to the fact that dispersion effects also militate against mid vowels. For the purposes of this paper, we capture the markedness of mid vowels through the constraint *E/*O given in (21).

(21) *E/*O: Mid vowels are prohibited.

The tableau in (22) shows the effects of the hierarchy of markedness constraints from (20) and (21) on vowels in the reduplicant. In tableau (22), we abstract away from the segmental form of the base of reduplication to bring out the effects of the constraint hierarchy on reduplicant vowels more clearly.

The tableau in (22) shows the violations of the markedness constraints incurred by vowels, given the feature specifications of the Chamorro system of six vowel underliers laid out in §1. All round, back and mid vowels including [a] are suboptimal in this system because they violate the high-ranking constraints *PL/LAB, *PL/DORS or *E/*O, respectively. Furthermore, [a] violates *PL/Cor and *PL/PHAR and is, thus, suboptimal in comparison to [i] which only violates *PL/Cor. Thus, the markedness constraint hierarchy in (22) renders [i] optimal to the exclusion of all other Chamorro vowel underliers. This result is only partially desirable, as shown by the sad face next to viable candidate (22) (e). It captures the idea that [i] emerges in the reduplicant because it is the least marked underlier vowel of Chamorro. However, this system cannot license the appearance of [a] properly. Thus, the constraint hierarchy in (22) must be enhanced by constraints that favour [a] in the correct environments. We propose that two faithfulness relations are needed to capture the appearance of [a] in the reduplicant correctly. The first one demands a match in vowel lowness between input root and reduplicative word; the second one demands a match in vowel lowness between reduplicant and reduplicative base. We address these faithfulness relations in turn in the following section.

4.3. Word faithfulness and base-reduplicant faithfulness

The data presented in this paper have shown ample evidence that reduplicant vowels in Chamorro may be more faithful to the input root than to the output reduplicative base. Thus, in continuative aspect reduplication mid vowels appear in the reduplicant with mid vowel roots even though the reduplicative base may contain a high vowel in these cases (cf. ['kmm?], 'to catch, to take' versus ['ko+kunru?], 'catching, taking'). Furthermore, we have seen that the low vowel [a] appears in the reduplicant of nominalising reduplication with low vowel roots even though the reduplicative output base may contain a local schwa (cf. ['katsa?], 'to husk a coconut' versus ['ka+katsa?], 'coconut husker'). This type of pattern in which the reduplicant morpheme appears to be more faithful to the input root than to the output reduplicative stem has been the subject of extensive discussion in the recent literature (Raimy & Isard 1997, Spaetl 1997, Struijke 1998, 2000; see also Yip 2000). To a significant extent, these works converge on the idea that, in order to capture the type of input-reduplicative word faithfulness displayed in Chamorro and other languages, reduplication must be understood as the fission of an input root where faithfulness to the input may be satisfied in the whole of the reduplicative output, rather than just the
reduplicative base. This is in sharp distinction to the full model of correspondence in McCarthy & Prince (1995) in which input-reduplicant faithfulness is captured through the cross-morphemic correspondence between input and reduplicant (see Raimy & Idsardi 1997 and Spaelti 1997 for arguments against this approach). For the purposes of this paper, we invoke Struijke’s (2000) model in which there are three correspondence relations as displayed in fig. 2.

![Diagram of Correspondence]

BR faithfulness in fig. 2, is analogous to BR faithfulness in McCarthy & Prince (1995). The difference lies in the conception of root faithfulness and word faithfulness. Root faithfulness is satisfied if an input element is matched in the reduplicative base. In contrast, word faithfulness (cf. MAX\textsubscript{wd} in (23) below) is satisfied if an input element is matched anywhere in the word, that is, either in the reduplicant or in the base. A more specific faithfulness constraint utilising this correspondence relation is exemplified in (24) (cf. Struijke 2000).

(23) MAX\textsubscript{wd}: Every segment in the input has some correspondent in the output word.
(24) ID-FEATURE\textsubscript{wd}: If a segment S is [αF] in the input, then some correspondent of S is [αF] in the output.

Given constraints of the general form in (24), it is clear that reduplicative items such as ['ka-katsɔ?] 'coconut husker' formed from the root ['katsɔ?] 'to husk a coconut' satisfy a constraint that demands matching low vowels in input and reduplicative word. This constraint is formulated in (25).

(25) ID-[PHAR]\textsubscript{wd}: If a segment S is [PHAR] in the input, then some correspondent of S is [PHAR] in the output.

The constraint in (25) demands that headedness by the feature [PHAR] in any given input segment must be matched as headedness by the feature [PHAR] in some segment of the output word. While this constraint is shown below to contribute significantly to the account of low vowel matching in Chamorro nominalising reduplication, it is obvious that this constraint captures only a subset of the input-word correspondence relations in Chamorro continuative aspect reduplication. For reasons of space, we leave the analysis of input-word correspondence in nominalising versus continuative aspect reduplication as a task for future

94
Fixed segmentism, markedness and faithfulness: Nominalising reduplication in Chamorro

research and, consequently, focus on the analysis of nominalising reduplication in the remainder of this paper.

Recall from tableau (22) in section 4.2. that the place-markedness hierarchy by itself favours fixed [i] to the exclusion of matching [a]. Thus, the task is to show how the proper ranking of the input-word correspondence constraint ID-[PHAR]_{wd} contributes to the account of faithfulness to the lowness of input vowels in nominalising reduplication.

(26) Input: RED+/katsaʔ/ *Pl/DOR ID-[PHAR]_{wd} *Pl/Cor *Pl/PHAR
a. [ka\textsubscript{Rec}+katsaʔ] * * *
b. [ki\textsubscript{Rec}+katsaʔ] * * *
c. [ka\textsubscript{Rec}+katsaʔ] **! * *

All candidates in (26) incur a violation of *Pl/DOR because of the faithful rendition of the final input low back vowel. However, the low back vowel in the reduplicant in (26) (c) is suboptimal because it incurs an additional, fatal violation of *Pl/DOR. Reduplicant [i] as in (26) (b) is suboptimal with low vowel roots because it incurs a fatal violation of the constraint ID-[PHAR]_{wd}. Thus, ID-[PHAR]_{wd} is violated in (26) (b) because the lowness of the first root vowel does not have a match anywhere in the reduplicated word. In sharp contrast, (26) (a) passes ID-[PHAR]_{wd} because the lowness of the first root vowel /a/ is matched through the lowness of the reduplicant vowel [a]. Note that the change along the horizontal dimension is forced in the reduplicant through the high ranking of the place-markedness constraint *Pl/DOR. In short, [a] is correctly favoured over [i] as the reduplicant vowel with low input vowels because it incurs the fewest violations of the high-ranking constraints *Pl/DOR and satisfies the constraint ID-[PHAR]_{wd}, thus overriding the unmarkedness of [i].

The constraint hierarchy motivated to this point is not able to account for the preference for the reduplicant vowel [a] in cases in which the low vowel of the root surfaces as a low vowel in the reduplicative base. Consider tableau (27).

(27) Input: RED+/napasi/ ID-[PHAR]_{wd} *Pl/Cor *Pl/PHAR
a. ['la\textsubscript{Rec}+napasi] *** ***!
b. ['la\textsubscript{Rec}+napasi] *** ***

If all low vowels of the input root are matched as low vowels in the reduplicative base, such candidates pass the constraint ID-[PHAR]_{wd}, leaving the decision between the reduplicant vowels [i] and [a] to the place-markedness constraints. Given that the place-markedness hierarchy favours [i] over [a], the incorrect result obtains with low vowel roots, as shown through the comparison of (27) (a) with (27) (b). To resolve this situation we invoke the general faithfulness constraint MAXBR (cf. McCarthy & Prince 1995).

(28) MAXBR: Every segment of the base has a correspondent in the reduplicant.

The constraint MAXBR demands that all segments in base and reduplicant must match. Consider now the tableau in (29).

(29) Input: RED+/napasi/ ID-[PHAR]_{wd} *Pl/Cor MAXBR *Pl/PHAR
a. [napasi] *** ***!
b. [napasi] *** *

Both candidates in tableau (29) pass the constraint ID-[PHAR]_{wd} because the low vowels in the input root appear as low vowels in the reduplicative base. Furthermore, both candidates
in (29) violate the constraint *PL/Cor in equal measure because they contain three front vowels each. Crucially, candidate (29) (b) fatally violates the constraint MaxBR because the vowel in the CV reduplicant in (29) (b) does not match the appropriate vowel in the base. Consequently, candidate (29) (a) correctly emerges as optimal even though its three low vowels incur more violations of the constraint *PL/Phar than its competitor (29) (b).

In the preceding paragraphs we have shown that the appearance of the reduplicant vowel [a] with low vowel roots must be accounted for through correspondence relations between the input and the reduplicative word and between the reduplicant and the base. It remains to be shown how [i] emerges in this system as the reduplicant vowel with roots and bases containing nonlow vowels.

4.4. The emergence of unmarked [i]

In this subsection we show that [i] emerges as the reduplicant vowel with nonlow vowel roots and bases because the general MaxBR constraint requiring faithfulness of all segments in the base-reduplicant relation is ranked quite low.

\[
(30) \begin{array}{|llll|}
\hline
\text{Input:} & *EJ*o & \text{ID-[PHAR]}_\text{WD} & *\text{PL/Cor} & \text{MaxBR} & *\text{PL/Phar} \\
\hline
\text{RED+} & \text{peska/} & & & & \\
\hline
\text{a. } & \text{pl}_{\text{Red}}+\text{peska} & * & ** & * & ** \\
\text{b. } & \text{pe}_{\text{Red}}+\text{peska} & **! & ** & *** & \\
\text{c. } & \text{pa}_{\text{Red}}+\text{peska} & * & ** & * & **! \\
\hline
\end{array}
\]

The reduplicant vowel [e] with roots or bases containing initial front mid vowels is correctly suboptimal with nominalising reduplication through the high ranking of the markedness constraint *EJ*o, as shown in candidate (30) (b). The comparison between candidate (30) (a) and (30) (c) shows that the decision between these two candidates cannot be made by the constraints ID-[PHAR]_WD, *PL/Cor or MaxBR. Both candidates pass ID-[PHAR]_WD because the input low vowel is represented faithfully in the output. Furthermore, both candidates violate *PL/Cor equally because they contain the same number of front vowels. Moreover, (30) (a) and (30) (c) violate MaxBR equally because the initial base vowel does not match the reduplicant vowel in either of the candidates. Thus, the decision is passed along to the low-ranked markedness constraint *PL/Phar. Candidate (30) (c) incurs more violations of this constraint than (30) (a) because the low reduplicant vowel in (30) (e) incurs a violation of *PL/Phar that is not present for the reduplicant vowel [i] in (30) (a). Thus, candidate (30) (a) correctly emerges as optimal. For the sake of completeness, tableaux for nominalising reduplication with the remaining input vowels are presented in (31) to (33).

\[
(31) \begin{array}{|llll|}
\hline
\text{Input:} & *\text{PL/Cor} & \text{MaxBR} & *\text{PL/Phar} \\
\hline
\text{RED+} & \text{li?i?} & & & \\
\hline
\text{a. } & \text{li}_{\text{Red}}+\text{li?i?} & *** & *! & * \\
\text{b. } & \text{la}_{\text{Red}}+\text{li?i?} & *** & *! & * \\
\hline
\end{array}
\]

\[
(32) \begin{array}{|llll|}
\hline
\text{Input:} & *\text{PL/Cor} & \text{MaxBR} & *\text{PL/Phar} \\
\hline
\text{RED+} & \text{gupu/} & & & \\
\hline
\text{a. } & \text{gl}_{\text{Red}}+\text{gupu} & * & * & *! \\
\text{b. } & \text{ga}_{\text{Red}}+\text{gupu} & * & * & *! \\
\hline
\end{array}
\]

\[3 \text{ Note that our analysis abstracts away from a number of issues in nominalising reduplication peripheral to our present concerns including the CV shape of the nominalising reduplicant and its stress-attracting nature.}\]
Fixed segmentism, markedness and faithfulness: Nominalising reduplication in Chamorro

The tableaux in (31)-(33) demonstrate that the reduplicant vowel [i] is always favoured over the reduplicant vowel [a] with roots and bases with nonlow vowels either because reduplicant [a] fatally violates MAXBR or because it fatally violates *PL/PHAR.

Given the discussion in this paper, the phonology of the Chamorro nominalising reduplicant vowel is accounted for through the constraint hierarchy given in (34).

(34) Summary of constraint hierarchy


Given high-ranking constraint(s) ensuring that consonants are copied faithfully in Chamorro reduplication, the high ranking of *PL/LAB, *PL/DORS and *E/*O eliminates round, back and mid vowels from the nominalising reduplicant. The proper ranking of ID-ΦHAR \_wD accounts for the fact that the reduplicant vowel is low with low vowel roots even if the adjacent base vowel is reduced to schwa. Given the feature specifications of Chamorro vowels in §1, the constraint *PL/ΦOR cannot decide between the reduplicant vowels [i] and [a]. The decision for a matching reduplicant vowel falls to MAXBR if the base vowels are low or high and front. In the remaining cases, *PL/ΦHAR favours [i] as the reduplicant vowel because [a] incurs a violation of *PL/ΦHAR that [i] does not.

5. Conclusion

The data from our fieldwork presented in this paper have contributed towards a complete description of Chamorro nominalising reduplication and have provided ample evidence for fixed segmentism in nominalising reduplication. Thus, the vowel in the prefixing reduplicant of nominalising reduplication is low front [a] if the first vowel of the input root or the output base is a low vowel, whereas the vowel in the reduplicant is [i] if the first vowel in the root or the base is a nonlow vowel. Furthermore, in line with the predictions of Alderete et al. (1999), the restriction of the vowel inventory of the Chamorro nominalising reduplicant to [i] and [a] has been shown to be matched as nonschwa vowel inventories in the phonology of other languages such as Margi, Iatmul and Abelam. In addition, in contrast to earlier descriptions, we have shown that the vocalism of the nominalising reduplicant is not caused by umlaut/vowel fronting and does not depend on the presence of the determiner l. Instead, the vocalism of the reduplicant in nominalising reduplication has been shown to be an integral part of the morphophonological process of reduplication. Furthermore, the vocalism of the base in nominalising reduplication has been shown to be influenced by Vowel Lowering and by umlaut induced by the reduplicant itself.

The fixed segmentism in Chamorro nominalising reduplication has been shown to be a case of markedness reduction with some interesting complexities. Markedness straightforwardly excludes back, round and mid vowels from the stress-attracting nominalising reduplicant although such vowels are part of the phonology of Chamorro elsewhere. However, the appearance of [a] versus [i] in the reduplicant depends on the height of the root or base vowel. Thus, the cross-linguistically marked low front vowel [a] appears if there is a local low vowel in the root or the base, whereas the unmarked vowel [i] appears only if there is no local low vowel in the input or in the base of reduplication. In this respect, the preference for reduplicant [i] is a case of the emergence of the unmarked in the sense of McCarthy & Prince (1994).

The appearance of [a] in nominalising reduplication has been shown to be cued by faithfulness to the reduplicative base and to the input to reduplication. Whereas the model proposed in Alderete et al. (1999) has been shown to be well-suited to account for the
markedness reduction observed in the fixed segmentism of Chamorro nominalising reduplication, their model cannot account for the fact that reduplicant vowels in Chamorro under certain conditions take their cue from the input to reduplication rather than the reduplicative output. We have proposed to resolve this issue by integrating recent models of correspondence theory, in particular Struijk (2000), into the markedness reduction framework for fixed segmentism. Thus, the proper ranking of a word faithfulness constraint accounts for the appearance of [a] in the nominalising reduplicant even if there is no local low vowel in the reduplicative output. Clearly, the relation between fixed segmentism, markedness and faithfulness needs to examined further in future work. Other issues arising from the present paper that are well worth further investigation include the account of umlaut and Vowel Lowering in the reduplicative base of nominalising reduplication, the analysis of stress in reduplication and the analysis of continuative aspect reduplication in Chamorro.

References


Authors' address:
Department of Linguistics
The University of Manchester
Oxford Road
Manchester M13 9PL
UK

email:
thomas.klein@man.ac.uk (http://lings.ln.man.ac.uk/Html/TK/)
meta.y.harris@man.ac.uk
An Interpretation of the Voice Affix /i-/ in Tagalog

Anja Latrouite and Ralf Naumann
Seminar für Allgemeine Sprachwissenschaft
University of Düsseldorf
Germany

1 Introduction

In Tagalog, the main dialect spoken in the Philippines, at least five voice affixes (VA) can be distinguished: the two Actor voice (AV) affixes /um-/ and /mag-/ that are both prefixes, and the three Undergoer voice (UV) affixes /-in/ (UV₁), /i-/ (UV₂) and /-an/ (UV₃), the first and third of which are suffixes whereas /i-/ is a prefix.

Voice Affixes

<table>
<thead>
<tr>
<th>ACTOR VOICE (AV)</th>
<th>UNDERGOER VOICE (UV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/um-, /mag-/</td>
<td>/-in/ (UV₁), /i-/ (UV₂), /-an/ (UV₃)</td>
</tr>
</tbody>
</table>

A (verb-) stem v combines with a VA to build a complex predicate. This combination is subject to certain restrictions, i.e., not each VA is admissible for each stem. Furthermore, both the argument structure (AS) of a stem and its interpretation possibly depends on the VA with which it combines. These phenomena must be explained by any theory of the VAs in Tagalog.

2 The Data

2.1 Transfer-Verbs

According to Foley 1976, the VA /i-/ indicates a nominal which is an intermediary in the transfer of the main action of the clause between the two other nominals. This characterization applies to stems like /bigay/ ('give'), /turo/ ('teach') and /tanim/ ('plant'), elements of which will be said to belong to the 'bigay'-type in the sequel.
Anja Latrouite and Ralf Naumann

| a. | I-b-in-igay | ng babae | ang liham | sa kapit-bahay |
|    | UV₂-bₙₑₙ-REALIS-give | NOM woman | GEN letter | DAT neighbour |
| The woman gave the letter to the neighbour |
| b. | I-t-in-uro | ng titser | ang Pilosopya | sa akin |
| UV₂-tₙₑₙ-REALIS-teach | GEN teacher | NOM philosophy | I.DAT |
| The teacher taught me philosophy |
| c. | I-tanim | mo | sa hardin | ang mga bulaklak! |
| UV₂-plant | you.GEN | DAT garden | NOM PL flower |
| Plant the flowers in the garden! |

For this type of stem only the UV affixes /i-/ and /-an/ are admissible. The UV /-in/ is excluded. There is no meaning variance (i.e., in the English translation the same verb is used) and the AS is the same for both UV affixes that are admissible.

Foley’s characterization of /i-/ does not distinguish different sorts of transfer. In particular, it does not account for Pittman’s (1966) observation that in Tagalog centripetal and centrifugal actions can be distinguished. Pittman applied this distinction to the interpretation of the AV affixes /um-/ and /mag-/.¹ But it can equally be used for an at least partial characterization of the UV affixes /i-/ and /-in/.

- the admissibility of /i-/ is related to centripetal action: /i-/ identifies the argument denoting the object moving away from the Actor as the subject of the clause
- the admissibility of /-in/ is related to centrifugal action: /-in/ identifies the object that is moved towards the Actor as the subject

An example for a stem that expresses a centrifugal action is /kuha/ (‘take’). In contrast to elements of the ‘bigay’-type, which express a centripetal action, /-in/ and /-an/ are admissible whereas /i-/ is excluded. The ‘kuha’-type is similar to the ‘bigay’-type in showing neither meaning variance nor changes of AS, depending on the VA with which a stem combines.

(3) a. Kun-in mo sa kaniya ang lapis |
Take-UV₁ you.GEN he.DAT NOM pencil
Get the pencil from him!

b. Kun-an mo siya ng lapis |
Take-UV₃ you.GEN he.NOM GEN pencil
Get a pencil from him!

¹See Latrouite (1999) for extensive discussion.
An Interpretation of the Voice Affix /i-/ in Tagalog

This difference with respect to the admissibility is not accounted for by Foley's characterization, according to which the 'bigay'- and the 'kuha'-type should both admit /i- / because there always is an intermediary (a transferred object).

2.2 Two Sorts of Transfer

In this section the two sorts of transfer, centripetal and centrifugal, will informally be defined. In the sequel the following convention is used: x = Actor; y = transferred object; z = destination/source of transfer (where it is required that z ≠ x, i.e. z is not identical to the Actor). Let Q be the property such that s ∈ Q(a)(b) just in case a has b at s, where 'has' is to be understood in a broad sense that not only covers possession but also knowledge (a knows b) and location (a is in/at b).

For a centripetal transfer one gets the following three possibilities that are admissible (α(e) is the beginning and ω(e) the end of an event that brings about the transfer).

\[
\begin{align*}
\text{α(e)} & \quad \text{ω(e)} \\
(i) & \quad \in Q(x)(y) \quad \notin Q(x)(y) \quad /\text{bigay}/ \text{(physical movement)} \\
(ii) & \quad \in Q(x)(y) \quad \in Q(x)(y) \quad /\text{turo}/ \\
(iii) & \quad \notin Q(x)(y) \quad \notin Q(x)(y) \quad /\text{tanim}/
\end{align*}
\]

Cases (ii) and (iii) do not involve any change between x and y because the condition is an invariant. E.g., if x teaches z Philosophy, then x knows Philosophy at the beginning of the event (otherwise x could not teach it). But x does not loose this knowledge in the sense that at the end of the event z knows Philosophy whereas x has lost his knowledge on that particular topic. Examples of case (iii) are events of type 'tanim' where the Actor does not have the flowers that he is going to plant in the garden at the beginning of the event but begins by taking them. Another way of executing events of this type are subsumed by case (i). The Actor already has the object to be planted at the beginning.

A centrifugal transfer is characterized by the remaining fourth possibility.

\[
\text{α(e)} \notin Q(x)(y) \quad \omega(e) \in Q(x)(y) \quad '\text{kuha}'-\text{type}
\]

2.3 Further types of Stems admitting /i-/ 

Even if one distinguishes between two sorts of transfer, this does not capture all generalizations concerning /i-/i-. Consider the following data.²

(i) there are stems that admit all three UV-affixes

/akyat/ (go up), /hagis/ (throw), /halo/ (mix)

²Many of these data are taken from Himmelmann 1987.
(4) a. Akyat-in mo ang kanya-ng kuwarto
Go-up-UV$_1$ you.GEN NOM he-LINKER room
Go up (upstairs) to his room
b. Akyat-an mo ang kaniya-ng kuwarto ng mga libro
Go-up-UV$_3$ you.GEN NOM he-LINKER room GEN PL book
Bring the books up (upstairs) to his room!
c. I-akyat mo ang mga libro
UV$_2$-bring up you.GEN NOM PL book
Bring the books up(stairs)

For /akyat/ one gets both meaning variance and change of argument structure: /-in/: transitive; /-i/, /-an/ ditransitive.

(5) a. Hagis-in mo siya ng sapatos
Throw-UV$_1$ you.GEN he.NOM GEN shoe
Throw the shoe at him!
b. Hagis-an mo ako ng sapatos
Throw-UV$_3$ you.GEN I.NOM GEN shoe
Throw a show to me/in front of me!
c. I-hagis mo sa akin ang sapatos
UV$_2$-throw you.GEN I.DAT NOM shoe
Throw a show at me!

In contrast to /akyat/, /hagis/ shows neither meaning variance nor change of AS.

(iiia) there are stems that, like /bigay/, admit only /i-/ and /-an/ but with MV and change of AS

(6) a. Hiwalay-an mo ang iba-ng pasyente
Leave-UV$_3$ you.GEN NOM other-LINKER patient
Leave (you) the other patients
b. I-hiwalay mo siya sa iba-ng pasyente
UV$_2$-separate you.GEN he.NOM DAT other-LINKER patient
Separate (isolate) him from the other patients!

(iiib) there are stems that, like /bigay/, admit only /i-/ and /-an/ and for which there is neither meaning variance nor a change of AS but which do not express a transfer. An example is /bukas/ ('open')

(7) a. I-bukas mo ang pinto
UV$_2$-open you.GEN NOM door
Open the door!
b. Biks-an mo ang pinto
UV$_3$-open you.GEN NOM door
Open the door!
An Interpretation of the Voice Affix /i-/ in Tagalog

(iii) there are stems that admit only /i-/ but neither /-in/ nor /-an/: /bagsak/ ('drop/slam down'), /ikot/ ('turn'), /tungo/ ('bend') and /takbo/ ('run', 'take s.th. on the run')

(8) a. B-um-agsak ang baso
   Bstem-AV/REALIS-drop NOM vase
   The vase fell/dropped
b. I-b-in-agsak ng bata ang baso
   UV2-bstem-REALIS-drop GEN child NOM vase
   The child slammed down the vase

(9) a. T-um-akbo ako
   Tstem-AV-run I.NOM
   I run
b. I-takbo mo ang gamot
   UV2-run you.GEN NOM medicine
   Run (away) with the medicine

These stems show both meaning variance and change of AS between AV- and UV-affix: with /um-/ they are intransitive, whereas with /i-/ they are transitive. Similarly to the 'bukas'-type, no third argument is realized such that there is no object that is transferred to a destination different from the Actor.

(iv) there are stems for which /-in/ and /li-/ are admissible but not /-an/ and which show neither meaning variance nor change of AS. Examples are /saksak/ ('stab') and Point-verbs like /puknok/ ('hit') and /hampas/ ('hit', 'beat')

(10) a. Saksak-in mo siya
    Stab-UV1 you.GEN he.NOM
    I stabbed him!
b. I-s-in-aksak ko sa kaniya ang lanseta
    UV2-sstem-REALIS-stab I.GEN he.DAT NOM folding knife
    I stabbed him with a folding knife

(v) there are verbs that allow all three UG affixes, yet for which /i-/ is admissible only if the subject denotes the instrument with which the action is executed.

(11) I-kain mo ng isda ang iyong kutsara
    UV2-eat you.GEN GEN fish NOM-this-spoon
    Eat the fish with the spoon

The stems belonging to either (i), (iiia) and (iv) can be interpreted as expressing a centripetal transfer if they combine with /i-/ . This is shown in (12) by relating the stems to the three cases characterizing a centripetal transfer and in (13) for the specific examples given above.

(12) a. /hagis/: the Actor has the transferred object at the beginning of the event but not at its end. This corresponds to the first case of a centripetal transfer
Anja Latrouite and Ralf Naumann

b. /akyat/, /saksak/: they are covered by case (ii) of a centripetal transfer. The Actor has the transferred object continuously from the beginning until the end of the event.

c. /halo/, /hiwalay/: these two stems belong to the third case of a centripetal transfer. The Actor has the transferred object y neither at the beginning nor at the end of the event (i.e., (s)he does not give away y to some destination).

(13) a. /hagis/: the object thrown is transferred from the Actor to the place (or person) to (at) which it is thrown (or the destination is left implicit).

b. /akyat/: the books are transferred to the destination (i.e. the room).

c. /saksak/: the object stabbed, e.g. a knife, is transferred from a place outside the body of person z into z’s body such that z functions as the destination.

d. /halo/: the stuff is transferred from some source location to the stuff it is mixed in as destination.

e. /hiwalay/: the patient who is separated from the others is transferred from the latter to some destination that is left implicit.

This interpretation is possible because the stem is always used ditransitively. From this it does not follow that whenever a stem is used ditransitively if it combines with /i-/ it expresses a centripetal transfer. Counterexamples to this generalization are stems like /kain/, listed in (v), where the third argument is the instrument that is used to carry out the action.

The stems discussed in (12) and (13) become problematic if one considers the admissibility with other UV-affixes and the possible changes both at the level of meaning and at the level of argument structure. E.g., if it combines with /-in/, the stem /akyat/ no longer expresses a transfer because it is used transitively and not ditransitively. This shows that admissibility with respect to this voice affix cannot be defined in terms of a centripetal action. They therefore constitute the same type of counterevidence as the stems in (iib) and (iii). The importance of these counterexamples can be taken to consist in showing that whenever a stem v expresses a transfer only for certain stems that are admissible for it but not for all stems, the (dynamic) meaning of v cannot be reduced to some sort of transfer but must be taken as something more general that subsumes a transfer as a special case. Such a generalization seems exactly what is needed in order to solve the problem that pose stems admitting /i-/ that do not express a transfer because they are used transitively.

The conclusion that must be drawn from the above discussion is that the notion of a centripetal transfer does not capture all uses of /i-/ if the notion of transfer is understood as involving three objects: a source, a destination and a transferred object. Counterexamples to this conception are stems discussed in (iib), (iii) and (v). In (14) the behaviour of /i-/ is summarized.

(14) The UV-affix /i-/ is admissible for a stem v if one of the following conditions is satisfied.

(i) an object y is transferred to some destination (location) that is not the Actor (‘bigay’-type, centripetal transfer); in this case the stem v is used ditransitively; the subject determined by /i-/ is the transferred object.
(ii) the stem does not express a transfer
two subcases must be distinguished
   a. the action inherently involves an instrument (/kain/); the stem v is again used
ditransitively; the subject determined is the instrument
   b. stems covered by (iib) and (iii); only in this case the stem v is used transi­
tively; the subject is the non-Actor argument

What is needed is a generalization of the notion of a centripetal Transfer that also captures
cases (14iia) and (14iib).

3 Dynamic Event Semantics (DES)

3.1 Changes as Objects and Changes as Transformations of States

Dynamic Event Semantics (DES), Naumann 1998, to appear, is based on the intuition that
non-stative verbs express changes. The intuitive notion of a change can be made precise in
two different ways. It is either (i) an object (action, event), or (ii) a transformation of state,
i.e., a state \( s \) at which a result \( Q \) does not hold is transformed into a state \( s' \) at which \( Q \) holds.
Two aspects of a transformation of state can be distinguished: (a) a result \( Q \) that is brought
about and (b) the way the result \( Q \) is brought about. This idea is made precise in DES by
having both an eventuality-structure \( E \) and a transition-structure \( S \).

A basic eventuality-structure \( E \) is a tuple \( \langle E, \Xi_E, \{P_v | v \in \text{VERB}\} \rangle \) such that
- \( E \) is the domain of events
- \( \Xi_E \) is the material part of relation on \( E \)
- each \( P_v \) is a subset of \( e \); \( P_v \) is the set of all events of type \( P_v \); e.g., for \( v = \text{eat} \), \( P_v \) is the
set of all eating events

A basic transition-structure \( S \) is a tuple \( \langle S, \prec_S \rangle \) such that
- \( S \) is the domain of states (or time points)
- \( \prec_S \) is the precedence (earlier) relation on \( S \)

The two structures are related by two total functions \( \alpha : E \rightarrow S \) and \( \omega : E \rightarrow S \) that assign
to each event \( e \in E \) its beginning- and end point, respectively. Together, \( \alpha \) and \( \omega \) determine \( e \)'s execution sequence \( \tau(e) = \{s \in S | \alpha(e) \preceq_S s \preceq_S \omega(e)\} \).
This sequence can be split into three parts corresponding to a nucleus-structure (Moens/Steedman (1988)).
There is the inception-point (IP), the development-portion (DP) and the culmi­nation-point (CP)
\( (\alpha(e) = s, \omega(e) = s') \).

\[
\begin{array}{cccc}
 s & - & - & - & - & - & - & s' \\
 IP & DP & CP & \text{Nucleus-Structure}
\end{array}
\]

\( IP \) (inception-point) = \( [s] \), \( DP \) (development-portion) = \( ]s, s'[ \), \( CP \) (culmination-point) = \( [s'] \)
3.2 Types of Results

Each event-type $P_v$ determines for each of its elements $e$ a set $Res(P_v, e)$ of results that $e$ can possibly bring about. Two principle types of results are distinguished: minimal and maximal ones. They are informally defined in (15) (For a formal definition, see Naumann/Latrouite 1999 and Latrouite/Naumann 1999b).

(15) a. minimal results possibly hold at intermediate states of the execution sequence of an event
b. maximal results only hold at the end point of an event (if they hold there at all)

The two types of results are illustrated by an event of type 'John eat a fish'. Minimal results are: (i) partial decrease of the mass of the fish and (ii) part of the mass of the fish is in John's stomach. Results which are maximal are: (i) the mass of the fish is zero (complete decrease of its mass) and (ii) the (complete) mass of the fish is in John's stomach. Minimal and maximal results together are the v-closed results. Intuitively, a result is v-closed if bringing it about on an initial stage $e'$ of an event $e$ that belongs to $P_v$ is sufficient for $e'$ to belong to $P_v$ too.

The types of results are temporarily ordered by $\leq_v$: if $Q \leq_v Q'$, this means that whenever $Q'$ is brought about on the execution sequence of an event $e \in P_v$, then $Q$ has already been brought about, i.e., $Q$ is brought about not later than $Q'$. Results that are distinct but which are brought about simultaneously, can be distinguished by implicational relations; this yields the more fine grained ordering $\leq_v^\ast$: $Q \leq_v^\ast Q'$ just in case $Q <_v Q'$ or $[Q =_v Q' \wedge Q' \leq Q]$.

3.3 Sorts of Event-Types

Minimal and maximal results do not exclude each other. The exact relationship depends on the sort of event-type to which $P_v$ belongs. Two basic sorts are distinguished:

(16) a. P-atomic event-types
   $$\forall P[P - ATOMIC(P) \iff \forall e[e \in P \to \neg \exists e'[prefix(e', e) \wedge e' \in P]]]$$

   an event-type $P$ is P-atomic just in case for each of its elements $e$ no proper initial stage (prefix) $e'$ belongs to $P$

b. non-atomic event-types
   $$\forall P[NON - ATOMIC(P) \iff \neg P - ATOMIC(P) \wedge \forall e[e \in P \to \alpha(e) <_S \omega(e)]]$$

   an event-type $P$ is non-atomic just in case none of its elements is point-like and there is at least one element which has a proper initial stage $e'$ that is of type $P$ too

Examples for P-atomic event-types are those corresponding to the stems /kuha/ and /bimagay/. Non-atomic event-types are those corresponding to stems of sort Accomplishment like /kain/, /kanta/ ('sing') and /bukas/.

For P-atomic event-types, the distinction between minimal and maximal results collapses. Each minimal result is maximal and vice versa. For non-atomic event-types minimal results are never maximal. These possibilities lead to the following definitions of further types of results in terms of the two basic ones.
An Interpretation of the Voice Affix /i-/ in Tagalog

(17) a. A result \( Q \) is \( w \)-maximal just in case it is both minimal and maximal
b. A result \( Q \) is \( s \)-maximal just in case it is maximal but not minimal
c. A result \( Q \) is \( w \)-minimal just in case it is minimal but not maximal

From what has been said it follows that P-atomic event-types determine only \( w \)-maximal results, whereas Non-atomic event-types (corresponding to non-stative stems) determine either only \( w \)-minimal results or \( w \)-minimal and \( s \)-maximal results. In Table 1 the types of results for three classes together with the ordering in terms of \( \leq_v^* \) are depicted.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>types of results determined</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCO</td>
<td>( w )-min ( \leq_v^* ) (( \leq_v )) ( s )-max</td>
</tr>
<tr>
<td>ACT</td>
<td>( w )-min</td>
</tr>
<tr>
<td>/kuha/, /bigay/</td>
<td>( w )-max (2); ( w )-max ( 1 ) ( \leq_v^* ) ( w )-max ( 2 )</td>
</tr>
</tbody>
</table>

Table 2 illustrates the assignment of results to different arguments whereas Table 3 gives examples of different types of results for various stems together with the assignment to arguments (Results that are minimal, \( non-\_v \), are brought about before \( w \)-minimal results and are not \( v \)-closed, i.e., bringing about only results of this type is not sufficient for an event to be of type \( v \)).

<table>
<thead>
<tr>
<th>Table 2</th>
<th></th>
<th>( x )</th>
<th>( y )</th>
<th>( z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCO</td>
<td>/kain/</td>
<td>( w )-min; ( s )-max ( 2 )</td>
<td>( w )-min</td>
<td>-</td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
<td>( w )-min</td>
<td>( s )-max ( 1 ); ( s )-max ( 2 )</td>
<td></td>
</tr>
<tr>
<td>sort</td>
<td>/kuha/</td>
<td>( w )-max ( 2 )</td>
<td>( w )-max ( 1 )</td>
<td>( w )-max ( 1 )</td>
</tr>
<tr>
<td>sort</td>
<td>/bigay/</td>
<td>( w )-max ( 1 )</td>
<td>( w )-max ( 1 )</td>
<td>( w )-max ( 2 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>( min_{_non-_v} )</th>
<th>( w )-min.</th>
<th>( w )-max. or ( s )-max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>eat an ( y )</td>
<td>initial</td>
<td>partial decrease of ( y )</td>
<td>mass of ( y = 0 ) (( s )-max ( 1 ))</td>
</tr>
<tr>
<td>push ( y )</td>
<td>actions</td>
<td>( y ) traverses a path</td>
<td>mass of ( y ) is in ( x )'s stomach (( s )-max ( 2 ))</td>
</tr>
<tr>
<td>( x ) takes ( y ) from ( z )</td>
<td>performed</td>
<td>-</td>
<td>( x ) has ( y ) (( w )-max ( 2 ))</td>
</tr>
<tr>
<td>( x ) gives ( y ) ( z ) (involving physical action)</td>
<td>by the</td>
<td>-</td>
<td>( z ) does not have ( y ) (( w )-max ( 1 ))</td>
</tr>
<tr>
<td>( x )</td>
<td>Actor</td>
<td>-</td>
<td>( z ) has ( y ) (( w )-max ( 2 ))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( x ) does not have ( y ) (( w )-max ( 1 ))</td>
</tr>
</tbody>
</table>
4 Generalizing the Notion of Transfer

From the perspective of DES a transfer, either a centripetal or a centrifugal one, is an event e that brings about a number of results that are determined by the event-type P, to which e belongs. E.g., for an event e of type ‘bigay’ the two principle (v-closed) results are that the Actor x ceases to have the Transferred Object y whereas the Destination z starts to have y at the end of the event (see Table 3). Both results are of type w-maximal and are therefore maximal. As these results are required to hold at the end point ω(e) of e, it follows that they only hold at that point of e’s execution sequence τ(e) and at no other point. This means that these results are brought about last. They are therefore maximal elements of Res(Pv, e) relative to the ordering ≤v. Results which satisfy this condition will be called strongest results. The exact definition is given in (18) (Q a set of results).3

\[ (18) \ \forall Q \forall Q'(\text{Strongest}_v(Q, Q) \iff Q \in Q \land \forall Q'(Q' \in Q \rightarrow Q' \leq_v Q)) \]

Maximal results are always strongest results. The inverse need not be true. E.g., an Activity-verb like ‘push’ in English or the stem /asín/ (‘salt’) in Tagalog determine only (w-) minimal but no maximal results. Consequently, the strongest results of Res(Pv, e) for v = push are (w-) minimal and not maximal.

Strongest results are maximal with respect to ≤v. Similarly, greatest results are maximal elements of Res(Pv, e) relative to the more fine-grained ordering ≤v.

\[ (19) \ \text{a result } Q \text{ is the greatest element of } Res(Pv, e) \text{ with respect to } \leq_v \text{ just in case it is a maximal element of } Res(Pv, e) \text{ with respect to } \leq_v \]

\[ \forall Q \forall Q'(\text{Greatest}_v(Q, Q) \iff Q \in Q \land \forall Q'(Q' \in Q \rightarrow Q' \leq_v Q)) \]

Greatest elements are in general unique. An exception is the stem /akyat/, which can define two greatest results. Furthermore, greatest elements are always strongest results. The inverse need not be true. This has to do with =v -equivalence: different results that are brought about simultaneously can all be strongest results but they can (possibly) be distinguished by ≤v. An example is given by w-maximal results. One has: w-max1 ≤v w-max2 and w-max1 =v w-max2. Both w-maximal results are strongest results but only the w-max2 result is a greatest element of Res(Pv, e). In Table 4 the distinction between strongest and greatest results is illustrated by means of some examples.

\[ ^3 \text{Note that according to (18) the definition of a strongest result is relativized to an event-type } P, \text{ because there is not a single relation Strongest but for each } v \in \text{VERB there is a corresponding relation Strongest}_v. \]
An Interpretation of the Voice Affix /i-/ in Tagalog

From Table 4 it follows that the characterizing result of a centripetal transfer, the Destination \( z \) has the Transferred Object \( y \), is expressed in DES as a particular type of result, namely as the greatest element of \( \text{Res}(P_v, e) \). Furthermore, the Transferred Object \( y \) is, in addition, assigned all v-closed results. E.g., in the case of an event \( e \) of type ‘bigay’ \( y \) is assigned both the w-maximal\(_1\) result (‘\( x \) does not have \( y \)’) and the w-maximal\(_2\) result (‘\( z \) has \( y \)’). When taken together, one gets (20).

\[
\begin{align*}
(20) & \quad \text{a. The Destination } z \text{ and the Transferred Object } y \text{ are both assigned a greatest result from } \text{Res}(P_v, e) \\
& \quad \text{b. the Transferred Object } y \text{ is assigned all v-closed results}
\end{align*}
\]

The condition characterizing \( y \) in (20) does not rely on the fact that the greatest element(s) of \( \text{Res}(P_v, e) \) is (are) also assigned to some other object \( z \). Thus, (20) is an instance of the more general scheme in (21).

\[
\begin{align*}
(21) & \quad \text{a. The object } y \text{ is assigned a greatest result from } \text{Res}(P_v, e) \\
& \quad \text{b. The object } y \text{ is assigned all v-closed results}
\end{align*}
\]

(21) differs from (20) in not requiring that the greatest result(s) from \( \text{Res}(P_v, e) \) be assigned to any other object besides \( y \). According to (21), a transfer (-event) \( e \) is characterized by the fact that there is an object participating in \( e \) that is assigned all v-closed results and (therefore) a greatest element from \( \text{Res}(P_v, e) \). This characterization not only applies to transfers, i.e., events involving a Source, a Destination and a Transferred Object, but also to the events denoted by the stems in (iib) and (iii) for which only two participants undergo a change. In Table 5 the v-closed results for the corresponding stems are given and in (22) the results are paraphrased in terms of a transformation of state.

\[
\begin{align*}
\text{Table 4} & \quad \text{strongest results} & \text{greatest results} \\
\text{/bigay/} & \begin{array}{l}
\text{\( z \) has \( y \)} \\
\text{not (\( x \) has \( y \))}
\end{array} & \begin{array}{l}
\text{\( z \) has \( y \)}
\end{array} \\
\text{/kuha/} & \begin{array}{l}
\text{\( x \) has \( y \)} \\
\text{not (\( z \) has \( y \))}
\end{array} & \begin{array}{l}
\text{\( x \) has \( y \)}
\end{array} \\
\text{/kain/} & \begin{array}{l}
\text{mass of } \text{\( y \) is 0} \\
\text{mass of } \text{\( y \) is in stomach of } \text{\( x \)}
\end{array} & \begin{array}{l}
\text{mass of } \text{\( y \)} \\
\text{is in stomach of } \text{\( x \)}
\end{array} \\
\text{/hagis/} & \begin{array}{l}
\text{\( y \) is at } \text{\( z \)}
\end{array} & \begin{array}{l}
\text{\( y \) is at } \text{\( z \)}
\end{array} \\
\text{/akyat/} & \begin{array}{l}
\text{\( x/y \) is at } \text{\( z \)}
\end{array} & \begin{array}{l}
\text{\( x/y \) is at } \text{\( z \)}
\end{array} \\
\text{/saksak/} & \begin{array}{l}
\text{\( y \) is in } \text{\( z \)}
\end{array} & \begin{array}{l}
\text{\( y \) is in } \text{\( z \)}
\end{array}
\end{align*}
\]

\[
\begin{align*}
\text{Table 5} & \quad \text{v-closed results} \\
\text{/bukas/} & \begin{array}{l}
\text{\( y \) is open}
\end{array} \\
\text{/bagsak/} & \begin{array}{l}
\text{\( y \) is falling (down) \( \rightarrow \) \( y \) traverses a non-empty path}
\end{array} \\
\text{/tungo/} & \begin{array}{l}
\text{\( y \) is in a bending state \( \rightarrow \) \( y \) traverses a non-empty path}
\end{array} \\
\text{/ikot/} & \begin{array}{l}
\text{\( y \) is turning (rotating) \( \rightarrow \) \( y \) traverses a non-empty path}
\end{array}
\end{align*}
\]

111
Anja Latrouite and Ralf Naumann

(22) a. /bukas/: the object $y$ that is opened, e.g. a door, is transferred from its initial position, it is closed, to its final position, it is open, i.e., its angle to the wall is no longer zero degree

b. /bagsak/: the object $y$ slammed down, e.g. a vase, is transferred from its initial position (say, standing at some place $z'$) to its final position (say on the ground)

c. /tungo/: the object $y$ is brought from a state in which it is straight into a bending state

d. /ikot/: the object $y$ is brought from a state in which it is at rest into a state in which it is no longer at rest but in which, e.g., it is turning

The intuition behind the generalization in (21) is the following. If there is a destination $z$ to which $y$ is transferred during the event $e$, this is expressed as a result: $z$ has (or possesses) $y$, which is the greatest result of $Res(P_v,e)$. If there is no destination, this is equally expressed in terms of a result: $y$ is in state $\phi$ at the end of $e$. Thus, $y$ can be said to be 'transferred' into a state $\phi$ in which it has not been at the beginning of the event, i.e., $y$ is transferred in a non-literal sense.

The problem that the generalization in (21) faces is that it is too broad. E.g., it equally applies to stems of the ‘kuha’-type. What is needed is a generalization of (20) that is more restricted than the generalization given in (21). (21) can be restricted in at least two ways. On the one hand, it is possible to impose restrictions on the type of result assigned to either $y$ or all participants of $e$. On the other hand, there are restrictions on which objects, besides $y$, are assigned the greatest element(s). These options will be considered in the next section.

5 Admissibility Condition for /i/-

In the preceding section it was shown that if an object $y$ is transferred to the destination $z$, then a greatest result is assigned to both $y$ and $z$: $Q(z)(y)$ for $s \in Q(a)(b)$ just in case $a$ has $b$ at state (point) $s$. This assignment to $y$ is independent of the sort of transfer, i.e. of whether the transfer is centripetal or centrifugal. Consequently, in a proper transfer the Transferred Object $y$ is always assigned a greatest element from $Res(P_v,e)$: $Q(a)(y)$ for some $a$ participating in $e$. For a centrifugal transfer, $a = x$ holds, i.e., $a$ is the Actor, whereas for a centripetal transfer $a$ is not identical to the Actor such that $a \neq x$ holds. Thus, the requirement in (23) is sufficient to exclude stems of the ‘kuha’-type.

(23) The Actor $x$ is not assigned a greatest element from $Res(P_v,e)$

If (23) holds, i.e., if $a \neq x$, the two sorts of transfer characterized by $\omega(e) \notin Q(x)(y)$ are captured. Now, $\omega(e) \notin Q(x)(y)$ is an instance of the following general scheme.

(24) for all $d$ s.t. $Q(...,d,....)$ with $Q \in Res(P_v,e)$ and $Q$ a greatest element: $d \neq x$

(24) requires that the Actor $x$ be assigned no greatest result. This requirement applies not only to Transfer-verbs but to any verb because it imposes only a condition on the Actor $x$. Consequently, also cases in which there is no transfer and in which $x$ is not assigned a greatest element are captured. This applies to stems like /bukas/, /ikot/ or /tungo/.
An Interpretation of the Voice Affix /i-/ in Tagalog

(23) restricts (21) because it simply adds a further condition that must be satisfied, besides those imposed by (21). It differs from (20) because no condition is imposed on a possible Destination z but rather on the Actor x. As each event has an Actor x but not necessarily a Destination z, (23) is a generalization of (20).

(23) cannot be the final version of the admissibility condition for /i-/ because it is too restrictive. It excludes terms that admit this VA. Examples are the stems /akyat/, /kain/ and /takbo/. In each case the interpretation of the stem assigns for a given event e \( \in \mathcal{P} \), a greatest element to the Actor x. They are therefore similar to stems like /kuha/ although they differ with respect to the admissibility of /i-/. Stems belonging to the ‘kuha’-type can be distinguished from stems like /akyat/ as follows. There is a result \( Q' \), namely the w-max, result, that is assigned to y but not x: z does not have y. As each w-maximal result is v-closed, it follows that there is a v-closed result that is assigned to y but not to x. The Actor x is the object that is assigned the greatest result besides the transferred object y which is assigned all v-closed results. Now consider the requirement in (25).

(25) If the greatest result(s) \( Q \) is (are) assigned to some object \( x' \) besides the object y that is assigned all v-closed results, each v-closed result \( Q' \) that is assigned to y (and some other participant \( x'' \)) is also assigned to \( x' \).

This condition is not satisfied for the ‘kuha’-type, as shown above: y is assigned the w-max, result but it is not assigned to x, although x is assigned the greatest element besides y. This condition is satisfied for the three stems that admit /i-/ although the Actor x is assigned the greatest element. In Table 6 the greatest element(s) together with the other v-closed results that are assigned to y and some other participant \( x' \) are shown.

<table>
<thead>
<tr>
<th>Table 6</th>
<th>greatest element</th>
<th>non-greatest elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>/kain/</td>
<td>the mass of y is in the stomach of x</td>
<td>part of the mass of y is in the stomach of x</td>
</tr>
<tr>
<td>/akyat/</td>
<td>x/y are at z</td>
<td>x/y move up towards z</td>
</tr>
<tr>
<td>/takbo/</td>
<td>x and y move together (only v-closed result)</td>
<td>x and y move together</td>
</tr>
</tbody>
</table>

The first two stems semantically determine w-minimal and s-maximal results. The s-maximal result that is the greatest element is assigned not only to y but also to the Actor x (and the Goal z in the case of /akyat/). The same assignment of objects also holds for the w-minimal result that is assigned to more than one object so that condition (25) is satisfied. For /takbo/ the greatest element is of type w-minimal, i.e., no maximal results are determined. This result is assigned to both x and y, i.e., condition (25) is again satisfied.

The intuition behind (25) is that if the Actor x is assigned a greatest element, it is required that (s)he be involved with the object y that is assigned this result too for all initial stages \( e' \) of e that are of type v. The condition in (25) not only applies to Transfer-verbs but to all verbs because y is not required to be transferred in some sense but is singled out by the requirement that it be assigned all v-closed results.

When taken together, one arrives at the admissibility condition for /i-/ in (26).
(26) The $\text{UV}_2$ affix /i-/ is admissible for a stem $\nu$ just in case either

a. the Actor $x$ is not assigned a greatest element from $\text{Res}(P_\nu, e)$ for $e \in P_\nu$, or

b. the Actor $x$ is assigned a greatest element $Q$ from $\text{Res}(P_\nu, e)$ for $e \in P_\nu$ and for each $y \neq x$ that is assigned $Q$ too all $\nu$-closed results assigned to $y$ (and at least one other object $x'$) are assigned to $x$ too.

6 Admissibility Condition for /-in/

If only stems belonging to the 'kuha'-type are considered, the admissibility condition in (27) could be used for the $\text{UV}_1$ affix /-in/.

(27) admissibility condition for /-in/ (1. version)

the VA /-in/ is admissible for a stem $\nu$ just in case

a. $P_\nu$ determines for each of its elements a maximal result

$$\forall e (e \in P_\nu \rightarrow \exists Q (Q \in \text{Res}(P_\nu, e) \land \text{max}_\nu(Q)))$$

b. the Actor $x$ of an event $e \in P_\nu$ is always assigned a greatest element of $\text{Res}(P_\nu, e)$

$$\forall e (e \in P_\nu \rightarrow \exists Q \exists d (Q \in \text{Res}(P_\nu, e) \land \text{Greatest}_\nu(Q, \text{Res}(P_\nu, e)) \land \text{Actor}(e) = d))$$

If $P_\nu$ determines for each of its elements a maximal result, it follows that the greatest elements are maximal results. (27b) can therefore be strengthened to: 'the Actor $x$ is always assigned a maximal result that is a greatest element of $\text{Res}(P_\nu, e)$'.

The problem that the admissibility condition in (27) faces was already alluded to above. There are stems $\nu$ that admit /-in/ but for which the Actor is not assigned a maximal result that is a greatest element of $\text{Res}(P_\nu, e)$. Three classes can be distinguished:

(28) a. stems like /kantala/ ('sing'), /basal/ ('read'), /halo/ ('mix') and /saksak/ ('stab')

b. Point-verbs like /hampas/ ('hit') and /puknok/ ('hit', 'beat')

c. stems like /hagis/ ('throw')

Stems $\nu$ belonging to the first class are characterized by the property that an event $e \in P_\nu$ is materialy incremental with respect to at least one object that is assigned a greatest element from $\text{Res}(P_\nu, e)$. This property is formally defined in (29).

(29) an event $e \in P_\nu$ is incremental with respect to an object $d$ participating in it just in case the following conditions hold $((e, e') \in R_B$ just in case $e$ is a proper initial stage (prefix) of $e'$; $\sqsubseteq_O$ is the (proper) material part relation on the domain $O$ of objects; $\Delta^*_\nu(e)(d)$ is true just in case $d$ participates in $e$)

a. $\forall e, e', d (e \in P_\nu \land R_B(e', e) \land \Delta^*_\nu(e)(d) \rightarrow \exists d' (d' \sqsubseteq_O d \land \Delta^*_\nu(e')(d']))$
b. $\forall d, d', e\{e \in P_v \land \Delta^v(e)(d) \land d' \subseteq d \rightarrow \exists e'\{R_B(e', e) \land \Delta^v(e')(d')\}\}$

The second class is characterized by the property in (30).

(30) The Destination $z$ does not have (possess) the Transferred Object $y$ at the end of the event; rather, $y$ only touches $z$ and produces a sound. Furthermore, the result does not continue to hold after the end of the event. In DES two sorts of results are distinguished: event-related and state-related results. A result $Q$ is event-related just in case it does not continue to hold after the end of an event that brought it about. A result $Q$ is state-related just in case it continues to hold after the end of an event that brought it about until another event undoes it (if there is such an event at all), (31).

(31) $\forall Q\forall e\{e \in P_v \land Q \in Res(P_v, e) \land \text{Greatest}_v(Q, Res(P_v, e)) \rightarrow \text{Event-Related}(Q)\}$

The characterizing property of verbs belonging to the third class, /hagis/, is given in (32).

(32) The Actor is never active until the end of an event of this type

(32) implies that for all results $Q$ that are determined by $P_{\text{hagis}}$ for an event $e$ belonging to this set and that are required to hold at the end point $\omega(e)$ of $e$ one has that $Q$ is not assigned to the Actor, (33).

(33) $\forall Q\forall e\{Q \in Res(P_v, e) \land \omega(e) \in Q \rightarrow \forall d(\Delta_v(e)(d) \rightarrow \neg \text{Actor}(e, d))\}$

The characterizing properties of the three classes must be accounted for in the admissibility condition of /-in/. Formally, this means that the second clause of the admissibility condition in (27) must be disjunctive. This yields (34) (the admissibility condition improves on that given in Latrouite/naumann 1999a/b).

(34) admissibility condition for /-in/ (final version)

the VA /-in/ is admissible for a stem $v$ just in case

a. $P_v$ determines for each of its elements a maximal result, and

either

bi. the Actor is always assigned a maximal result that is the greatest element with respect to $Res(P_v, e)$

or

bii. the Actor is never assigned a maximal result, and either (i) at least one object that is assigned a greatest element from $Res(P_v, e)$ is incremental (an Incremental Theme) or (ii) the Actor is never active until the end of the event

115
The first clause excludes activity-verbs like /asin/ ('salt') as well as stems like /takbo/ ('run', 'x takes y on the run'). Clauses bi and bii exclude stems like /bigay/, /bukas/, /ikut/, and /tungo/ because the Actor is never assigned the greatest element from Res($P_v, e$) and the participants that are assigned the greatest elements do not satisfy any of the conditions mentioned in bii; a stem like /hiwalay/ is excluded because the Actor $x$ is not always assigned the greatest element (for ditransitive uses; for transitive uses, with meaning 'leave', $x$ is assigned the greatest element) and the clauses of bii do not apply to it; a stem like /bagsak/ is excluded because it does not satisfy clause bi and neither of the conditions mentioned in bii; in particular, the third condition is excluded because for intransitive uses the Actor is involved until the end of the event.

7 The Determination of the Subject

7.1 The Subject determined by /i-/ 

Consider the set $S_{v,e}$ of participants of an event $e \in P_v$ that is determined by the following conditions.

(35) An object $d$ is an element of the set $S_{v,e}$ just in case

- a. $d$ is assigned a greatest result from Res($P_v, e$)
- b. $d$ is temporally minimal with respect to condition (a)
- c. $d$ is not globally minimal

(36) a. An object $d$ participating in an event $e \in P_v$ is temporally minimal with respect to a condition $Q$ relative to the participants of $e$ just in case $d$ satisfies $Q$ and for each $d'$ participating in $e$ that satisfies $Q$ too (i) $d$ and $d'$ are assigned the same results or (ii) the least result $Q$ relative to the ordering $\leq_v$ that is assigned to $d \leq^*_v$ preceeds the least result relative to the ordering $\leq^*_v$ assigned to $d'$ or (iii) $d$ and $d'$ agree on the least result relative to $\leq^*_v$ that is assigned to them and there is a least result $Q$ that is assigned to $d'$ but not to $d$

b. An object $d$ participating in an event $e \in P_v$ is globally minimal just in case (i) it is assigned the least result from Res($P_v, e$) relative to the ordering $\leq_v$ and (ii) if there is an object $d'$ participating in $e$ that satisfies this condition too, then there is an antecedent event $e'$ of $e$ such that $d$ is temporally minimal relative to $d'$ with respect to condition (i) relative to $e'$

The first clause requires that $d$ be assigned a greatest result from Res($P_v, e$). This condition need not single out a unique object participating in $e$, witness many of the examples that have been discussed so far. If, on the other hand, a unique participant is singled out, it is determined as the subject by /i-/ This is the case for stems like /bukas/, /baksak/, /ikut/ and /tungo/. In each case the object that satisfies clause (a) also satisfies clauses (b) and (c). Clause (b) is vacuously satisfied because only one object satisfies clause (a). Clause (c) is satisfied because the object that is assigned the greatest element is not the Actor, who is always globally minimal. Thus, if only those stems admitting /i-/ are considered for which the
greater elements(s) is (are) assigned to a unique object, the set $S_{v,e}$ can be taken to uniquely determine the subject that is determined by $/i-/$.

If the greatest element(s) is (are) not assigned to a unique object, two cases can be distinguished: either the Actor $x$ is one of those objects or the Actor $x$ is not among the objects that are assigned a greatest element. If the second possibility holds, i.e., if the Actor $x$ is not assigned a greatest result, the stem expresses a centripetal transfer. This means that the greatest result is assigned both to the Transferred Object $y$ and the Destination/Source $z$. This is the case for stems belonging to the 'bigay'-type, /hagis/, /halo/, /hiwalay/, /saksak/ and Point-verbs like /hampas/. The Destination/Source $z$ is characterized by being assigned only the greatest element whereas the Transferred Object $y$ is assigned further results which precede the greatest element. That means that the Transferred Object $y$ is temporally minimal relative to $z$ with respect to the condition of being assigned a greatest result. This condition is expressed in clause (ii), which, therefore, excludes the Destination/Source $z$. As the Transferred Object is in addition not the Actor, who is excluded by clause (iii), it follows that for this case too the set singles out a unique object that is determined as the subject by $/i-/$.

The last case to be considered is given by the condition that the greatest element of $Res(P_v,e)$ is assigned to more than one object, one of which is the Actor. This is satisfied for the stems /akyat/, /takbo/ and /kain/. For /akyat/ the Destination $z$ is also assigned a w-minimal result ('$x/y$ traversed a non-empty path towards $z$'). Yet, there is a further w-minimal result, '$x/y$ traversed a non-empty path', that is not assigned to $z$. As a consequence, $z$ does not satisfy the minimality condition expressed in clause (ii). Contrary to the other cases, no single object is singled out so far because both the Actor $x$ and the Transferred Object $y$ satisfy the first two clauses. The Actor $x$ is excluded by clause (iii) because the Actor is always globally minimal. Thus, for /akyat/ too $S_{akyat,e}$ determines a unique element that is denoted by the nominative argument if this stem combines with $/i-/$. The stem /takbo/ is similar to /akyat/ relative to the first two clauses. No unique participant $d$ of $e$ is singled out. The first two clauses determine the Actor $x$ as well as the object $y$ that is taken with by $x$. Similarly to the stem /akyat/ the third clause excludes $x$ such that only $y$ is an element of $S_{takbo,e}$, which is determined as subject by $/i-/$. For /kain/ the first condition yields both the Actor and the object eaten (the Incremental Theme). The minimality condition in (ii) excludes the latter (i.e. the Incremental Theme) so that only the Actor $x$ is determined by the first two clauses. The latter is then excluded by clause (iii). As a consequence, for /kain/ $S_{kain,e}$ is the empty set, i.e., no object is singled out. Yet, for /kain/ neither the Actor nor the Incremental Theme is determined as subject by $/i-/$. This is exactly what is required.

From the above discussion it follows that for each stem admitting $/i-/$ that does not determine the instrument as subject the set $S_{v,e}$ is a singleton containing the object that is denoted by the nominative argument. Only for the stem /kain/ no object is determined, so that in this case the instrument is singled out as denotation of the nominative argument. When taken together, this yields the (37).

(37) The subject that is determined by $/i-/$ for a stem $v$ is that participant $d$ of an event $e \in P_v$ such that

a. either $d$ is the unique element of $S_{v,e}$

117
Anja Latrouite and Ralf Naumann

b. or, if $S_{v,e}$ does not single out a unique participant of $e$, i.e., if $S_{v,e}$ is not a singleton, $d$ is the instrument

7.2 The Subject determined by /-in/

If one considers only transfers of the sort expressed by stems of the ‘kuha’-type, the subject determined by /-in/ is the Transferred Object $y$, similar to stems belonging to the ‘bigay’-type when they combine with /i-/ . The two VAs differ with respect to stems for which they are both admissible and for which they are used ditransitively. Examples are /hagis/, /saksak/ and Point-verbs like /puknok/ . If they combine with /i-/ , the subject is the Transferred Object $y$ and not the Destination $z$. For the VA /-in/ , the subject is the Destination $z$ and not the Transferred Object $y$. The stem /akya/ is similar to these three stems, although the argument structures differ (/-in/ : transitive; /i-/ : ditransitive). Whereas for /-in/ it is the Goal that is determined as subject, it is the transferred object which is singled out by /i-/ .

As was shown in the preceding section, the difference between the Transferred Object $y$ and the Destination/Goal $z$ consists in the number (or types) of results that are assigned to them. The Transferred Object $y$ is always assigned all $v$-closed results whereas this is in general not the case for $z$. A $v$-closed result is not necessary a strongest result, witness $w$-minimal results that are $v$-closed but not strongest if the event-type also determines $s$-maximal results.

If one considers strongest results instead of $v$-closed ones, a Transferred Object is always assigned all strongest results, as Table 7 shows.

<table>
<thead>
<tr>
<th>Table 7</th>
<th>strongest</th>
<th>assigned to</th>
<th>non-strongest</th>
</tr>
</thead>
<tbody>
<tr>
<td>/kuha/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>transitive</td>
<td>$x$ has $y$</td>
<td>$x, y$</td>
<td></td>
</tr>
<tr>
<td>ditransitive</td>
<td>$x$ has $y$</td>
<td>$x, y(2), z$</td>
<td></td>
</tr>
<tr>
<td>/hagis/</td>
<td>$y$ is at $z$</td>
<td>$y, z$</td>
<td>$y$ traversed a path</td>
</tr>
<tr>
<td>/saksak/</td>
<td>$y$ is in $z$</td>
<td>$y, z$</td>
<td>$y$ traversed a path</td>
</tr>
<tr>
<td>/puknok/</td>
<td>$y$ is at/on $z$</td>
<td>$y, z$</td>
<td>$y$ traversed a path</td>
</tr>
</tbody>
</table>

The ‘kuha’-type differs from the other stems in the following respect. For /kuha/ $y$ is either the unique object that is assigned all strongest results (ditransitive use) or it is temporally maximal with respect to this condition (transitive use). If the subject is determined by the condition that it (i) be assigned all strongest results and (ii) be temporally maximal with respect to this condition, this yields exactly the correct prediction.

If, on the other hand, this criterion is applied to the other stems in Table 7, it is the Destination $z$ that is determined because $y$ is assigned a further, non-strongest result: $y$ traverses a non-empty path. But this is again what is required because $z$ is singled out as subject by /-in/. The difference to the case of /kuha/ is the following. There is only one strongest result and it is assigned to the Transferred Object $y$ and the Destination $z$, and not to the Actor $x$, as in the case of /kuha/ (transitive use). Therefore, the maximality condition singles out $z$ and not $y$ because $y$ is assigned a non-strongest result that is not assigned to $z$.

For other stems admitting /-in/ the above criterion can be used too.
Table 8 illustrates this for the stems /akyat/ and /kain/. For the former, x and y are both assigned all strongest results but x is temporally minimal because it is assigned a result that does not hold for y such that y is temporally maximal. A similar argument holds for stems like /sunod/ ('obey' if it combines with /-in/). For /kain/, y is already uniquely determined by the first condition because it is assigned both strongest results whereas x is assigned only one. An analogous argument applies to Achievement-verbs like /patay/ ('kill') and stems like /halo/ ('stir' if it combines with /-in/) and /kanta/ ('sing'). Thus, the above condition can be taken as the criterion which determines the subject for /-in/. It is repeated in (38).

(38) The subject that is determined by /-in/ for a stem v is that participant d of an event e ∈ P, such that

a. d is assigned all strongest results from Res(P, e)

b. d is temporally maximal with respect to condition (a)

References


In this paper I will examine various aspects of the nominal clause of Niuean (a VSO Oceanic language of the Tongic subgroup as classified by Pawley, 1966, 1967). In Section 1 I will provide a basic analysis of non-derived Niuean nominal clauses, as described by Seiter (1980), Wilson (1989) and Massam and Sperlich (to appear). In Section 2 I will focus on derived nominalizations, particularly on the case system found therein, from a Minimalist perspective. Although the facts are somewhat ambiguous, as shown in Section 3, I will advance the hypothesis that Niuean, while an Ergative/Absolutive (E/A) language at the sentential level, has an intransitive Nominative structure in the nominal clause. This is of interest, because it has often been claimed, most recently and in detail by Alexiadou (1999), that nominal clauses have an ergative structure in a wide range of N/A languages, including Greek, English, Romance and Slavic languages, and Hungarian. In addition, in the typological study of Koptjevskaja-Tamm (1993) this type of nominalization is found only in SVO Nominative/Accusative (N/A) languages.

1. The Structure of the Niuean Nominal Clause

1.1. Basic Word Order

The unmarked word order of a Niuean DP is shown in (1), with an example in (2).

(1) Order of elements

[CaP Det/# N Mods Dem Gen]

(2) (CaP Det/# N Mods Dem Gen)

\[e \text{ tau } \text{ mena } gahua \text{ nā } \text{ hauu}\]

Abs/Com Spec/PI thing work that you (Gen)

"those tools of yours" (Seiter.117b.45)

The first element in the clause is a portmanteau element which marks both the case of the DP and whether it is proper/pronominal (henceforth termed simply proper), or common. In (3) I provide a paradigm of these particles which I term CaP (for Case, [+/- Proper]). (Seiter 1980 and Clark 1976 each make different analyses of these items.)

---

*I would like to particularly thank Wolfgang Sperlich as this work has grown out of joint work with him. In addition, I thank Marian Klamer and other AFLA VII organizers and audience members, the University of Toronto Syntax Group, Susan Bejar, William Foley, and Hitay Yukseker for various sorts of help with this work. All errors are mine. This work was supported by a research grant: SSHRCC (#410-97-0493). Data for this paper comes principally from Sperlich (1997), Seiter (1980) and field notes.*
Diane Massam

(3) Niuean CaP elements

<table>
<thead>
<tr>
<th>Common Prep</th>
<th>Case</th>
<th>Art</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abs</td>
<td></td>
<td>e</td>
</tr>
<tr>
<td>Erg</td>
<td>h</td>
<td>e</td>
</tr>
<tr>
<td>Gen</td>
<td>h</td>
<td>e</td>
</tr>
<tr>
<td>Loc</td>
<td>h</td>
<td>e</td>
</tr>
<tr>
<td>Goal</td>
<td>ke</td>
<td>e</td>
</tr>
<tr>
<td>Source</td>
<td>mai</td>
<td>e</td>
</tr>
<tr>
<td>Topic</td>
<td>ko</td>
<td>e</td>
</tr>
<tr>
<td>Ben</td>
<td>ma</td>
<td>e</td>
</tr>
<tr>
<td>Comit</td>
<td>mo</td>
<td>e</td>
</tr>
<tr>
<td>Instr</td>
<td>aki</td>
<td>e</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proper Prep</th>
<th>Case</th>
<th>Art</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abs</td>
<td></td>
<td>a</td>
</tr>
<tr>
<td>Erg</td>
<td>e</td>
<td>a</td>
</tr>
<tr>
<td>Gen</td>
<td>h</td>
<td>a</td>
</tr>
<tr>
<td>Loc</td>
<td>i</td>
<td>a</td>
</tr>
<tr>
<td>Goal</td>
<td>ki</td>
<td>a</td>
</tr>
<tr>
<td>Source</td>
<td>mai</td>
<td>a</td>
</tr>
<tr>
<td>Topic</td>
<td>ko</td>
<td>a</td>
</tr>
<tr>
<td>Ben</td>
<td>ma</td>
<td>a</td>
</tr>
<tr>
<td>Comit</td>
<td>mo</td>
<td>a</td>
</tr>
<tr>
<td>Instr</td>
<td>aki</td>
<td>a</td>
</tr>
</tbody>
</table>

*It is unclear whether the i on the goal and source prepositions is the Case marker or part of the preposition.

Although there are many as yet unsolved issues with the CaP element, it can be seen that there are three main types of DP: unmarked, marked for case, and prepositional. The prepositional cases divide into those which take marked case complements and those which take unmarked case complements. In addition, there are more case distinctions in the proper series than in the common series, and the proper article has a more irregular distribution than the common article.

The second element in the clause is another portmanteau item which marks specificity and number. The basic paradigm as described by Seiter (1980) can be laid out as in (4).

(4) A. [+Specific] [+/-Definite]    Plural     Dual       Singular  No #
    tau       nā       Ø         N/A

B. [-Definite] [+/-Specific]        Plural     Dual       Singular  No #
    falu a   N/A      taha      N/A

C. [-Specific]                      Plural     Dual       Singular  No #
    N/A      N/A      N/A       ha

*It is unclear whether the i on the goal and source prepositions is the Case marker or part of the preposition.

Although there are many as yet unsolved issues with the CaP element, it can be seen that there are three main types of DP: unmarked, marked for case, and prepositional. The prepositional cases divide into those which take marked case complements and those which take unmarked case complements. In addition, there are more case distinctions in the proper series than in the common series, and the proper article has a more irregular distribution than the common article.

The second element in the clause is another portmanteau item which marks specificity and number. The basic paradigm as described by Seiter (1980) can be laid out as in (4).

(4) A. [+Specific] [+/-Definite]    Plural     Dual       Singular  No #
    tau       nā       Ø         N/A

B. [-Definite] [+/-Specific]        Plural     Dual       Singular  No #
    falu a   N/A      taha      N/A

C. [-Specific]                      Plural     Dual       Singular  No #
    N/A      N/A      N/A       ha

The head noun follows the specificity/number marker, and it is in turn followed by modifiers, a demonstrative (classed here as another modifier), and a genitive phrase.

and ArtP (CaP) heads coalesce, as do the heads of DP and the #P most of the time (but see below, Section 1.b). Some problems here are that the nature of the position that the noun moves to is not clear, and it seems that the NP never has internal structure.

(5) Nominal structure

1.2. Prenominal Genitives (aka Possessive Preposing/Incorporation)

As well as appearing clause finally, the genitive can appear between the CaP and the number particle (or the N if there is no number particle), as shown in (6). The lack of the determiner on the right side of the arrow will be explained below. Examples are given in (7) and (8).

(6)  CaP Det/# N Mods Gen  --->  CaP Gen a  # N Mods

(7)  a. e leo  ha Sione  
     Abs  voice  Gen  Sione  
     "Sione's voice"

   b. (e) ha Sione a leo  
      Abs  Gen  Sione  a  voice  
      "Sione's voice"

(8)  a. Ko e haana a tau hui  
     CaP*  3sgGen  "a"  Pl  foot  
     "his feet"

   b. Ko e hana male  
      TopCaP  his  house  
      "his house" (Sperlich.104)

(7b) shows that the preposed proper genitive is followed by the ligature item a. (8a) shows a preposed pronoun, also with a, which can be compared with (2). (8b) shows a shortened form of the pronoun which can appear prenominally, and it does not appear with a.

When the genitive is preposed, the DP receives a definite interpretation, as shown in (9). This means that a preposed structure will never appear with an overt determiner marking specificity, but just with an optional number marker (see (4)). Sperlich (1997) states that (8b) receives an even more definite interpretation than (9b).
A preposed genitive must be either a pronoun or a proper name (true also, interestingly, in German and Icelandic: Alexiadou 1999:113). Common noun genitives may not prepose. (See Massam and Sperlich, to appear for a discussion of how this preposing is treated in the literature on Polynesian.) Preposing can be accounted for by analyzing the genitive to be in the specifier of DP, as in (10) and considering the ligature a to be in the head of DP. Thus we see that Niuean has shifted preposing from the Tongic pattern of pronominal cliticization (Chung, 1973, Wilson, 1982) to an XP movement which includes proper genitives as well as pronominal ones. The ligature item might be related to the proper article a. If so, the proper/pronoun constraint on preposing follows from a properness agreement relation between head of DP and specifier of DP. The example in (8b) is analogous to the Tongan cliticization rule, where only short forms of the pronouns can appear in the head of DP, thus usurping the determiner a.

This provides us with the outline of the basic Niuean clause structure.

2. The Case of Nominalized Clauses

2.1. Nominalization in N/A languages (Alexiadou, 1999)

Alexiadou (1999) examines nominalizations in a variety of languages, including Greek and English. She concludes that "aspects of nominal syntax are closely related to patterns of ergativity" (p.2) and that "the light verb included in process nominals does not project an agent" nor does it assign accusative case (p.109). She provides the following structure.
Niuean Nominalization

(11) DP
    ^
   D0 the
      FP [= #P in a nominalization]
    /
   AP FP
     /
    F0 AspP
      /
     Asp'
       /
      Asp0 vP
        \
         v LP
            \
           vDESTROY Comp (= theme)

She considers the nominal to be a lexical root (L) which only acquires nominal status by virtue of appearing with nominal functional projections (cf. the concept of precategorial roots as in Foley, 2000). The lexical head moves to FP, which in a nominal would be #P, to be in the domain of the determiner and the object moves to specifier of AspP to check genitive case. (See Alexiadou, 1999 for full detail). The principal aspect of her analysis for us is that she argues that nominalized clauses have a deficient light verb, which does not assign accusative case, nor an external theta role. In this way, she claims, the light verb is an unaccusative light verb (see also den Dikken and Sybesma 1998, Harley and Noyer 1998, Marantz, 1997). Because the external theta role is not assigned to specifier of light v, it can appear only as an oblique (12a), a possessor (12b) or not at all (12c). This clears the way for the object to get the one available structural case - genitive - marked with of in English.

(12) a. the destruction of the city by the Romans
    b. the Romans' destruction of the city
    c. the destruction of the city

Alexiadou develops the idea, often proposed in the literature, that in N/A languages nominalized clausal structure is analogous to unaccusative, passive, and in particular, ergative clause structure, and she argues therefore that nominal clauses are ergative in structure. In her view, ergative languages have deficient light verbs also, which do not assign external theta roles, nor accusative case. For this reason, external arguments in ergative languages must appear in prepositional or oblique cases such as locative, genitive, or not at all. Alexiadou also points out that ergative arguments also share another property with by-phrases in nominal clauses: they are secondary cases in that they are only possible if there is a theme expressed.

Given the hypothesis that nominalizations are ergative in N/A languages, it is interesting to ask what case patterns nominalizations show in ergative languages (see Koptjevskaia-Tamm, 1992 for many other examples). On the one hand, since ergative languages already have a deficient light verb (but see below for a different view), we might expect that in an ergative language the case system of the nominal phrase will be identical to the verbal phrase. On the other hand, we might find a sort of antipassive situation (as discussed by Silverstein, 1986). The argument which targets and eliminates the external case/argument in a N/A language, might well target and eliminate the internal case/argument in an E/A language. In the next section, I
will argue that it is essentially the latter situation we find in Niuean, though as I will demonstrate in Section 3, the data remain somewhat ambiguous on this point.

2.2. Nominalization in an E/A language

Examples of nominalized clauses appear in (13).

(13) a. Ne tāmate e Tofua e kuli
   Pst kill Erg Tofua Abs dog
   "Tofua killed the dog."

b. ke he tāmate e Tofua e kuli
   GoalCaP kill Erg Tofua Abs dog
   "(about) Tofua’s killing the dog" (Seiter.82a.119)

(14) a. e tele haaku i a ia
   Abs kick meGen Loc Art him
   "my kicking him" (Seiter.89b.121)

b. e pākia haaku he pilu nā
   Abs injured meGen on knife that
   "my being injured on that knife" (Seiter.83b.119)

c. e fano he tagata ia ki Niu Silani
   Abs go Gen man that to New Zealand
   "that man’s going to New Zealand" (Seiter.84b.119)

Nominalizations in Niuean can be formed by simply replacing the Tense-Aspect-Mood particle (TAM) with a CaP particle, as in (13a,b). The rest of the clause is unchanged. Another option is shown in (14a,b,c) where, as well as the TAM becoming a CaP, we also see case changes. In particular, in (14a) the argument which would be ergative in a verbal clause is expressed as genitive in the nominal clause, and the argument which would be absolutive in the verbal clause is expressed as locative in the nominal clause. Finally we see two examples of intransitive clauses in (14b,c), one semantically active and one semantically passive. In both of these, the argument which would have been absolutive in the verbal clause appears in the genitive case.

I will henceforth employ the "ASO" terminology of Dixon (1979, 1994). I will use A to refer to the agent of a transitive clause, S to refer to the single argument of an intransitive clause and O to refer to the other argument of a transitive clause. We can see that A and S pattern together in taking genitive case in the nominalized clause, hence, that Niuean exhibits an nominative pattern in the nominalized clause, as seen in (15). This was observed by Seiter (1980) as evidenced in his naming the locative case in nominalized transitive clauses "fake accusative". He states (p.302) that this case is bizarre, since it is not clear why absolutive could not be used to mark the object, as in verbal clauses. (He provides a historical discussion which we will not review here, except to note that the Niuean locative is morphologically the same as the accusative in other Polynesian languages.)
Niuean Nominalization

Table 1 (see Alexiadou, 1999)

<table>
<thead>
<tr>
<th>English clause</th>
<th>Nominalization</th>
<th>Niuean clause</th>
<th>Nominalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Nom</td>
<td>{PP,Poss}</td>
<td>Erg</td>
<td>Gen</td>
</tr>
<tr>
<td>S: Nom</td>
<td>Gen</td>
<td>Abs</td>
<td>Loc</td>
</tr>
<tr>
<td>O: Acc</td>
<td>Gen</td>
<td>Abs</td>
<td>Gen</td>
</tr>
</tbody>
</table>

In providing an analysis of the Niuean nominalized clause, it is necessary first to show that Alexiadou’s characterization of ergativity does not fit Niuean. There are two principal reasons why it does not. First, the ergative DP in Niuean does not act as an oblique or prepositional argument for any operation, such as wh-extraction, quantifier float, raising, etc. Second, the [Pred/Erg/Abs] word order of Niuean does not allow for an analysis in which absolutive moves to check case in a position higher than the ergative DP (such as in Specifier of TP, for example), which is what we would expect if the ergative were an oblique case within VP. Instead, I propose that Niuean has a rich, fully functional transitive light verb which in a transitive clause checks case (abs) against the internal argument and also assigns an external theta role along with an inherent case to its specifier position (Woolford, 1997, Ura, 1998). This is schematized in (15).

(15) Transitive Clause

```
\(\text{vP} \begin{array}{c}
\text{A} \\
  \text{vP} \\
    \text{v'} \\
      \text{vO} \\
        [\text{abs}] \\
          [\text{ag/erg}] \\
            \text{V} \\
              \text{O}
\end{array}
\text{VP}
```

What is different about Niuean (vs English), then, is that the agent receives case in situ and that absolutive is “super strong” and must be checked immediately, prior to merge of the external object, violating Chomsky’s (1995) claim that Merge will always occur before Move. (Another option is that the object moves after merge of the subject, but that it "tucks in" (Richards 1997). A transitive vmax, then, has two specifiers, one by move and one by merge. Intransitive clauses are as in (16).

(16) Intransitive Clause

```
\(\text{vP} \begin{array}{c}
\text{v'} \\
  \text{v0} \\
    [\text{abs}] \\
      \text{V} \\
        \text{S}
\end{array}
\text{VP}
```

The intransitive light verb here is lacking the external case/theta package but retains absolutive case, which causes the internal argument to undergo object shift. For unergatives,
this same analysis might hold, or the single S argument could be merged directly into specifier of the light verb to maintain the unaccusative/unergative distinction (for which I have found no evidence in Niuean). Either way, all S arguments check absolutive case in specifier of vmax.

We now turn to nominalization clauses, where the light head is deficient, just as it is in Alexiadou's analysis. In Niuean, the light verb, more properly termed a light noun since it has such a different feature composition, is as in (17). Its deficiency, compared to the verbal counterpart lies in the absence of absolutive case. Also, in the light noun, the external argument is not tied to a particular theta role, unlike the ergative case.

(17) Transitive Nominalization

(18) Intransitive Nominalization

One assumes here that the absence of absolutive case rules out the possibility for a direct object. There is no way for the internal argument to get direct case, hence it must appear in oblique locative case (parallel to the external argument appearing as a by-phrase in English nominalizations). But the question remains why it is not possible for the object to undergo object shift to receive genitive case, thereby forcing the external argument to be inexpressible as a direct argument. This is because genitive, unlike absolutive, is not "super strong", but merely strong, thus merge of the external argument will precede movement of the internal argument. When the external argument is merged, it checks genitive case. In intransitive nmax clauses, if there is no external theta role to be assigned, the internal argument is called upon to undergo move to check genitive case. Transitive nominalizations in Niuean are thus unergative-like in that the light noun assigns an external theta role and case, but there is no direct object, no accusative case. Instead, the object appears as a locative oblique. Alternatively we could view locative as accusative, as Seiter does, but since it has the morphology of locative, I consider it oblique. The analysis of the case of O as oblique renders Niuean nominalization different from other Polynesian languages, and also different from other ergative languages. According to Koptjevskaya-Tamm (1992), the pattern of genitive A,S and oblique O is limited to SV O N/A languages.
3. Two Unsolved Problems

The above analysis is interesting in cross linguistic terms and it accounts for the Niuean data presented so far. There are two crucial problems remaining however, which I cannot fully explain as yet. First, it is not always possible for an A argument to be expressed as genitive, in particular, a common A argument may not be expressed as a genitive. Second, it is possible, under certain circumstances, for the O argument to be expressed as a genitive argument. This latter situation is shown in (19), which, according to Seiter (1980) is a "formal" register construction. In (19), the O is genitive, and the A is ergative.

(19) e kotofa haaku e lautolu ke fakamatala
   Abs choose meGen Erg they Sbjntv speak
   "my being chosen by them to speak" (Seiter.87a.120)

Each of the above facts might provoke a completely different analysis from the one presented in this paper, more in line with that of Chung (1973) who claimed that no A argument, only S arguments can be genitive in Niuean. In other words, an analysis which rules out A as Genitive is half right and an analysis which rules in A as Genitive is half right. The first would be an E/A analysis, and the second, an N/A analysis. Also, if the construction in (19) is fully admitted into the paradigm, then O can be Genitive, also suggestive of an E/A pattern.

It is not directly apparent that a common A argument cannot be ergative, since common ergative and common genitive are both marked with he. Nonetheless we know, as Seiter (1980) demonstrates, that the case of he kuli "dog" in (20) is ergative and not genitive, as the O argument in this clause must have absolutive and may not have locative case. Absolutive partners with ergative, not genitive case.

(20)a. e gagau he kuli kɔ a au
   Abs bite Erg dog that Abs me
   "That dog's biting me." (Seiter.91a.121)

b. *e gagau he kuli kɔ i a au
   Nom bite Gen dog that Loc Pers me
   ("That dog's biting me.") (Seiter.91b.121)

The full range of possibilities in Niuean is thus as shown in (21). Note that if the genitive in a nominalized clause is proper, it can undergo preposing, just as a genitive in a non-derived clause.

(21) A_proper can be genitive
    A_common cannot be genitive
    S can be genitive
    O can be genitive but is marked, formal

    A_proper can be preposed
    S_proper can be preposed
    O cannot be preposed
    A,S_common cannot be preposed

I do not have an explanation for why common A arguments cannot be genitive. Seiter suggests it is not possible because the genitive he marking in a nominalized clause would be
ambiguous between common ergative and common genitive. It does seem suggestive that the
two case markings are identical, which is not the case for proper ergative (e) and proper genitive
(ha), or common or proper absolutive (e, a) and common or proper genitive (he, ha). But it is
difficult to see how this could be formalized.

As for the situation with O being realized as genitive, as in (19), I depend here on Seiter's
observation that this construction is a marked one, restricted to a formal register. If it is a frozen
sentence type, we can explain why the genitive O, even if proper, cannot be preposed.

4. Conclusion

In conclusion, if we put aside the two cases discussed in the preceding section, it appears
that Niuean, an E/A language in the verbal clause, switches to an intransitive Nominative system
in the nominalization system. In this, it appears to exhibit a reverse situation of that presented
for N/A languages such as English, by, for example, Alexiadou (1999). Many N/A languages
are said to switch from a N/A system in the verbal clause to an unaccusative system in the
nominalization system. Niuean is also unusual in the typology developed in Koptjevskaya-
Tamm in that it is a VSO E/A language exhibiting a pattern of nominalization found by her only
in SVO N/A languages.

References

Alexiadou, Artemis. 1999. On the Syntax of Nominalization and Possession: Remarks on
Patterns of Ergativity. Doctoral dissertation, University of Potsdam.
Bittner, Maria and Ken Hale. 1996. The Structural Determination of Case and Agreement.
Linguistic Inquiry 27:1-68.
Zealand.
den Dikken, M. and Rint Sybesma. 1998. Take Serials Light up the Middle. Ms., CUNY and
Leiden University.
Foley, William. 2000. Categorial Change in Oceanic Languages: First Contact on the North
University Press.
Harley Heidi, and R. Noyer. 1998. Licensing in the Non-lexicalist Lexicon: Nominalizations,
Marantz, Alec. 1997. No Escape from Syntax: Don't Try a Morphological Analysis in the
Privacy of your own Lexicon. Ms., MIT.
Massam, Diane and Wolfgang Sperlich. to appear. Possession in Niuean. submitted to
Language Typology and Universals (Sprachtypologie und Universalienforschung) special
volume on Possessive Markers in Central Pacific Languages.
Pawley, Andrew. 1966. Polynesian Languages: a Subgrouping based on Shared Innovations in
Niuean Nominalization


Valence changing clitics and incorporated prepositions in Teop
(Oceanic, Bougainville, PNG)
Ulrike Mosel & Jessika Reinig
University of Kiel

1. Introduction

Our paper examines valence changing clitics and incorporated prepositions in Teop, an Oceanic language of the Meso-Melanesian Cluster which is spoken on the north east coast on Bougainville. For practical reasons, the investigation will be strictly synchronic.

2. The structure of the verb complex

Teop has a highly complex verb phrase which besides the nucleus and TAM markers can contain the following elements:

- a discontinuous negation
- an incorporated noun
- a serial verb
- a directional particle
- several adverbs
- a valence changing clitic or an incorporated preposition
- the 3rd person plural object marker.

The boundaries of the verb phrase are defined by TAM markers which hold the initial and the final position. The final TAM marker is a portemanteau morph which indicates imperfective aspect and person. Typical examples are the following:

(1)  E bubuu na hio ihuana nana bona nahu kaukau.
      ART granny REAL sit wait for IMPF:3SG ART pot sweet potato

      ‘Granny is sitting waiting for the pot of sweet-potatoes.’ (pr. 4.27)

Example (1) shows the TAM markers na and nana which enclose the serial verb construction hio ihuana ‘sit and wait for’.

In example (2)

(2)  Ean sa paa rake haa bata maa nom a beiko te-naa.
      2SG NEG PAST like NEG at the same time DIR IMPF:2SG ART child PREP 1SG

      ‘At the same time you did not like my child.’ (Ioanna)

we find:
- the discontinuous negation sa ... haa.
- the past tense marker paa
- the nucleus rake ‘like’
Ulrike Mosel & Jessika Reinig

- the adverb bata ‘at the same time’
- the directional marker maa ‘hither, towards me or the deictic centre’,
- and nom, the imperfective aspect marker of the second person singular. This form is also used for first person singular, exclusive plural, and the second person plural.

The two examples also show the basic word order of transitive clauses, namely:

subject + verbal predicate + object

3. Valence of verbs and verb complexes

As Margetts (1999) has recently shown for the Oceanic Papuan Tip language Saliba, we have to distinguish between the valence of verbs and the valence of verb complexes, because the valence of the verb and the verb complex may be different. In Teop, for instance, an intransitive verb can be combined with an applicative clitic or an incorporated preposition which makes the verb complex transitive. On the other hand, an intransitive verb complex can be formed by the combination of a transitive verb with an incorporated object.

Teop has intransitive, transitive and ditransitive verbs and verb complexes. An example for a ditransitive verb is hee ‘give’ as in the following example:

(3) Eam toro hee e beera bona kahoo guu.
2PL must give ART chief ART head pig
‘You must give the chief the pig’s head.’ (TD hee)

Here the verb complex toro hee ‘must give’ is also ditransitive. The semantic roles of the two objects of ditransitive verbs and verb complexes are not confined to Recipient, Benefactive or Addressee and Patient as in (3). In our data we also found several examples of activity verbs whose first and second object denote a patient and an instrument as in (4)

(4) Miriki a beiko bona ta naono.
strike ART child ART piece wood
‘Strike the child with a stick.’ (TD miriki)

The core arguments of intransitive, transitive and ditransitive predicates do not need to be expressed if the participants can be identified from the context.
4. Articles

Noun phrases start with an article. The paradigm of articles is rather complex and cannot be described here. For the present paper it is sufficient to know that there are - among others -

1. the basic articles a, e and o as in (1) e bubuu 'granny' and in (2) a beiko tenaa 'my child' and
2. the object articles bona, bene and bono as in (1) bona nahu kaukau 'the pot of sweet-potatoes'.

The object articles are used with object noun phrases provided that the object follows the verb complex, and that the subject is a third person singular as in example (1). Otherwise the object noun phrase has the basic article as in (2) where the subject is a first or second person. Which of the basic articles a, e, o or the object articles bona, bene, bono is selected depends on the noun class of the noun. The article selection rule also applies for the first object of ditransitive predicates. Depending on the person of the subject, the object takes either the basic article or the object article. The second object, however, is always marked by the object article; for example bona kahoo guu in (3) and bona ta naono in (4).

5. The valence changing clitics ni and me/mi

The valence of verb complexes can be increased by the clitics ni and me/ mi which never occur outside the verb complex. As will be shown in a minute, both clitics can be compared to applicatives in other languages according to the definition given by Baker (1996):

(5) Applicatives are standardly characterized as constructions in which a derivational morpheme is attached to the verb resulting in the verb taking a new nominal as its surface object. Thus, the applicative form of an intransitive base verb acts like a transitive verb, while the applicative form of a transitive base verb acts like a ditransitive verb. This new 'applied' nominal expresses a thematic role that otherwise would have to be expressed obliquely or not at all. (BAKER 1996: 427)

(6) ni applicative clitic marking an object whose semantic role is determined by the lexical semantics of the verb  
me/mi applicative clitic marking a comitative object
5.1. The applicative clitic ni

The clitic ni transitivises intransitive verb phrases. Example (7)

(7) A moon na gavagava nana.
    ART girl REAL angry IMPF:3.SG
    'The girl is angry.' (TD gavagava)

shows an intransitive verb complex. In example (8)

(8) E bubuu na gavagava ni nana bene Sovavi.
    ART granny REAL angry APPL IMPF:3.SG ART Sovavi
    'Granny is angry with Sovavi.' (pr. 3.19)

the verb complex na gavagava nana is transitivised by ni and governs the object bene Sovavi.

Our data do not provide any example in which ni makes a transitive verb complex ditransitive, but there are examples in which ni increases the semantic transitivity in the sense of Hopper & Thompson’s (1980) notion of transitivity. For instance,

(9) ato v.t., touch s.th., ato ni v.t., hold onto s.th.

In (10)

(10) ... eove he ato bona metatono.
    3SG CONJ touch ART eel
    '... he touched the eel.' (pr. 5.25)

the eel is only touched, whereas the object in (11)

(11) Ato vakikisi ni a taba.
    hold strongly APPL ART thing
    'Strongly hold onto the thing' (for instance, a wriggling fish you have in your hand).
    (TD ato)

is hold against its resistance. Another example is

(12) rahi v.t., pull (a string, fishing line), haul (a canoe),
    rahi ni v.t., catch (fish with a fishing line) (TD rahi)

This corresponds to Hopper and Thompson’s (1980) notion of transitivity as a graduable property of clauses. They observed that clauses expressing a high degree of affectedness of the object are often also formally more transitive than those with less affected objects.

The semantic role of the applied object is directly related to the lexical semantics of the verb and seems to be predictable. It expresses for example
the CONTENT with verbs of speech and thought
  moroko 'speak' moroko ni 'speak about'

the CAUSE with verbs of physiological reactions
  dadana 'shake, shiver' dadana ni 'shake, shiver because of'
  mate 'die' mate ni 'die of'
  tagune 'wake up' tagune ni 'wake up from'

the STIMULUS/TARGET of psychological verbs:
  mararae 'be happy' mararae ni 'be happy about'
  naabu 'be afraid' naabu ni 'be afraid of'

the INSTRUMENT of activity verbs
  kiki 'play soccer' kiki ni 'play soccer with (the ball of ...)'

There do not seem to be constructions with oblique arguments which would correspond to the transitive constructions with ni. Thus 'to speak about', 'shiver because of', 'die of' etc. can only be expressed by the applicative construction. In other words, the ni-constructions do not serve as a means of promoting oblique arguments to core arguments.

5.2. The applicative clitic me

The applicative clitic me/ni is similar to ni, the only difference being that the semantic role of the applied object is always CONCOMITANCE. Typical examples are found in

(14) tei (itr.) 'live, stay, be' tei me/ni (tr.) 'be with, have s.th.'
  toho (itr.) 'play' toho me/ni (tr.) 'play with s.o.'

Examples (15) and (16) show the contrast between the intransitive construction of tei and the transitive construction of tei ni.

(15) Na tei nana te-a inu a runaa.
    REAL be IMPF:3SG PREP-ART house ART small
    'He stays in a small house.' (TD tei)

(16) A beiko na tei mi nana bono kaakan ...
    ART child REAL be with IMPF:3SG ART worm
    'The child has worms...' (TD tei)

As with ni, we do not have any examples of di-transitivisation.
6. The incorporation of prepositions

The valence of verb complexes can be increased by the incorporation of one of three prepositions. These prepositions are:

(17) **ki / ka-** ‘for, to’ (recipient, benefactive, addressee)

**kahí** ‘from’ (source)

**suku** ‘because of’ (cause)

Prepositional phrases introduced by **ki** can function as a predicate in non-verbal sentences and as adjuncts as in (18) and (19).

(18) **A- maa toogori ki bene teije?**

**ART-PL toogori for ART who**

'The toogori (fish) are for whom?' (pr. 3.57)

(19) **Enam na von a sosepene a voon ki bene iaa.**

**EXC REAL buy ART saucepan ART new PREP ART Mum**

'We bought a new saucepan for Mum.' (elicited)

When the prepositional phrase functions as a predicate or an adjunct, the noun phrase governed by **ki** has the object article. In (19), the **PATIENT NP a sosepene a voon ‘a new saucepan’** is the direct object and the **RECIPIENT ki bene iaa ‘for Mum’** is the adjunct. According to the article selection rule, the **PATIENT NP** has the basic article and the **RECIPIENT NP** has the object article. But when **ki** is incorporated, the **RECIPIENT** becomes a direct object as in (20):

(20) **Enam na von ki ma-e iaa bona nahu a voon.**

**EXC REAL buy for DIR-ART Mum ART pot ART new**

'We bought Mum a new pot.' (TD nahu)

Here the **RECIPIENT NP (e iaa)** has the basic article because it is the direct object and the subject is not a third person. The **PATIENT NP bona nahu** is the second object and has the object article.

The examples (19) and (20) show that incorporation of **ki** promotes the adjunct to a direct object and at the same time demotes the former direct object to a second object. Thus the incorporation of **ki** here makes a transitive verb complex ditransitive. An example of transitivisation is given in (21) and (22):

(21) **Na ahehe nana.**

**REAL sing IMPF:3G**

'She is singing.'

(22) **Na ahehe ki nana bene Satoka ...**

**REAL sing for IMPF:3SG ART Satoka**

'She is singing for Satoka.' (pr. 4.19)

The examples (23) to (26) show the incorporation of the prepositions **kahí** and **suku**.
Valence changing clitics and incorporated prepositions in Teop

In (23) kahi ‘from’ introduces an adjunct, whereas in (24) it is incorporated.

(23) Huriki a rokoroko vaarau a mate kahi bona hanana.
remove ART frog DEM ART dead from ART road
'Remove the frog from the road.' (TD huriki)

(24) E Kakato na rosin kahi bata maa-na bona otei.
ART Kakato REAL run from simultaneously DIR- IMPF:3SG ART boy
'Kakato is running away from the boy towards me.' (TD kahi)

Similarly, the preposition suku ‘because of’ introduces an adjunct in (25) and is incorporated in (26).

(25) A maa ta vaan na aun kahi bona maa vaan te-ori suku bona puana.
ART PL people REAL leave from ART PL village PREP-3PL because.of ART war
'The people left their villages because of the war.' (TD suku)

(26) A beiko moon na arata suku nana bene sina-na-e to kare mate koa.
ART child woman REAL grieve because.of IMPF:3SG ART mother-POSS.3SG REL recently die just
'The girl is grieving because of her mother who just recently died.' (TD suku)

7. Conclusion

To conclude, there are two different kinds of applicatives in Teop. The transitivising particles ni and me which only occur within the verb complex and the incorporated prepositions ki, kahi and suku. The clitics ni and me/mi transitivise intransitive verb complexes, and ni increases the semantic transitivity of clauses. The preposition ki makes intransitive verb complexes transitive and transitive verb complexes ditransitive. Our paper also showed that it is very useful to distinguish between the valence of verbs and the valence of verb complexes.

<table>
<thead>
<tr>
<th></th>
<th>with intransitive verbs</th>
<th>with transitive verbs</th>
<th>effect on the clause</th>
</tr>
</thead>
<tbody>
<tr>
<td>ni</td>
<td>transitivisation</td>
<td>increase of semantic</td>
<td>the only way to express a certain state of affairs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>transitivity</td>
<td></td>
</tr>
<tr>
<td>me</td>
<td>transitivisation</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>ki</td>
<td>transitivisation</td>
<td>ditransitivisation</td>
<td>change of grammatical relations</td>
</tr>
</tbody>
</table>

Notes

1 Thanks to Ruth Saovana-Spriggs who taught us Teop and provided all data, i.e. the elicited sentences, the primers and the texts which she taped during fieldwork in Bougainville in 1994. This fieldwork was funded by a small Australian Research Council grant.
References
Emotion Predicates and Grammatical Functions in Indonesian

Simon Musgrave.

Department of Linguistics and Applied Linguistics

The University of Melbourne

1. Introduction

This paper deals with a group of predicates in Indonesian and their status in regard to the syntactic transitivity system of that language. These predicates are part of a larger group which are characterised by their syntactic behaviour and by their semantics. Syntactically, these predicates appear to be intransitive in that they do not take the transitive verb prefixes men- and di-, but they can have a second argument which is coded as a prepositional phrase. Semantically, with one exception, these predicates denote emotional and cognitive states. Typical examples are the predicates puas 'content, satisfied' and salut 'respect, admire':

1. Dan mereka juga sudah puas dengan kasus wanita yang hilang itu?
   and 3pl also PERF content with case woman that 'And were they also satisfied with the case of the woman who disappeared?' (SDPGS: 128)

2. ia bangga dan salut akan kesadaran Sandy
   3sg proud and respect about NOM.aware.NOM Sandy
   'she was proud and respected Sandy's attention' (PYD:19)

Some of these predicates can also appear in a construction where the second argument, which I will refer to often as the stimulus (of the emotion or cognition), is coded as a simple NP. That is, the clause has what look like two direct arguments as seen in the following examples:

3. Kamu lupa rumahku?
   2sg forget house.1sg
   'You have forgotten my house?' (SDM: 96)

---

1 Examples in this paper which are not taken from texts reflect the judgments of M. Umar Muslim and Katerina Sukamto, for whose patience I am very grateful. I am also grateful to Helen McKay for providing me with numerous examples drawn from the corpus of Indonesian journalism which she has collected. Several discussions with Peter Austin and one with Bill Foley were important in shaping this material, as was the feedback of an audience at The University of Melbourne particularly that of Nick Evans and Rachel Nordlinger. Remaining errors are my own responsibility.

2 The following abbreviations are used in glosses: 1,2,3 - 1st, 2nd, 3rd person, sg - singular, pl - plural, APPL - applicative, EX - exclusive, FUT - future, INT - intensifier, LOC - locative preposition, NEG - negation, NOM - nominalizer, PERF - perfect aspect, PROG - progressive aspect, RED - reduplication, REL - relative clause marker. The verbal prefixes men- and di- are left unglossed throughout. The abbreviations used for sources of examples are explained at the conclusion of the paper.
4. *Ia kuatir suratnya tidak sampai*  
3sg fear letter.3 NEG arrive  
'She was afraid her letter would not arrive.' (ES: 313)

5. *Aku takut wanita yang aku cintai ternyata*  
1sg fear woman REL 1sg love.APPL apparently  
tidak mencintai ku  
NEG meN.love.APPL.1sg  
'I am afraid of the woman that I love, apparently she doesn't love me.' (PYD: 12)

6. *Cuma aku tidak suka omongannya yang sok suci,*  
only 1sg NEG like gossip.3 REL as.if pure  
'I just don't like his gossip that takes the moral high ground.' (PYD: 87)

7. *Saya nggak senang wartawannya*  
1sg NEG like journalist.3  
'I don't like journalists.' (McKay)

However, as will be shown in detail below, such clauses are not syntactically transitive. The question this paper seeks to answer then is what grammatical function is assigned to the stimulus argument in a clauses such as examples 3 to 7?

I attempt to answer this question using the theoretical resources of Lexical Functional Grammar (LFG). This framework assumes a constrained list of grammatical functions (GFs): subject (SUBJ), object (OBJ), oblique (OBL₉) and second object (OBJₑ). The last two GFs are subscripted with a theta because in the case of OBJₑs their thematic role is assumed to be restricted within any particular language (Bresnan and Kanerva 1989), and in the case of OBL₉s their thematic role is indexed by the adposition or case marker which licenses them. These GFs are not configurationally defined. They are elements of f(functional)-structure, a structure which is distinct from but parallel to c(onsituent)-structure (Bresnan, to appear: 55-63). Syntactic properties other than configurational ones are therefore of primary interest in this investigation.

The group of predicates of interest here includes the following words:

---

³ The possessive clitic *-nya* is used here to indicate that the referent is identifiable.
Emotion Predicates in Indonesian and Grammatical Functions

8. berang angry, irate mimpi dream
bosan bored mirip resemble
gila crazy, obsessed peduli care about, pay attention
kangen miss, long for percaya believe
kasih love sayang love, pity
kasihan love, pity senang happy, like
kuatir fear suka like
lupa forget takut fear
marah angry yakin sure

This list is not exhaustive; rather it includes just those words for which I have clear examples from texts. Also, several words are left out of this list which might have a claim to be included. These words (benci 'hate', ingat 'remember' and percaya 'believe') all occur in the construction of examples 1 and 2. But they also all have related transitive verbs which are not derived with an applicative suffix (membenci, mengingat and mempercaya). Therefore it is not possible to say whether clauses with these predicates and a NP stimulus are examples of the construction seen in examples 3 to 7, or whether they are rather examples of the unprefixed transitive verb construction exemplified in example 15 below. Given this unclarity, it is preferable to omit these predicates from my discussion.

The translations given above suggest that many of these words have an adjectival character. Whether Indonesian has adjectives as a lexical category or not is a disputed question (see Cumming 1991 and Mahdi 1998 for differing views), and the evidence is particularly unclear for the words under consideration. For example, ter- prefixation derives adjectival superlatives for most of the words with which it is possible (e.g. tergila 'most crazy', termarah 'most angry'), but some verbal derivatives occur also (e.g. terlupa 'forget accidentally'). Indonesian allows all major phrasal categories to function as the predicate of a clause, that is verbs (see examples such as 12), nouns, prepositions and putative adjectives can all head the predicate constituent:

9. Ini keputusan saya
   this NOM.decide.NOM 1sg
   'This is my decision.'

10. Mereka di Jakarta sekarang
    3pl LOC Jakarta now
    'They are in Jakarta now.'

11. Mencari pekerjaan di kota tidak begitu mudah
    menN.find NOM.work.NOM LOC city NEG like that easy
    'Finding work in the city isn't very easy.'

It is uncontroversial that all of these predicate types assign a SUBJ grammatical function, and a maximally general account of Indonesian syntax must be based on the assumption
that if they assign other clause-level grammatical functions, these should also be part of the
inventory provided by our theory⁴. Therefore, the question at issue here remains the same
under either categorial analysis. I will leave the question of category open here, and will
continue to use the neutral term *predicate* in this paper.

The data presented in this paper does not represent standard Indonesian. Many of the
examples are drawn from contemporary, popular novels, one set in Jakarta (SDM) and one
set in Surabaya (PYD). The majority of the characters in both novels are educated people,
and in conjunction with the judgments of the native speakers with whom I have worked,
this data may be said to represent the careful speech of educated Indonesians⁵.

Section 2 of the paper examines the evidence that constructions such as those in examples
2 to 7 are distinct from normal transitive clauses in Indonesian, and shows that the stimulus
is neither a SUBJ nor an OBJ. Sections 3 and 4 examine evidence which might show that
the stimulus is an OBLₑ or an OBJ₀ respectively, concluding that the second possibility is
to be preferred. Section 5 discusses some theoretical consequences of this analysis and
compares these to the consequences of instead claiming that bare NP stimuli in Indonesian
are examples of some grammatical function not yet recognised by LFG.

### 2. Transitive Clauses - SUBJ and OBJ

The overwhelming majority of transitive verbs in Indonesian can appear in four types of
clause:

12.  *Dia membaca buku itu*
S/he read the book.

13.  *Buku itu dibaca (oleh) Ali*
The book was read by Ali.

14.  *Buku itu saya baca*
The book, I read.

15.  *Dia baca buku itu*
S/he read the book.

Various restrictions apply to the constructions exemplified in 12 to 15 which need not
concern us here. Clauses with emotion predicates look the same as the type exemplified in
15, but the other clause types are not possible:

---

⁴ There is also the possibility that clauses such as examples 9, 10 and 11 should be analysed as containing a
zero copula, a possibility which I ignore here.

⁵ The two native speakers whose judgments are reported here are both Javanese. The issue of whether the
phenomena discussed here are affected by the speaker's first language (if that is not Indonesian) remains for
future research.

The type of clause illustrated in example 13 is not possible with a first person actor, but substitution of a third person actor does not make this construction possible with this predicate:

17. "I am fed up with monotonous life in Surabaya." (PYD: 180)


The type exemplified in 14 is not possible either, although the direct manipulation of example 16 yields a sentence which does have a reading:


'This monotonous life in Surabaya, I am fed up with it.'

This clause requires an intonation distinct from that associated with clauses of the type seen in example 14 with a pause after the initial NP which is a topicalised constituent, and the addition of an auxiliary verb to the sentence makes the distinction absolutely clear. In a construction of the type in example 14, the actor pronoun remains adjacent to the main verb even when an auxiliary or modal is included, but in emotion predicate clauses, the actor pronoun appears to the left of an auxiliary:

20. Buku itu akan saya baca.

book that PROG 1sg read

'I will read that book.'


'This monotonous life in Surabaya, I will be fed up with it.'


This evidence shows that these predicates are not syntactically transitive.

This status alone may be enough to show that the stimulus argument of an emotion predicate is not an OBJ. If OBJ is defined as the non-SUBJ argument of a transitive verb, then stimuli are not objects because the predicate that they are associated with is not syntactically transitive. A more stringent test requires OBJs be able to become the SUBJ of a related clause, typically a passive. Leaving aside the issue of whether the related clause types in Indonesian should be described as passive, it is clear that stimuli also fail this test as seen in example 18. Many of the emotion predicates derive true transitive verbs with an applicative suffix, and a comparison of their properties with the basic predicates is instructive. The applicativised verbs appear with both meN- and di- prefixes, with the alternatives used to make an argument available for relativisation as in the following examples:

6 I assume that, as SUBJ and OBJ are syntactic concepts, syntactic transitivity is the relevant notion for this argument.
23. *Kamu punya ibu yang menyayangimu.*

2sg have mother REL meN.love.APPL.2sg

'You have a mother who loves you.' (SDM: 84)

24. *Untuk menemui anak yang begitu disayanginya.*

for meN.meet.APPL child REL so di.love.APPL.3

'In order to meet a child who is so loved.' (SDM: 208)

The contrast in syntactic behaviour between the derived transitive verb which does have an OBJ and the emotion predicate is clear.

The possibility that the stimulus is the SUBJ of its clause can also be eliminated for various reasons. Firstly, it would be surprising that an argument alternated between being an OBL and a SUBJ with no verbal morphology. Secondly, it can be seen from example 23 that the experiencer is the SUBJ of the corresponding applicative verb with meN-prefix, the clause type which corresponds to an active clause. Applicativization does not affect SUBJs, therefore the experiencer should be the SUBJ of the basic predicate. A clause can only have one SUBJ, therefore the stimulus cannot be one. Finally, Indonesian has a third person pronoun, *ia*, which is restricted (more or less) to appear in SUBJ position. This pronoun can occur as experiencer of an emotion predicate clause:

25. *Ia kuatir suratnya tidak sampai.*

3sg fear letter.3 NEG arrive

'She was afraid her letter would not arrive.' (ES: 313)

Once again, the fact that a clause can only have one SUBJ forces the conclusion that the stimulus cannot be SUBJ.

This section has demonstrated that emotion predicates are not transitive verbs in Indonesian. Therefore, a bare NP stimulus following such a predicate cannot be an OBJ. And the (remote) possibility that such arguments are SUBJs has also been eliminated.

3. **OBL**

Having established that a NP stimulus of an emotion predicate cannot be assigned the GFs SUBJ or OBJ, I now turn to examine the arguments as to whether such elements might be classified as OBL. The following pair of examples, which are adjacent in the text in which they occur, suggests that there is no clear semantic difference between the two constructions for speakers:


father love to.2 Risa

'Father loves you, Risa'.


Risa also love father

'Risa loves Father too.' (SDM: 293)
Therefore, it might be possible to argue that the stimulus in example 27 and similar clauses is still an OBL, but one with non-standard coding. There are two pieces of evidence which suggest that this would be a wrong conclusion.

Firstly, there is once more a definitional issue. Indonesian is not a case-marking language, and therefore the obvious distinction between core and non-core arguments is the presence of prepositions. This is not a definitive criterion however. The literature includes examples both of NPs with the coding properties of obliques being treated as direct arguments, and of NPs with the coding properties of direct arguments being treated as obliques. Examples of the first type are animate, referential objects in Spanish, which are preceded by the preposition a (Hopper & Thompson 1980) and various subjects in Icelandic with ‘quirky’ case-marking (Zaenen, Maling & Thrainsson 1985). An example of the second type is the use of accusative case in Icelandic to mark some temporal adjuncts (Smith 1996). Such examples show that coding properties alone are not sufficient to establish the syntactic status of an argument. Behavioural tests are necessary also, and I now turn to two such test which show that the stimulus of example 27 and similar clauses does not have the properties of an oblique argument.

The first way in which NP stimuli behave like direct arguments is that a quantifier can be floated from them. This is not possible with a PP stimulus, or a PP which is an adjunct:

   child.RED that like sugar.RED that all
   'All the children like the sweets.' OR
   'The children like all the sweets.'

29. Anak-anak itu suka dengan gula-gula itu semuanya
   child.RED that like with sugar.RED that all
   'All the children like the sweets.' NOT
   'The children like all the sweets.'

30. Orang-orang Sasak datang dengan anak-anaknya semuanya
   man.RED Sasak come with child.RED.3 all
   'All the Sasak people came with their children.' NOT
   **'The Sasak people came with all their children.'**

In each case, the quantifier can be read as having floated from the SUBJ, and evidence to be presented below will show that other direct arguments also have this property. But only with the NP complement of an emotion verb can the quantifier be read as having floated from the other NP in the clause; when that NP is within a PP the quantifier cannot be construed with it.7

The second piece of evidence comes from extraction facts. Indonesian subjects can always be extracted leaving a gap and this generalisation is agreed on by all sources. The facts regarding other arguments of transitive verbs are in dispute and will be discussed below. For emotion predicates, I am only aware of one discussion of extraction in the literature,

---

7 One of my consultants allows both readings of example 29, but the second reading seems to be marginal and is strongly rejected by the other consultant.
Stephens (1970), and this source claims that extraction of NP stimuli is grammatical. This judgment is shared by the native speakers I have consulted. Gapped extraction from PPs is completely impossible however, neither preposition-stranding nor pied-piping is available in Indonesian. The status of the PP involved, selected or not, is not relevant:

31. *gula-gula yang anak-anak suka dengan itu
   sugar.RED REL child.RED like with that
   'the sweets that the children like'

32. *gula-gula yang anak-anak suka dengan itu
   sugar.RED REL child.RED like with that

33. *gula-gula dengan yang anak-anak suka itu
   sugar.RED with REL child.RED like that

34. *anak-anak yang orang Sasak datang dengan itu
   FOR: 'the children that the Sasak people came with'

Clearly, the behavioural properties of NP stimuli are different from those associated with the grammatical function OBL in Indonesian. It is not possible to maintain an analysis which treats these constituents as obliques with unusual coding properties. Therefore, it seems that the only possible analysis within LFG's inventory of grammatical functions is to treat the stimuli as OBJes and I turn to an examination of the evidence for this position in the following section.

Although not part of the main argument being developed here, it is of interest to note that even when stimuli are coded as PPs, in the construction of examples 1 and 2, they have some properties which are not typical of obliques in general. In this construction, for any one predicate various prepositions are possible introducing the second argument:

35. tak percaya akan diri sendiri
    NEG believe about self self.INT
    'lack self-confidence' (E&S: 422)

36. Kau kira sekarang orang akan percaya dengan
    2sg guess now man FUT believe with
    bualmu itu?
    boasting.2 that
    'Do you think anyone will believe your boasting now?' (PYD: 200)

37. Kami tidak percaya kepada mereka.
    1pl.EX NEG believe to 3pl
    'We don't believe (in) them.' (McKay)

---

8 Kana (1986) discusses emotion predicates in some detail (see also below), but does not consider extraction.
9 Extraction with a resumptive pronoun is possible: see Sneddon (1996):289-291.
38. *Surat itu dikirim kepada wanita itu
letter that di.send to woman that
('The letter was sent to the woman.') (Chung 1976: ex.63a)

In the case of derived ditransitive verbs such as that in example 39, the same thematic argument would be OBJ in a related clause with an underived, meN-prefixed verb and has access to SUBJ position in the related clause with a di-prefixed verb:

40. Surat itu dikirim kepada wanita itu
letter that di.send to woman that
'The letter was sent to the woman.'
with a PP stimulus and the comparison between the two types is not in fact direct on this point.\(^{11}\)

In the previous section, examples were given which showed that a quantifier could float from a SUBJ. This is also possible from OBJS and from OBJ\(_{0}\)s:

41. \textit{Saya mukul anak-anak itu kemarin semuanya}  
\hspace{1cm} 1sg hit child.RED that yesterday all  
\hspace{1.5cm} 'I hit all the children yesterday.'

42. \textit{Saya memberinya hadiah itu semuanya}  
\hspace{1cm} 1sg me\textit{N}. give gift that all  
\hspace{1.5cm} 'I gave her all the presents.'

The descriptive generalisation is that quantifiers can float from any direct argument in Indonesian. And as previously shown, a quantifier can also float from an NP stimulus (example 28, repeated here):

43. \textit{Anak-anak itu suka gula-gula itu semuanya}.  
\hspace{1cm} child.RED that like sugar.RED that all  
\hspace{1.5cm} 'All the children like the sweets.' OR  
\hspace{2cm} 'The children like all the sweets.'

In this case, the stimuli behave like direct arguments; they are not SUBJ or OBJ, so OBJ\(_{0}\) is the remaining possibility.

Many descriptions of Indonesian state that extraction (relative clause formation and question formation) is only possible with SUBJs. The facts are a good deal more complicated than this as soon as anything other than the prescriptive standard is investigated; a more complete account is given by Voskuil (1996: Ch 8). For the purposes of this discussion, I will assume that all direct arguments can be extracted in the absence of verb affixation\(^{12}\). Thus an OBJ can be extracted from a clause of the type seen in example 15 (auxiliary added to disambiguate the construction)\(^{13}\):

44. \textit{buku yang dia akan baca itu}  
\hspace{1cm} book REL 3sg FUT read that  
\hspace{1.5cm} 'the book that she will read'

As for OBJ\(_{0}\)s, judgments vary. As might be expected, such arguments cannot be extracted when the verb carries a prefix, and no native speaker I have consulted will allow extraction when the verb is derived with an applicative suffix. In the case of undervield, unprefixed

\(^{11}\) I ignore here the issue of applicative derivations in Indonesian which take a transitive verb and derive another transitive verb with a different meaning e.g. \textit{memergok} 'catch someone by surprise', \textit{memergoki} 'catch someone redhanded'.

\(^{12}\) The construction seen in example 14 is a special case: despite the lack of a verb affixes, only the SUBJ can be extracted. Musgrave (in preparation) offers a tentative account of why this should be so.

\(^{13}\) Michael Ewing (p.c.) informs me that he has observed numerous examples of such constructions in recorded conversation of educated Indonesians.
ditransitive verbs, judgments vary. At least some speakers will permit extraction of the second object from clauses of this type:

45. Saya akan beri dia buku itu
    1sg FUT give 3sg book that
    'I will give him the book.'

46. buku yang saya akan beri dia itu
    book REL 1sg FUT give 3sg that
    'the book that I will give him'

As demonstrated above (example 31), this type of extraction is also possible with a NP stimuli:

47. orang yang saya suka itu
    person REL 1sg like that
    'the person that I like'

On this test, the NP stimulus possesses the syntactic properties of a direct argument more clearly than OBJs. I have demonstrated that NP stimuli cannot be plausibly assigned any of the other three GFs recognised by LFG, and that their syntactic properties are very similar to those of OBJs. Within the theory as currently formulated, this would seem to be the only possible analysis. The only alternative is to add a GF to the inventory, and the final section of this paper examines which of these two possibilities is preferable from a theoretical point of view.

5. Theoretical Implications

The data presented in the previous sections show clearly that the NP stimuli associated with emotion predicates in Indonesian are direct arguments of the predicate, but they are not OBJs as usually understood. Previous studies have ignored this issue, or presented unsatisfactory accounts. Stephens (1970) aims only to draw attention to the existence of what he terms a class of 'pseudo-transitive' verbs in Indonesian, and offers no account of the syntax of the clauses in which they occur. Kana (1986: 283-289), working in the Relational Grammar framework, analyses such arguments as initial 3s or Locatives, without appearing to consider these as genuinely different possibilities. The possibility that such arguments are Locatives, presumably a variety of oblique, seems to contradicted by the evidence presented in section 3 of this paper. This analysis must also explain the fact that out of the range of prepositions which are used to introduce the stimuli of emotion verbs, locative prepositions are noticeably absent. In my database, the basic locative preposition di never occurs with these predicates. I have two examples using the locative noun atas 'top surface' alone; it is normally used in association with di. Otherwise, the common prepositions used code movement towards (pada, kepada) or accompaniment (dengan, sama). Kana's alternative, that the NP stimuli are initial 3s, is essentially the position argued for in this paper.

---

14 This is not true for the verb ajari 'teach' for my primary consultant. I have no explanation of this fact.
In classic LFG, as presented in Kaplan and Bresnan (1982), OBJes are called OBJ-2s and it is explicitly stated that this grammatical function can only be assigned if the function OBJ has already been assigned. Clearly, the analysis argued for above is not possible within this scheme. However, in LFG as currently practised (Bresnan, to appear), it is not clear that the constraint still holds. Grammatical functions are now decomposed into feature bundles and all have equal status from this point of view. The differences in occurrence of the various GFs can be ascribed to the operations of Lexical Mapping Theory (LMT) on the lexical entries of the verbs of a language, rather than being stated as direct constraints on GFs.

Further consideration of LMT suggests both a way of thinking about the properties of Indonesian emotion predicates which makes them seem less strange, and a good reason for preferring the analysis of NP stimuli as OBJes to the alternative. On the first point, LMT is based on the decomposition of GFs into the features [+/-object] and [+/-restricted]. The predicates in question can have a second argument which is either an OBL or an OBJ. These form a natural class in LMT, both having the feature [+r]. So the proposed analysis reduces to claiming that this class of predicates lexically specifies that their second argument is [+r]; then some predicates have a second lexical entry with the additional information that they license a second direct argument. The operations of LMT will do the rest of the work, assigning the GF OBL or in the first case and OBJ in the second case. LMT treats all four GFs as equal in essence, and from this point of view the relative unimportance of the GF OBJ in the grammars of languages is surprising. The present analysis might be viewed then as adding empirical support to the view of GFs implicit in the LMT.

On the second point, it might be suggested that the evidence presented here justifies proposing a more fine-grained analysis of non-subject arguments in Indonesian than is allowed by LFG. The syntactic properties of NP stimuli and OBJes are not identical, and the considerable overlap might be due only to the fact that both are types of direct argument. There might be more than three types of non-subject argument in the language. However, the LMT offers a strong theory-internal argument against adopting such an analysis. Two binary-valued features allow exactly the four GFs assumed by the theory, and introducing a new GF would disrupt this aspect of the theory. To do so must mean that the two features used thus far do not exhaust the relevant information, and another feature would be required. This in turn would mean that rather than adding one GF, at least two would have to be added, on the assumption that the new feature would be dependent on one of those already assumed. If the new feature was independent of the existing ones, then four new GFs would be added to the system. To my knowledge, there is no empirical support for reducing the constraints on the theory in this way. In the absence of such evidence, I conclude that it is preferable to retain the current inventory of GFs, and to analyse the NP stimulus associated with an emotion predicate in Indonesian as an OBJ.

A slight change of viewpoint suggests an alternative viewpoint which may be worth exploring. The picture of GFs which most of us are familiar with is very definitely based on the transitive clause - there are SUBJs, OBJs and OBLs and anything else is rather awkward to deal with. The data presented here certainly do not exhaust the complications of GFs in Indonesian (some additional problems are discussed in Arka and Manning (to appear)), and the best picture of Indonesian syntax may in fact be one which sees the major
oppositions as being between SUBJ, other direct arguments and OBLs with OBJ of a transitive clause being the awkward element to deal with. While the arguments presented here are valid for this limited data, a different treatment may be appropriate for more complete data and Musgrave (to appear) aims to provide such a treatment.

Sources of Examples
McKay - Corpus of Indonesian journalism collected by Helen McKay

References
Bresnan, Joan W. (to appear) Lexical-Functional Syntax To be published by Blackwell.
Clefts vs. pseudo-clefts in Austronesian
Ileana Paul
MIT/UQAM

1 Introduction

The goal of this paper is two-fold. First, I argue that the cleft in many Austronesian languages is best analyzed as a (kind of) pseudo-cleft. More precisely, clefts are copular constructions with a headless relative in subject position. Second, I show that the special focus interpretation associated with these clefts derives quite simply from the semantics of the headless relative (a definite description). (1) illustrates a typical example from Malagasy, a Western Austronesian language with VOS word order.

(1)   [ I Sahondra ] [no nanapaka ity hazo ity].
     Sahondra   NO PST.AT.cut  this tree this
     ‘It was Sahondra who cut this tree.’
     (lit.) ‘The one who cut this tree was Sahondra.’

The focused element, *i Sahondra*, is the matrix (copular) predicate. The clause-final subject is the headless relative *no nanapaka ity hazo ity* ‘the one who cut this tree’. The connection between clefts and pseudo-clefts has a long history in generative linguistics, beginning with observations by Jespersen (1928). Much research in the 60’s and 70’s explored this connection (e.g. Akmajian (1970); Pinkham and Hankamer (1975)). More recently, Percus (1996) has proposed that cleft sentences in English are derived from a structure with a headless relative in subject position. Whether or not this analysis is correct for English, I will show that their arguments apply very elegantly to the Austronesian data. That clefts are in fact pseudo-clefts in Austronesian is not a new claim. Other researchers have come to the same conclusion about Chamorro (Chung (1998)), Madurese (Davies (2000)), Malay (Cole, Hermon and Aman (to appear) (henceforth CHA)), Maori (Bauer (1991)), Palauan (Georgopoulos (1991)) and

---

1 I would like to thank Saholy Hanitriniaina for judgements on the Malagasy data and participants at AFLA VII for comments. Funding for this research was made possible by a post-doctoral fellowship from FCAR and by a SSHRC grant to Anna-Maria di Sciuolo.
2 I am using the term “pseudo-cleft” somewhat liberally here. Since Higgins (1973), pseudo-clefts in English are defined as having a free relative in subject position:
   (i) What Sahondra did was cut the tree.
   In the languages discussed in this paper, the subject is either a headless relative or a free relative. I do not take this difference to be important for present purposes.
3 Although I argue that the correct analysis of Austronesian “clefts” involves a pseudo-cleft, I will generally translate with cleft sentences. I choose clefts as in English, they are more natural and they bring out the properties I wish to discuss more clearly.
What has not been previously noted is the connection between the syntactic structure and interpretation. This paper does not make any claims about the universality of this analysis of clefts. In fact, both a movement analysis and a pseudo-cleft analysis may be possible (as is argued for English by Pinkham and Hankamer (1975) and for Chamorro by Chung (1998)). This paper, however, only motivates the pseudo-cleft analysis, based on structural and interpretive arguments. In other words, I leave open the possibility that some clefts may indeed be generated via movement.

2 Structure

Since the bulk of the data in the paper are from Malagasy, I give a brief introduction to Malagasy syntax. Malagasy is a Western Austronesian language spoken in Madagascar. The unmarked word order, which is fairly rigid, is VOS, illustrated in (2), where the subject is marked with a dotted underline.

(2) a. Nanapaka ity hazo ity tamin’ny antsy i Sahondra.
PST.AT.cut this tree this PST.P.GEN.DET knife Sahondra
’Sahondra cut this tree with the knife.’

b. Notapahin’i Sahondra tamin’ny antsy ity hazo,ity.

PST.TT.cut.GEN.Sahondra PST.P.GEN.DET knife this tree this
‘This tree was cut by Sahondra with the knife.’

c. Nanapahan’i Sahondra ity hazo ity ny.antsy.
PST.CT.cut.GEN.Sahondra this tree this DET knife
‘The knife was used by Sahondra to cut the tree.’

(2) also illustrates the basic voice alternations: Actor Topic (AT), Theme Topic (TT) and Circumstantial Topic (CT). Simplifying somewhat, the different voices promote different elements to subject: agents, themes and obliques, respectively. Similar voice paradigms may be found in other Austronesian languages, with varying degrees of complexity.

The core data I will consider in this paper are clefts, as in (1), repeated in (3).

(3) I Sahondra no nanapaka ity hazo ity.

Sahondra NO PST.AT.cut this tree this
‘It was Sahondra who cut this tree.’

On the surface, (3) involves the “fronting” of the subject, which is followed by a particle no. In Malagasy, as in most western Austronesian languages, only subjects and (certain) obliques may be focussed (Keenan (1972)). In order to focus an internal argument, passive is used. This restriction is shown by the contrast between the ungrammatical (4a) and the grammatical (4b).

(4) a. *Ity hazo ity no nanapaka i Sahondra.

this tree this NO PST.AT.cut Sahondra
‘It was this tree that Sahondra cut.’

4 The pseudo-cleft analysis of Malagasy is suggested by Pearson (1996: fn17).
CLEFTS VS. PSEUDO CLEFTS IN AUSTRONESIAN

b. Ity hazo ity no notapahin’i Sahondra.
   this tree this NO PST.TT.cut.GEN.Sahondra
   ‘It was this tree that was cut by Sahondra.’

Adjects, however, freely extract, whatever the voice on the verb, AT, TT, or CT.

(5) a. Tamin’ny antsy no nanapaka ity hazo ity Sahondra.
   PST.P.GEN.DET knife NO PST.AT.cut this tree this Sahondra
   ‘It was with the knife that Sahondra cut this tree.’

   b. Tamin’ny antsy no notapahin’i Sahondra ity hazo ity.
   PST.P.GEN.DET knife NO PST.TT.cut.GEN.Sahondra this tree this
   ‘It was with the knife that this tree was cut by Sahondra.’

   c. Tamin’ny antsy no nanapahan’i Sahondra ity hazo ity.
   PST.P.GEN.DET knife NO PST.CT.cut.GEN.Sahondra this tree this
   ‘It was with the knife that Sahondra cut this tree.’

Note that in (5c), the adjunct is first promoted to subject and then clefted. I will not provide an explicit analysis of the extraction restriction (which applies to all instances of A-bar movement) in this paper (but see section 2.3 for some discussion). Instead, I focus on the structure and interpretation of clefts.

As already seen above, typical examples of clefts involve NPs and PPs. In certain cases, apparent VP clefts occur.

(6) Mihinana akoho no mitsangana Rasoa.
   AT.eat chicken NO AT.stand Rasoa
   ‘It’s while eating chicken that Rasoa stands.’

In (6), the VP mihinana akoho ‘eating chicken’ acts like an adverbial modifier. The exact categorial status is therefore unclear.

Similar examples can be found in Chamorro, Madurese, Malay, Maori, Palauan and Tagalog:

(7) a. Pära si Jessie pära u-fina’gasi i kareta
   FUT Jessie FUT AGR-PASS.wash the car
   ‘It will be Jessie by whom the car will be washed.’ (Chung 1998)

   b. Siti se entar daq Jakarta.
   Siti REL go to Jakarta
   ‘It was Siti who went to Jakarta.’ (Davies 2000)

   c. Apa yang Ali beli?
   what that Ali buy
   ‘What did Ali buy?’ (CHA)

   d. Ko Hone i kite i te taahae.
   EQ John see DO the thief
   ‘It was John who saw the thief.’ (Bauer 1991)
e. Ng-Basilia a mengaus er tia el tet. [Palauan]
AGR-Basilia R-weave P DEM L bag 'It’s Basilia who’s weaving this bag.' (Georgopoulos 1991)

f. Si Bing ang gumawa ng sapatos na iyon [Tagalog]
NOM Bing NOM AV.PERF.make GEN shoe LNK that 'It was Bing who made those shoes.' (Kroeger 1993)

In all cases, the focussed element appears clause-initially.

It would be tempting to analyze these data as involving movement to the specifier position of a focus projection.

\[
\text{(8) } \begin{array}{c}
\text{FP} \\
\text{XP}_i \rightarrow \text{F'} \\
\text{focus} \rightarrow \text{F}^* \\
\text{IP} \\
\end{array}
\]

Such approaches have been proposed for focus movement in Hungarian (É. Kiss (1998)) and Italian (Rizzi (1997)). I argue, however, that in all the above languages the “cleft” in is in fact a pseudo-cleft. The basic structure is given in (9). The focussed element is an XP predicate and the presuppositional clause is a headless relative in subject position.\(^5\)

\[
\text{(9) a. } \begin{array}{c}
\text{I } \\
\text{Sahondra no nanapaka ity hazo ity.} \\
\text{Sahondra NO PST.AT.cut this tree this} \\
\text{(lit.) ‘The one who cut this tree was Sahondra.’} \\
\end{array}
\]

\[
\text{b. } \begin{array}{c}
\text{IP} \\
\text{I} \\
\text{XP} \\
\text{focus} \rightarrow \text{OP}_1 \ldots \text{ti} \\
\end{array}
\]

I first show that the focussed element patterns with predicates. I then turn to the presuppositional clause and provide arguments for a headless relative clause in subject position.

\(^5\) For the purposes of this paper, I assume the simple clause structure in (9), which is based on the analysis of V-initial word order in Guilfoyle, Hung and Travis (1992). Other analyses involve some kind of predicate fronting (e.g. V’ or VP movement to a specifier position), but the differences are not crucial to the story presented here. Of the languages discussed, Madurese and Malay are not predicate-initial. Therefore, clefting does involve movement of the predicate to some clause-initial position. I discuss the difference between predicate-initial languages and the SVO languages in section 2.1.
2.1 Focus=predicate

In the above cited languages, the focussed XP does have predicate-like properties. As a first observation, most of these languages are verb-initial. The null hypothesis is therefore that the clause-initial focus is a predicate. There are also language-specific arguments that support this analysis. For example, in Chamorro, the focus can be immediately preceded by a tense/aspect/mood particle (see (7a)). The Palauan example in (7e) shows that the focus takes the subject agreement marker ng. Thus in these languages, verbal markers such as agreement and tense associate with the focussed element.

In Malay the interrogative focus particle kah provides evidence for the predicate status of focus (Kader (1976)). Simplifying somewhat, kah can appear on elements in the predicate, but not on the subject (boldface indicates focus).

(10) a. Fatimah kata Siti membeli buku itukah semalam?
    Fatimah say Siti bought book that-Q yesterday
    ‘Did Fatimah say that Siti bought that book yesterday?’ (CHA (40))

b. * Fatimah kata Siti-kah membeli buku itu semalam?
    Fatimah say Siti-Q bought book that yesterday (CHA (41))

If the clefted element is indeed a predicate, we expect kah to be able to appear, as the data in (11) show to be the case. Crucially, (11b) is grammatical even though the wh word corresponds to the subject of the clause.

(11) a. Apakah yang Ali beli?
    what-Q that Ali buy
    ‘What did Ali buy?’ (CHA (46))

b. Siapakah yang datang?
    who-Q that came
    ‘Who came?’ (CHA (47))

The grammaticality of kah on the clefted word is evidence that the focussed element originates in predicate position, rather than subject position.

Similar data can be found in Madurese (Davies (2000)). The emphatic particle jhoj can appear in the predicate, as shown by the examples below.

(12) a. Ali a-barriq jhoj buku jhuwa daq Siti!
    Ali AV-give EMPH book that to Siti
    ‘Ali gave the book to Siti!’

b. Ali a-barriq buku jhuwa daq Siti jhoj!
    Ali AV-give book that to Siti EMPH
    ‘Ali gave the book to Siti!’

If the emphatic particle appears on the subject, however, the subject must be in the clefted position.

(13) Ali jhoj *(se) a-barriq buku jhuwa daq Siti!
    Ali EMPH REL AV-give book that to Siti
    ‘Ali is the one who gave the book to Siti!’
Just as in Malay, the clefted element patterns with predicates.

Turning now to Malagasy, most pivots are DPs or PPs, which are also possible as matrix predicates.

(14) a. Any Antananarivo no mpetraka i Ketaka.
    there Antananarivo NO AT.live Ketaka
    'It's in Antananarivo that Ketaka lives.'

b. Any Antananarivo i Ketaka.
    there Antananarivo Ketaka
    'Ketaka is in Antananarivo.'

(14a) shows a clefted PP and (14b) a PP as a matrix predicate. Consider the parallel DP examples below.6

(15) a. Mpianatra no mamaky teny.
    student NO AT.read word
    'It's students who are reading.'

b. Mpianatra i Ketaka.
    student Ketaka
    'Ketaka is a student.'

The data in (14) and (15) show that the elements that occur in the clefted position can function independently as predicates. Moreover, a clefted DP can be negated, unlike argument DPs and like predicates (verbal or other). Thus the pivot in (16a), Rasoa, can take negation (marked bytsy) and patterns with the nominal predicate in (16d), mpianatra 'student'. (16b,c) show that as arguments, DPs cannot be negated.

(16) a. Tsy Rasoa no nanoroka an-dRakoto.
    NEG Rasoa NO PST.AT.kiss ACC-Rakoto
    'It's not Rasoa who kissed Rakoto.'

b. * Nanoroka tsy an-dRakoto Rasoa.
    PST.AT.kiss NEG ACC-Rakoto Rasoa

c. * Nanoroka an-dRakoto tsy Rasoa.
    PST.AT.kiss ACC-Rakoto NEG Rasoa

6 With DPs, there arises a difference between simple predicates and clefts. Standard DP predicates cannot be definite, while definite DPs can clearly cleft. Hence the examples in (i) below contrast with (15).

(i) a. Ny mpianatra no mamaky teny.
    det student NO AT.read word
    'It is the students who are reading.'

b. * Ny mpianatra i Ketaka.
    DET student Ketaka
    'Ketaka is the student.'

This difference clearly requires some explanation. To do so, however, would involve a complete discussion of nominal predication in Malagasy, something I will not undertake in this paper.
CLEFTS VS. PSEUDO CLEFTS IN AUSTRONESIAN

d. Tsy mpianatra Rasoa.
   NEG student Rasoa
   'Rasoa is not a student.'

Similarly, the pivot may be preceded by the verbal particle *toa* 'seems'. As shown in (17a), *toa* normally precedes the predicate.7

(17) a. Toa nanoroka an-dRakoto Rasoa.
   seem PST.AT.kiss ACC-Rakoto Rasoa
   'Rasoa seems to have kissed Rakoto.'

b. Toa Rasoa no nanoroka an-dRakoto.
   seem Rasoa NO PST.AT.kiss ACC-Rakoto
   'It seems to be Rasoa who kissed Rakoto.'

The above examples indicate that the pivot has a similar distribution to predicates.8

Summing up, a range of data show that the focussed element in a cleft patterns with predicates. I take this as evidence in favour of the structure in (9), where the focus appears in the matrix predicate position. Recall that treating the focus as a predicate is consistent with the predicate-initial word order of these languages. The only exceptions are the Javanic languages, such as Madurese and Malay, which are SVO. In order to account for the SVO languages, I must stipulate that in these cases, focus movement obtains. In other words, the focus element is generated in a predicate position and then moves to the specifier of a functional projection that dominates the subject position. An analysis along these lines is proposed in CHA.

(18)

Note that this movement may be due to theme-rheme considerations and is clearly not required in the verb-initial languages.

7 Other adverbs that patterns with *toa* are *tena* 'really' and *tokony* 'should'.
8 Note that in clefts, there are two potential positions for both *tsy* and *toa*: preceding the focussed element or preceding the embedded verb (with a clear difference in interpretation). Compare (ia) with (16a) and (ib) with (17b).

(i) a. Rasoa no *tsy* nanoroka an-dRakoto.
   Rasoa NO NEG PST.AT.kiss ACC-Rakoto
   'It was Rasoa who didn’t kiss Rakoto.'

b. Rasoa no *toa* nanoroka an-dRakoto.
   Rasoa NO seem PST.AT.kiss ACC-Rakoto
   'It was Rasoa who seemed to kiss Rakoto.'

Hence it is not simply the case that *tsy* and *toa* are clause-initial particles.

161
2.2 Presuppositional clause=headless relative

Let us now examine the presuppositional clause in more detail. In (9), the presuppositional clause has the structure of a headless relative. The following subsections show that in a wide range of Austronesian languages, the cleft marker is a relative clause marker or another nominal marker.

2.2.1 Malay

The data below illustrate the identity between the relative marker and the cleft marker in Malay.

(19) a. Siapa yang kau nampak?
   who that you see
   ‘Who do you see?’ (CHA (36))

b. Yang kau nampak Siti (-lah).
   that you see Siti (foc)
   ‘The one you see is Siti.’ (CHA (23))

c. [DP buku [CP yang John beli ]]
   book that John bought
   ‘the book that John bought’ (CHA (33))

(19a) is a clefted question, marked by yang (glossed as ‘that’). (19b) illustrates yang as a headless relative marker. That yang can be used for regular (headed) relative clauses is shown in (19c). Thus the yang in clefts marks the presence of a headless relative in subject position. Davies (2000) cites parallel data from Madurese, which uses the relative/focus marker se (see (7b)).

2.2.2 Palauan

In Palauan, the morpheme a is an all-purpose DP marker and precedes the presuppositional clause in a cleft. The data in (20) are from Georgopoulos (1991). (20a) illustrates the standard predicate-initial word order while (20b,c) are clefts. Note the presence of a in all examples.

(20) a. [predicate Ng-mekelekolt] [subject a ralm ].
   AGR-cold R-water
   ‘The water is cold.’

b. [predicate Ng-Basilia ] [subject a mengaus er tia el tet ].
   AGR-Basilia R-weave P dem L bag
   ‘It’s Basilia who’s weaving this bag.’

c. [predicate Ng-te’a ] [subject a kileld-ii a sub ]?
   AGR-who R-pf-heat-3s soup
   ‘Who heated up the soup?’

As described by Georgopoulos, a always occurs before a DP (but not before names, pronouns or demonstratives). She concludes that since the presuppositional clause in (20b,c) is marked with a, it is a nominal in the subject position.
2.2.3 Tagalog

Richards (1998) cites the following data in favour of a pseudo-cleft structure for Tagalog clefts. He points out that *ang* usually marks nominals in topic position, as in (21a), where *lalaki* 'man' has been topicalized. (Richards calls "topic" what I refer to as the subject position.) Clefts, on the other hand, obligatorily involve *ang* placed immediately before the verb, as illustrated in (21b).

(21) a. Bumili ang lalaki ng tela.  
AT.buy T man GEN cloth  
'The man bought cloth.'

b. Sino ang bumili ng tela.  
who T AT.buy GEN cloth  
'Who bought cloth?'

The distribution of *ang* is explained if the string of words following the clefted element (*sino* 'who' in (21b)) is a headless relative in topic (subject) position.

2.2.4 Malagasy

The Malagasy "focus marker" *no* is somewhat mysterious by comparison with the above languages as it is not used elsewhere in the language. Hence this type of headless relative does not surface except in clefts. Free relatives in argument positions use *izay*, as illustrated in (22a). Headed relative clauses such as in (22b) are also marked with *izay* (although it is often optional).

(22) a. Hahazo karama be izay miasa mafy.  
FUT.AT.get salary big REL AT.work hard  
'Whoever works hard will make lots of money.'

b. ny vehivavy (izay) miasa mafy  
DET woman (REL) AT.work hard  
'the woman who works hard'

Other headless relatives are marked with the determiner *ny*.

(23) ny miasa mafy  
DET AT.work hard  
'the ones who are working hard' or 'the event of working hard'

I simply stipulate that the headless relative marked by *no* is restricted to certain copular

---

9 Clearly unrelated is the past tense marker *no*. Another use of *no*, likely related to the focus construction is in the first clause of an *if...then* statement, when the second clause expresses a cause.

(i) Izaho no tsa tonga, nisy raharaha nalehako.  
ISG.NOM NO NEG arrive PST.AT.exist business PST.a.gone.ISG.GEN  
'If I didn't come, it's because business called me elsewhere.'

Finally, *no* appears in certain SVO contexts, where the subject is an indefinite pronoun.

(ii) Na iza na iza (no) tsa mamafa lalana dia voasazy.  
or who or who (NO) NEG AT.sweep road TOP voa.punish  
'Whoever doesn't sweep the road will be punished.'

In (ii), *no* is optional and in fact some speakers prefer to omit it. In regular clefts, *no* is obligatorily present.
clauses. Note that the headless relatives in other languages do not seem to have this restriction and may appear in any argument position.

Coordination in Malagasy points toward a DP structure for the presuppositional clause. More precisely, it is evidence against treating the cleft construction as fronting to a specifier position of a functional category in the CP structure (à la Rizzi (1997)). Under such an analysis, the focus constituent moves to the specifier position of [FocusP] and no is the head of FocusP or some other projection. The remainder of the clause is simply material in IP. This structure is shown in (24).

(24) \[
\text{[FocusP Rasoa \[ no \[ IP nijinja vary t\_i \].]}
\]
\[
\text{Rasoa NO PST.AT.harvest rice }
\]
\[
\text{‘It was Rasoa who was harvesting rice.’}
\]

Consider now coordination. Malagasy has two main types of coordinating conjunction: ary and sy. The former conjoins clauses (IP or CP), while the latter is for smaller constituents, for example VP or DP (or heads).

(25) a. 
\[
\text{Miteny ny mpampianatra ary mihaino ny mpianatra.}
\]
\[
\text{AT.speak DET teacher and AT.listen DET student}
\]
\[
\text{‘The teacher speaks and the students listen.’}
\]
b. 
\[
\text{Misotro sy miloka izy.}
\]
\[
\text{AT.drink and AT. play 3.NOM}
\]
\[
\text{‘He drinks and plays.’}
\]

Crucially, sy (and not ary) is used to coordinate presuppositional clauses.

(26) 
\[
\text{Rasoa \[ no nijinja vary \] sy/*ary \[ no nanapaka bozaka\].}
\]
\[
\text{Rasoa NO PST.AT.harvest rice and NO PST.AT.cut grass}
\]
\[
\text{‘It was Rasoa who harvested rice and cut grass.’}
\]

First, the datum in (26) indicates that the string of no and whatever follows (i.e. the presuppositional clause) is a constituent. Second, (26) shows that the presuppositional clause is not IP or CP since ary is ungrammatical. Thus we have evidence against the analysis briefly sketched in (24).

Due to the similarities between clefts and relative constructions, I will assume that no is a nominal marker. The two trees in (27) illustrate two possible structures for the headless relative in question. Under this analysis, no is either a determiner, as in (27a), or a complementizer, as in (27b).

---

10 Many other Austronesian languages have a relative marker similar in form to the Malagasy no (either nu or ano). For example, in Sundanese, an Indonesian language, the relative clause marker is nu (Hardjadiibrata (1985)). It is therefore likely that the Malagasy no is a historical remnant. Malzac (1960) mentions that certain Malagasy grammarians believe no to be diachronically related to the determiner ny. He does not provide any references, however.
Since the precise position of *no* is not crucial to my analysis, I will not attempt to distinguish between these two possibilities (see footnote 10 for some speculation on the category of *no*).

### 2.3 Headless relatives

In this section, I address certain aspects of the interpretation of the headless relative clause in the proposed structure of the cleft. In a cleft where the pivot corresponds to the subject, the headless relative means something like ‘the one/thing who/that...’.

(28) a. I Bakoly no manapaka bozaka.
   Bakoly NO AT.cut grass
   (lit.) ‘The one who is cutting grass is Bakoly.’

b. Bozaka no tapahin’i Bakoly.
   grass NO TT.cut.GEN.Bakoly
   (lit.) ‘The thing that Bakoly is cutting is grass.’

c. Ny antsy no anapahin’i Bakoly bozaka.
   DET knife NO CT.cut.GEN.Bakoly grass
   (lit.) ‘The thing that Bakoly is cutting grass with is the knife.’

The voice on the verb will determine which argument is interpreted as the external argument. An agent with AT, as in (28a), a theme with TT, as in (28b), or some oblique with CT, as in (28c).

Recall the restrictions on clefting mentioned in section 2. The absence of object clefts is a direct consequence of the headless relative structure proposed. It is not possible to relativize an object, hence the corresponding object cleft is impossible. In other words, the ungrammaticality of the cleft (29a) is related to the ungrammaticality of the relative clause in (29b).

(29) a. * Bozaka no manapaka i Bakoly.
   grass NO AT.cut Bakoly
   ‘It’s grass that Bakoly is cutting.’
For an object to be relativized, the verb must have TT morphology, as shown in (29c). The same is true for object clefts, as can be seen by the contrast between (28b) and (29a). Thus the restriction on clefting reduces to the restriction on relativization.

Recall, however, that adjuncts can be clefted without being promoted to subject. This is illustrated in (30a), which has a PP pivot and an AT verb. (30b), on the other hand, shows that adjuncts cannot be relativized with AT morphology.

(30)  
a. Amin'ny antsy no manapaka bozaka i Bakoly.
   P.GEN.DET knife NO AT.cut grass Bakoly
   'It is with a knife that Bakoly is cutting grass.'

b. * ny antony izay manapaka bozaka i Bakoly
   DET reason REL AT.cut grass Bakoly
   'the reason why Bakoly is cutting grass'

Clearly we cannot simply relate the availability of clefts to the grammaticality of relatives. Moreover, in (30a) the headless relative cannot mean 'the one who is cutting grass' for two reasons. First, the agent of cutting (Bakoly) is expressed within the relative. Second, if clefts have an equative structure, it is somewhat odd to equate a PP with a nominal referring to an individual.

Before providing an analysis of (30a), I note that these types of adjunct clefts are not common in the Austronesian languages discussed in this paper. CHA explicitly discuss this point with reference to Malay. They point out that the lack of adjunct clefts correlates with certain gaps in the interpretation of headless relatives.

(31)  
a. ?? Ke mana yang kau pergi?
   to where that you go
   'Where are you going?'

b. ?? Yang aku pergi (ialah) ke Kuala Lumpur.
   that I go is to Kuala Lumpur
   'Where I am going is to Kuala Lumpur.'

The impossibility of the adjunct cleft in (31a) is related to the impossible meaning for the headless relative in (31b). On the other hand, in the closely related language Madurese, adjunct clefts are grammatical and headless relatives may have the interpretation disallowed in Malay.

To account for the grammaticality of adjunct clefts, I suggest that the headless relative is interpreted as an event nominal (like a gerund). In other words, (30a) means 'The event of Bakoly cutting grass was with a knife'. Invoking the event reading allows a range of obliques to appear in an AT cleft, as is in fact the case. I therefore conclude that these constructions are not strictly equative, but rather copular in a looser sense.

This account may at first appear stipulative. In fact, however, zero nominals in Malagasy freely have either an event or an individual interpretation. Both readings are illustrated in (32).
CLEFTS VS. PSEUDO CLEFTS IN AUSTRONESIAN

(32)  
(a) Faly ny manapaka bozaka.
happy DET AT.cut grass
'The ones who are cutting grass are happy.'

(b) Sarotra ny manapaka bozaka.
difficult DET AT.cut grass
'Cutting grass is difficult.'

(c) Mihira ny tiana.
AT.sing DET TT.love
'The loved ones are singing.'

d. Mahafinaritra ny tiana.
AT.happy DET TT.love
'Being loved is pleasant.'

In (32a), the zero nominal clearly receives an individual reading. (32b), on the other hand, is a gerund-like zero nominal. (32c,d) shows that the individual and the gerund readings are possible for passive nominals as well as active ones. It is therefore not unreasonable to suggest the event and the individual readings are also available for the headless relatives in cleft constructions.

Another possible analysis of adjunct clefts involves movement. For examples such as (30a), the adjunct moves to a clausal focus position, rather than being generated as a predicate. The structure is shown in (33).

(33)  
[ [ Amin'ny antsy Ji no [ manapaka bozaka ti i Bakoly ].
P.GEN.DET knife NO AT.cut grass Bakoly
'It is with a knife that Bakoly is cutting grass.'

A similar dual analysis of clefts is suggested by Pinkham and Hankamer (1975) for English and by Chung (1998) for Chamorro. By invoking movement for adjunct clefts, we can maintain a strictly equative analysis for subject clefts. In other words, by complicating the syntax (two alternate derivations), we simplify the semantics. This solution, however, does not provide any explanation for why clefting of adjuncts is ruled out in certain languages (such as Malay) and not in others (like Madurese). More importantly, a purely syntactic explanation does not capture the correlation between the interpretation of headless relatives and the possibility of adjunct clefts.

2.4 Summary

In this section, I have provided syntactic arguments for analyzing the cleft in a range of Austronesian languages as a copular construction. The focus is in fact the matrix predicate and the presuppositional clause is a headless relative clause in the subject position. The final two sections of this paper investigate the interpretation of clefts and how this interpretation relates to the proposed structure.

3 Interpretation

Turning to the semantics of cleft constructions, it has long been noted that they are associated
with a certain interpretation: existential presupposition and exhaustivity (Halvorsen (1978)). This interpretation is also apparent in Malagasy.

3.1 Existence

First, a cleft (it is x that P) presupposes that there is some individual x that has the property P (Ex s.t. Px is true). Hence the presupposition of the cleft in (34b) is that someone painted houses. This clearly contradicts (34a), which asserts that no one painted houses.

(34) a. Tsy misy olona nandoko trano...
   NEG exist person PST.AT.paint house
   "No one painted houses..."

   b. # ... noho izany dia tsy i Koto no nandoko trano.
      because that TOP NEG Koto NO PST.AT.paint house
      "... therefore it wasn’t Koto who painted houses."

A similar conflict arises in (35), which presupposes that someone is painting houses while simultaneously asserting that no one is.

(35) * Tsy na iza na iza no mandoko trano.
   NEG or who or who NO AT.paint house
   * "It’s no one who is painting houses."

Just as in English, Malagasy clefts carry existential presuppositions.

3.2 Exhaustivity

Second, a cleft expresses exhaustive identification. É. Kiss (1998) provides tests for the exhaustive reading. Consider first question-answer pairs. The answers in (36b,c) to the question in (36a) have different meanings.

(36) a. Nandeha taiza ianao?
    PST.AT.go PST.where 2SG.NOM
    "Where did you go?"

   b. Nandeha tany Ambositra aho.
      PST.AT.go PST.there Ambositra 1SG.NOM
      "I went to Ambositra."

   c. Tany Ambositra no nandeha aho.
      PST.there Ambositra NO PST.AT.go 1SG.NOM
      "It was to Ambositra that I went."

(36b) does not exclude the possibility that I went to other places as well as Ambositra. The cleft construction in (36c), however, is an exhaustive answer; Ambositra is the only destination.

11 For this section, I rely mainly on data from Malagasy. Data from Madurese cited in example (40) at the end of this section suggest that the interpretive effects that I identify are common to clefts in other languages.
Similarly, consider the following pairs.\footnote{É. Kiss attributes this test to Szabolcsi (1981). The judgements in (37) do not change if the verb is in AT. I use TT in (37) to provide minimal pairs with (38), where TT is necessary to allow clefting of the logical object.}

(37)  
\begin{align*}
\text{a.} & \quad \text{Novidin'i Bakoly ny satroka sy ny kiraro.} \\
& \quad \text{PST.TT.buy.GEN.Bakoly DET hat and DET shoe} \\
& \quad \text{\textquoteleft Bakoly bought a hat and shoes\textquoteright.} \\
\text{b.} & \quad \text{Novidin'i Bakoly ny satroka.} \\
& \quad \text{PST.TT.buy.GEN.Bakoly DET hat} \\
& \quad \text{\textquoteleft Bakoly bought a hat.\textquoteright.}
\end{align*}

(38)  
\begin{align*}
\text{a.} & \quad \text{Ny satroka sy ny kiraro no novidin'i Bakoly.} \\
& \quad \text{DET hat and DET shoe NO PST.TT.buy.GEN.Bakoly} \\
& \quad \text{\textquoteleft It\'s a hat and shoes that Bakoly bought.\textquoteright.} \\
\text{b.} & \quad \text{Ny satroka no novidin'i Bakoly.} \\
& \quad \text{DET hat NO PST.A.T.buy.GEN.Bakoly} \\
& \quad \text{\textquoteleft It\'s a hat that Bakoly bought.\textquoteright.}
\end{align*}

As in the English equivalents, the sentence in (37b) is a logical consequence of the one in (37a). On the other hand, (38b) is not a logical consequence of (38a). In fact, (38b) contradicts (38a). Hence, the cleft construction in (38) passes the test of exhaustivity.

Exhaustivity is further illustrated with the following test. Due to the assertion of exhaustivity, there are certain distributional restrictions on the elements that can appear in the focus position.

(39)  
\begin{align*}
\text{a.} & \quad \text{* Bakoly koa no nandeha tany Ambositra.} \\
& \quad \text{Bakoly also NO PST.AT.go PST.A.T.go PST. there Ambositra} \\
& \quad \text{\textquoteleft It was also Bakoly who went to Ambositra.\textquoteright.} \\
\text{b.} & \quad \text{* Na ny mpianatra votsavotsa aza no nahazo isa tsara.} \\
& \quad \text{or DET student weak even NO PST.AT.get number good} \\
& \quad \text{\textquoteleft It was even the weak students who got good grades.\textquoteright.} \\
\text{c.} & \quad \text{* Na iza na iza no mahavita izany.} \\
& \quad \text{or who or who NO PST.A.T.done that} \\
& \quad \text{\textquoteleft It\'s anyone who can do that.\textquoteright.}
\end{align*}

These elements appear to have some semantic clash with exhaustive identification. For example, the import of koa \textquoteleft also\textquoteright in (39a) is to assert that going to Ambositra is true of some other individual as well as Bakoly. Hence the meaning of this ad-verb conflicts with the assertion of exhaustivity, which leads to the ungrammaticality of (39a). Since DPs with these semantic features are not permitted in a cleft, we see that the cleft position is associated with a particular interpretation, in this case exhaustivity.

Although most of the information I have on other Austronesian clefts does not discuss these interpretational facts, Davies (2000) provides similar data from Madurese. He shows that the clefted position is incompatible with the adverb ghia \textquoteleft too, also\textquoteright.
The data in the section have shown that Malagasy clefts are associated with a certain interpretation, parallel to English clefts. Clefts carry two presuppositions: existence and exhaustivity. I now explain why these two presuppositions are present, drawing on the structure of clefts proposed in section 2. In other words, I link the interpretation to the underlying presence of a definite description.

4 Solution

I account for the particular focus interpretation of clefts by invoking the headless relative structure. Recall the structure proposed for clefts in Austronesian.

Crucially, the subject DP is a headless relative (‘the one who...’). Headless relatives are definite descriptions and therefore have the same presuppositions as definite descriptions. These are precisely the same presuppositions as those exhibited by clefts.

It is well known that definite descriptions presuppose the existence of the individual described. Moreover, definite descriptions presuppose (or entail) that there is exactly one referent as described (Strawson (1950)). In other words, the pattern identified in the previous section is paralleled by (42)-(43) below, which contain overt definite descriptions. Compare the ungrammaticality of (35) (repeated as (42a)) with (42b). In (42b), the definite description ‘the one who is painting houses’ presupposes the existence of someone painting houses. If this description is predicated of ‘no one’, ungrammaticality results.

(41)  a.  I Sahondra no nanapaka ity hazo ity.
       Sahondra NO PST.AT.cut this tree this
       (lit.) ‘The one who cut this tree was Sahondra.’

       b.  I*
           IP
           DP
           I* XP DP CP
           focus Ø OPi...ti
           I Sahondra no nanapaka ity hazo ity

The ungrammaticality of (35) (repeated as (42a)) with (42b). In (42b), the definite description ‘the one who is painting houses’ presupposes the existence of someone painting houses. If this description is predicated of ‘no one’, ungrammaticality results.

(42)  a.  * Tsy na iza na iza no mandoko trano.
        NEG or who or who NO AT.paint house
        * ‘It’s no one who is painting houses.’

 170
CLEFTS VS. PSEUDO CLEFTS IN AUSTRONESIAN

b. * Tsy na iza na iza ny (olona) mandoko trano.
   NEG or who or who DET (person) AT.paint house
   * ‘The one who is painting houses is no one.’

Similarly, compare (43a) with (43b). The adverb koa ‘also’ is incompatible with the uniqueness reading that arises in a definite description in (43b) and with the exhaustivity of the cleft in (43a).

(43) a. * Rasoa koa no mandoko trano
   Rasoa also NO AT.paint house
   * ‘It’s also Rasoa who is painting houses.’

b. * Rasoa koa ny (olona) mandoko trano.
   Rasoa also DET (person) AT.paint house
   * ‘The one who is painting houses is also Rasoa.’

If we analyze clefts as containing a definite description, these parallels in interpretation are accounted for in a simple manner.

5 Conclusion

In this paper, I have proposed a structure for clefts where the focussed element is the matrix predicate and the presuppositional clause corresponds to a headless relative in subject position. I have argued that this analysis of Austronesian clefts accounts both for their structural properties and their interpretation. The focussed element patterns with predicates and the presuppositional clause patterns with complex DP subjects. It is the headless relative, a definite description, that induces the focus interpretation associated with clefts. Note that the headless relative structure proposed is quite common in the languages discussed. In other words, the Austronesian languages under consideration allow productive zero nominal formation, consisting of the simple addition of a determiner or other nominal marker to a predicate. Thus the present analysis of clefts meshes with the syntactic structure of Austronesian languages and is not dependent on special focus features in the syntax or movement to particular focus projections. Instead, the focus reading arises independently and with minimal additional assumptions.

References


A Non-Linear Analysis of Vowel Harmony and
Vowel Harmony Blocking in Pendau

Phil Quick
Australian National University and the Summer Institute of Linguistics

1. INTRODUCTION

Vowel harmony is without a doubt the most interesting phonological feature in Pendau (and is found in many if not all of the languages in the Western Austronesian Tomini-Totoli language group, northern Central Sulawesi, Indonesia).1 Vowel harmony in Pendau is a type of prosodic alignment where certain features of one vowel in the stem spreads to certain prefixes (it can spread in up to three prefixes from right to left) and one infix. The /ø/ vowel of a prefix may front to either /a/ or /e/ depending on the features of the first vowel of the root or stem.2

Kenstowicz (1994:347) provides a definition for vowel harmony (carefully distinguishing it from umlaut and other similar processes):

Vowel harmony is a phonological state in which the vowels in a given domain share or harmonize for a particular feature. It differs from other processes affecting adjacent vowels (e.g., umlaut) in that typically all of the vowels of the language participate in the harmonic constraint. In addition, the harmony applies in an essentially unbounded fashion, affecting all the relevant vowels within the domain (typically the word). Virtually any of the common features used to distinguish among vowels have been discovered to seat a harmonic system, including vowel height, backness, rounding, nasality, and pharyngeal opening or [ATR]. Vowel harmony exhibits many of the "action-at-a-distance" properties displayed by tone.

This paper begins by introducing the Pendau data which will give representative examples of vowel harmony for each of the harmonic prefixes. After this the analysis will begin with a brief classical generative analysis of the vowel harmony phenomena showing that the features back and rounding can be collapsed into one rule, but that this rule and this approach to the data is unsatisfactory (especially with the harmony blocking environments). The major part of the presentation utilizes autosegmental, feature geometry, and lexical phonology theories to describe both the vowel harmony process and the environment in which vowel harmony is blocked. Although initially it appears that vowel harmony is blocked by the sequence of two consonants there is data that also shows vowel harmony spreading past two consonants (such as sa-n-tanga "half"). This analysis resolves this by showing that nasal assimilation and nasal deletion block vowel harmony from spreading at the cavity tier. Since the Well Formedness Condition (WFC) asserts that lines cannot be crossed, then by replacing the BR feature with the oral node this vowel harmony problem is resolved elegantly. Finally this paper will briefly put Pendau into a context with other Austronesian languages which have vowel harmony, concluding with the suggestion that vowel harmony can help understand some historical problems in Sulawesi languages.

---

1 This analysis has directly or indirectly benefitted from discussions over the years with: Avery Andrews, René van den Berg, John Bowden, Donald Burquest, Mike Cahill, Timothy Friberg, Chaz Mortensen, Steve Parker, Rebecca Quick. Part of this paper was originally presented in the RSFAS linguistics seminar series in 1996 with a similar title, and I also want to thank that seminar's participants for comments and feedback. Abbreviations: AV active voice, DIR directional, CAUS causative, DISTR distributive aspect, HPS harmonic prefix set, IR irrealis, IV inverse voice, NV non-volitional, ONE numeral one prefix, POSS possessive, RE realis, RSLTV resultative ST stative, STEM stem former, TZ transitivizer.

2 See Topping 1968 and Latta 1972 for discussion of vowel fronting harmony in Chamorro which is the reverse of this, that is the first vowel of the root or stem fronts to the vowel of the preceding prefix.
2. VOWEL HARMONY DATA

A list of the harmonic prefixes (with the underlying formatives) is provided in (1).

(1) Harmonic Prefix Set (HPS)

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Category 1</th>
<th>Category 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>mong-</td>
<td>AV/IR</td>
<td>ro-</td>
</tr>
<tr>
<td>nong-</td>
<td>AV/RE</td>
<td>mo-</td>
</tr>
<tr>
<td>pong-</td>
<td>STEM</td>
<td>no-</td>
</tr>
<tr>
<td>song-</td>
<td>ONE (verb)</td>
<td>so-</td>
</tr>
<tr>
<td>'o-</td>
<td>RSLTV or POSS</td>
<td>po-</td>
</tr>
<tr>
<td>-ong-</td>
<td>DISTR</td>
<td></td>
</tr>
</tbody>
</table>

Most of the harmonic prefixes are verbal prefixes (note that there are other verbal prefixes which are not harmonic, e.g. me-/pe- dynamic verb class, mo-/no- denominal verb class, etc.). The so- prefix is a numeral prefix used on classifiers and measure nouns. The 'o- prefix is used to possess a noun or in combination with other verbal prefixes (see Quick in prep. for more details). In addition to the harmonic prefixes, there is also one harmonic infix -ong- which marks distributive plurality (since most of the harmonic affixes are prefixes, I will include the infix under the label Harmonic Prefix Set).

Example (2) shows vowel harmony prefixes with numerals, classifiers, measure nouns, nominalized classifiers, stative verbs, stative intensification (in some cases the vowel harmony prefix is in combination with other affixes).

(2) Numerals:

<table>
<thead>
<tr>
<th>numeral</th>
<th>meaning</th>
<th>numeral</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>soung</td>
<td>'one'</td>
<td>sa-gatus</td>
<td>'one hundred'</td>
</tr>
<tr>
<td>so-mpulu</td>
<td>'one ten'</td>
<td>se-ribu</td>
<td>'one thousand'</td>
</tr>
</tbody>
</table>

Stative Verbs (mo- stative irrealis prefix):

<table>
<thead>
<tr>
<th>verb</th>
<th>meaning</th>
<th>verb</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ma-pangkat</td>
<td>'tall'</td>
<td>me-empeng</td>
<td>'short'</td>
</tr>
<tr>
<td>me-ide</td>
<td>'small'</td>
<td>mo-onda</td>
<td>'hot'</td>
</tr>
<tr>
<td>mo-oge</td>
<td>'large'</td>
<td>me-menyong</td>
<td>'cold'</td>
</tr>
<tr>
<td>me-itong</td>
<td>'black'</td>
<td>ma-paris</td>
<td>'difficult'</td>
</tr>
<tr>
<td>me-meas</td>
<td>'white'</td>
<td>ma-lamor</td>
<td>'easy'</td>
</tr>
<tr>
<td>mo-doda</td>
<td>'red'</td>
<td>mo-mbosi</td>
<td>'good'</td>
</tr>
<tr>
<td>mo-bulung</td>
<td>'green, blue'</td>
<td>mo-boat</td>
<td>'heavy'</td>
</tr>
<tr>
<td>me-ri</td>
<td>'yellow'</td>
<td>ma-nggaang</td>
<td>'light'</td>
</tr>
<tr>
<td>ma-dantang</td>
<td>'long'</td>
<td>ma-nggaang</td>
<td>'light'</td>
</tr>
</tbody>
</table>

Stative Intensification ('o- σ-o-):

<table>
<thead>
<tr>
<th>expression</th>
<th>meaning</th>
<th>expression</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>'a-sa-sanang</td>
<td>'to be exuberent'</td>
<td>sanang</td>
<td>'happy'</td>
</tr>
<tr>
<td>'a-pa-pangkat</td>
<td>'to be the most high'</td>
<td>pangkat</td>
<td>'tall'</td>
</tr>
<tr>
<td>'o-to-tou'</td>
<td>'the end of it all'</td>
<td>tou</td>
<td>'finish'</td>
</tr>
<tr>
<td>'a-ga-garang</td>
<td>'the most loved'</td>
<td>garang</td>
<td>'love'</td>
</tr>
<tr>
<td>'e-de-dea'</td>
<td>'uncountable'</td>
<td>dea</td>
<td>'many'</td>
</tr>
<tr>
<td>'a-ga-gau'-ong</td>
<td>'have the most deeds'</td>
<td>gau</td>
<td>'event, activity'</td>
</tr>
</tbody>
</table>
Classifiers:

- sa-lai: 'one thread-like, hair-like object'
- sa-mata: 'one eye, one sharp-pointed object'
- se-mpe'a: 'one flat hard object'
- sa-ngkayu: 'one snake-like item, tubular'
- so-ngkolo: 'one cut or shaped heavy object'
- so-bua: 'one fruit-like object (larger)'
- se-ilas: 'one ring-like object'
- sa-dampe: 'one seed-like object'
- sa-nta'u: 'one fruit (durian, coconuts, etc.), or 4-legged animals'

Nominalized (Quantified) Classifiers:

- pe-sa-mata-ong: 'one each of a sharp-pointed type object'
- pe-so-bua-ong: 'one each of a fruit-like object'
- pe-se-ilas-ong: 'one each of a ring-like object'

Measure Nouns:

- se-insang: 'one time'
- so-ndoung: 'one evening'
- se-mbengi: 'one night'
- se-eleo: 'one sun, one day'
- se-mmggu: 'one week'
- so-bungkus: 'one package'

Infaxes are given in example (3) below and the infix is shown within brackets. The distributive -ong- infix occurs immediately after the first consonant of the word (this is in contrast with the telic aspectual infix -um- which is inserted after the first consonant of the root).

(3) Distributive Infix -ong-

<table>
<thead>
<tr>
<th>Inflixation</th>
<th>Root</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>n[eng]po-duling=omo</td>
<td>duling</td>
<td>'lie down'</td>
</tr>
<tr>
<td>n[eng]teule=mo</td>
<td>teule</td>
<td>'return home'</td>
</tr>
<tr>
<td>m[eng]omung-komung</td>
<td>omung</td>
<td>'bring, carry'</td>
</tr>
<tr>
<td>m[eng]inum</td>
<td>inum</td>
<td>'drink'</td>
</tr>
<tr>
<td>m[eng]lolo</td>
<td>lolo</td>
<td>'search'</td>
</tr>
<tr>
<td>n[eng]lampa</td>
<td>lampa</td>
<td>'walk, travel'</td>
</tr>
</tbody>
</table>

What is most revealing is a comparison of the Inverse Voice Irrealis prefix (IV/IR) with the Active Voice prefixes (AV/IR or AV/RE). For our purposes here I will only provide the Irrealis form of the Active Voice since the Realis prefix performs exactly the same way as the Irrealis form. Compare the three sets in (4). In (4a) the roots have vowel initial roots, and in (4b) the roots have voiceless consonant initial roots, and in (4c) the roots have (mostly) voiced consonant initial roots. The prefix ro- (IV/IR) is subject to vowel harmony in all instances, whereas the prefix mong- (AV/IR) only undergoes vowel harmony when it occurs with a vowel initial root.

1 Also see Rubino in prep for using these three sets as a phonological exercise.
Two possible ways of analyzing this blocking are apparent: that it is blocked by underlying consonant sequences, or that it is blocked by the assimilatory processes that also occur in (4b)-(4c) In §6 I will argue for an assimilation blocking account.

3. THE CLASSICAL GENERATIVE APPROACH

This section begins the description of vowel harmony in Pendau by using the classical generative model. This approach is instructive in demonstrating precisely what the harmonic prosody is and also shows the inherent weakness of a linear treatment of vowel harmony.

If the first vowel in a stem is /a/ then the /o/ is fronted to the same features as /a/, that is [-rd]. The features [-hi, +bk] are redundant and implied in the remainder of the rule.

The following rule (5) could be formulated (ignoring the more complicated environments for the time being).

\[
\begin{align*}
V \rightarrow V & \quad \#_C \quad (C) + V \\
\quad [-rd] \quad [-rd] & \\
\quad [+rd] \quad [-hi]
\end{align*}
\]
A Non-linear Analysis of Vowel Harmony and Vowel Harmony Blocking in Pendau

The generative rule can be translated like this: A mid back vowel becomes unrounded when it is in a(n) (open or closed prefix) syllable preceding a contiguous morpheme whose (first) vowel is unrounded.

When the first vowel in a stem is /e/ or /i/ then the /o/ is fronted to the same features as /e,i/ that is [-bk]. Here the important redundant features are [-hi,-rd], as shown in (6).

(6)
\[ V \rightarrow V / \# C \_ (C) + V \]
\[ \begin{bmatrix} +rd \\ -hi \end{bmatrix} \]
\[ [-bk] \]

The two rules above shows that in Pendau the vowel harmony agrees in unrounding and backing of [-hi] vowels. When a word has the high back vowel [u] in the initial position no rules are needed since [o] and [u] agree in backness. To capture this natural process we can collapse the two rules above into the rule in (7). In addition we will add the important condition that consonant sequences block the vowel harmony, as there does not seem a convenient way to notate that only one consonant is allowable at a time for vowel harmony to occur. Further it should be noted that the vowel in the stem closest to the prefix is the vowel to which the prefix vowel harmonizes. The specific prefixes are abbreviated as the Harmonic Prefix Set (HPS) as there are other prefixes which do not allow vowel harmony. The word initial boundary is not identified as there are cases where vowel harmony may spread iteratively up to three prefixes.

(7)
\[ V \rightarrow V / C \_ (C) + (C) V \]
\[ \begin{bmatrix} +rd \\ -hi \end{bmatrix} \begin{bmatrix} -rd \\ \alpha bk \end{bmatrix} \]
\[ [\text{HPS}] \]
\[ \text{Word Stem} \]

**Harmonic Prefix Set (HPS): Condition: A consonant sequence will block the rule.**

4. Feature Geometry and Vowel Underspecification

The feature geometry in this analysis will follow the general consensus as presented in Kenstowicz (1994). The description of Pendau phonology is enhanced by understanding the theoretical model of feature geometry. Feature geometry models a close relationship between phonology and phonetics. The basic model of feature geometry is given in figure 1 below.

The vowels in Pendau can be underspecified as displayed in figure 2. Since a vowel chart can be displayed in several possible ways, the fact that the /o/ is the underlying vowel in vowel harmony and is used in openess was a determining factor in developing this particular underspecification configuration (originally suggested to me by Steve Parker personal communication).

---

4 I am indebted to Tim Friberg for pointing out to me the unrounding process in an early analysis.
Listed below are the redundancy rules for Pendau vowels. The post-lexical complement rules derive the /o/ in epenthesis. These also apply to the underlying /o/ when vowel harmony is blocked (see §6).

(8) Lexical Rules Post-Lexical Complement Rules
[-back]  [-rd]  \emptyset  \rightarrow  [-high]
[+low]  [-rd]  \emptyset  \rightarrow  [+back]
\emptyset  \rightarrow  [-low]

5. **An Autosegmental Analysis Of Vowel Harmony In Pendau**

Vowel harmony in Pendau is a type of prosodic alignment where certain features of one vowel in the stem spreads to vowels in some prefixes. Vowel harmony has been found to operate in much the same manner as tone cross-linguistically. The Autosegmental Theory can be applied in the analysis of Vowel Harmony and is more revealing and elegant than the Classical Generative approach. Autosegmental theory assumes that there is more than one tier and that each tier may operate independently of the other. The theory assumes a set of universal Well Formedness Conditions and each specific language may have its own specific conditions in addition.
5.1 Pendau Vowel Harmony Data and Analysis (Lexical [Level 2])

In Pendau the harmonizing feature is back-round harmony. There are several competing views about the location of the consonant node and the vocalic node. Quoting Kenstowicz (1994:473):

In the Clements model, the Vocalic node branches into V-PI and Aperture. This groups vocalic rounding and backness together and isolates height — a natural acoustic parsing since height is primarily reflected in the first formant and backness and rounding in the second formant. On the other hand, the articulator-based Halle-Sagery model groups backness and height as Dorsal dependents and isolates rounding as a Labial dependent. Odden (1991) discusses a number of cases that appear to support the former partitioning.

Vowel harmony in Pendau may contribute to an understanding of this current issue, as we will see shortly. Clements and Hume (1995:227) further states that back and rounding features can be eliminated in their model:

A further innovation of this model is that the features [labial], [coronal], and [dorsal], occurring under the V-place node in vocoids, are sufficient, by themselves, to distinguish place of articulation in vowels, and replace the traditional features [back] and [round]. In order to fulfill this new and expanded role in the theory, they must be redefined in terms of constrictions rather than articulator movements as such.

For this analysis I will begin by using BR as a useful heuristic to initially demonstrate that each word stem or word base carries the preassociated feature of the back-round node (BR). Accordingly the following steps would be followed to harmonize the vowel, see example (9) below. First in (9a) the correct preassociation with the Vowel Harmony Tier is drawn. Secondly, since vowel harmony spreads from the root or stem to the prefix the association spreads from right to left (see 9b).

(9) a. [BR]  
    ro + alap  

b. [BR]  
    ro + alap  →  ra-alap

IV/IR-take

The rules for associating the vowel harmony spreading in Pendau can be stated in these three steps:

Step one  Every root/stem has [BR] as a preassociated feature which is associated to the first vowel of the stem/root.

Step two  If nasal assimilation or glottal assimilation takes place, then vowel harmony is blocked (vowel harmony is blocked in assimilation processes due to the Well Formedness Condition that association lines do not cross, see discussion in §6).

Step three Prosodic vowel alignment spreads from right to left, so an association line is attached to the vowel(s) of the prefixes. This occurs in level two of the lexical module, and since level two is cyclic, the vowel alignment spreads consecutively to harmonic prefixes to its left.

1 Odden 1991 describes back-round as a formal part of vowel geometry.
When the preassociated vowel harmony tier already has the same BR value, as in (10), there is no change (notice that the BR restricts the vowel from harmonizing with [+high]).

(10)

\[
\begin{array}{c}
\text{[BR]} \\
m\text{ong}+\text{un}d\text{ur} \rightarrow \text{mong-undur} \\
\text{AV/IR-sing}
\end{array}
\]

Example (11) shows the BR unrounding the \text{o} to become a low \text{a}, in effect fronting it.

(11)

\[
\begin{array}{c}
\text{[BR]} \\
m\text{ong}+\text{ab}u\text{t} \rightarrow \text{mang-abut} \\
\text{AV/IR-weed}
\end{array}
\]

Example (12) further illustrate the process where the back rounded vowel becomes a front vowel.

(12)

\[
\begin{array}{c}
\text{[BR]} \\
\text{re-}+\text{reken} \rightarrow \text{re-reken} \\
\text{IV/IR-count}
\end{array}
\]

Example (13) illustrates that despite mixed vowels in the root the primacy must be on the first vowel of the stem or root base in deciding which vowel has the preassociation, or last prefix to undergo vowel harmony where there are multiple harmonic prefixes.

(13)

\[
\begin{array}{c}
\text{[BR]} \\
\text{re-}+\text{piyor} \rightarrow \text{re-piyor} \\
\text{IV/IR-twist}
\end{array}
\]

Pendau has several types of derivational prefixes which attach and become the stem of the word as far as vowel harmony is concerned. The vowel in this prefix becomes the vowel to preassociate the features for the vowel harmony/alignment tier.

(14)

\[
\begin{array}{c}
\text{[BR]} \\
\text{re-}+\text{pe-guru-i} \rightarrow \text{re-pe-guru-i} \\
\text{IV/IR-STEM-learn-DIR}
\end{array}
\]

Examples (15)-(16) demonstrates that vowel harmony predictably spreads to more than one possible prefix. Example (15) shows harmony of two prefixes, and example (16) shows harmony spreading into three prefixes. The language specific spreading from right to left still holds true for a more complicated affix situation. Examples (15)-(16) demonstrates how vowel harmony applies iteratively as each affix applies vowel harmony in its turn.
A Non-linear Analysis of Vowel Harmony and Vowel Harmony Blocking in Pendau

6. ASSIMILATORY PROCESSES THAT BLOCK VOWEL HARMONY

6.1 Nasal Assimilation (Lexical [Level 1 and 2]) and Voiceless Consonant Deletion (p, t, k, s) (Lexical [Level 2])

In Pendau, as in many other Austronesian languages the final nasal assimilates to a point of articulation homorganic with the initial consonant of the stem/root to which it is attached. Example (17) illustrates nasal assimilation to the following voiced obstruent (the affricate [dʒ] assimilates the nasal to the same point of articulation as the voiced coronal obstruent [d] below, e.g. ponjaong ‘the sewing place’).

(17) nasal assimilation before voiced stops

\[
\begin{array}{c|c|c|c}
\text{Labial} & \text{Coronal} & \text{Dorsal} \\
\hline
\text{Oral} & \text{Oral} & \text{Oral} \\
\hline
\eta + b & \eta + d & \eta + g \\
\hline
\text{SP} & \text{SP} & \text{SP} \\
\hline
\hline
\end{array}
\]

In example (18) it is assumed that voiceless obstruents are deleted because of
underspecification of voicing (or possibly due to some other spreading motivation which isn’t specifically germane to the vowel harmony topic).

(18) *nasal assimilation before voiceless obstruents* 6

<table>
<thead>
<tr>
<th>labial</th>
<th>coronal</th>
<th>dorsal</th>
<th>coronal</th>
</tr>
</thead>
<tbody>
<tr>
<td>oral</td>
<td>oral</td>
<td>oral</td>
<td>oral</td>
</tr>
<tr>
<td>( \eta + p )</td>
<td>( \eta + t )</td>
<td>( \eta + k )</td>
<td>( \eta + s )</td>
</tr>
<tr>
<td>SP</td>
<td>SP</td>
<td>SP</td>
<td>SP</td>
</tr>
<tr>
<td>[+nas]</td>
<td>[+nas]</td>
<td>[+nas]</td>
<td>[+nas]</td>
</tr>
</tbody>
</table>

vl. obstr. deletion


Classifiers and some measure nouns are other word classes prefixed by the harmonic prefix *so- 'one'. Many classifiers and measure nouns have a problematic nasal that occurs before an obstruent on some words (see (18) for some examples). There are two possibilities for the source of this nasal. One is that the nasal is diachronically a genitive linker no longer productive (or marginally productive), and the second alternative is that they are part of the word stem inherently (it appears that there are instances of both). Some words clearly have environments without the nasal when used as a noun.

(19) se-m-pea one flat hard object so-ndoung one night
    sa-n-ta'u one four-legged animal so-m-buland one month
    sa-n-jangkayu one snake-like object se-m-biti one time
    sa-n-janggang one handspan sa-m-paa one cluster of s.t.
    se-n-si'u one cubit so-m-buli one branch of s.t.
    sa-n-tanga one-hand so-m-bo'a one plate of s.t.
    so-ng-kolo one cut/shaped heavy object

The vowel in *so- 'one' always follows vowel harmony regardless of the consonant sequence, which is contradictory to previous examples which shows that *mong-* does not take vowel harmony when the root or stem begins with a consonant. However, if we assume that *g-* is a productive prefix, then by assigning it to level 1 and nasal assimilation to level 1 (L1), then it becomes a new stem or "lexical item" which can then allow vowel harmony to occur in level 2 (L2) without the vowel harmony blocking that occurs from nasal assimilation (see (20)).

(20) [[so][\( \eta \)][p\( e \)]]  Underlying representation
    [[so][m\( p \)\( e \)]]  L1 nasal assimilation, bracket erasure
    [sem\( p \)e\( a \)]  L2 vowel harmony, bracket erasure, resyllabification
    sem\( p \)e\( a \)  surface representation

---

6 The phoneme *e* does not delete in all lexical words, and in some words it appears to be optionally deleted. When the *s/ is not deleted the preceding nasal becomes the alveolar *la*. Nasals preceding /s/ behave irregularly in many other Western Austronesian languages (e.g. Indonesian, etc.).
6.2 Vowel Harmony Blocked by Nasal Assimilation (Lexical [Level 2])

We have seen that the nasal assimilation is a kind of spreading. Similarly we have seen that vowel harmony spreads. In this section I demonstrate how nasal assimilation blocks the vowel harmony from spreading, since the WFC states that association lines cannot cross (see McCarthy 1988, Kenstowicz 1994, Goldsmith 1990, etc.). Current work on feature geometry appears to be moving towards a consensus that the vowels and consonants have a node on the same tier. We will now dispense with the heuristic usage of [BR] and replace it with the oral node. For example Kenstowicz (1994:469) says:

Attaching the V-PI node as a subtree under C-PI predicts that rules spreading a V-PI feature from one vowel to another across an intervening consonant will block when the consonant is specified for the relevant feature as a secondary but not as a primary articulation. Although this prediction remains to be systematically investigated, there are cases on record that point in the right direction.

This provides theoretical motivation for the vowel harmony blocking in the case of active voice consonant sequences (contrast to example (18) in §6.1 above where vowel harmony crosses consonant sequences—the lexical phonology module is needed to explain this apparent inconsistency of the phonology, see this discussion in §6.1 and §7). In (21)-(22) the dotted lines show assimilation of features taking place. After nasal assimilation has taken place, vowel harmony cannot apply now because in the current cycle the association lines cannot be crossed. So vowel harmony is blocked by assimilation of a consonant in the same cycle where vowel harmony would normally apply.

(21) oral oral

\[
\begin{array}{c}
\text{m o n} + \text{b a l i} + a' \\
[+\text{nas}] \\
\text{mom-bali-a'} \\
\text{AV/IR-move-BEN}
\end{array}
\]

(22) oral oral

\[
\begin{array}{c}
\text{m o n} + \text{p a r e sa} \\
[\text{nas}] [\text{Coronal}] (p, t, k, s) \\
\text{mom-[p]aresa} \\
\text{AV/IR-check}
\end{array}
\]

6.3 Nasal deletion before Sonorants (Lexical [Level 2])

Nasal deletion occurs according to the Obligatory Contour Principle (OCP). The OCP states that "adjacent identical elements are prohibited (see McCarthy 1988:88)" Any word root or stem with an initial sonorant consonant will invoke this rule when preceded by a nasal prefix. When the nasal assimilation occurs, the OCP is met and the initial consonant is deleted. Vowel harmony is still blocked, as occurs when nasals precede obstruents, as shown in (23)-(25).
(23)  

\[ \text{oral oral} \]  

\[ \text{mo} \eta + \eta \quad \text{o} \eta \eta \text{o} \]

\[ \text{[nas]} \quad \text{[nas]} \]

\[ \rightarrow \quad \text{nasal assimilation} \]

\[ \text{vowel harmony blocked} \]

(24)  

\[ \text{oral oral} \]  

\[ \text{mo} \eta + \eta \quad \text{a} \eta \text{gir} \]

\[ \text{[nas]} \quad \text{[lat]} \]

b.  

\[ \text{oral oral} \]  

\[ \text{mo} \eta + \eta \quad \text{a} \eta \text{gir} \]

\[ \text{[lat]} \quad \text{[lat]} \]

\[ \rightarrow \quad \text{mo(} \eta \text{)}-\text{a} \eta \text{gir} \]

\[ \text{AV/IR-ceremonial washing} \]

OCP causes lateral deletion

(25)  

\[ \text{oral oral} \]  

\[ \text{mo} \eta \quad \text{ampu} \eta \]

\[ \text{[nas]} \quad \text{[son]} \]

\[ \rightarrow \quad \text{nasal assimilation} \]

\[ \text{vowel harmony blocked} \]
6.4 Glottal Assimilation (Lexical [Level 2])

Autosegmental phonology once again provides motivation for selecting the glottal stop over the voiceless velar stop [k] as underlying where they alternate. Diachronically it is clear that the glottal stop in Pendau, in many words (if not all), is a reflex of *k (see for example Kaili-Pamona languages, Martens in prep.). Placing an underlying *k for glottal stops does not work since there are minimal pairs which show contrast and there are clear instances of [k] in many other analogous environments. Following an autosegmental approach I will show that the simplest approach is to posit the glottal stop as underlying in the cases where it alternates with [k], and that when [k] appears after the velar nasal it is due to a spreading of a point of articulation to the adjacent glottal stop. The rule presented in (26) demonstrates how the point of articulation assimilates to the nasal voicing, and then delinks the constricted glottis point of articulation resulting in the back velar stop [k].

(26)

\[
\begin{array}{c}
\text{oral oral} \\
\text{mo(ŋ)-rampung} \\
\text{AV/IR-burn} \\
\emptyset
\end{array}
\]

OCP causes r-deletion

The paradigm in (27) illustrates that all of the data can be accounted for under this viewpoint. Underlying k can be contrasted with underlying ? in the environment of active voice prefixes. Underlying ks will delete like the other obstruents in its natural class (p, t, k, s) when preceded by the nasal segment (voiceless obstruent deletion rule).7 Whenever a [k] is found in the surface following a nasal segment (in this particular boundary) it has to be from an underlying glottal stop. This can also be verified by other affixes which end in a vowel. The glottal stop remains a glottal stop intervocally.8

---

7 This phonological process is not isolated to Pendau. The same process that assimilates the glottal stop to the voicing in Pendau occurs in the neighboring Kaili-Pamona group. In Da’a and Ledo (Kaili languages) voiceless obstruents assimilate the voicing of an affixed nasal. It is also clear that the two sets of data reflect two different historical stages of Pendau. Those with k probably reflect borrowing at a later stage, as Bill Foley pointed out during the AFLA7 conference.

8 Glottal stops appear to have a different range in the Kaili languages. Many cognates words between Da’a and Pendau for example contrast ks and glottal stops, as in kawu and ‘ayu “wood, tree” respectively. The Da’a data is from Barr’s lexicon database, personal communication, and Barr 1990; also compare Ledo data in Ghani Hali 1990.
Lexical phonology splits the phonology into two modules: the lexical and the post-lexical modules. Phonological processes within the lexical module apply in a cyclic manner. In figure 3 I have proposed three levels for the Pendau lexical phonology.

![Figure 3. Levels in Pendau Lexical Phonology](image)

The Strict Cycle Condition and the Elsewhere Condition both apply at all levels within the lexical module. The Elsewhere Condition allows disjunctive phonological rules to apply the more specific or idiosyncratic rules when there is a choice between two rules that would apply. The Strict Cycle Condition allows phonological processes to be recycled such as vowel harmony and resyllabification as additional affixes are combined into a word. The Strict Cycle Condition is a complex formulation that has solved complex problems in phonology and solves an apparent conflict in Pendau phonological rules as well. Goldsmith provides a simplified statement of the formulation (1990:223; see Kenstowicz 1994:208 for a more formal statement):

...a rule must apply to phonological material at the first chance – the first cycle – or else for ever hold its peace, and must never return to that earlier cycle to have an effect.
Each morphological non-root formative (i.e. affixes) is associated with a particular level and is added at the appropriate level for the word formation process, and not at the beginning of the entire lexical module. All lexical derivations go through each level (see Kenstowicz 1994:214). Each level has phonological rules that interact cyclically with each applicable process for that particular string sequence (the Strict Cyclicity Rule and Elsewhere Condition constraining the overall procedures). After all phonological processes have applied for one level than the bracket erasure convention is applied on those formatives which have been affected. The bracket erasure invokes resyllabification where that is applicable. The output of each level is a "lexical item" (as defined by Kenstowicz 1994:214). Later levels must interact with the new string output as a whole unit.

Earlier levels in the word formation process are assigned affixes closer to the root. Level 1 in Pendau has the more idiosyncratic processes such as nasal ligature assimilation as well as the pa- causative, the gu- reflexive, the infixation of the -um- aspectual infix, etc. These are all affixes close to the root. Nasal assimilation is the only feature changing phonological process in level 1 (notice that it also applies in level 2; phonological processes may apply in adjacent levels, see Katamba 1993:140).

Level 2 has the greatest number of affixes assigned to it and also has the greatest number of phonological processes which can apply. The affixes assigned to level two are pV(C)- stem formers, causatives po₁- and po₂-, the infixation -ong- and the reduplication types 1 and 2. Feature changing phonological processes which apply here are vowel harmony, nasal assimilation, glottal stop assimilation, voiceless consonant deletion, the application of the obligatory contour principle (OCP), and template copying for reduplication types 1 and 2. Some processes such as vowel harmony are iterative as they may apply to more than one affix.

Examples of word formation processes at level 2 are illustrated in example (28). The first cycle deals with the root (see Kenstowicz 1994:208-209), but none of these phonological processes are applicable to the root (or stem). Cycle 2 applies vowel harmony and nasal assimilation processes on the words that will be output as momaresa ‘check’ and mangabut ‘weed, clear’. Blanks (indicated by dashes) are rules inapplicable at that cycle for those formatives. Vowel harmony and nasal assimilation have to be on the same level. If nasal assimilation applied at a previous level to vowel harmony then vowel harmony would be blocked in all words (not just those with underlying consonant sequences) since the bracket erasure convention would erase the affix boundary from the word base, and vowel harmony could never apply. Sample derivations of nasal assimilation, vowel harmony (and blocking), and voiceless consonant deletion are provided in (28).

(28)

<table>
<thead>
<tr>
<th>Cycle 1 (Level 2)</th>
<th>Nasal Assimilation</th>
<th>Vowel Harmony</th>
<th>VI. Cons. Deletion</th>
<th>Glottal Stop Assim.</th>
<th>Glide Formation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[moŋ] [paresa]</td>
<td>‘check’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[moŋ] [abut]</td>
<td>‘clear, weed’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>________________________</td>
<td>___________________</td>
<td>_______________</td>
<td>___________________</td>
<td>___________________</td>
<td>________________</td>
</tr>
<tr>
<td>________________________</td>
<td>___________________</td>
<td>_______________</td>
<td>___________________</td>
<td>___________________</td>
<td>________________</td>
</tr>
<tr>
<td>________________________</td>
<td>___________________</td>
<td>_______________</td>
<td>___________________</td>
<td>___________________</td>
<td>________________</td>
</tr>
<tr>
<td>________________________</td>
<td>___________________</td>
<td>_______________</td>
<td>___________________</td>
<td>___________________</td>
<td>________________</td>
</tr>
</tbody>
</table>

187
In Pendau not all voiceless consonants are deleted in apparently the same environment, however the application of the Strict Cycle Condition (SCC) easily resolves this issue. The sample derivation in (29) demonstrates that the harmonic stative causative prefix po-, remains unaffected by the voiceless consonant deletion rule. The roots *lalo* ‘deep’ and ide ‘small’ are causativized and mean ‘deepen’ and ‘decrease’ respectively. Cycle 2 can only refer to information between the [po] [ide] boundaries, and so only vowel harmony applies. Cycle 3 can only refer to information between the [mom] [pe] and the nasal assimilation applies, but then nasal assimilation blocks the vowel harmony from applying (see §6). Voiceless consonant deletion cannot apply because the segments in [pe] were already affected in an earlier cycle (cycle 2).

(29)

- [mom] [paresa] (blocked) [man] [abut]
- [mom] [aresa] [manabut]
- [momoaresa]
- [mompa] [lalo’] ‘deep’ [mom] [po] [lalo’] [ide] ‘small’
- [mompa] [lalo’] [mompe] [lalo’] [ide]
- [mom] [pa] [lalo’] [mom] [pe] [lalo’] [ide]
- [mom] [pa] [lalo’] (blocked) [mom] [pe] [lalo’] (blocked) [mompalalo’] [mompeide]

The derivation in (30) shows the interaction between level one and level two and demonstrates the need for assigning the causative prefix pa- to a different level than the other causative prefixes. Since the voiceless consonant in pa- deletes it cannot be in the same level as the po- causative prefix, since the SCC would prohibit the deletion of this consonant. This supports the lexical phonology model where bracket erasure at the end of a level, level 1 in this case outputs in essence a “lexical item” and enters the next level as a
A Non-linear Analysis of Vowel Harmony and Vowel Harmony Blocking in Pendau

unit. This will then allow the initial voiceless consonant to delete as it does in words like paresa ‘check’ (see the derivation in (30)). The causative prefix pa- changes the lexical meaning of guru ‘learn’ to paguru ‘teach’ and inang ‘eat’ to painang ‘feed’.

(30)

<table>
<thead>
<tr>
<th>Cycle 1 (Level 1)</th>
<th>Cycle 1 (Level 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasal Assimilation</td>
<td>Nasal Assimilation</td>
</tr>
<tr>
<td>Bracket Erasure</td>
<td>Vowel Harmony</td>
</tr>
<tr>
<td></td>
<td>VI. Cons. Deletion</td>
</tr>
<tr>
<td></td>
<td>Glottal Stop Assim.</td>
</tr>
<tr>
<td></td>
<td>Glide Formation</td>
</tr>
<tr>
<td>[mon] [paguru]</td>
<td>[mom] [painaŋ]</td>
</tr>
<tr>
<td>(blocked)</td>
<td>(blocked)</td>
</tr>
<tr>
<td>[mom] [aguru]</td>
<td></td>
</tr>
<tr>
<td>[momaguru]</td>
<td>[momainaŋ]</td>
</tr>
</tbody>
</table>

Going back to level one will provide the motivation for explaining why vowel harmony is blocked by nasal assimilation when there are consonant sequences for the active voice prefixes mong-/nong- but vowel harmony is allowable for certain cases such as for classifiers (see §6.1). So a combination of theoretical insights from lexical phonology and autosegmental phonology provide a powerful, motivating and elegant solution to the vowel harmony blocking problem.

8. Vowel Harmony in Western Austronesian Languages

Vowel harmony exists in several Western Austronesian languages, however there has never been a diachronic overview of the historical relationship of this interesting phonological phenomenon or a systematic study or survey. It appears that vowel harmony may have been an innovation that occurred at some point in the history of these languages. Languages in Sabah, Malaysia are noted to have vowel harmony. Hurlburt (1988) mentions briefly vowel harmony in Eastern Kadazan where the root vowels are affected by affixes (i.e. the root vowel(s) harmonize to a specific vowel in a prefix or a suffix). Boutin (1988, 1994) mentions vowel harmony in Bonggi (also called Banggi) which has some similarities to Pendau and Balantak. Kroeger (1994) mentions vowel harmony in Kimaragang, and it is clear from the pronoun and verbal systems in Sabah that there are a lot of similarities between some Sabah languages and of languages in Central Sulawesi. Balantak, a Saluan language in eastern Central Sulawesi has vowel harmony in prefixes and suffixes (see Busenitz 1991, 1994). Lauje and Tialo have similar vowel harmony as Pendau (Whatley nd. and 1984; Himmelmann 1991; Yoshimura in prep.). Vowel harmony in the Tomini-Tolitoli group is a striking phonological characteristic of the verbal system of many if not all of these languages. In light of indications that this is a significant group characteristic, the historical and comparative implications demands a close look at the data.
and a detailed investigation over these Western Austronesian languages. Historical
comparativists should welcome an additional phonological feature that will lend another
criterion to reconstruction and increase the accuracy of their methodology.

Van den Berg has proposed a verbal reconstruction for Proto-Celebic (1991, 1995; see
figure 4 below; also compare Mead 1997, 1999). Vowel harmony phenomena should
provide a more detailed explanation for the Proto-Celebic verbal marking system. Some
readjustment of his proposal will be required in light of the major use of vowel harmony in
languages have verb classes marked by ma-/mo-/me-, some of which appear to have
semantic or grammatical correlates (notably transitivity), but usually with a large degree of
randomness. Data from the Tomini languages suggest that these classes may have
developed from an earlier partial vowel harmony system.” In the Kaili-Pamona languages
it is clear that vowel harmony has littered the prefix system in the Da’a and Ledo (Kaili
languages) and left it non-productive (see Barr 1988 and Evans personal communication).
It is hard to imagine that vowel harmony did not precede the Kaili languages at an earlier
proto stage which overarches Kaili and the Tomini-Tolitoli groups (again compare figure
4, since the Proto-Celebic is essentially the same as the Kaili language affixes). Noting
that vowel harmony appears in several Western Austronesian languages is more than
curious. It suggests that either vowel harmony was an important part of the protolanguage
or that there is as yet some unexplained phonological motivation (or both?) similar to
tonogenesis that creates this phenomenon. The rapid growth of phonological theory has
already made an impact in our understanding of the processes of vowel harmony.

<table>
<thead>
<tr>
<th>Proto-Celebic</th>
<th>Pendau</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active Voice</strong></td>
<td></td>
</tr>
<tr>
<td>mo-/no-</td>
<td>mong-/nong-</td>
</tr>
<tr>
<td>ma-/na-</td>
<td>(mang-/nang-,</td>
</tr>
<tr>
<td>me-/ne-</td>
<td>meng-/neng-)</td>
</tr>
<tr>
<td><strong>Dynamic</strong></td>
<td></td>
</tr>
<tr>
<td>Ø-/Ø-</td>
<td>me-/ne-</td>
</tr>
<tr>
<td>me-/ne-</td>
<td></td>
</tr>
<tr>
<td>mo-/no-</td>
<td></td>
</tr>
<tr>
<td><strong>Stative</strong></td>
<td></td>
</tr>
<tr>
<td>Ø-/Ø-</td>
<td>mo-/no-</td>
</tr>
<tr>
<td>ma-/na-</td>
<td>(ma-/na-, me-</td>
</tr>
<tr>
<td>mo-/no-</td>
<td>/ne-)</td>
</tr>
</tbody>
</table>

Figure 4. Some Proto-Celebic Verbal Prefixes Compared with the Cognate Pendau Prefixes
(parentheses indicate Pendau Vowel Harmony; Proto-Celebic Affixes based on van den Berg 1995)

This discussion has provided extensive data of vowel harmony in Pendau. The large
number of prefixes that are harmonic provides an interesting showcase for vowel harmony.
Theoretical implications of vowel harmony as described for Pendau should provide further
fodder for phonological theoreticians as well as for historical and comparative researchers.
REFERENCES


Mead, David. 1997. Reflexes of *aken across Central and Southeastern Sulawesi*.
Introduction

In this paper, it will be shown that the interaction between Kayne’s Binary Branching [as in Kayne (1981) and implemented for Malagasy in Randriamasimanana (1999c)] and the (nominal vs non-nominal) nature of the element that can show up under Infl(ections for tense/aspect) [as outlined in Randriamasimanana (1997)] will determine whether a null element is allowable — iff Infl does not comprise a nominal element— or not allowable — iff Infl comprises a nominal element— in the external subject position.

The above account of the distribution of null subjects in Malagasy has several consequences and puts to the fore the crucial importance of the Specifier-Head relationship in the grammar of this language. The major purpose of this paper is four-fold:

(i) to outline the specific subsystem in which the tense and aspect system of Malagasy interacts with both Binary Branching and the minimalist type of phrase structure proposed in Koizumi (1995) and shown to be relevant for Malagasy in Randriamasimanana (1999a) to yield the basic/building block structures found in this Malayo-Polynesian language;
(ii) to show specifically how more complex constructions involving for instance, motion verbs can be said to derive from ‘mergers’ of pre-existing, actual, surface binary structures of Malagasy, taking into account the crucial distinction between arguments and adjuncts established in Randriamasimanana (1999b) and
(iii) where the process of incorporation, which is triggered by the positive value for the verbal atomic feature [ +/- CONTROL] associated with the higher verb plays a crucial role in the formation of such complex structures in Malagasy, as outlined in Randriamasimanana (1998) and discussed in some depth in Randriamasimanana (1986: 29-74).
(iv) As many of the issues raised above are illustrated in many of the examples found in Keenan (1999), some of this author’s Malagasy sentences will be used to show precisely how Binary Branching applies to this language, forcing a Small Clause analysis of some of the complex Malagasy illustrations.

This paper is organised as follows: In Part I, some justification for a Binary Branching analysis of Malagasy sentences will be proposed, based on a Small Clause analysis of predicates involving lexical causative verbs like ‘kill’ (section A), motion verbs (section B) as well as di-
transitive verbs (section C). Part II will initially distinguish between discourse-based null subject phenomena and linguistic utterance-based null subject phenomena before looking into the nature and function of Malagasy AGR(eement). Part III will subsequently propose an account of the distribution of linguistic utterance-based null subject phenomena in terms of the nominal or nominal nature of the Inflections for tense and aspect outlining the crucial importance of the Spec-Head relationship.

As a starting point, we will assume the following type of tree diagram reproduced from Randriamasimanana (1998: 304), where a distinction is made on Figure 1 between a lexical item projection like V as opposed to a functional head projection like Infl(ections) for tense/aspect. In addition, Figure 2 gives a summary of the distribution of Empty Categories of type 2 (see sections D, F, J and K below for relevant details) in Malagasy.

\[
\begin{array}{c}
\text{Inflmax} \\
\text{Infl} \\
\text{Head} \\
\text{Infl} \\
\text{Vmax} \\
\text{Complement} \\
\text{Specifier} \\
\end{array}
\]

Where head = lexical = \{V, P, N, A\}; head = functional = \{tense, aspect, agreement\}; NP = DP or bare noun

Figure 1: X-Bar Theory and Tree Geometry

Summary of the Distribution of Empty Categories

\[
\begin{array}{c}
\text{Inflmax} \\
\text{Infl} \\
\text{Infl} \\
\text{Head} \\
\text{Complement} \\
\text{Specifier} \\
\end{array}
\]

(i) AGR/Nom.Tense............Overt NP
(ii) [- Finite]................empty

Randriamasimanana (1998: 304)

Nominal = AGR, tense-markers like no (past) and ho (future).
Not nominal = tense-markers like n- (past) and h- (future).
Part I: Malagasy and Binary Branching

Section A: Lexical causative verbs.

In order to understand the necessity for a Binary Branching type of analysis a la Kayne (1981) for Malagasy, we need to look at lexical causative verbs like ‘kill’ in utterances of the following kind:

(1) N-amon o t sy n-aha-faty i Paoly.
Past-kill not past-cause-dead art. Paul
Lit: 'Paul killed but did not cause (someone) to die.'
i.e. English: ‘Paul tried to kill (someone), but did not manage to.’

Randriamasimanana (1999.a: 513)

The continuation ‘but did not cause (someone) to die’ is perfectly grammatical and acceptable in Malagasy since there is absolutely no contradiction involved in the entire sequence shown in (1). The reason for this is that the first verb n-amon o ‘kill’ is derived from a combination of two different predicates, i.e. a higher verb anao ‘do’ compressed into an and a lower predicate vono ‘kill’. This lower predicate can be part of a Small Clause made up of a root vono and an empty subject; whereas the higher verb an will take a tense-marker, in this case the past tense-marker n. That the lower predicate can often be part of a Small Clause with an empty subject can be illustrated with the following utterance originating from a recent Malagasy newspaper:

(2) Tsy mbola te rak a hono.
Not yet born be-said Empty subject Nonverbal
predicate
"X is said to be not yet born!"

Randriamasimanana (1998)

Where teraka ‘be born’ is a root serving as a predicate in a so-called nonverbal construction, hence the label Small Clause. Such a distinction between so-called nonverbal constructions and verbal constructions was established as far back as Rajaona (1972) for Malagasy. Essentially the distinction hinges on the fact that Malagasy verbal predicates take a tense-marker, whereas nonverbal predicates can only accommodate aspectual markers to the exclusion of tense-markers.

As far as the second part of the utterance shown in (1) is concerned, the second predicate is a combination of the higher causative verb aha ‘cause’ and the root predicate faty ‘dead’. This causative predicate carries an entailment of whatever is asserted in the lower Small Clause, which can be represented thus:

(3) Faty ‘dead’ Empty subject
Nonverbal predicate
Thus the higher verbs used in (1) describe two different situations in Malagasy: While the higher
anao for the first verb compressed into an refers to the inception of some activity described by
the verb, the higher verb aha for the second verb refers to the completion of the activity being
described. In both instances, the lower predicate is a root word which can be part of a Small
Clause, as defined above with reference to Rajaona (1972).

Section B: Motion verbs.

An analysis of motion verbs in Malagasy will also require utilisation of the notion Small
Clause along with consideration of verbal atomic features such as CONTROL inherent in the
relevant higher verb. This essentially means that all motion verbs in this language will have to be
analysed in terms of a higher and a lower verb along lines sketched above for lexical causatives
like ‘kill’. Thus

(4) N-an-deha t-any Antsirabe i Paoly
past-prf-go perf-to Antsirabe art Paul
[ + CONTROL] Argument
'Paul went to Antsirabe.'

(5) N-an-deha i Paoly
past-prf-go art Paul
'Paul went'

(6) T-any Antsirabe i Paoly
perf-to Antsirabe art Paul
'Paul went to /was in Antsirabe.'

Randriamasimanana (1999b.)

The (complex) motion verb shown in (4) can be decomposed into a higher verb, as in (5), and a
lower verb, as in (6). Note that the higher verb n-an-deha ‘went’ takes a past tense-marker,
whereas the lower verb takes an aspect-marker t-. The main reason why this morpheme t- is
analysed as an aspect-marker and not as a tense-marker is due to the fact that the higher verb
contains the atomic feature [+CONTROL], which will trigger incorporation of the embedded
Small Clause into the higher verb, thus turning the constituent t-any Antsirabe into an argument
of the higher verb and not into a mere adjunct. This distinction between an argument of the verb
and an adjunct was already noted although not accounted for in Rabenilaina (1985).

It was shown in Randriamasimanana (1999b) that when the higher verbal predicate does
not take a positive value for this atomic feature, i.e. [-CONTROL], then the embedded Small
Clause will simply remain an adjunct and will not become an argument of the higher verb:

(7) N-i-petraka t-any Antsirabe i Paoly.
past-prf-stay perf-at Antsirabe art Paul
[ +/-CONTROL ] Argument/Adjunct
'Paul stayed at Antsirabe.'
Thus in (7) the verb ‘n-ipetraka’ ‘stayed’ in Malagasy is ambiguous between a + or a -CONTROL interpretation in that it could mean either that ‘Paul took an active part in selecting his place of residence’ or simply that ‘Paul just happens to be living at this particular location’. In its [-CONTROL] interpretation, the constituent made up by t-any Antsirabe will remain a mere adjunct of the higher verb. The independent evidence that shows that indeed it remains an adjunct is provided in (8): When the constituent is moved to the front, nothing happens to the higher verb, it remains in the active voice. On the other hand, when the higher verb has a positive value for the same feature, as shown in (9), then when the constituent is moved to the front, the higher verb must be passivised otherwise an ungrammatical sequence will ensue.

Section C: Di-transitive verbs.

It is not only lexical causative verbs like ‘kill’ and motion verbs which require a Small Clause analysis. Di-transitive verbs also do. Consider the following:

(10) N-anome an'i Jeanne ilay boky i Paoly
    'past-give DO art. J. the(previous mention) book art. Paul'
    'Paul gave Jeanne the (previously mentioned) book.'

(11) An' i Jeanne ilay boky. = Small Clause S.
    Nonverbal
    Predicate article J. article book
    'The (previously mentioned) book belongs to Jeanne.'

Where the nonverbal construction an'i Jeanne ilay boky shown in (11) does have an independent existence as a clause of its own: In this utterance an is a nonverbal predicate in that it certainly cannot accommodate a tense-marker – just like other Small Clauses of Malagasy. In (10) the same word is tentatively analysed as a case-marker. However such an analysis will not do since if we replaced the constituent an'i Jeanne with another constituent like ilay olona ‘the (previous mention) person’, the putative case-marker will not have to appear at all. In fact, it will now be purely optional:

(12) N-anome (an) ilay olona ilay boky i Paoly
    'past-give (DO) art. person the(previous mention) book art. Paul'
    'Paul gave the person the (previously mentioned) book.'
If on the other hand, we analyse *an* as a nonverbal predicate, an explanation for the optionality of the nonverbal predicate *an* is readily available in Malagasy. When the word is present within the utterance, there is no particular problem since it will mean something like ‘belong’. If it is absent from the utterance, then we obtain a slightly different kind of Small Clause, a subtype described in some depth in Rajaona (1972):

(13) \[ Tsena \ y ny\ olo\ na. \]
    Market the people
    ‘People hold market.’

Rajaona (1972)

Where the word *tsena* ‘market’ serves as a nonverbal predicate and where the constituent *ny olo\ na* ‘the people’ will be the grammatical subject so that the thematic relationship obtaining between the two elements is one involving possession.

While up to this point the difference between a case-marker analysis and a Small Clause analysis may still not be decisive with respect to the utterances shown in (10) and (12), here is a case where it does make a crucial difference. The following case of di-transitive verbs comes from Keenan (1999: 34).

(14) \[ N-anolotra \ vary \ ho \ an'ny \ vahiny \ t-amin'ny \ lov\ ia \ vaovao \ aho \]
    Past-hand rice to the guest past-prep the dish new I

The intended meaning for (14) is: ‘I presented rice to the guest on the new dishes.’ However, the Malagasy sentence means something entirely different, i.e. ‘I presented rice (which was destined) for the guest on the new dishes’ with a relative clause kind of meaning.

One major assumption inherent in (14) analysed as meaning ‘I presented rice to the guest on the new dishes’ is that the particle *ho* (just like the particle *an* in sentence (10) is a case-marker. Unfortunately to the native speaker this word *ho* is not a case-marker, it means something like ‘which was destined for’ in that it invariably forms the beginning of a relative clause making more precise the meaning of the headnoun *vary* ‘rice’ to which it is attached.

By contrast, with a Small Clause analysis of (14), we will simply leave the particle *ho* out altogether, obtaining the following two possibilities, both involving Small Clauses:

(15) \[ N-anolotra \ [ vary \ ny \ vahiny ] \ [ t-amin'ny \ lov\ ia \ vaovao ] \ aho. \]
    Past-hand the guest rice past-prep the dish new I I

(16) \[ N-anolotra \ [ ny \ vahiny \ vary ] \ [ t-amin'ny \ lov\ ia \ vaovao ] \ aho \]
    Past-hand the guest rice past-prep the dish I I

Both (15) and (16) mean: ‘I presented rice to the guest on the new dishes.’ In (15), we have two Small Clauses, the first with a nonverbal predicate *ny vahiny* ‘the guest’; the second with a
prepositional predicate comprising a past tense-marker t- indicating that this constituent is a mere adjunct to the higher verb. In (16), we also have the inverse word order within the first Small Clause.

Part II: Malagasy and Null Subjects

As a preliminary, it is essential to distinguish between discourse context-based null subject phenomena, on the one hand and linguistic utterance-based null subject phenomena, on the other hand, in Malagasy. We will refer to the first category as type 1 null subjects and to the second category as type 2 null subjects. In what follows, we will concentrate exclusively on the nature and function of the latter category.

Section D: Discourse context-based null subjects.

The relevant cases allow access to the identity of the referent from the immediate context of the given utterance. Reference could be either to the speaker/writer as in (17) or to the interlocutor as in (19). Note by contrast that foreigners tend to have an overt grammatical subject even when none is required, as in (18).

(17) Date: Fri, 05 Mar 1999 12:19:59 -0500
From: JR<r@magma.com>
To: R<uzsmav@uni-bonn.de>
CC: fbra2@bigfoot.com

*M-amerina ny arahaba ho an'i Mia sy ny rehetra* [1st S]

pres-renew the greeting for art Mia and the all Empty

‘(I) renew my greetings to Mia and to everybody…’

(18) Subject: Valin’dresaka hoan’i Rina Ralison
Date: 1997/10/09
Author: DD d@math.u-strasbg.fr

*Faly m-iarahaba anareo rehetra aho.*

Happy pres-greet you-plural all I

‘I am happy to greet you all!’

(19) Date: Thu, 24 Dec 1998 06:52:17-0500
From: CRab114103.442@compuserve.com
Sender: CRab114103.442@compuserve.com
To: Charles Randriamasimanana

*M-anao ahoana indray* [1st S in message]

Pres-do how once-more Empty?

‘How are (you)?’
Section E: Linguistic utterance-based null subjects.

As will be seen in Part III, there is a direct correlation between the distribution of Malagasy AGR(eement) and the obligatory presence of the grammatical subject; whereas its absence coincides with the possibility of an empty subject. In other words, with type 2 null subjects, there is a link between the structure of the linguistic utterance and the distribution of empty subjects, specifically there is a crucial relationship obtaining between the inflections for tense/aspect and the like and the grammatical subject within the utterance.

What will first be illustrated below is the crucial relationship between the specifier and its head, and in particular, Malagasy AGR(eement) and the specifier, as this phenomenon has not been extensively explored and discussed in the literature. As noted in Randriamasimanana (1997), AGR imposes a singular vs plural meaning on the grammatical subject of the clause, thus highlighting the crucial importance of the spec-head relationship in this language. Relevant illustrations go as far back as the nineteenth century and include examples from Abinal & Malzac (henceforth A & M) (1888) as well as from contemporary sources:

(20) *Ireo m-iady ireo ny zanakao.*
AGR pres- fight AGR the children-of-yours

**Plural verb plural**

'Voilà vos enfants qui se battent.' (A & M 1888: 282)

From French to English: 'Your children are there, fighting.'

Literally: 'Your children are fighting—as we can see for ourselves.'

(21) *Io tamy io ny zanako.*
AGR coming AGR the child-of-mine

**Singular verb singular**

'Voici mon enfant qui vient.' (A & M 1888: 281)

Eng.: 'Here is my child coming.'

Literally: 'My child is coming—as you can see for yourself.'

(22) To: fbra2@bigfoot.com
Date: Mon, 18 Jan 1999 00:38:14 -0500
From: crazafi@juno.com (Carol M Razaf...)

**Ireto manaraka ireto ny valin'ireo ohabolana 11-20...**

**AGR verb AGR the answers**

These pres-follow these the answer-of-those proverbs 11-20

'The answers to proverbs 11 to 20 are following...'

In (20) from (A & M. 1888: 282) and (22) from an e-mail dated January 1999, we have an illustration of a plural AGR(eement), i.e. one of a number of Malagasy deictics such as *io* 'this-singular-near the hearer', *ireo* 'these-plural-near the hearer' and *ito* 'this-singular-near the speaker' and *ireto* 'these-plural-near the speaker' which morphologically encodes the singular vs the plural, which shows up under inflections along with tense and aspect. The grammatical number which appears on AGR will affect the semantic interpretation of the grammatical subject:
Thus, in (20) and (22), because AGR is plural, the subject has to be interpreted as plural; whereas in (21), where AGR is singular, the subject must be interpreted as singular. By contrast if AGR was to be left out in, for instance, (20) or (22), the relevant subject could then receive either a singular or a plural interpretation, depending on the extralinguistic context of situation.

**Section F: Nature and Function of Malagasy AGR.**

In Malagasy, the absence of morphology corresponding to AGR, tense/aspect indicates a context-dependance of temporal location along the time axis whereas the presence of AGR, tense/aspect morphologically signals overt linguistic encoding of location along the time axis. In the first type of situation, location within the speaker here and now is usually assumed, in other words, 'de re' is assumed to coincide precisely with 'de dicto' and as a direct consequence, no overt indication of time location within the utterance is required since the relevant bit of information is recoverable from the immediate context; whereas in the second, a fundamental and basic distinction between 'de dicto' and 'de re' is assumed and as a result, an overt indication of time location within the utterance is absolutely indispensable since in such a case, there is no possibility of recoverability of the missing pieces of information.

In light of the system described above, the following revised meaning is proposed for the already published sentence below from Keenan (1976: 257):

(23) *j-sasan-dRasoa ny lambda.*
    ṭpass-be-washed-by Rasoa the clothes
    'The clothes are washed by Rasoa.'

This sentence, which contains a zero tense-marker should mean 'The clothes are being washed by Rasoa' and not as originally claimed by the author. Some evidence pointing in that direction comes from the following electronic message:

(24) To: fbra2@bigfoot.com
    Date: Mon, 1 Feb 1999 01:10:24 -0500
    Subject: Ohabolana 21-30(V), 31-40(L), 1-10(F)
    From: Carol M Razafi... <crazafi@juno.com>

    y-arabaina daholo indray ianareo rehetra,
    Pass-saluted all once you-plural all
    Literally: "You are all being saluted (by me)!"

Where it is quite clear that there is a correlation between the zero-tense morpheme and location within the speaker/writer here and now. This contrasts with the presence of an overt tense-marker, as in the following sentence adapted from E. Keenan (1976: 255):

(25) *M-anasa lamba amin'ity savony ity Rasoa.*
    Pres-wash clothes with this soap this Rasoa
    Translated as 'Rasoa is washing clothes with this soap.'
Which actually means ‘As a rule, Rasoa washes clothes with this soap’ and not as originally translated. In fact, in order to convey the meaning ‘Rasoa is washing clothes with the soap’, one has to say:

(26) *ity m-anasa lamba amin’ny savony ity Rasoa.
    AGR pres-wash clothes with the soap AGR Rasoa
    ‘Rasoa is washing clothes with the soap.’

Where the singular deictic ity ‘this’ functions as an AGR inside the inflections for tense/aspect and not accompanying the noun savony ‘soap’, as in

(27) M-anasa lamba amin’ity savony ity Rasoa.
    Pres-wash clothes with this soap this Rasoa

Sentence (27) is a perfect sequence except that its meaning has nothing to do with the original translation provided under (25) but rather with something like: ‘Hey, Rasoa, go and wash clothes with this soap.’ Note that this new interpretation is in line with the kind of analysis proposed for lexical causatives in section A above, where the higher verb an derived from the compression of anao ‘do’ in (27) will refer to the inception of the activity described by the verb; furthermore, the overt tense-marker m- indicating the present tense does not strictly locate the event within the speaker here and now. See Randriamasimanana (1985) for an analysis of this phenomenon.

Part III: An Account of the Distribution of Null Subjects

At this stage, we will look into the distribution of type 2 null subjects. This is linked to the distribution of AGR. In fact, the distribution of AGR coincides with the obligatory presence of the grammatical subject; whereas its absence coincides with the possibility of an empty subject. However, as far as tense projections are concerned, the nominal or non-nominal nature of the head is crucial.

Section J: AGR and Control Structures.

As outlined in Randriamasimanana (1998), AGR only shows up in a matrix clause and usually cannot show up in a structure embedded under a Control predicate:

(28)a. N-itady [... h-ajaka ] Ravoniarisoa ...
    Past-seek [... fut-dominate EC] Ravoniarisoa
    "Ravoniarisoa sought to dominate."

    AGR AGR

(29)a. M-angataka anao aho [h-ondra ity any amin-dRama]
    Pres-ask you I [... fut-take this to pre-Rama EC]
"I ask you to take this to Rama."
From Rajaona 1969, TN, p 65, lines 57-60.

b. *M-angataka anao aho [ity h-itondra ity any amin-dR.-]
AGR       AGR

In (28), we have a subject control verb, whereas in (29) we have an object control verb—which poses a problem for our Binary Branching analysis for Malagasy. Research into this question is ongoing and we will postpone presentation of relevant data which will argue for this specific aspect of our analysis. The only point being made is that there is absolutely no possibility whatsoever of having AGR and an empty subject in a structure embedded under a Control predicate, as suggested by the ungrammaticality of both (28)b and (29)b.

Section K: Nominal vs Non-nominal Nature of Malagasy Tense-markers.

Malagasy has two distinct series of tense-markers for the future and the past tenses: Thus, the morpheme h- indicates the future in a structure with an active voice verb, but another morpheme, ho will have to be used to indicate the future in a sequence with the passive voice; likewise for the past-tense, the morpheme n- indicates the past in an active voice sequence, but the relevant morpheme is no- for the passive voice.

One first contrast in behavior between the two series of tense-markers is apparent in the following pair of utterances:

(30) N-ikasany h-andeha i Paoly.
past-intend comp fut-go Empty deic Paul
"Paul intended to leave."

Randriamasimanana (1997: 488)

(31) Tia-ko ho entina ilay fiara.
be-liked-by-me fut be-driven the car
"I would like to drive the (previous mention) car."

In (30), the embedded predicate is in the active voice and as a result, the relevant future tense-marker is the bound morpheme h-; as a further consequence, there is an empty subject in the embedded structure. By contrast, in (31) the embedded structure is in the passive voice and the relevant future tense-marker is the independent morpheme ho; as a consequence of this, an overt subject shows up in the subordinate clause. Indeed in (31), the constituent ‘ho entina ilay fiara’ is a sentential subject of the passive matrix verb.

Another contrast in behavior between the two parallel series of tense-markers arises from the lexical vs functional nature of the projection. Thus, with a projection involving a purely lexical item, there is a case feature to assign, whereas with a purely functional one, there seems to be no presence of a case feature at all.

(32) No-kasa-in’ i Paoly ho entina ilay fiara.
past-intend-by deic Paul future be-taken the car
"Paul intended to take the (previous mention) car."

(33) N-ikasa (ny) h-itondra ilay fiara i Paoly.  
Past-intend (comp) fut-drive the car EC art. Paul  
'Paul intended to drive the (previous mention) car.'

(34) N-andefa entana ho azy i Paoly.  
past-send parcel part. him-DO deic Paul  
"Paul sent a parcel which is for him/her."

(35) N-andrama-n' i Paoly no-loko-ina ny tran.  
past-try-by deic Paul past-paint-by the house.  
"Paul tried to paint the house."


In (34), we have the purely lexical item *ho* as head of construction with the meaning of something like 'which is destined for' and its complement, the pronoun accompanying it, must be in the accusative form of the third person pronoun *azy*. This contrasts with the situation in (33), where the head of construction is the tense-marker *h*- in the embedded clause, a form which is not purely lexical. On the other hand, in (32) the head of the embedded clause is the future tense-marker *ho*, which is exactly the same as the purely lexical item in (34). It looks therefore as though the independent morpheme *ho* in (32) behaves more like a purely lexical item like the one in (34); as a consequence of this, it is legitimate to assume that it carries a case feature with it. Finally, in (35) we have an illustration of the behavior of the independent morpheme for past tense *no*, which is exactly the same as for its future tense counterpart *ho*: Indeed *no* like *ho* requires an overt subject in the embedded structure *no-loko-ina ny tran* even though this form appears to be well on its way to becoming a clitic to the verb and acquiring the status of a verbal prefix.

Section L: Non-nominal Nature of Complementizer.  
In essence in (34), we had a purely lexical item projection of the particle *ho*. That contrasts with the behavior of the (apparently) same item serving as a complementizer introducing an embedded equative type of clause, as argued for and illustrated in Randriamasimanana (1986: 562-563) and explained in Randriamasimanana (1997: 491):

(36) M-ihevitra azy ho m-ahay - i Paoly.  
pres-think him comp pres-intelligent Empty deic Paul  
"Paul considers himself intelligent."

(37) M-itady ho babena - i Paoly.  
pres-seek comp lift-pass. Empty deic Paul  
"Paul wants to be picked up."


In (36), the embedded clause comprises an equative type of structure (as opposed to a nonequative type or specifically one which involves Control as defined in Randriama-simanana (1986: 29-74)). The relevant complementizer which accompanies such a clause is the function
word *ho* precisely. Note that this function word does not seem to be accompanied by any apparent case feature as an empty subject is permissible in the subordinate clause in both (36) and (37).

Section M: Other Consequences of Spec-Head Relationship.

As the inflections domain within a clause covers projections corresponding to AGR, tense and aspect respectively and since there is a privileged relationship obtaining between inflections as head of the entire sentence and its specifier, it is no big surprise that the semantic interpretation of the subject is influenced by what appears under inflections.

In this section, the co-variation in semantic interpretation of the grammatical subject with the voice, tense and aspect showing up under inflections will be succinctly described. Subsequently it will be shown that Malagasy has recourse to movement to the front as a strategy to avoid an anomalous interpretation of the subject.

The relevant constructions involve a passive voice verb with different inflections for tense and aspect on them. There exist four basic cases to consider:

A. *No...in(a)* passive typically indicates a **PUNCTUAL** aspect. As a result of this, the subject is usually interpreted as an entity affected as an unanalyzed whole.

\[(38) \text{No-didi-ndRabe ny mofo.} \quad \text{Punctual meaning}\]
\[
\text{pst-root-pass-byRabe the bread}\]
\[
\text{‘The (whole) bread was cut by Rabe.’}\]

B. *...in(a)* passive typically describes an **ONGOING ACTIVITY**, as in (24) above. Typically no overt reference to the referent of the subject is required, which is retrievable from the extralinguistic context of situation.

C. With *an...in(a)* circumstantial voice, we usually have a **DURATIVE** aspect, hence partitive interpretation of the subject since its referent is only partially affected.

\[(39) \text{N-an-didi-andRabe ny mofo.} \quad \text{Partitive meaning}\]
\[
\text{pst-active-root-circ-byRabe the bread}\]
\[
\text{‘(Some of the) bread was cut by Rabe.’}\]

D. With *a...in(a)* passive, there is the notion of **INCEPTIVE/BALLISTIC** aspect. Only referents that contain this particular feature, i.e. which only requires an initial impulse will be able to appear in the grammatical subject position.

\[(40) \text{N-a-tsipin-dRabe ny rano.} \quad \text{Ballistic meaning}\]
\[
\text{Past-pass-throw-by-Rabe the water}\]
\[
\text{‘The water was thrown away by Rabe.’}\]

What appears under subsection C is particularly relevant for a newly published Malagasy sentence reproduced immediately below:
According to the system outlined in C, the subject Rasoa in (41) should receive a partitive reading since the circumstantial form of passive shows up under inflections along the same lines as in example (39), which also contains a circumstantial voice verb. However, it is not clear what it would mean to claim that 'part of Rasoa was bought the hat by Rabe!'

Presumably sentence (41) derives from something like:

(42) N-ividy (an) ilay satroka ho an-dRasoa Rabe.  
Pst-prf-root (DO) the hat for DO-Rasoa Rabe  
'Rabe bought the (previously mentioned) hat for Rasoa.'

Instead of (41), a native speaker would say either the a or the b sequence below:

(43)a No-vidy n-dRabe ho an-dRasoa ilay satroka.  
PUNCTUAL SPECIFIC  
'Was bought by Rabe for Rasoa the hat.'

b. y-vidy nRabe ho andRasoa ilay satroka.  
ONGOING-buy-byR for Rasoa the hat  
'Is being bought by Rabe for Rasoa the hat.'

The first problem which arises with regard to sentence (41) relates to the incorporation of the constituent ho an-dRasoa since it is not obvious that the verb nividy 'bought' unequivocally has a [+Control] meaning. This is important since as shown in section B, the presence of a positive value for this atomic feature is a sine qua non condition for triggering incorporation of this constituent into the higher verb. And if incorporation does not take place, then the circumstantial voice passive as used in (41) is simply illegal in Malagasy since only an argument of the verb can be promoted to subject, but not an adjunct. But even assuming that such an interpretation of the verb as used in (41) was possible, we are still faced with the thorny issue posed by the partitive interpretation of the subject.

Section N: Obligatory Movement.

At this stage, the question that arises is the following: What happens if the inflections comprise anyone of the tense/aspect elements described above and if the resulting semantic interpretation of the subject is anomalous? Besides the strategy yielding the alternative sentences shown in (43)a and (43)b above, there is another strategy, which consists in moving the affected subject into the front position in the sequence.
Thus, in (44) the higher verb n-i-petraka can receive a [ + CONTROL ] interpretation so that the constituent t-any Antsirabe can be assumed to have been incorporated into the higher verb. The evidence for this comes from the circumstantial form of the passive on the higher verb, as in (45). However, it is not possible to keep the newly derived subject t-any Antsirabe in the external subject position, as demonstrated by the ungrammaticality of (46): The newly derived subject will have to be compatible with a partitive reading, as is usually the case for subjects accompanying a circumstantial passive form, as illustrated in sentence (39) above. To avoid such an anomalous interpretation, the newly derived subject will have to move out of the range of the projection of inflections, ie. Away from Spec and towards the front.

Conclusion

The main purpose of this paper was to use Kayne 1981 ‘s principle of Binary Branching in conjunction with the concept of Small Clause—in addition to basic principles proposed in Chomsky (1981, 1982 and 1986b)—to account for the distribution of null subjects in Malagasy. As a starting point, we assumed a kind of tree diagram such as in Randriamasimanana (1998: 304), where there is a distinction between the projection of a lexical item and that of a function word. Sections A, B and C show that there are advantages to a Binary Branching analysis of even lexical verbs, motion verbs and di-transitive verbs in Malagasy: In section A relative to lexical causatives, it was shown that such an analysis sheds some light on the semantic interpretation of sentences like (1), where the Malagasy verb ‘kill’ simply does not entail that the patient is dead and that to account for such a reading, we need an analysis of the lexical causative into a higher and a lower verb —this will presumably necessitate a revision of the nature of the projection of lexical verbs as represented on Figure 1 along lines sketched in Koizumi (1995); section B involving motion verbs highlights the crucial importance of the positive value for the atomic feature CONTROL associated with the higher verb in the process of incorporation of the lower structure into the higher verb, as in (4), whereas section C illustrates the superiority of an analysis based on the twin notions of Binary Branching and Small Clause as opposed to a
treatment of certain Malagasy particles as representing case-markings from the main verb, as proposed in (10).

Sections D and E distinguish between discourse context-based null subject phenomena, i.e. type 1 null subjects from linguistic utterance-based null subject phenomena, i.e. type 2 null subjects as tentatively summarized on Figure 2. This then paves the way for some understanding of the nature and function of Malagasy AGR in section F in terms of the singular or plural interpretation of the grammatical subject depending on what shows up under the inflections for voice, tense and aspect. The partial description of the distribution of type 2 null subjects described in sections I and K enables us to account for the distribution of empty subjects in terms of the nominal or non-nominal nature of the inflections for tense: This language has a series of parallel tense-markers for at least the future and for the past tense. The selection of the relevant series is dependent upon the verbal voice on the embedded structure as well as the selectional requirements of the higher verb since the latter will often dictate the voice to be utilized in the embedded position, as illustrated in sentence (31), for instance. This is of crucial importance in a language like Malagasy since as outlined in Randriamasi-manana (1999a: 37), it has a very high frequency for passive in texts. One reason why this is so has to with the notion of barriers as proposed in Chomsky (1986b) and illustrated for Malagasy in Randriamasimanana (2000: 274-276). Finally, one of the major consequences of the special relationship between the inflections head of the clause and its specifier is that if there is a clash between inherent features of the subject and those contained in the head inflections, then the subject will have to move out of the the domain range of the head, i.e. to the front within the sentence, as illustrated in (45). This kind of move fits in well within a minimalist framework such as the one proposed in Radford (1997).

The overall picture that emerges from the above is that all Malagasy verbs have to be analyzed as comprising a higher as well as a lower predicate, that the higher verb is compatible with a tense-marker, but that the lower predicate is only compatible with an aspectual marker. This suggests that the inflections projection in Malagasy should be exploded into separate projections of tense as well as aspect along lines sketched in Pollock (1989): Aspect as defined in Comrie (1976) will be the closest to the root or radical of the verb, followed by tense -- also as defined in Comrie (1985) -- which will have a projection of its own; sitting on top of the previous two projections, we will have an AGR (element) projection, whose presence requires an explicit grammatical subject in Malagasy. This contrasts with the behavior of tense-markers some of which may be nominal in nature whereas others are not nominal in nature, with direct consequences on the possibility of a null subject.

References


Randriamasimanana, Ch. 1998. Evidence for Null Subjects in Malagasy. Papers from


Word Order Variation and Topic Continuity in Atayal

Der-Hwa V. Rau
Providence University
E-mail: dhrau@pu.edu.tw

1. Introduction

Previous studies on word order variation have indicated that a correlation exists between the frequency of each word order type in a given language and the factors predicting the choice of one order or another on a particular occasion in that language. This correlation can be established by quantitative approaches to discourse analysis. However, the majority of the text counts have not considered possible skewing of the data, failed to distinguish the truly important factors from the minor ones, and did not consider the possibility of interactions among factors. In this study, a multivariate analysis is used to address these problems.

This study attempts to investigate how discourse functions account for the VS/SV word order variation in Atayal, an Austronesian language spoken in Taiwan, based on a variation approach to discourse analysis (Myhill 1992, Schiffrin 1994). GoldVarb 2.0 (Rand & Sankoff 1990) is used to analyze all the factors accounting for the word order alternation.

A pilot study (Rau 1997b) of word order variation in Atayal indicated topicworthiness was the only factor that could account for word order variation. However that analysis was based only on one narrative text with 204 tokens. Although one cannot put too much weight on the results of the pilot study due to the idiosyncratic nature of a single text, the methodology points to a promising direction for future study on word order variation. Thus the current study utilizes four times as much data collected from the Mstbon community, where Squilq Atayal is spoken. The results are parallel to those obtained for Ute (Givón 1983) and Chamorro (Coorman 1983) using a different method.

The organization of the paper is as follows. The brief introduction is followed by a literature review of previous discourse studies on word order variation and studies on the grammar and discourse of the Atayalic languages. After the methodology of the study is laid out and the analytical categories are defined and illustrated with examples from the texts, the results and a discussion of the factors accounting for word order alternation between VS and SV are presented.

1 For detailed information on the social stratification of the community, phonological variation, and sound change, please refer to Rau (2000).
2. Literature review

2.1. Word order variation

Previous discourse studies on referring terms have shown sequential dependencies across clauses (Ariel 1990, Fox 1987, Givón 1989, Tomlin 1987). The first mentions of a referent tend to be indefinite and explicit, while the second mentions tend to be definite and less explicit. The main functional factor underlying this distribution is information status and accessibility (Ariel 1990, Chafe 1987, Prince 1981).

Textual evidence from Chamorro (Cooreman 1983, 1992), Tagalog (Fox 1985), and Ute (Givón 1983), among others, suggest that word order variation can be explained in terms of referential continuity and thematic continuity. The unmarked VS order correlates highly with referential continuity and is more common in paragraph medial and final position, whereas the marked SV order is used to mark topic discontinuity and is more commonly found when the thematic unity of the paragraph is disrupted.

Myhill (1992:164-215) reviews the results of a number of studies of word order variation in individual languages, focusing on the order of subject, object and verb, and proposes the following two generalizations on sequencing. First, “temporal sequencing”, or an equivalent for Hopper’s (1979) “foregrounding”, correlates strongly with VS order in strongly VS languages. This finding echoes Hopper’s (1987:471) statement, “Verb-initial clauses narrate, noun-initial clauses describe.” Second, a general discourse principle in languages with a high proportion of VS order can be summarized with the following statement: “Putting the most important new information at the beginning of the clause.” His findings are consistent with Herring’s (1990) universal pragmatic word order pattern, which claims that marked or contrastive topics (or shifted topics) tend to go in initial position in all languages. Other studies of basically verb-initial languages (Payne 1995, Longacre 1995) also suggest that preverbal position is associated with arguments which are more or less “unexpected”.

Cumming (1997) offers an explanation for the association between the discourse functions of predicate-initial order and patient-prominent syntax in Western Austronesian languages. Specifically, parameters related to the semantics of events (including perfective aspect and telicity) and parameters related to the semantics and pragmatics of noun phrases (including definiteness and referentiality of the patient) are shown to correlate in several languages, corresponding to Hopper & Thompson’s (1980) construct of “high discourse transitivity”. In order to ensure the comparability of future studies, Cumming proposes two checklists listing factors of discourse functions and morphosyntactic features relevant to the analysis of constituent order alternations.

Quakenbush (1992) draws attention to the variation of word order by discourse type. He claims that the incontrovertible verb-initial status of Austronesian languages is
established by narrative-based discourse analysis. However, different text types were found to present differing “basic orders”, defined as “textual frequency” according to Hawkins’ criteria (1983:13). The basic verb-initial word order in Agutaynen narrative texts presents a mirror image of that in expository texts. In other words, the “subject-like nominal” (the ang NPs) precede the verb more often in expository texts. Longacre (1982:484) attributes this alternation to the topic-comment structure of expository discourse. In reaction to the narrative bias tradition of discourse analysis, Cumming (1997) also calls for future research to take up questions of constituency and argument structure in conversational discourse.

In summary, in strongly VS languages, the unmarked VS order is associated with topic continuity, whereas the marked SV order is associated with topic discontinuity. Two types of quantitative methodology have been proposed by Givón (1983) and Myhill (1992). The former measures degrees of topicality by counting the referential distance, persistence (decay), and potential interference/ambiguity of the third person referents in narratives. The latter, on the other hand, applies statistical tests to distinguish the truly important factors from the minor ones and considers the possibility of interactions among factors in word order variation.

2.2. Word order in the Atayalic languages

Previous studies on the grammatical structure of the Atayalic languages agree on the primarily verb-initial status of the language group but differ in terms of the definitions of S and O and “basic” word order. In the following review of these studies, the discussion begins by summarizing the generalizations on word order of different dialects and ends with a comparison of the two major views.

2.2.1. Squiliq Atayal

According to Egerod’s (1966) description, nouns usually follow verbs. However, if the sentence contains a preverb (including nonconjugatable quasi-verbs, verb particles, modal adverbs, and auxiliaries) with or without a main verb, the pronouns immediately follow the preverb. He did not take an explicit position on the issue of basic word order.

Chen & Lin (1985) propose that the general word order in Atayal is subject follows verb and object follows subject. Here the S and O refer to semantic subject and object.

Rau (1992) describes the word order in terms of the order of subject and predicate, an equivalent of Halliday’s (1985) theme and rhyme, or the topic-comment structure. The predicate normally comes before the subject in the unmarked cases. However when the subject precedes the predicate, special attention is drawn to the subject. The subject may be followed by the particle ga or a pause to form a topic separate from the rest of the
sentence. In other words, the often-used commentary construction in Atayal, as suggested by Egerod (1988), exhibits the form of ‘as for A, (there is) B’, where A is the theme, B the rheme. In a discourse analysis of Atayal narrative, Rau & Grimes (1994) claim that the basic order in Atayal is VSO, where S and O represent the typical semantic agent and patient of a transitive verb respectively. We suggest word order variation in Atayal may correlate with the referentiality and definiteness of the Os. In other words, non-referential indefinite Os generally co-occur with VOS order while referential definite Os co-occur with VSO order. Our findings support Hopper and Thompson’s “high discourse transitivity” as mentioned previously.

Huang (1993) claimed VOS is the basic order of Wulai Atayal. The S refers to a semantic agent rather than a grammatical subject. She noticed a correlation between Atayal word order and its voice system. VOS order was found to generally co-occur with verbs marked by -m- affixes whereas VSO order with verbs affixed with -un, -an, or s-.

However, in a later frequency count of m-clauses and non-m clauses, Huang (1994) observed that non-m clauses had a higher frequency in texts than m-clauses and claimed that non-m clauses are canonical transitive clauses both syntactically and semantically. This seems to indicate she has changed her view of the canonical word order from VOS to VSO since the latter is associated with the more frequent non-m clauses.

2.2.2. C’uli’ Atayal

Li’s (1995, 1996) studies of Mayrinax and Skikun syntax claim that verbs (or predicates) generally occur in sentence initial position and are followed by the subject and the object. His S and O refer to semantic subject and object.

In their accounts of Mayrinax syntax, Mei (1994), Chang (1995), and Huang (1995) all point to the same canonical word order in Mayrinax Atayal, i.e., VOS. The S and O are no longer semantic agent and patient respectively, but rather the grammatical subject and object.

2.2.3. Seediq

Both S. Huang and Su investigated the Nakahara dialect of Seediq (i.e., Tkdaya). Huang’s (1997) findings can be summarized into the following five points. First, the basic word order in Seediq is VOA. The definition of the valency roles follows Dixon (1979). A and O refer to the typical agent and patient of a transitive verb and S the single argument of an intransitive verb. Second, the preferred argument structure (PAS) in Seediq discourse appears to display what Du Bois (1987) termed one Lexical Argument Constraint and the Non-Lexical A Constraint. Third, Seediq displays ergative patterning in both grammatical and pragmatic dimensions of PAS. However, S=A links outnumber
S=O links in terms of topic continuity. Fourth, word order and focus in Seediq are interdependent and mutually predictive. The unmarked word order for Agent Focus (AF) is VOA, while that for Non-Agent Focus (NAF) is VAO. Finally, Seediq was not found to be what Cooreman, Fox and Givón (1984) termed “discourse ergative” since the measure of topic continuity did not exhibit any difference between AF and NAF.

Su (1997a) observes the word order in Seediq is VOS, or VSO when bound pronouns occur. The participant in focus is generally found in sentence-final position and the semantic roles of the participants are mostly determined by word order. Here the S refers to a grammatical subject. She (1997b) further argues that the pivots in Seediq display a grammaticization process changing the topic nominals into the subject-like ones.

Chang (1997) proposes that Seediq is a verb-initial (or predicate-initial) and subject-final language. But pronominal subjects must attach to the sentence-initial verbal elements. Again, his use of subject refers to a grammatical subject.

2.2.4. VSO, VOS or SVO?

The difference in the two proposed basic word orders, i.e., VOS vs. VSO in the Atayalic languages, on the surface seems to lie in how S and O are defined. However, when examined closely, the two views actually point to the same conclusion. If S and O refer to semantic roles of typical agent and patient in a transitive clause, the basic word order is VOS (or VOA) or VSO (or VAO). The former is the unmarked order for Agent Focus while the latter for Non-Agent Focus. Notwithstanding the word order variation, the focused element occurs in the sentence final position. If, on the other hand, S and O are defined as grammatical subject and object respectively, the canonical word order is VOS, where S is the pivot, the grammatical subject, or the focused element. In other words, the two interpretations do not really differ in predicting where the verb and the focused element occur. They both agree on the verb or predicate initial status of the language group and the sentence final position of the focused element.

Although the basic word order of Atayal has been well established, very little investigation has been conducted on the word order alternation between VS and SV, except for Rau’s preliminary study (1997). In fact, the preposed topics/subjects are a common feature in not only the Atayalic languages but in Formosan languages in general (Starosta 1988). Thus the current study aims to fill in the gap by investigating the discourse factors that have a significant effect on word order variation in Atayal. S and V here refer to the grammatical subject NPs (or the focused argument) and predicate respectively.
3. Methodology

3.1. Research questions and hypotheses

The hypothesis tested in this paper is that word order variation (pre-verbal vs. post-verbal subject) in Atayal can be accounted for by the following factors: the information status and topic continuity of the subject, the grammatical roles (i.e., agent, patient, beneficiary, etc.) of the subject and voice alternations, nominalization of the subject derived from dependent clauses, and the gender and age of the speakers.

First, the word order variation between pre-verbal and post-verbal subject in Atayal is hypothesized to correlate with the information status and topic continuity of the subject. In terms of the information status of the subject, we would predict post-verbal subject to be expected while the preverbal subject unexpected, based on the findings of Cooreman (1992) and Hopper (1987), as discussed in 2.1. Furthermore, following Givón’s hierarchy, as presented as follows in (1), the syntactic devices at the top involve topic continuity and less surprise, while those closer to the bottom involve less topic continuity and higher surprise.

(1) zero anaphora > unstressed /bound pronouns or grammatical agreement > stressed/independent pronouns > right dislocated NP’s > simple definite NP’s > left dislocated NP’s > indefinite NP’s > Y-movement > cleft/focus constructions

In other words, VS order is associated with previous mentions, paragraph medial and final, definiteness, referentiality, inexplicitness, more topicworthiness, foregrounding, and topic continuity, while the SV order is associated with first mentions, paragraph initial, indefiniteness, non-referentiality, explicitness, less topicworthiness, backgrounding, and topic discontinuity.

Second, the grammatical roles (i.e., agent, patient, beneficiary, etc.) of the subject and voice alternations (active, ergative, etc.) are correlated with word order variation. Previous studies indicate co-occurrences between VOS order and AF or active and VSO order with NAF or ergative, but no study has investigated the relationship between voice alternations and the pre-verbal S.

Third, nominalization of S derived from dependent clauses is tested for its relationship with word order variation because nominalization is sensitive to the demands of discourse, as pointed out by Hopper and Thompson (1980, 1984). Nominalization might be associated with a background rather than a foreground.

Finally, the sex and age factors are tested for preferences of word order.

3.2. Data

The data consist of 10 narrative texts of the Squilq dialect of Atayal, spoken in
Mstbon (Rui-yan in Chinese) community, in Ren'ai Village, Nantou County, Central Taiwan. The texts were tape-recorded and transcribed by Pastor Batu Temu, a native speaker of Atayal in his 30s, who was trained to do the interviews and transcribe the data. He was raised in the community and has established good relationships with the local residents. The distribution of the demographic background of the speakers is presented as follows in Table 1.

Table 1. Demographic background of the 10 speakers

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lawa Naway</td>
<td>F</td>
<td>82</td>
</tr>
<tr>
<td>Sayung Nawa</td>
<td>F</td>
<td>74</td>
</tr>
<tr>
<td>Temu Nakaw</td>
<td>M</td>
<td>73</td>
</tr>
<tr>
<td>Temu Syat</td>
<td>M</td>
<td>67</td>
</tr>
<tr>
<td>Temu Piling</td>
<td>M</td>
<td>66</td>
</tr>
<tr>
<td>Kumu Tusang</td>
<td>F</td>
<td>58</td>
</tr>
<tr>
<td>Walis Nabu</td>
<td>M</td>
<td>54</td>
</tr>
<tr>
<td>Abey Toci</td>
<td>F</td>
<td>45</td>
</tr>
<tr>
<td>Ciwas Nabu</td>
<td>F</td>
<td>39</td>
</tr>
<tr>
<td>Terow Sasan</td>
<td>M</td>
<td>35</td>
</tr>
</tbody>
</table>

The texts were divided into 22 paragraphs and 840 clauses. A paragraph boundary is determined by a change of plot, topic, location, or time. Following S. Huang (1997), the clause is defined as consisting of a verb (or predicate nominal or adjective) and its core argument NPs, but excluding oblique nominals.

All the preverbs (tense/aspect/mode markers, negators, adverbs) are coded as V, following Starosta’s (1988:555) AMV (auxiliaries as main verbs) analysis and Jeng’s (1997:213-281) classification of Bunun verbs.

Both the occurrence of S and zero anaphoras were counted. The position of zero anaphoras were coded as following the verb because all the bound pronouns occur after the verb.

Following Young & Bayley’s (1996:261) coding procedure for VARBRUL, the dependent variable is coded in factor group 1, then the co-occurring contextual features are coded in factor groups 2-15. Each factor within a factor group is assigned a unique one-character identifier in parentheses. Each factor group should be orthogonal or independent from all others and all the factors in each factor group are exhaustive (Guy 1993).

3.3. Coding

The following is a list of the analytical categories examined in this study. The
coding sheet is first presented, followed by definitions and examples from the texts to illustrate the coded categories.

3.3.1. Coding sheet

(1) Word order variation: VS (v), SV (s)
(2) First mentions: first-mentions (f), recent second mentions (s), resumptive (r), text-structure boundary (b).
(3) Position within paragraph: initial (i), medial (m), final (f).
(4) Definiteness: definite (d), indefinite (i).
(5) Referentiality: referential (r), non-referential (n).
(6) Explicitness: simple NP (i), modified NP (e), zero (z).
(7) Referentiality: first person pronoun (f), second person pronoun (s), third person pronoun (t), proper name (p), human NP (h), non-human animate NP (a), inanimate NP (i).
(8) Temporal sequencing: background (b), foreground (f).
(9) Topic particle in SV clauses: ga (g), none (n), non-applicable (/).
(10) Gender: male (m), female (f).
(11) Age: 30's (a), 40's (b), 50's (c), 60's (d), 70's (e), 80's (f).
(12) Nominalization of S derived from dependent clauses: yes (y), no (n).
(13) Grammatical roles of S: Agent (a), Patient (p), Beneficiary (b), Locative (l), Instrumental (i), Conveyance (c).
(14) Pronouns: free personal pronoun (f), bound personal pronoun (b), demonstrative pronoun (d), interrogative pronoun (i), zero (z), non-applicable (/).
(15) Voice alternations: active (a), antipassive (t), ergative (e).

3.3.2. Analytical categories: definitions and examples

(1) Word order variation

S can occur after the verb, such as “I” and “the fire” in (1a), or before the verb, such as “my father” and “my mother” in (1b).

---

2 Abbreviations:
1PGEX=first person plural genitive exclusive, 1PGIN=first person plural genitive inclusive, 1PG=first person genitive, 1SD=first person singular dative, 1SG=first person singular genitive exclusive, 1PNEX=first person plural nominative exclusive, 1SN=first person singular nominative, 1SNF=first person singular nominative free, 1SD=first person singular dative, 2SG=second person singular genitive, 3SD=third person singular dative, 3SG=third person singular genitive, 3SNF=third person singular nominative free, AF=agent focus, ASP=aspect, BF=benefactive focus, CF=conveyance focus, CON=conjunction, FUT=future, GEN=genitive marker, IF=instrument focus, IRR=irrealis, LF=locative focus, NEG=negation, NOM=nominative, PAR=particle, PERF=perfective, PF=patient focus, PN=personal name, TOP=topicalization.
(1a) m'abi saku ru ini muyen ktay puneq qasa ru,
AF-sleep 1SN CON NEG 1PGEX see-LF fire that CON "I fell asleep and did not see the fire."
(1b) yaba maku ga wan mhoqin cipoq saku na.
father 1SG TOP PERF AF-die small 1SN yet "My father died when I was still young."

(2) First mentions
S is coded as "first-mention" if it has not been mentioned previously in the discourse, "recent second mention" when S has been mentioned in the previous clause in the discourse, "resumptive" when S is being reintroduced into the discourse after a lapse, or "text-structure boundary" when S has been mentioned in the previous discourse, but a paragraph boundary intervenes since the prior mention.

In the following example (2a), first person singular pronoun "I" is illustrated in terms of its coding. The free nominative form kuzing is coded as "first mention" at the beginning of the narrative in (1). The short bound nominative ku is coded as "recent second mention" in (2). After a paragraph boundary, the long bound nominative form saku is coded as "text-structure boundary". After a lapse, the long bound nominative form saku is coded as "resumptive" in (9) when the S is reintroduced into the discourse. Finally, the short bound nominative ku is again coded as "recent second mention". The occurrences of short vs. long forms seem to follow the principle of iconicity. The short form encodes inexplicitness and recent second mention while the long form explicitness, first mentions, and reintroduction into the discourse.

(2a) (1) kuzing hiya ga lalu maku ga Temu Nakaw.
1SNF 3SNF TOP name 1SG TOP PN "I, my name is Temu Nakaw."
(2) yaba ku na Batu Bokusi.
Father 1SN GEN PN Pastor "I am Pastor Batu's father."
(3) ru kawas maku hiya ga pitu pgan ciwan.
CON age 1SG 3SNF TOP seven ten three "I am 73 years of age."
[Cluses 4-6, Paragraph boundary]
(7) nwah saku smi rusa nxan,
PERF.AF-go 1SN AF-put trap life "In the old days, I went to set traps."
(8) ru pitu bingi lga, say ta mlaw,
    CON seven day PAR go-LF.IRR 1PG AF-check
    "After seven days, we were going to check (it)."
(9) mha saku.
    Thus 1SN
    "I did that."
(10) mtuleq ku sasan,
    AF-rise 1SN morning
    "I rose in the morning."

(3) Position within paragraph
    S may occur in the paragraph initially, medially, or finally.

(4) Definiteness
    S may be definite or indefinite depending on whether it is identifiable to the hearer.
    In Atayal, nouns preceded by qutux "one" or without any determiners are coded "indefinite".

(5) Referentiality
    S may be referential or non-referential depending on whether it is identifiable to the speaker.

(6) Explicitness
    S may be coded as a simple NP, as qulih "fish", modified NP, as qulih raran "fish in the past", qulih misuw qani hiya "fish nowadays", or zero.

(7) Topicworthiness
    The term "topicworthiness" is equivalent to what typologists used to call "animacy".4
    Our categories correspond with Silverstein’s Animacy Hierarchy (1976), except that the kin term is grouped with the human NP instead of proper names.5 S may occur as first person pronoun, second person pronoun, third person pronoun, proper name, human NP, non-human animate NP, or inanimate NP.

---

3 Whaley (1997) suggests that "animacy" is a misnomer because it is only one of the parameters that are reflected in the so called "Animacy Hierarchy". The other parameters are sociocentric orientation, empathy, and definiteness.
4 The result of the pilot study indicated that proper names might be associated with SV order. Therefore, proper names were separated as a factor to be tested in this study.

220
(8) Temporal sequencing

S may occur in foregrounded or backgrounded clauses. Foregrounding refers to the thread or backbone of the discourse, which is ordered in temporal sequence and frequently associated with non-agent focus, whereas backgrounding refers to the clauses that merely assist, amplify, or comment on the speaker's goal and are not ordered with respect to each other, frequently associated with agent focus.

(9) Topic particle in SV clauses

S is marked with a special particle ga for topicalization, reserved for pre-predicate elements, as in (1b) and (2a-1), or none, as yaya nanak "only my mother" without the following topicalizer ga in (9a).

(9a) trang cipok myan ga⁵, uka qa yaba, yaya nanak maki.
when small 1SG PAR no that father mother only AF-exist
"From my childhood, I had no father. Only my mother was alive."

In SV order, S may be an L-dislocated NP or a Y-moved NP. In the L-dislocation construction, S precedes a complete clause, with a pronominal element cross-referencing S in the clause, as in (9b). In the Y-movement construction, on the other hand, S precedes a predicate, without a pronominal element cross-referencing S in the predicate, as in (9c).

(9b) maki qutux ryax,
AF-exist one day
sami yangu mu,
1PNEX younger brother's wife
nwah myan⁶ sbayux tmubux pagay.
1SG PERF.AF-go 1PGEX share AF-plant rice
"One day, my younger brother's wife and I, we went to share the work in rice planting."

(9c) yaqa yaya anay uku mpanga.
that female let-LF.IRR 1SG AF-carry
"That female (pig), let me carry."

(10) Gender

The 10 speakers were equally divided between males and females.

(11) Age

The 10 speakers were divided into six age groups: 30's, 40's, 50's, 60's, 70's, 80's.

⁵ This is a homophonous particle ga meaning "when, after", functioning as a conjunction for adverbial clause.
⁶ The S is treated the same way as A and differently as O in discourse.
(12) Nominalization of S derived from dependent clauses
S may occur in nominalized construction derived from dependent clauses, as in (12a).
(12a) cipok alay qu miyuk uzi qa qoqoq qga,
Small very NOM AF-enter also that spear that-PAR
"The stabbing of the spear was not deep either."

(13) Grammatical roles of S
The grammatical roles of S may be either Agent (13a), Patient (13b), Beneficiary (13c),
Locative (13d), Instrumental (13e), or Conveyance (13f).

(13a) ru ana ga ulung su gmalu utux kayan.
CON but PAR fortunate 2SG AF-bestow god heaven
"But fortunately the Heavenly God bestowed grace."
(13b) qngun ku nya ptehuk qalang.
support-PF 1SN 3SG FUT-arrive village
"He supported me to get home."
(13c) hazi ku nya sgalu uzi.
maybe 1SN 3SG BF-bestow also
"Maybe He bestowed grace to me."
(13d) biqan ku nya kinlokah.
give-LF 1SN 3SG strength
"He gave me strength."
(13e) ciwan sbiq maku hiyan.
three IF-give 1SG 3SD
"I gave him three (fish)."
(13f) ana ga cyux ku nya snaga kraya.
but PAR ASP 1SN 3SG CF-wait above
"But he was awaiting me above."

(14) Pronouns
Pronouns are coded as free personal pronoun, bound personal pronoun, demonstrative
pronoun, interrogative pronoun, or zero. The categorization of pronouns were
constructed to test Givón's scale of topic continuity (Givón, 1983a:17).
The system of personal pronouns in Atayal is presented in Table 2 (cf. Rau 1992, and
Huang 1993).
Table 2. Atayal Pronominal System

<table>
<thead>
<tr>
<th>Person/Number</th>
<th>Nominative</th>
<th>Genitive</th>
<th>Dative</th>
<th>Free</th>
</tr>
</thead>
<tbody>
<tr>
<td>1S</td>
<td>saku-ku</td>
<td>maku-mu-amu-aku-uku</td>
<td>knan</td>
<td>kuzing-kun</td>
</tr>
<tr>
<td>2S</td>
<td>su</td>
<td>su</td>
<td>sunan</td>
<td>isu</td>
</tr>
<tr>
<td>3S</td>
<td>hiya</td>
<td>nya</td>
<td>hiyan</td>
<td>hiya</td>
</tr>
<tr>
<td>1P (inclusive)</td>
<td>ta</td>
<td>ta</td>
<td>itan</td>
<td>ita</td>
</tr>
<tr>
<td>1P (exclusive)</td>
<td>sami</td>
<td>myan-muyan</td>
<td>sminan</td>
<td>sami</td>
</tr>
<tr>
<td>2P</td>
<td>simu</td>
<td>mamu</td>
<td>smunan</td>
<td>simu</td>
</tr>
<tr>
<td>3P</td>
<td>hga-lhga</td>
<td>nha</td>
<td>hgan</td>
<td>hga-lhga</td>
</tr>
<tr>
<td>1SG+2SN=</td>
<td>misu</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The following examples illustrate interrogative pronouns (14a), demonstrative pronouns (14b), and zero (14c).

(14a) son ta nanu la?
    go-PF 1PGIN what PAR
    "What are we going to do?"

(14b) qani ga ciriqan nanu yaba su.
    here TOP catch-LP what father 1SG
    "Here is what your father caught."

(14c) nyux mtakuy ru nyux mu sr'un ra.
    ASP AF-fall CON ASP 1SG hold down-PF PAR
    "After (it) fell, I was holding (it) down."

(15) Voice alternations: active (a), antipassive (t), ergative (e).

S may occur in active constructions, including one-argument intransitive clause (13a), existential sentence (15a), equational sentence (15b), and adjectival predicates (15c), antipassive constructions (15d), or ergative constructions, including patient focus (13b), benefactive focus (13c), locative focus (13d), instrumental focus (13e), and conveyance focus (13f).

(15a) ru yasa qa uka yaba mu la.
    CON therefore that no father 1SG PAR
    "Therefore, I have no father."

(15b) lalu maku ga Lawa Naway.
    Name 1SG TOP PN
    "My name is Lawa Naway."
Der-Hwa Victoria Rau

(15c) yaqih balay qnxan myan.
bad very life 1PGEX
"Our life was very difficult."

(15d) yaya maku qutux Imanga knan.
Mother 1SG one AF-raise 1SD
"My mother raised me as a single parent."

4. Results and discussion
4.1. An initial VARBRUL run

In this section, we simply present the results of the preliminary run. Detailed discussions on the final results are deferred to the next section on recodings and subsequent runs.

Our initial VARBRUL run shows the VS order (750/840, or 89%) is the unmarked word order. Ergative or non-agent focus (36%) occurs more frequently than anti-passive (13%). But the active construction (including one-argument agent focus, existential sentences, equational sentences, and adjectival predicates) is more frequent than either the ergative or the anti-passive construction, constituting 50% of all the voice alternations.

VS order seems to be associated with previous mentions, paragraph medial and final positions, definiteness, referentiality, inexplicitness, more topicworthiness, foregrounding, and topic continuity, if we compare all the percentages of the factors affecting VS order in each factor group. The factors with the highest percentages associated with VS order are recent second mentions (97%), paragraph medial position (90%), definite (91%), referential (91%), simple NPs (94%), no nominalization of S (89%), first person (96%), foregrounding (93%), and bound personal pronoun (100%).

On the other hand, SV order seems to be associated with first mentions, paragraph initial, indefiniteness, non-referentiality, explicitness, less topicworthiness, backgrounding, and topic discontinuity, if we compare all the percentages of the factors affecting SV order in each factor group. The factors with the highest percentages associated with SV order are text-structure boundary (42%), first mentions (35%), paragraph initial position (52%), indefinite (26%), non-referential (27%), modified NP (26%), nominalization of S (14%), inanimate NP (21%), backgrounding (17%), and free personal pronoun (59%).

In terms of social factors, male speakers (91%) seem to have a preference for VS order while female speakers (16%) seem to have a preference for SV order. The oldest group has a preference for SV order, but this number is based only on the speech of one female speaker in her 80's.

However, our quantitative analysis certainly does not stop here because a closer look reveals that only paragraph initial position and independence of personal pronoun indicate
slightly higher percentages of SV order than VS order. Besides, the knockout factors have to be recoded, and the independence of the factor groups and the information on goodness-of-fit have to be determined before the results can be interpreted. Thus the factors were recoded for subsequent VARBRUL runs.

4.2. Recoding

Several recodings and subsequent VARBRUL runs have brought us to recode our factor groups as follows:

(1) Word order variation: VS (v), SV (s).

(2) First mentions: first mentions (f), recent second mentions (s), resumptive (r), text-structure boundary (b).

(3) Explicitness: modified NP (m), simple NP and zero (s).

(4) Pronouns: zero and bound personal pronoun (b), free personal pronoun (f), demonstrative pronoun (d), interrogative pronoun (i), non-applicable (/).

4.3. One-level binomial analysis

The results of the one-level binomial analysis are shown in Tables 3.

<table>
<thead>
<tr>
<th>Group</th>
<th>Factor</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) First mentions</td>
<td>First mentions</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Recent second mentions</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Resumptive</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Text-structure boundary</td>
<td>0.11</td>
</tr>
<tr>
<td>(2) Explicitness</td>
<td>Modified NPs</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Simple NPs and zero</td>
<td>0.61</td>
</tr>
<tr>
<td>(3) Pronouns</td>
<td>Bound personal and zero</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>Free personal</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Demonstrative</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>Interrogative</td>
<td>0.59</td>
</tr>
</tbody>
</table>

The input probability, that is the likelihood that the post-verbal rule (VS order) will operate irrespective of conditioning factors, is 0.81. First mentions, explicitness and independence of the pronouns are the three factor groups that affect word order variation in Atayal. In terms of first mentions, recent second mentions (0.80) and resumptive (0.73) favor VS order, while first mentions (0.43) and text-structure boundary (0.11) favor SV order. As for explicitness, simple NPs and zero (0.61) favor VS order, whereas modified NPs (0.39) favor SV order. Finally, in terms of the independence of the pronouns, bound
personal pronouns and zero (0.97) and interrogative (0.59) favor VS order, while free personal pronouns (0.06) and demonstrative pronouns (0.26) favor SV order. The results indicate that information status of the S and topic continuity affect word order variation in Atayal.

The three factor groups and all the factors are independent because the error values are mostly below 1.5, the total Chi-square value (5.1041) is within the limit (6.490), and the Chi-square per cell figure (0.2686) is also very low. The result of the Chi-square test indicates goodness-of-fit and the least likelihood that interaction exist among factors.

4.4. Step-up/step-down binomial analysis

A step-up/step-down binomial analysis was conducted to test whether all the three factor groups in the last run contribute to the observed pattern of variation. The probability level is 0.05. The best stepping up and stepping down runs selected all three factor groups. In other words, first mentions, explicitness, and independence of pronouns have significant effect on word order variation in Atayal.

4.5. Discussion

Our hypothesis that word order variation is affected by the information status and topic continuity of the subject is supported. First mentions, explicitness, and independence of pronouns have significant effect on word order variation in Atayal. Recent second mentions, resumptives, simple NPs, bound personal pronouns, zeroes, and interrogative pronouns favor the VS order whereas text-structural boundary, first mentions, modified NPs, free pronouns, and demonstrative pronouns favor the SV order.

The fact that interrogative pronouns occur more frequently in VS order needs explanation. In WH-questions, the interrogative pronouns functioning as subject in action-oriented clauses do not change order, as illustrated in (4-1), (4-2).

(4-1) son ta nanu la?
    go-PF 1PG what PAR
    "We are going to do what?"
(4-2) kutan inu, mama?
    cut-LF where uncle
    "Cut where, uncle?"

The interrogative pronouns occur pre-verbally in cleft/focus construction or state-oriented clauses, such as (4-3).
Factors that were not found to have any significant effect on the choice of VS or SV include position within paragraph, definiteness, referentiality, topicworthiness, temporal sequencing, grammatical roles, voice alternations, nominalization of the subject, and social factors of gender and age. However, the division of paragraphs, the assignment of definiteness, and the decision on foregrounding and backgrounding are yet to be worked out in a more objective way.

5. Conclusion

The study confirms that word order variation is associated with the information status and topic continuity of the Subject. VS order is associated with old information and topic continuity while SV order is associated with new information and topic discontinuity. In particular, recent second mentions, resumptives, simple NPs, bound personal pronouns, zeroes, and interrogative pronouns favor the VS order whereas text-structural boundary, first mentions, modified NPs, free pronouns, and demonstrative pronouns favor the SV order.

By using a multivariate analysis, GoldVarb 2.0, to distinguish the truly important factors from the minor ones and remove interactions among factors in the design, this study addressed the problems that previous studies of word order variation based on text counts have generally failed to do.

References


AUSTRONESIAN FEATURES IN A LINGUISTIC AREA

Ger P. Reesink, Leiden University

Linguistic areas are, in general, defined by diffused traits shared by languages belonging to more than one, more or less well-defined (genetic) family (see, among many others, Campbell 1998). In the New Guinea area (including adjacent islands) a distinction is made between Austronesian (AN) and Non-Austronesian (NAN) languages. The latter are often referred to as Papuan, without any implication regarding their genetic homogeneity (Foley 1986), but sometimes lists of characteristic features are given (e.g. Wurm et al. 1975; Foley 1998).

The NAN languages of the Bird’s Head peninsula and Halmahera in the Indonesian provinces Papua (Irian Jaya) and Maluku do not show many of the typical Papuan traits (Reesink 1996, 1998). Are they perhaps Austronesian? Framing the question thus would lead us on a search for “platonic essences” which may not be available, when virtually all features are “negotiable and contestable” in an area with centuries-old contacts (Foley 1998:515-16). But, in order “to sort out more clearly the typological vs. genetic affiliations in the large Papuan-Austronesian Sprachbund” (Bradshaw 1998), it would be helpful to determine the origin of the shared features. I will try to determine the essential origin of a few features present in Biak and some other AN languages in the area, which are also found in (some of) what I call West Papuan languages. Notice that for the purpose of this paper I do not focus on the distinction between the AN families CMP and SHWNG, nor do I intend any commitment to the existence of a genetic grouping called West Papuan Phylum. For the time being, it is enough to assume a genetic relationship for what we call Austronesian, based on extensive cognate sets between the languages thus defined, while a possible overall genetic relationship between West Papuan languages remains to be proven. The term Papuan, then, I use rather loosely to refer to the NAN languages of the area.

This paper will discuss the spread of a few basically AN traits, of which two are mainly syntactic: (i) SVO word order; (ii) the N+G order for possessive constructions; (iii) one is morphosyntactic: the Inclusive-Exclusive opposition for first person plural; (iv) one is a morphophonological feature: Ca(C)-reduplication to form an Instrument-like noun from a verb (Blust 1998). Conversely, two semantico-syntactic features, which are arguably Non-Austronesian, will be considered: (v) morphological form and syntactic position of the negative, and (vi) the function of two verbal adjuncts, which change the meaning (and sometimes valency) of the main verb.

1. Constituent order

The West Papuan languages have for the most part an SVO order, which is quite unusual for Papuan languages. The only other group with this feature are the Torricelli languages. As Voorhoeve (1994:656) points out, this order is most likely due to the influence of neighboring AN languages. We assume that originally the West Papuan languages all had the “canonical” Papuan order SOV, because this is still present in the Northeast Halmahera languages. An independent spontaneous switch from SVO to SOV
order is not likely, as also Voorhoeve reasoned. Although there are many cases of AN languages that have adopted an SOV order and a distinction between medial and final verbs, such as Takia (Ross 1993), or even a switch-reference mechanism, as is claimed for Dami (Roberts 1997:192), there are no neighboring languages from which North Halmahera languages could have borrowed this order.

On the Bird’s Head, the southern Bird’s Head languages (claimed to belong to the Trans New Guinea Phylum) maintain the canonical Papuan SOV order, even though these languages exhibit yet another aberrant feature, subject and object prefixes on the verb. This is shared with Marind languages to the east and North Halmahera to the west.

Neither the SOV order, nor the object prefixes are very stable features in the Halmahera languages. The languages on the small islands west of Halmahera and the west coast of the main island have had the most contact with AN speakers for many centuries. This contact has apparently been instrumental in the erosion of object prefixes in Ternate, Tidore and West-Makian, while these languages, plus Sahu on the west coast have adopted the SVO order. Similarly, in the south Bird’s Head languages, the SOV order is leaking to an SVO order, as some examples of Inanwatan suggest (De Vries 1996:120-121).

Thus, we assume that West Papuan languages originally had the canonical Papuan SOV order, and that most of these languages have changed their basic order under influence from contact with AN languages. Whether the other languages of the Bird’s Head ever had object prefixes is doubtful, but possible, given the attested erosion in some of the North-Hamahera languages.

2. Possessive construction

The so-called ‘reversed genitive’ has long been discussed as a possible diagnostic for the nature of the languages in the area (Van der Veen 1915:92; Voorhoeve 1994:658; Discussion on the AN-LANG site, May 1998). I am not advocating that G+N order can be used to make any statement about the nature of a language. I do think, though, that in the linguistic area under discussion, the spread of possessive orders says something about language contact and diffusion in one direction or the other. It seems significant that a language like Biak has only postnominal possessors, both for inalienable and alienable possession. The possessive pronouns in Biak are rather complex (see also Van Hasselt 1905), indicating number and gender (animate-inanimate; only for plural) of the possessed, as in (1). Only body-part terms and some kinship terms seem to have remnants of AN possessive suffixes, as in (2). The $b$ in 2SG form is phonetic.

(1) $inokən$ $ai-di-ne$ 
   bag 1SG-POS-this
   ‘my bag’

   ($fno$ $ai-ja$)
   si.ch 1SG-POS
   ‘my sister’s child’

   $inokən$ $ai-su-ine$
   bag 1SG-DU-this
   my two bags

   ($fno$ $ai-e-su(-ya)$)
   si.ch 1SG-POS-DU-TOP
   my two nephews/nieces

   $inokən$ $ai-na-ne$
   bag 1SG-PL-this

   ($fno$ $ai-e-si(-ya)$)
   si.ch 1SG-POS-PL-TOP
   232
Austronesian features in a linguistic area

<table>
<thead>
<tr>
<th>English (my bags)</th>
<th>English (my nephews/nieces (my sister’s children))</th>
</tr>
</thead>
<tbody>
<tr>
<td>bru-ri head.3SG-ART</td>
<td>sn-a-ri mother.3SG-ART</td>
</tr>
<tr>
<td>my/his/her head</td>
<td>my/his/her mother</td>
</tr>
<tr>
<td>bru-m(b)-ri head-2SG-ART</td>
<td>sn-a-m(b)-ri mother-2SG-ART</td>
</tr>
<tr>
<td>your head</td>
<td>your mother</td>
</tr>
</tbody>
</table>

The suffix *ri* on singular body-part and kinship terms (body-parts which always come in pairs do not have *ri*, but *si*, as *mgasi* ‘my, his, her eye(s)’, *mga-m- si* ‘your eye(s)’), which I gloss as article, seems to have some connection with the possessive marker in other AN languages of the area. Waropen distinguishes factual possession and expected possession (Held 1942:46-46; 119-124). The first notion employs –*(r) i* : *ra-i* ‘1SG-POS’, *a-ri* ‘2SG-POS’, *(i)-ri* ‘3SG-POS’, etc., the second is expressed by –*na*: *ra-na* ‘1SG-POS’, *a-na* ‘2SG-POS’, *(i)-na* ‘3SG-POS’, etc., so that a subtle difference can be expressed, as in (3).

(3) a. Ra-wu ra-na bin-o
  1SG-take 1SG-POS woman-CLIT.
  I take a wife (the woman still has to become mine)

(3) b Ra-wu ra-i-bin-gha
  1SG-take 1SG-POS-woman-ART
  I took a wife (she has already become my wife) (Held 1942:121)

Just as Waropen, other (genetic) relatives Taba (Bowden 1998:271-282) and Wandamen contrast with Biak in that they have prenominal possessors, although in Wandamen (or Windesi) “[…] sometimes we also find the normal Austronesian order, e.g. *anio Maranipasiat* ‘the house of the Maranis’; *rawa-tawai-pai* ‘the snake’s skin’ (*rawa* = ‘skin’)" (Cowan 1955:47). This alternative AN order is not given in the Wandamen phrase book (Ramar et al. 1983:27):

(4) Yohan nie anio wai
  John his house a
  John’s house.

Compare this with postnominal clitics on inalienable nouns, and preposed clitics on alienable nouns in Bandanese (Collins & Kaartinen 1998:535) and other CMP languages.

Now, in the majority of the West Papuan languages the regular order is: possessor prefix on inalienables and pre-nominal possessor + possessive pronoun for alienables. In a language like Maybrat (5) the inalienable pattern is maintained, but the alienable possession is modeled after the post-nominal AN construction (not necessarily implying this comes directly from Biak):

(5) Tfo ro-Yan y-atia
  machete POS-Yan 3SG.M-father
In Hatam both pre- and post-nominal possessors are possible on alienable nouns, that is, both are attested in unelicited narrative material, (6) and (7). Inalienable possession is always expressed by means of a prefix, as in *di-cig* '1SG-father', *a-cig* '2SG-father', etc.

(6) *Munggwom ji-de-nya*
   child 2PL-POS-PL
   Your children.

(7) *Nyen-de andigpoi-nya*
   1PL-POS old.man-PL
   Our parents.

There are both so-called Papuan, as Hatam, and AN languages, as Mor (Laycock 1978:300), that allow both orders. These facts suggest that G+N is basically a Papuan feature, and that its presence in a language which has an overwhelming AN nature is due to diffusion. Conversely, N+G is basically an AN feature and its presence in a language which is Papuan equally suggests the result of language contact.

3. Inclusive-Exclusive opposition

The inclusive-exclusive opposition for first person plural (and dual) pronouns is a rather stable AN feature. It yields a global cline, showing complete absence in the far west (Europe, Africa) and nearly 100% presence in the far east (Australia). I am not sure which of Nichols' three interpretations (Nichols 1992:278) will turn out to be the correct one: (i) the lack originated in the far west and spread to the east, reached New Guinea but not Australia; (ii) the diversity tree: the Australia-New Guinea discontinuity goes back to second stage: the upper Paleolithic and Mesolithic time of circum-Pacific spread and differentiation; (iii) zero frequency in the far west and 100% frequency in Australia, both result from initial frequencies of 50%. (ii) and (iii) are mutually compatible (p. 279).

The lack of this opposition can be said to be 100% in New Guinea. The only Papuan languages which do employ the distinction are those adjacent to more recently arrived AN languages (about 4.000 BP). Conversely, there are a few AN languages in New Guinea which have discarded the opposition (Ross 1988:131), presumably due to Papuan contact, I think.

It would appear then, that in the area of the Moluccas and New Guinea (specifically the Bird’s Head), the I/E opposition is clearly of AN origin and it has dispersed to (West) Papuan languages through contact. All the West Papuan languages have a clear I/E opposition, except the three more centrally located languages in the Bird’s Head, Maybrat, Abun and Mpur. Hatam seems to make the distinction in verbal (and inalienable nominal) prefixation: the 3SG prefix is used for exclusive, the 3PL for inclusive, but no opposition is found in the free pronouns. In contrast, adjacent languages Meyah and Sough, as well as the western languages, such as Moi and Tehit, exhibit the opposition wherever it is feasible, including the very robust dual paradigm.
4. Ca(C)-reduplication

In his survey of Ca-reduplication in AN languages, Blust (1998:49) raises the question whether the CaC-template in Taba (Bowden 1998:91-94; 207) and similar facts in South Halmahera language are perhaps products of an independent history, and not related to the proposed Proto-Austronesian pattern of deriving instrumental nouns from verbs. Whatever the outcome of this question, it seems rather safe to claim that Ca(C)-reduplication is a strong diagnostic for Austronesian languages. Blust (1998:50) admits that the semantic function of Ca-reduplication is not always unequivocally instrumental. And Bowden (1998:94) recognizes a 'plurality of action' as a second function in Taba, for forms such as K-sang-sung um '1SG-RED-enter house' meaning 'I entered many houses'. Although a thorough analysis of similar forms in Biak is not yet possible, it seems that it behaves somewhat like its relative Taba. A number of CaC-CVC items can be found in Van Hasselt’s Numfor dictionary (1947), at least one of which can be interpreted as instrumental, kankun ‘fireplace’ from kun ‘burn, cook’. Others are not so clear: bisar ‘to be hungry’ and bas-bisar ‘famine’. Yet others fall more easily in the category of ‘plurality of action’ or agree with the general intensifying function of any reduplication in many languages: kenem ‘to live’ > kan-kenem ‘life’, or i-kan-kun ‘3SG-RED-cook’ from kun ‘cook, burn’, which my informant translated as ‘she is cooking’ (dia sedang masak). From Lex van der Leeden I have the information that the ubiquitous Bird’s Head word for ‘clothing’ sa!Suil (or [sdsun], [sasun] etc. depending on the individual language) originate in Ma’ya, the AN language of Salawati, where the verb sun means ‘to enter’ (also present in Biak for o-i-s-i-un ‘sun 3SG-goes in’ = ‘the sun is going down’).

From a Wandamen wordlist (Henning et al. 1991) I have extracted only one possible example, saso ‘bellows’ from so ‘to blow’, which is also found in Waropen (Held’s dictionary gives so ‘blow’ and saso-ri ‘bellows’).

A similar pattern is found in other AN languages of the region, like Ambai (Silzer 1983), where not only a, but also other vowels, e and i are used, and Keiese (Geurtjens 1921). The latter does have Ci(C)-, rather than CaC-, for intensification: sa ‘wrong’ > si-sa ‘very wrong’; waruk ‘sprinkle’ > wir-waruk ‘squander, waste’. Geurtjens (p. 47) claims that if the stem vowel is i, the reduplicated form contains a and follows: wil ‘move’ > wil-wal ‘move all the time’; but this seems rather unlikely.

Biak seems to use another template as well: C1C1V... > C1-a C2-a... , as in: frur ‘to make’ > f-ara-rur ‘work’ (as verb and as noun), and possibly f-aya-yer ‘dance’ from fyer ‘dancing women, similar to the scratching of chickens’ (Van Hasselt 1947:84-85; verb used in Luk15:25); mkak ‘to fear’ > m-aka-kak ‘fear’; pyar ‘float’ > p-aya-yar ‘anchoring place’ (Hasselt 1947:173). Ambai has a variant: fobera ‘to pull’ > fo-ba-bera ‘to keep pulling’; mirisin ‘to be happy’ > mi-ra-risin ‘to be very happy’ (Silzer 1983:58). These variants are probably due to some interaction with fossilized prefixes.

It seems likely that more focussed research would turn up more instances of Ca(C)-reduplications in the AN languages of the area, some with instrumental meaning, others with general durative intensification of the basic event.

No such reduplication template is present in any of the (West-)Papuan languages, with one possible exception. Tidore (Van Staden 2000) has a similar reduplication in which the vowel is either a copy of the stem vowel or o rather than the invariant a, noted by Blust. Possibly, Tidore has adopted this feature, as there are many other signs for heavy AN
influence, due to its and Ternate’s prolonged exposure to AN speakers, far more than is the case for other North Halmahera languages.

5. Form and position of negative adverb

The canonical place of the negative adverb is pre-predicate, both in SVO AN languages, like Tetun (Van Klinken 1999:228), Leti (Van Engelenhoven 1995:213), and in SOV Papuan languages, where the consequence of negation is often a total or partial reduction of person-number and tense categories of the verbal inflection, as in Sentani (Cowan 1965:22, Hartzler 1994).

The behavior of the negative in the West Papuan languages and the AN languages in this area is quite different. In all these languages the position of the negative adverb is rather strictly clause-final, or at least post-predicate in a language such as Moi, where it may move through the clause, if its scope needs to be narrowed. Even its actual phonemic form is quite wide-spread throughout the area. It is /fa/ in Mor (Laycock 1978:300), Wandamen, Biak and (ni)wa in North Halmahera languages (Van der Veen 1915:98), and wo(mo) in one dialect of Waropen, while the Napan dialect has te (Held 1942: 80). As just one example, consider the Biak sentence in (8), in which the first word no doubt contains the negative element as well: /fa-pe ‘not-?’ = ‘but’.

(8) /fa-pe w-ak-fluk neknek mkun oso fa but 2SG-with-give goat young one for
ya-fiuk-i d-ak-marisen kiker manibofay-e-si /fa
1SG-give-3SG 3SG-with-happy with friend 1SG-POS-PL not
But you have not given me a young goat that I (could) give it and have fun with my friends.

The West Papuan languages Mansim and related Hatam have a regular sound correspondence word-final –ar in Mansim and –ig in Hatam, attested in the negative adverbs bar and big, respectively. None of the other West Papuan languages have this form. West-Bird’s Head Moi dan might be related through metathesis and additional d-, but clearly Maybrat fe, Abun nde, Mpur jan, Meyah guru, Sough ero are not. Nor is AN Ambai kaka ‘not’ related; this seems more a reduplicated reflex of PAN *ta.

It is perhaps significant that Abun has a pre-predicate negative element as well, which is yo. It is tempting to relate these two items to the negative adverbs, divided over the two languages of Makian: nde = te in the Austronesian Taba and yo = yo in Papuan West-Makian (with its endonym Moi). Notice, that just as /fa/, the form te is found in both AN and West Papuan languages, the latter presumably a reflex of PAN *ta.

Which direction the diffusion of the basic form bar-fa-wa has taken is not easily determined. My hypothesis is that it is ‘originally’ West Papuan, and that Biak,

1 Laycock p. 289 mentions that /b/ definitely contrasts with /w/, but that these sounds are often difficult to distinguish, while [b] only seems to occur following a bilabial nasal. Similarly, in Biak /b, w, and b seem to contrast but in many contexts they fluctuate, as in the orthography of the Biak NT.
Austronesian features in a linguistic area

Wandamen, Mor have borrowed the form from Mansim, the original language along the coast around Manokwari, and/or Tidore.

In most Papuan languages the negative occurs in pre-predicate position on the last verb of a sentence. Some Papuan languages have a sentence-final negative with different semantic import from the canonical pre-predicate one. For example, Sentani has, in addition to to the verbal negator which is ə or homorganic vowel preceding the verb stem plus postclitic i on a verb stem devoid of any tense/person-number suffix, the form ban to negate non-verbal material. But this can operate on a clause with a fully inflected final verb, as in (9), which Hartzler (1994:59) characterizes as ‘psychological negation’: ‘but there is just this small thing’. I would interpret this sentence as conveying ‘it is not the case that I intend for you to do something for me’, parallel to the non-verbal negation of (10), meaning ‘It is not a true word’.

\[
(9) \text{Rabuhi-re mokonsele-re ban}
\]
\[
\text{something-to you:will:do:for:me-purpose not}
\]
\[
\text{I don’t intend for you to do anything for me [but there is just this one small thing...]}\]

\[
(10) \text{At hele ban}
\]
\[
\text{word true not}
\]
\[
\text{It’s not true, untrue word.}
\]

Perhaps, the clause-final negatives in Dani (and related languages) may suggest that it is an alternative strategy of Papuan languages. In these languages, the negative does not need to be sentence-finally, but it is always post-predicate, as in:

\[
(11) \text{At-en wam watik lek}
\]
\[
\text{3s-by pig hit.Iterative.participle not}
\]
\[
\text{He did not kill a/the pig. (Bromley, p.c.)}
\]

\[
(12) i... nykky lek lakeikhatek
\]
\[
\text{water consuming not they.normally.go}
\]
\[
\text{They travel without drinking. (Bromley 1981:250)}
\]

The strict clause-final position is an innovation of the languages of Maluku and the Bird’s Head. Strictly clause-final negative adverbs are typologically highly unusual (Horn 1989:447-462). It cannot easily be traced to either AN or Papuan languages outside this area. It seems safe to say that it is linked closer to Papuan than to Austronesian languages, just as the ‘reversed genitive’ is due to Papuan influence on the later arrived AN languages.

6. Verbal adjuncts

A rather striking feature of eastern Bird’s Head languages is the behavior of two adverbial adjuncts, which intensify the action expressed by the verb and/or increase its valence. The semantic function of these adjuncts is not easily captured. They appear to

\[\text{Payne (1985:226) notes that a few Chadic languages have solely a final negative marker in a similar configuration of S-V-O-Neg.}\]
follow immediately the main verb but in some cases a nominal object may intervene. What is of interest here is the striking similarity, not only in function, but also in some instances in actual form, of two Biak elements. So far, I am not aware of equivalents in other AN languages, and thus, until such evidence shows up, my hypothesis is that Biak has adopted such constructions from the languages of the eastern Bird's Head. Let me first illustrate the adjunct that means something like 'press' or 'hold onto with force', which may express a durative aspect in Hatam (13) and Sougb (14). The actual forms are:

<table>
<thead>
<tr>
<th>Language</th>
<th>Adjunct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meyah</td>
<td>keingg</td>
</tr>
<tr>
<td>Sougb</td>
<td>deb(-in)</td>
</tr>
<tr>
<td>Hatam</td>
<td>kep</td>
</tr>
<tr>
<td>Biak</td>
<td>epən</td>
</tr>
</tbody>
</table>

(13) Noni coi kep dimbou dini
s/he enter keep.onto door this
S/he always enters this door.

(14) Dan d-eigton deb
I 1SG-sit keep.onto
I'm sitting

The meaning of 'press, with force' is possible in conjunction with other verbs in Hatam and Sougb. It is illustrated by Meyah (15) and Biak (16). My informant gave yaref epen in (16) as the involuntary analog of the voluntary y-ores epen 'I-stand onto' which he explained as 'saya berdiri dan tindis' = 'I trample on'.

(15) Esin keingg anggur efek
poke onto grape juice
Smash/squeeze grape.

(16) Y-aref epən mangkoko kapu
1SG-step onto chicken shit
I stepped in chicken shit

The other adjunct is even more difficult to translate. It is invariably translated by the Malay term pele, which means something like 'block, screen, shield off', and which in conjunction with a position verb like 'stand' means 'guard' or 'protect', as Meyah Ot jong efsa 's/he guards his/her child'. The forms of this adjunct in the four languages are:

<table>
<thead>
<tr>
<th>Language</th>
<th>Adjunct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meyah</td>
<td>jong</td>
</tr>
<tr>
<td>Sougb</td>
<td>dongwo</td>
</tr>
<tr>
<td>Hatam</td>
<td>ser</td>
</tr>
<tr>
<td>Biak</td>
<td>user</td>
</tr>
</tbody>
</table>

3 For further data on Hatam, see Reesink (1999); on Meyah, see Gravelle (1998). Sougb and Meyrh are also presented in Reesink (to appear).
The Biak item is given in isolation as [user], but in context it gets reduced to [usr] or [us], the latter form being the one given for Numfor dialect in Van Hasselt (1915:20). In the Biak New Testament it occurs as si-ya-user kiker wa-ya ‘they followed with the boat’ (Yoh.21:8), while si-ya-wa by itself means ‘they follow, chase’, similar to Meyah osku mem ‘s/he trails a bird’ and osku joug mem ‘s/he hunts a bird’. All of these languages have something like ‘sit + adjunct’ for ‘to protect, guard’, with extended meaning ‘comforting’ the bereaved, as illustrated for Hatam (17) and Biak (18).

(17) Yoni i-gwam ser ni-ngon ti  
they 3PL-sit ‘keep.out’ 3SG-heart NOM  
They are comforting him.

(18) S-kain us ena mar-mar  
3PL-sit ‘protect’ 3PL RED-die  
They sit down with the corpse.

The same adjunct with the verb ‘to see’ conveys some ‘scrutiny’, as Biak (19), Hatam (20) and Soughb (21) show.

(19) Ya-mam us ro namane resari  
1SG-see ‘protect’ REL things first  
I will study these things first.

(20) Dani di-ngat ser srad  
I 1SG-see ‘block.off’ letter  
I check the letter.

(21) Dan d-ed-eiya dougwo ind-an srat  
I 1SG-go-see ‘block.off’ 1SG-POS letter  
I go check my letters

For Abun Berry and Berry (1999:26-28) described a very similar function for the grammaticalized variant of the benefactive marker wa. Compare (22) with the Hatam and Biak examples for the action of joining bereaved people in their sorrow, presumably by sitting with them in order to block off evil influences. Meyah lacks this possibility for joug (Gravelle, p.c.)

(22) Noru ne ye ke-wa Lamber nombrok  
night DET 3INDEF sit-TRS Lamber morning  
That night they guarded Lamber’s (body until) the morning.

My preliminary conclusion is that Biak has adopted a basically eastern Bird’s Head template for two adverbs which seem to have been grammaticalized to increase the valence and/or affect of a verb. In turn, the north-western Bird’s Head language Abun may have calqued this construction from Biak visitors along the north coast.
7. Conclusion

A great number of sources give evidence that there have been extensive interlinguistic contacts in the area of Maluku and Papua for many centuries, at least since the arrival of the Europeans in the 16th century, but most likely these contacts (and migrations) had existed a long time before that date.

It seems likely that some of the speakers of (West) Papuan languages had moved from the Bird’s Head westward, to Halmahera (Voorhoeve 1984; Wurm et al. 1975). Some 3500 years BP Austronesians arrived in the area, some of them directly to Maluku and dispersed there (the CMP languages), others to Halmahera and in and around the Cenderawasih Bay (SHWNG group), from which a further group split off to become the ancestors of the Oceanic subgroup of Austronesian languages (Ross 1995:85). It is not clear whether the AN speakers moved from Halmahera to the islands around the Bird’s Head or vice versa.

Given the distribution of the AN traits in the West Papuan languages, one might entertain the hypothesis that these languages are reflexified AN languages. This possibility is ruled out, I believe, because it cannot explain the clearly abundant AN vocabulary in all the AN languages of the area (pronoun sets, with reflexes of PAN Inclusive and Exclusive pronouns, number systems, basic vocabulary, reflexes of PAN derivational prefixes *pang- and *mang- for actives/causatives and intrasitives/processes) and its absence in the West Papuan languages.

In fact, the wildly divergent vocabularies of the West Papuan languages do not allow a fruitful investigation into their mutual genetic relationship, except for some local groups, like the Western Bird’s Head languages Moi, Tehit, Moraid and Seget, and the two groups in the eastern Bird’s Head: Mansim and Hatam are clear relatives and so are Meyah, Moskona and Sough. But to what extent these groups are related to each other and to the ‘isolates’ Abun, Maybrat and Mpur remains dubious. Indeed, the pronoun sets of the West Papuan languages do suggest an ancient relation but one would like to have some further evidence. More than a few tentative cognates have not as yet been identified.

In the present paper I have tried to identify at least some features which may be related to one or the other group. The main difficulty in this discussion is that the term AN can be used to refer to a genetic group, whereas the term (West) Papuan for the moment lacks such denotation. The SVO order can be ascribed to AN influence on the West Papuan languages, whereas the word orders of the possessive constructions in the various languages point to a mutual influence.

The so-called ‘reversed’ Genitive-Noun, as present in AN Taba, Wandamen, and the local variants of Malay can be traced to Papuan languages, while the AN order N+G is present as an alternative in some of the NAN languages of the Bird’s Head, presumably motivated by other AN languages such as Biak.

The I/E opposition has infiltrated all the West Papuan languages, with the exception of three (or four) languages located in the center of the Bird’s Head, from west to east: Abun, Maybrat, Mpur, and (marginally) Hatam.

Quite a different fate befell the AN reduplication strategy of Ca(C)- deriving instrument-like nouns or indicating some durative aspect. It has not really been able to affect the West Papuan languages. Perhaps the very restricted spread of this AN trait (only Tidore, possibly Ternate?) as opposed to the wide-spread diffusion of the I/E opposition and the syntactic orders of SVO and N+G is due to (i) a sociological factor, such as degree and time of contact, and (ii) a psychological factor such as extent to which a certain trait is
Austronesian features in a linguistic area

part of the *Inner Form* of a language: the more deeply embedded morphological features are less easily borrowed.

The other two features discussed suggest that the AN languages have undergone Papuan influences. The final negative, most likely of Papuan origin, is attested throughout the area, while the typical function of two verbal adjuncts seems to be more local: originating in the eastern Bird’s Head languages and spread to Biak (of course, the absence of evidence for this feature in other AN languages is not evidence of absence – it may well turn up in other AN languages as well; as it seems to have found its way into Abun and Maybrat, possibly via Biak).

The mixing of linguistic features can be ascribed to centuries-old intermingling of people. The longstanding supremacy of the Sultanates of Ternate and Tidore, consisting of West Papuan speakers, with their AN speaking vasals of the Raja Empat islands and Biak, was instrumental in the movements of people from around the Bird’s Head to other areas. The intruders had all kinds of trading contacts along the coasts of the Bird’s Head. Within the Bird’s Head many people changed locations, either through intermarriage or because they were traded as slaves. These migrants brought along their own morphosyntactic configurations, which easily spread through the region. At the same time, each linguistic community held a strong sense of identity, expressed in a great diversity of vocabularies.

In this paper I have presented some global indications of some traits, meant as a start for further detailed comparison of many other features, such as the pronominal and spatial deictic systems, the expression of TAM categories, and other bound-like morphological material. Together with an inventory of the possible cognate sets, however small they may turn out for the West Papuan languages, such a comparison might lead us to answers to the remaining questions.

Whether the West Papuan languages indeed form another family of languages, apart from the proposed Trans New Guinea Family, see Campbell (1998:166) for the suggestion to avoid terms other than Family, remains a matter of further research. That is, there are two questions: To what extent form the West Papuan languages a coherent (genetic) group? And secondly, is such a group ultimately related to the TNG Family? And, notwithstanding my earlier dismissal of the hypothesis that (some) West Papuan could be relexified AN languages, it should not be completely ruled out.

References


Ger Reesink


Irian Jaya: Universitas Cenderawasih.


to appear. *Languages of the Eastern Bird's Head*.

Austronesian features in a linguistic area


Nominalization in Rukai and Amis*

Li-May Sung
Graduate Institute of Linguistics
National Taiwan University
limay@ccms.ntu.edu.tw

1. Introduction

In this paper my primary purpose is to examine the nominalization process in two Formosan languages in Taiwan, Rukai (Budai) and Amis (Changpin, Central Amis). I will argue that in both languages, special nominalizing elements, similar to -ing in English, are employed to make a verbal element into a nominal one. The nominalized phrase as a whole behaves externally as a noun phrase because it occupies the typical NP position and it can take a nominal case marker, but internally it behaves as a VP because the verb may take a direct object, assigning accusative case to it.

2. A Sketch of Rukai and Amis Grammar

Before discussing the nominalization process in details, I will consider some general properties of Rukai and Amis, with particular reference to those that are relevant to the discussion in this paper. Both Rukai and Amis are predicate-initial languages and the basic word order is VOS or VSO. This is illustrated in (1) and (2).

(1) Nouns as Predicates

Rukai

a. [mangoDaDekai] [np ka saLabo]
   Rukai Nom SaLabo
   “Salabo is a Rukai.”

Amis

b. [nv wawa nira] [np kaku]
   Neu child his I
   “I am his child.”

(2) Verbs as Predicates (verb-initial)

Rukai

a. ma-dalame ki LaiLai ka saLabo

* I would like to thank Zeitoun Elizabeth and Shuan-fan Huang for their comments and suggestions. Thanks are also due to my Rukai informants SaLabo kaDesengane, LaiLai kaDesengane, Walialane kaDesengane and Amis informants ‘ofad kacaw, lakaw piyaw, panay kacaw, for providing the data on which this paper is based. If not otherwise mentioned, the data used in this paper come from my own field notes.

1 Rukai, according to Li (1973), includes six major dialects, Tanan, Budai, Labuan, Maga, Tona, and Mantauran.

2 Amis, according to Tsuchida (1982), includes five major dialects, Sakizaya, Northern Amis, Tavalong-Yataan, Central Amis and Southern Amis.

3 In this paper I employ the following abbreviations in glosses:

| AF | agent focus                   | Neg | negative marker | 1 | first person |
| PF | patient focus                 | Past| past tense      | 2 | second person|
| LF | locative focus                | Fut | future tense    | 3 | third person |
| IF | instrument focus              | NonFut | nonfuture tense |   |             |
| Nom| Nominative                    | Present | Present tense |   |             |
| Acc| Accusative                    | Perf| Perfective     |   |             |
| Gen| Genitive                      | Rel | relative marker|   |             |
| Obl| Oblique                       | Neu | neutral case marker | | |
| Fin| finite marker                 | NonFin | nonfinite marker | | |

245
"SaLabo likes LaiLai.”

"Panay likes Mayaw.”

Sentences in (1) are the so-called equational sentences in NP-(be)-NP pattern and the copular verb, be, is not overtly lexicalized on the surface. In Rukai and Amis, nominal case markers precede the case-marked noun phrase and in most cases they are obligatory. As Table 1 shows, in Rukai ko or ka can be used as nominative, accusative and even locative case markers. While two arguments in a sentence are both marked by the same case markers, word order would be crucial in distinguishing objects from subjects. That is, the word order would be strictly VOS.

Table 1: Nominal Case System in Rukai

<table>
<thead>
<tr>
<th>Nominative</th>
<th>Accusative</th>
<th>Locative</th>
<th>Genitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>ko</td>
<td>ko</td>
<td>ko</td>
<td></td>
</tr>
<tr>
<td>(-visible, +distance ± human)</td>
<td>(-visible, +distance - human)</td>
<td>(generic)</td>
<td></td>
</tr>
<tr>
<td>ka</td>
<td>ka</td>
<td>ka</td>
<td></td>
</tr>
<tr>
<td>(+visible, -distance ± human)</td>
<td>(+visible, -distance + human)</td>
<td>(place name)</td>
<td></td>
</tr>
<tr>
<td>ki</td>
<td>ki</td>
<td></td>
<td>(+specific, + human)</td>
</tr>
<tr>
<td>(+specific, +human)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 lists Amis nominal case marking system. Compared to Rukai, because of Amis’ rich case markings, the word order is relatively free in Amis.

Table 2: Case Markers in Amis (Huang 1995:226)

<table>
<thead>
<tr>
<th>nouns</th>
<th>numbers</th>
<th>Neutral</th>
<th>Nominative</th>
<th>Locative/ Accusative</th>
<th>Genitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common</td>
<td>u</td>
<td>ku</td>
<td>tu</td>
<td></td>
<td>nu</td>
</tr>
<tr>
<td>Proper</td>
<td>Singular</td>
<td>ci</td>
<td>ci</td>
<td>ci .. an</td>
<td>ni</td>
</tr>
<tr>
<td></td>
<td>Plural</td>
<td>ca</td>
<td>ca</td>
<td>ca .. an</td>
<td>na</td>
</tr>
</tbody>
</table>

Like most of the Formosan languages, pronouns in Rukai can be roughly divided into two sets, namely, bound pronouns and free pronouns, as shown in Table 3. Bound

---

4 In some Formosan languages, the noun predicate in the pseudo-cleft construction need to be preceded with a “neutral case marker” in languages like Amis as argued by Huang (1994, 1995), or preceded with a “noun classifier” in languages like Paiwan and Kavalan as argued by Tang, Chang and Ho (1998) and Chang, Tang and Ho (1998). Following the line of Chang et al.’s analysis, Liu (1999) re-examines the traditional case marking system in Amis and argues that the so-called “case markers” should be analyzed as morphological complexes, composed of a case marker and a noun classifier.

(i) Paiwan (Tang, Chang and Ho, 1998: p. 337)

*(ti) kai a aicu

TI Kai Nom this.

“This is Kai.”

(ii) Amis (Huang 1994)

ci ufad kura mi-namun-ay a tamdaw.

Neu Ufad that-Nom AF-drink-water-AY Lin man

“That man that is drinking the water is Ufad.”
Nominalization in Rukai and Amis

Pronouns are further divided into nominative and genitive forms. A bound pronoun, if represents the subject, has to be suffixed to the verb and form one unit, rather than remains sentence-finally.

Table 3: Pronominal System of Rukai (Budai) (adapted from Chen, 1999: p. 10)

<table>
<thead>
<tr>
<th>Person</th>
<th>Plurality</th>
<th>Visible/Inclusive</th>
<th>Free</th>
<th>Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Accusative</td>
<td>Nominative</td>
</tr>
<tr>
<td>1</td>
<td>Singular</td>
<td></td>
<td>nakoane</td>
<td>-(C)ako, -naw</td>
</tr>
<tr>
<td></td>
<td>Plural</td>
<td>+ Inclusive</td>
<td>mitaane</td>
<td>-ta</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Exclusive</td>
<td>navane</td>
<td>-nay</td>
</tr>
<tr>
<td>2</td>
<td>Singular</td>
<td></td>
<td>mosoane</td>
<td>-so</td>
</tr>
<tr>
<td></td>
<td>Plural</td>
<td></td>
<td>nomiane</td>
<td>-nomi</td>
</tr>
<tr>
<td>3</td>
<td>Singular</td>
<td>+ Visible</td>
<td>iniane</td>
<td>-ini</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Invisible</td>
<td>liniane</td>
<td>-lini</td>
</tr>
<tr>
<td></td>
<td>Plural</td>
<td>+ Visible</td>
<td>liniane</td>
<td>-lini</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Invisible</td>
<td>liniane</td>
<td>-lini</td>
</tr>
</tbody>
</table>

Table 4 summarizes Amis personal pronouns. There are four sets of personal pronouns, indicating nominative, accusative, genitive and possessive. All of them are free forms.

Table 4: Personal Pronouns in Amis (Liu, 1999: p.18)

<table>
<thead>
<tr>
<th></th>
<th>Nominative</th>
<th>Accusative/Locative</th>
<th>Genitive</th>
<th>Possessive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1S</td>
<td>kaku</td>
<td>takuwanan</td>
<td>aku</td>
<td>maku</td>
</tr>
<tr>
<td>2S</td>
<td>kisu</td>
<td>tisuwanan</td>
<td>isu</td>
<td>misu</td>
</tr>
<tr>
<td>3S</td>
<td>ciNra</td>
<td>ciNranan</td>
<td>nira</td>
<td>niNra</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ciNraan</td>
<td>niNra</td>
<td>niNra</td>
</tr>
<tr>
<td>1PI</td>
<td>kita</td>
<td>kitanan</td>
<td>ita</td>
<td>mita</td>
</tr>
<tr>
<td></td>
<td></td>
<td>titanan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1PE</td>
<td>kami</td>
<td>Tamiyanan</td>
<td>niyam</td>
<td>niyam</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2P</td>
<td>kamu</td>
<td>Tamuwanan</td>
<td>namu</td>
<td>namu</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tamuwanan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3P</td>
<td>caNra</td>
<td>CaNranan</td>
<td>naNra</td>
<td>naNra</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CaNraan</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As was discussed by Li (1973), Kuo (1979), Zeitoun (1997a, b), Zeitoun et al. (1996, 1997) and Chen (1999), Rukai, unlike most Formosan languages which have focus marking system, exhibits an active/passive dichotomy, similar to English. This is illustrated in (3-4). In an active construction such as (3), an agent is marked as the subject, while in a passive construction such as (4), the patient or the theme is marked as the subject. The prefixes w- and ki- indicate active and passive voices respectively. As argued in Zeitoun et al. (1996, 1997), Rukai exhibits a bipartite tense system in which future is distinguished from nonfuture, rather than a tripartite one in which past contrasts with present and future as in English.

---

5 As pointed out by Zeitoun et al. (1996, 1997), the past and present interpretation in Rukai usually depends on the occurrence of aspectual affixes (-nga, verbal reduplication) or temporal adjuncts.
(3) Active Construction in Rukai
w-a-Lomai ki LaiLai ka/ko saLabo.
Act-Fin-beat Ace LaiLai Nom SaLabo
"SaLabo beat LaiLai."

(4) Passive Construction in Rukai
ki-a-Lomai ki saLabo ko LaiLai.
Pass-Fin-beat Ace SaLabo Nom LaiLai
"LaiLai is beaten by SaLabo."

Amis, different from Rukai, employs so-called verbal "focus" markings to indicate the voice, as was discussed in Wu (1994, 1995), Huang (1994), Lin (1995) and Liu (1999) and many others. The traditional term "focus" here refers to the semantic relationship (agreement) established between a verb (the focus affix) and the subject NP, which is termed as theta-agreement by Mei (1994) and Holmer (1996). In other words, in Amis the verb has to agree with the theta role of the subject noun phrase, which could be an Agent, a Patient/Theme, a Location, a Beneficiary or an Instrument. This is illustrated in (5-8):

(5) Agent Focus in Amis
mi-la'op kura wawa ci panay-an. (Liu 1999: p. 19)
AF-chase that.Nom child Acc Panay
"The child is chasing Panay."

(6) Patient Focus in Amis
mi-futiq-an nura wawa kuni kafutiqan (Liu 1999: p. 21)
PF-sleep-PF that.Gen child this. Nom bed
"That child has slept on this bed."
(lit. "This bed has been slept by that child.")

(7) Locative Focus in Amis
pi-cirah-an isu tu kuwa ku kureN (Liu 1999: p. 23)
Pl-pickle-LF 2Sg.Gen Acc papaya Nom urn
"You pickled the papayas in the urn."
(lit. "The urn is the place where you picked the papayas.")

(8) Instrument Focus in Amis
sapi-tiwas ni arik ku acam tuya qedo. (Liu 1999: p. 25)
IF-hook Gen Arik Nom bamboo that.Acc. mouse
"Arik hooked that mouse by a bamboo stick."
(lit. "The bamboo stick is the tool that Arik used to hook that mouse.")

3. Nominalization
3.1 Rukai

I shall now turn to discuss the process of nominalization. I will first examine Rukai.

6 Comparing the passive of Rukai (i) to that of English in (ii), the agent phrase in the passive is not "suppressed" as an adjunct as it is in English argued by Bresnan (1982), Shibatani (1988), Baker et al (1989) and many others.

(i) ki-a-Lomai ki saLabo ko LaiLai.
Pass-Fin-beat Ace SaLabo Nom LaiLai
"LaiLai is beaten by SaLabo."

(ii) Mary was beaten by John.
3.1.1. Pseudo-cleft Construction
3.1.1.1. the nominalizer -∅

In Rukai, only some pseudo-clefts and complement clauses optionally undergo nominalization. Here I will first discuss the pseudo-clefts. The so-called pseudo-cleft sentences7 as shown in (9) employ the equational construction discussed earlier.

(9) Pseudo Cleft Construction in Rukai
a. [NP moni] [NP ko w-a-Lomai ki LaiLai]
   Moni Nom Act-Fin-beat Acc LaiLai
   “The one who beat LaiLai is Moni.”

b. [NP moni] [NP ko ki-a-Lomai ki LaiLai]
   Moni Nom Pass-Fin-beat Acc LaiLai.
   “The one who is beaten by LaiLai is Moni.”

In (9a), the focussed element Moni is a caseless noun predicate and the presupposed clause w-a-Lomai ki LaiLai is in subject position, preceded with a nominative case marker ko. Examining carefully, the presupposed clause in the pseudo-cleft construction is a complex NP that contains a null head. This is well-known as the headless relative clause. We indicate the head by the symbol e in the head position. The empty head and the gap in the relativized clause t are coindexed. Example (9a) can be partially represented as:

(10) [NP ko [CP [IP [INF inf-as V [VP [SPEC VP t, Nom Lomai ki LaiLai ] ]] [N N e ] ]]]

Consider another pseudo-cleft sentence (11), which is different from (9a).

(11) a. [NP LaiLai] [NP ko Lomai ki/*ka/*ko saLabo]
   LaiLai Nom NonFin-beat Gen/*Nom (Acc)/Nom (Acc)
   ‘The one who saLabo beat is LaiLai.’

b. [NP LaiLai] [NP ko Lomai-ini/-li/*-əako]
   ‘The one who he/l beat is LaiLai.’

The presupposed clause in the pseudo-cleft (11) contains a non-finite verb Lomai, followed either with a case-marked personal proper noun ki saLabo in (11a) or a genitive/possessive bound pronoun ini in (11b). In (11a), saLabo cannot take ka or ko and it can only take ki. If the agent is represented by a genitive bound pronoun as in (11b), the genitive has a subject-like role, different from the modifier-like interpretation in (12). That is, (11b) indicates an expression that the event of beating has occurred.

(12) a. laimai ki saLabo
   ‘SaLabo’ clothes’

b. laimai-ini
   ‘his clothes’

As is well known from gerundive constructions in English and infinitival NPs in Italian, examples shown in (13) have been argued to be predicates of event-like entities. Both cases have been argued to undergo nominalization of verbs. Gerundives in English and infinitival NPs in Italian are analyzed as nominal IPs, head by -ing and infinitival morphemes -ere/-are, which take VP as its complement.

---
7 Ya-yin Chang (1998) is the first one who claims that the so-called cleft in Tsou is in fact a pseudo-cleft. And this perhaps is true for all Formosan languages. Also see Chung-lian Chang (1996) and Yung-li Chang (1997) for the discussion of cleft constructions in Seediq and Kavalan.
There is, however, an important difference between Italian and English. While with English nouns in the of-phrase in (13b) may correspond to the object of the related verb, with Italian, nouns in the of-phrase in (13d-e) can never correspond to the object of the related verb, but only to the subject.

The structure of the examples of Rukai in (11) is pretty similar to that of infinitival NPs of Italian. The verb in (11) is non-finite and the noun in the presupposed clause can only correspond to the subject of the verb. Following this line, I propose that (11a) has the underlying structure (14), no matter whether one adopts the traditional NP hypothesis or Abney (1987) kind of DP hypothesis. Which of the NP hypothesis or DP hypothesis should be adopted is not my main purpose in this paper.

(11) (repeated)

a. [NP LaiLai]
   LaiLai Nom NonFin be at Gen/*NonFin- beat Gen/*Nom (Acc)/*Nom (Acc)
   'The one who SaLabo beat is LaiLai.'

In (14), the IP is embedded under the scope of the nominalizer -φ. Given the nominal nature of IP_{[+N]}, the verbal complex head [I_{[+N]} + V] after V^o moving to INFL cannot assign nominative case to the Spec of IP, where the agent phrase salabo was supposed to move to. Instead, genitive case marking applies, and we have the output (11a). Note that we have mentioned earlier the presupposed clause ko Lomai ki salabo is in fact a headless relative clause, in which the original object t_i is relativized. The infinitival phrase Lom.ai ki

In (14), the IP is embedded under the scope of the nominalizer -φ. Given the nominal nature of IP_{[+N]}, the verbal complex head [I_{[+N]} + V] after V^o moving to INFL cannot assign nominative case to the Spec of IP, where the agent phrase salabo was supposed to move to. Instead, genitive case marking applies, and we have the output (11a). Note that we have mentioned earlier the presupposed clause ko Lomai ki salabo is in fact a headless relative clause, in which the original object t_i is relativized. The infinitival phrase Lom.ai ki
**Nominalization in Rukai and Amis**

*saLabo* as a whole behaves externally as a noun phrase taking a case-marker *ko* because it occupies the typical NP position, but internally it behaves as a VP because the verb may take a direct object, assigning accusative case to it.

A similar process of nominalization occurs optionally in the existential construction with the verb *kaDua* 'does not exist' in (15).

(15) Existential Construction in Rukai

```
kaDua ko ka-dalame-li kai ki lamaliali
don't exist Nom NonFin-like-my these Acc women
```

'That I like these women does not exist.' (Lit: I don't like these women.)

Following the DP analysis of Abney (1987), Szabolcsi (1987, 1989), Stowell (1989, 1991) and Longobardi (1994), Siloni (1997) argues that DP is important not only as a functional projection of nominal expressions, but also its unique selection of non-tensed nominal complements, making D as the equivalent of C which must be associated with a tense operator. All the cases of English gerundives, Italian nominalized infinitives, French reduced relatives and Hebrew semi-relatives are all treated similarly in Siloni (1997), by assuming that these relevant verbal forms do not bear tense features. This line of analysis seems to account for the nominalized infinitives in Rukai. I will show, however, in the following section, that Siloni’s proposal is arguably too strong for the nominalization involving the other nominalizing morpheme *-ane* in Rukai.

### 3.1.1.2. the nominalizer *-ane*

In addition to the nominalizer *-ə*, Rukai has another nominalized pseudo-cleft employing the nominalizer *-ane*, as exemplified in (16). While *-ane* is used, in most cases, it is used for the theme or the location, but not the agent.

(16) bava ka ta-ongolo-ane-li

```
wineNom NonFut-drink-ane-1.Sg.Gen
```

'The one that I drank is the wine.'

In addition to the prefix *ta-* in (16) there are other prefixes *a-, Li-* as shown in (17-18), which are used together with the nominalizer *-ane*. All these prefixes, as discussed by Kuo (1979) and Chen (1999), indicate an event which happens at any time other than the future, or in the immediate future or in the distant future. Then Siloni’s claim of DP’s selection of non-tensed operator is too strong for the case of the nominalizer *-ane*. A similar case is also found in Amis. This will be discussed when we proceed to the Amis data in section 3.2.

(17) bava ka a-ongolo-ane-li

```
wineNom Fut-drink-ane-1.Sg.Gen
```

'The one that I will drink is the wine.'

(18) bava ka Li-ongolo-ane-li

```
wineNom Fut-drink-ane-1.Sg.Gen
```

'The one that I will drink is the wine.'

In a serial verb construction as shown in (19), the matrix verb (not the embedded verb), placed immediately under the scope of the NOM, moves up to nominal IP to take up the nominalizer *-ane*.

(19) ngoDaDeaDeka-ane ko a-paadagil-ane kifiafingale ki vavalake.

```
Rukai Nom Fut-begin learn Gen child
```

"The language that the child will begin to learn is Rukai."
3.1.2. Complement Clauses
3.1.2.1 the nominalizers -ø and -ane

In addition to the nominalized pseudo-cleft, some of the clausal complements in Rukai undergo the nominalization optionally. This is illustrated in (20). Verbs like waDeDeDele ‘watch’ in (20a) take a nominalized infinitive head by the nominalizer -ø morpheme. Other verbs like kyasease ‘grateful’, masamali ‘surprised’ can either take a finite clausal complement lead by a complementizer alaka ‘that’ as in (20b), or take a nominalized IP head by -ane as in (20c-d). Example (20a) has the underlying structure (21a) and example (20d) has (21a):

(20)

a. w-a-DeDeDele ka saLabo ko sina-sinaw-li ko laimai
   Act-Fin-watch SaLabo Acc Act-Fin-Red.wash Acc clothes
   “SaLabo is watching me washing the clothes.”

b. kyasease-nako alaka paralobo-so nakwan. (Kuo 1979: p. 56,)
   grateful-I helped-your me
   “I am grateful that you helped me.”

c. kyasease-nako ko ta-paralob-ane-so nakwan. (Kuo 1979: p. 56,)
   Helping-your
   “I am grateful for your helping me.”

d. masamali-ako ko/ka ta-Lomao-ane-(ini) ki saLabo ki LaiLai
   “I am surprised by SaLabo’s beating LaiLai.”

(21)

a.

```
V'  
  |  V  
  |  DP (or NP)  
  |  Nom  
  |  IP [+N]  
  |  Spec  
  |  I [+N]  
  |  VP  
  |  -li  
  |  V'  
      |  NP
      |  sina-sinaw ko/ki laimai
```
Let's examine the structure (21b). The whole complement clause ta-Lomai-ane-(ini) ki saLabo ki LaiLai is nominalized and is case-marked by ko/ka. Similar to the case in (14), the agent subject, saLabo, due to the nominal nature of IP, takes a genitive case marker ki, rather than a nominative case marker ka or ko. The verb, Lomai, assigns an accusative case to the direct object LaiLai. The agent subject, saLabo, has to precede the object, LaiLai, which is strictly VSO order.

3.1.3. Lexical Nominalization

In Rukai, a large of number of verbs, involving the syntactic nominalization of the morpheme -ane, has been so widely used that they are lexicalized and become nouns. Some are listed in (22). They no longer show any verbal properties. They are referential and can therefore appear in argument positions. They can also be modified by relative clauses. Thus they should be analyzed lexically as nouns and they do not involve a syntactically verbal projection.

(22)
a. bengeLaı “flower” babengeLaı-ane “garden”
b. ngoDaDekai “Rukai people” ngoDaDekaı-ane “Rukai language”
c. bolo “teach” ta-kiboLoboLo-ane “school”
d. sinaı “wash” ta-sina-sinav-ane (Chen 1999: p. 16)
“a place where people wash clothes”
e. mubanava “bathe” ta-mubana-banav-ane (Chen 1999: p. 16)
“a place where people take a bath”
f. Dipon “Japanese people” DiDipong-ane “Japanese language”

3.2. Amis
3.2.1. Relative Clauses and Pseudo-cleft Constructions
3.2.1.1. the nominalizers -ay and -an (-ə)

I shall now turn to Amis. As I have mentioned earlier, Amis is different from Rukai in that Amis employs verbal focus markings to indicate the voice. It is a four-way focus marking system in Amis, compared to the active/passive dichotomy in Rukai. With respect to the nominalization in Amis, as was discussed in Liu (1999), only relative clauses (including headless relative clauses) and pseudo-clefts undergo nominalization and it is obligatory. In addition, the relative clauses in Amis (and in most Formosan languages)
are strictly subject to the so-called subject-sensitivity, which is firstly claimed by Keenan (1976). That is, only the grammatical subjects of relative clauses can be relativized. For example, it is possible to relativize the subject wawa ‘child’ (the agent), but not the object panay (the patient) in the Agent Focus Construction (5). On the other hand, in the Patient Focus construction (6), it is possible to relativize the subject kafutiqan ‘bed’ (the theme), but not the object wawa ‘child’ (the agent).

(5) Agent Focus in Amis (repeated)
mi-la’op kura wawa ci panay-an. (Liu 1999: p. 19)
AF-chase that.Nom child Acc Panay
‘The child is chasing Panay.’

(6) Patient Focus in Amis (repeated)
mi-futiq-an nura wawa kuni kafutiqan (Liu 1999: p. 21)
PF-sleep-PF that.Gen child this. Nom bed
“That child has slept on this bed.”
(lit. “This bed has been slept by that child.”)

In Amis, any relativized construction obligatorily undergoes a process of nominalization, as argued by Lin (1995) and Liu (1999). This is exemplified in (23-25). Similar to the case in Rukai, internally the verbal clause still keeps the verbal property of assigning accusative case to the direct object, but after Y* moving to INFL, and then to the nominalization projection of -ay and -a/-an, the whole IP phrase behaves externally as a noun phrase. The underlying structure of example (25b) can be roughly represented as (26). The verbal focus markings would be decisive in which of the nominalizers the verb should take. The verb in the agent focus construction has to take the morpheme -ay, while the verb in the patient focus, locative focus and instrument focus constructions will take the morpheme -a/-an.

(23) Relative Clause in Amis
a. ma-’osi ku Nohah aku tura [[[ma-tawa’-ay
[AF-hate Nom boy:friend 1S.Gen. that-Ace AF-laugh-AY
takuwanan ] (a) [mi-kilim-an isu IP)] migutiNay NP] (Liu 1999: p. 71)
1Sg.Acc Link fisherman]
‘My boy friend hates the fisherman who is laughing at me.’

b. ma-olah kura fa’inayan tuya [[[pi-kalaN-an
AF-ask Nom-that girl Acc-that LF-fetch:crabs-LF 2
aku ____IP] (a) [mi-kilim-an isu IP)] riyar i NP] (Liu 1999: p. 70)
1Sg. Link seashore
“The girl likes the seashore where I fetched crabs.”

(24) Headless Relative Clause in Amis
a. ma-’osi ku Nohah aku tura [[[ma-tawa’-ay
[AF-hate Nom boy:friend 1S.Gen. that-Ace AF-laugh-AY
1Sg.Acc

‘My boy friend hates the one who is laughing at me.’

PF-bite Gen PF-seek-PF 2Sg. Nom-this pig
“The thing which was looked for by you bit this pig.” (Liu 1999: p. 71)

(25) Pseudo-cleft Construction in Amis
a. [creNP uni wawa], [np ku mi-lamlam-ay tu matu’away____]
Neu-this child Nom AF-mingle-AY Acc old:man
“It is this child that is mingling with the old men.” (Liu 1999: p. 99)

b. [creNP ura lutuk] [np ku pi-’eli-an ni rekar __]
Nom-that mountain Nom LF-weed-LF Gen Rekar
“The place where Rekar weeded is that mountain.” (Liu 1999: p. 104)
Nominalization in Rukai and Amis

(26) 
\[ [\text{NP} (\text{or DP}) \text{ uni wawa}] [\text{NP} (\text{or DP}) [\text{NP} \text{ tu matu'away}]]] \emptyset (\text{N head}) \]

Remember that I have argued earlier that the claim of DP’s selection of nontensed complement proposed by Siloni (1997) is too strong for the case of nominalization of -a\text{ne} in Rukai. Here the Amis data provide another piece of argument against Siloni’s proposal. As is widely discussed in the literature of Formosan languages, the verbal focus (voice) morphology also bears overtones of tense. Agent Focus in general suggests nonpast tense and Patient, Locative, or Instrument Focus suggests past tense. In the case of nominalization in Amis, the co-occurrence of verbal focus markings with -a\text{y} and -\text{an} suggests once again Siloni’s claim is too strong.

With respect to the nominalizing morpheme -\text{a}, one may suggest that it is -a\text{an} morpheme and since the -a\text{an} morpheme is the same as the patient or locative focus affixes -a\text{an}, two -a\text{ans} merge morphologically as one. Here I am not in a position arguing in favor of either analysis.

In Amis serial verb construction as in (27), only the matrix verb takes the nominalizing element. This is similar to the case in Rukai.

(27) serial verb construction in Amis
   ci mayaw ku ma-talaw-a\text{y} (a) mi-pacuk tu fafuy.
   Nom Mayaw Nom AF-afraid-AY kill Acc pig
   “The one who is afraid of killing pig is Mayaw.”

Look at the Amis verbal negator ca’ay in (28). ca’ay can be optionally pronounced as cai in any declarative construction. But in the pseudo-cleft such as (28), ca’ay but not cai can be used. Obviously, ca’ay, as a verbal negator, has taken the nominalizing element -a\text{y}.

(28) ca’ay vs. cai
   a. ci panay anuca ci mayaw ku mi-fanaw-a\text{y} tu kaisin.
      Nom Panay or Nom Mayaw Nom AF-wash Acc dishes
      “Is the one who washed the dishes Panay or Mayaw?”
   b. ci panay anuca ci mayaw ku ca’ay/’ca’ pi-fanaw tu kaisin.
      Nom Panay or Nom Mayaw Nom Neg PF-wash Acc dishes
      “Is the one who doesn’t wash the dishes Panay or Mayaw?”

3.2.1.2. Lexical Nominalization

Similar to the case in Rukai, a group of fixed elements in Amis, derived from certain verbs, constitutes an instance of lexical nominalization. These are lexically analyzed as nouns.

(29) Amis (Liu 1999, p. 51)
   a. \text{mi-fut\text{i}N} ‘to fish’; \text{mifu}tiNay ‘fisherman’
   b. \text{mi-tild} ‘to study’; \text{mitiliday} ‘student’
   c. \text{ma-sakero} ‘to dance’; \text{masakero}’ay ‘dancer’
   d. \text{ma-tayar} ‘to work’; \text{matayaray} ‘worker’
   e. \text{ma-liNad} ‘to till’; \text{maliNaday} ‘farmer’
   f. \text{r-um-adiw} ‘to sing’; \text{maradiway} ‘singer’

(30) Amis
   a. \text{mi-cudad} ‘study’ \text{pi-cudad-an} ‘school’
   b. \text{mi-holol} ‘chat’ \text{pi-holol-an} ‘place where people get together’
Li-May Sung

c. mi-ngingoy ‘bathe’
d. mi-tangtang ‘cook’
e. ma-futiq ‘sleep’

(ta-holol-an) ‘person who we chat with’
pi-ninguy-an ‘bathroom’
pi-tangtang-an ‘kitchen’
ka-futiq-an ‘bed’

4. Conclusion

To summarize, in Table 5, I provide a brief summary comparing the nominalization process in Rukai and Amis. First, since Li (1973), Rukai has been argued to differ drastically from Amis and other Formosan languages in that Rukai does not display a four-way focus system. Instead, its voice system is based on an active/passive dichotomy. Second, nominalization is very productive in Rukai and it takes place optionally in syntactic constructions such as pseudo-clefts and clausal complements. Compared to Rukai, the nominalization in Amis is strictly limited to the relativized constructions. And it is obligatory. Third, Rukai is not that different from Amis with respect to the nominalization process. The nominalizer, -ø, of Rukai, is used in the case that the noun inside the nominalized phrase corresponds to the agent subject of the verb, parallel to the nominalizer -ay in Amis. The nominalizer -ane of Rukai is used in the case that the noun inside the nominalized phrase corresponds to the location or the theme of the verb, parallel to the nominalizer -an in Amis. The claim here further supports the conclusion argued by Zeitoun (1999) that Rukai shares a number of identical morphosyntactic processes with the other Formosan languages.

Table 5: Characteristics of Nominalization

<table>
<thead>
<tr>
<th></th>
<th>Rukai</th>
<th>Amis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice system</td>
<td>Passive/active dichotomy</td>
<td>Focus system</td>
</tr>
<tr>
<td></td>
<td>a four-way focus system</td>
<td></td>
</tr>
<tr>
<td>Optional nominalization in pseudo-clefts</td>
<td>Obligatory nominalization in relative clauses and pseudo-clefts</td>
<td>--</td>
</tr>
<tr>
<td>Optional nominalization in complement clauses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two nominalizers</td>
<td></td>
<td>Two nominalizers</td>
</tr>
<tr>
<td>-ø</td>
<td>agent subject</td>
<td>-ay</td>
</tr>
<tr>
<td>-ane</td>
<td>location; theme; patient (very limited)</td>
<td>-an (-ø)</td>
</tr>
</tbody>
</table>

256
Nominalization in Rukai and Amis

References:


1. Introduction

Templatic effects are widely observed in many languages of the world. Such effects are usually attributed to various domains of morphology; in particular, such effects are strongly associated with reduplicative morphology. Templatic effects may also be a more general property of a language, in which case the language as a whole may be characterized as templatic. It is such cases that I wish to draw attention to in this paper, where I examine a particular type of templatic effect in the Austronesian language Mukah Melanau.

Mukah Melanau is a member of the Northwest Borneo group of Austronesian languages, and is spoken on the northern central coast of Sarawak. This language has been described in the most depth in work of Blust, who performed fieldwork several times on Mukah Melanau and whose descriptions and (historical) analyses appear in Blust (1988, 1997). The data I will consider here are taken from these works.

The data I focus on involve allomorphic alternations in the realization of the active and passive voice markers in Mukah. To briefly lay out the direction of the paper, in section 2 I provide the data, showing that the active and passive markers in the languages have drastically different allomorphy which can be quite easily described. This explanation is given in section 3, where I argue that the allomorphy is predictable based on relatively simple assumptions about featural specification. This move not only allows for a satisfying account of the allomorphy but also relates to the representation of the underlying forms of the morphemes in question. Section 4 provides the bulk of the analysis. I first analyze cases of prefixing allomorphy, after which I turn to cases involving simple ablauting allomorphy. Here I show that the ablauting allomorphy is an effect of fixed prosody, arising in cases where featural preservation is not at stake. In section 5, I turn to the cases Blust refers to as "compound ablaut", which show an unexpected behavior with respect to a subclass of forms. These data, I argue, are evidence for a particular markedness constraint that forces two segments that share place to coalesce into a single segment. Finally, section 6 offers a conclusion.

2. The Phenomenon

The data I focus on concern three forms of verbs in Mukah. The basic form is an unaffixed stem, which may be either a verb or a noun. As a verb, it is usually interpreted as imperative (Blust 1997). There are two morphologically complex forms that are considered here as well: the active and passive forms. These forms present an intriguing allomorphy, which is conditioned by the phonological shape of the base of affixation, and can be divided into two principal surface patterns. The first of these patterns is a prefixational allomorphy: when the first vowel of the stem is any vowel other than schwa ([ε]), we find the prefixed allomorph (mε- or nε- if the stem is consonant-initial; m- or n- if the stem is vowel-initial):

(1) Affixation to consonant-initial verbal bases with full vowel

<table>
<thead>
<tr>
<th>Unaffixed</th>
<th>Active</th>
<th>Passive</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. bilêm</td>
<td>məbilêm</td>
<td>nəbilêm</td>
<td>'blacken'</td>
</tr>
<tr>
<td>b. gutîn</td>
<td>məgutîn</td>
<td>nəgutîn</td>
<td>'cut with scissors'</td>
</tr>
</tbody>
</table>
(2) Affixation to vowel-initial bases with full vowel

<table>
<thead>
<tr>
<th>Unaffixed</th>
<th>Active</th>
<th>Passive</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. anjit</td>
<td>marjit</td>
<td>nanjit</td>
<td>‘anger’</td>
</tr>
<tr>
<td>b. ituŋ</td>
<td>mituŋ</td>
<td>ituŋ</td>
<td>‘count’</td>
</tr>
<tr>
<td>c. ulin</td>
<td>mulin</td>
<td>nulin</td>
<td>‘rudder’</td>
</tr>
</tbody>
</table>

The second allomorph occurs when the first vowel of the stem is schwa [ə]. In these cases, we find ablaut: the passive is signaled by *i*, and the active by *u*, in the first syllable:

(3) Affixation to verbal bases with schwa: u-ablaut (active) vs. i-ablaut (passive):

<table>
<thead>
<tr>
<th>Unaffixed</th>
<th>Active</th>
<th>Passive</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. gaga</td>
<td>guga</td>
<td>giga</td>
<td>‘chase away’</td>
</tr>
<tr>
<td>b. gagat</td>
<td>guqat</td>
<td>giqat</td>
<td>‘gnaw, moth’</td>
</tr>
<tr>
<td>c. kəkay</td>
<td>kukay</td>
<td>kikay</td>
<td>‘rake’</td>
</tr>
<tr>
<td>d. kəkut</td>
<td>kukut</td>
<td>kikut</td>
<td>‘excavate’</td>
</tr>
<tr>
<td>e. ləpək</td>
<td>lupək</td>
<td>lipək</td>
<td>‘fold’</td>
</tr>
<tr>
<td>f. ləpəw</td>
<td>lupəw</td>
<td>lipəw</td>
<td>‘pick’</td>
</tr>
</tbody>
</table>

The remaining data of interest illustrate what Blust (1997) refers to as compound ablaut: in these cases the active voice not only has the *u* associated with normal ablaut, but in addition the resulting verb form begins with *m*-.

Notice that all the stems to which compound ablaut applies begin with a voiced or voiceless labial plosive.
Fixed Prosodic Effects in Austronesian: An Optimality-Theoretic Account

(4) Affixation to labial-initial verbal bases with schwa: m- plus u-ablaut (active) vs. i-ablaut (passive)

<table>
<thead>
<tr>
<th>Unaffixed</th>
<th>Active</th>
<th>Passive</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. babah</td>
<td>mubah</td>
<td>bibah</td>
<td>'split (stative)'</td>
</tr>
<tr>
<td>b. babed</td>
<td>mubed</td>
<td>bibed</td>
<td>'tie'</td>
</tr>
<tr>
<td>c. bunu?</td>
<td>munu?</td>
<td>binu?</td>
<td>'kill'</td>
</tr>
<tr>
<td>d. papah</td>
<td>mupah</td>
<td>pipah</td>
<td>'hit, whip'</td>
</tr>
<tr>
<td>e. papak</td>
<td>mupak</td>
<td>pipak</td>
<td>'a whip'</td>
</tr>
</tbody>
</table>

3. Full vs. featureless vowels

A crucial step toward understanding the alternation between forms that exhibit prefixation of the active or passive morpheme, as opposed to those which exhibit ablaut, is to recognize that the forms undergoing ablaut all contain a schwa in their initial syllable. In any form that contains any other vowel in the initial syllable we find a prefixed allomorph. This correlation, I believe, provides strong support for several constraints that are operative in an optimality-theoretic account of these facts.

Another crucial point is the input to these processes. As far as the verbal stems are concerned, I assume that their input is identical to their unaffixed surface form. The issue of input for the prefixes, however, is complicated by the attested allomorphy. I posit an abstract underlying form for each prefix as follows:

(5) Active and Passive morphemes

<table>
<thead>
<tr>
<th>Active</th>
<th>Passive</th>
</tr>
</thead>
<tbody>
<tr>
<td>/mu-/</td>
<td>/ni-/</td>
</tr>
</tbody>
</table>

These morphemes are underlyingly more abstract in order to explain their attested surface forms. Consider first the forms where the active or passive morpheme is prefixed; that is, in cases where the stem contains a full vowel in the first syllable. In these cases, the vowel of the prefix always surfaces as schwa. Assuming that Mukah words have a trochaic foot aligned at the right edge of the word, and following Blust (1997), this is due to a constraint called Prepenultimate Neutralization, which reduces any unfooted vowel to schwa. Naturally, this constraint requires further phonetic motivation, but its effects are visible throughout the language: in the data at hand, no exceptions to this constraint exist. As such, it is considered here as an undominated constraint:

(6) **Prepenultimate Neutralization (PPN)**

Unfooted syllables do not license vowel place features.

Thus, when the prefix mu- is attached to a two-syllable stem, the vowel of the prefix is reduced to schwa. This results in a violation of a correspondence-theoretic constraint on featural identity (McCarthy & Prince 1995):

Although I do not deal with this issue here, this neutralization can be related to observations concerning prosodic prominence: stress in Mukah falls on the penultimate syllable, unless that syllable contains schwa, in which case stress is final (Blust 1988). Prepenultimate position is never prosodically prominent, so contrast in vowel quality is not maintained here.
The two constraints interact in such a manner that PPN must outrank IDENT, as illustrated in the following tableau:

<table>
<thead>
<tr>
<th>/muqutin/</th>
<th>PPN</th>
<th>IDENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. muqutin</td>
<td>!</td>
<td>*</td>
</tr>
<tr>
<td>b. maquti</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

This basic constraint interaction explains the cases of prefixational allomorphy in the Mukah active and passive verbal paradigms. From here, we now move on to the more complicated instances of ablaut, where I argue that such cases involve the effects of requirements on prosodic alignment resulting in fixed prosody.

4. Fixed prosody and its interaction with faithfulness

Let's now consider vowel-initial stems, whose affixed forms involved deletion of an affixal vowel. An example is the form ituQ, whose passive form is nitug. Note that such a case results in a bisyllabic output. This illustrates our first case of fixed prosody, whereby a derived form must conform to a certain output shape. In this case, the fixed output shape is two syllables, and is enforced through a constraint on maximal word size. This constraint is called SYLLABLEALIGNMENT:

(9) SYLLABLEALIGNMENT (σALIGN)

Every syllable must be aligned to some edge of the prosodic word.

σALIGN may be considered an extension of the notion Hierarchical Alignment, as formalized in earlier work by Ito, Kitagawa, & Mester (1996:242). Hierarchical Alignment is defined as follows:

(10) Hierarchical Alignment (Ito, Kitagawa, & Mester 1996:242)

Every prosodic constituent is aligned with some prosodic constituent, containing it.

∀Peat1 ∃Peat2 [Peat2 ⊃ Peat1 & Align (Peat1, Peat2)]

where Peat stands for a prosodic category.

The essential insight of Hierarchical Alignment is to disallow any structure that involves more than two instances of a prosodic category x contained within a prosodic category y. In their analysis of Japanese zuija-go, Ito, Kitagawa, & Mester assume that this constraint applies at the foot level, such that every foot must be aligned to some edge of a prosodic word. In this analysis of Mukah fixed prosody, I propose to extend this constraint such that it may apply between non-adjacent levels of prosodic structure. Thus, rather than aligning foot edges to word edges, σALIGN demands that syllable edges be aligned to word
edges. This approach is also taken in Ussishkin (in preparation) to account for similar fixed prosodic effects in Semitic verbal morphology.

\(\sigma\text{-ALIGN}\) assesses a violation for every candidate containing more than two syllables, since any syllable not at the edge of the word will not be aligned to a word edge. Crucially, examples of verb stems beginning with a vowel show that the constraint MAX must be dominated. This constraint, familiar from correspondence theory (McCarthy \& Prince 1995), appears below.

\[\text{(11) MAX}\]

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

The interaction between the two relevant constraints is illustrated here for the derivation of the passive form nulin ‘rudder’.

\[\text{(12) nulin ‘rudder, passive’}\]

\[
\begin{array}{|c|c|c|}
\hline
\text{ni-ulin} & \sigma\text{-ALIGN} & \text{MAX} \\
\hline
\text{a. noulin} & \ast! & \ast \\
\text{b. nulin} & \ast & \ast \\
\hline
\end{array}
\]

Candidate (a), which satisfies PPN (at the cost of violating IDENT), crucially violates the constraint \(\sigma\text{-ALIGN}\). Candidate (b) satisfies \(\sigma\text{-ALIGN}\), at the cost of violating lower-ranking MAX. One potential argument against this ranking logic could be formulated as follows: since candidate (a) violates IDENT, it could be claimed that this form shows only that IDENT outranks MAX, and that there is no evidence for the constraint \(\sigma\text{-ALIGN}\). However, clearer evidence for this constraint and its ranking with respect to MAX is available by looking at forms that exhibit ablaut. For instance, consider the passive form kikut. The input to this surface form is ni-kakut. The following tableau again illustrates the interaction between \(\sigma\text{-ALIGN}\) and MAX.

\[\text{(13) kikut ‘to be excavated’}\]

\[
\begin{array}{|c|c|c|}
\hline
\text{ni-i-kakut} & \sigma\text{-ALIGN} & \text{MAX} \\
\hline
\text{a. ni-i-kakut} & \ast! & \ast! \\
\text{b. ki-kut} & \ast! & \ast! \\
\hline
\end{array}
\]

I return to such forms shortly; for now, it suffices to state that they clearly show that the constraint on fixed prosody outranks the faithfulness constraint MAX.

At this point, some further elaboration on the faithfulness constraints involved in prefixing allomorphy is required. Let us look more closely at the example neguiti. We have already seen how the optimal form is chosen when the two relevant constraints are PPN and IDENT. The following tableau recapitulates this, and also shows how the candidates fare with respect to \(\sigma\text{-ALIGN}\) and MAX:

\[\text{(13) neguiti}\]

\[
\begin{array}{|c|c|c|}
\hline
\text{ni-i-kakut} & \sigma\text{-ALIGN} & \text{MAX} \\
\hline
\text{a. ni-i-kakut} & \ast! & \ast! \\
\text{b. ki-kut} & \ast! & \ast! \\
\hline
\end{array}
\]

\[\text{Crucial correspondence relations in the input and output representations appearing in the following tableaux are indicated with subscripted numerals.}\]
Given this tableau, we may now attempt to establish further rankings among the four constraints seen so far. Given that candidate (c) is chosen over candidate (b), the constraint MAX must dominate the constraint IDENT, since the optimal output violates IDENT, but not MAX. Candidate (b), on the other hand, violates MAX, and satisfies IDENT. Since, as shown above, σ-ALIGN must dominate MAX, by transitivity it must move higher in the ranking as well. Therefore, we rerank these constraints to give the following ranking:

(15) Ranking among constraints

<table>
<thead>
<tr>
<th></th>
<th>PPN</th>
<th>σ-ALIGN</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDENT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAX</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ-ALIGN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPN</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This result is troublesome, however, because it fails to account for a crucial fact: with the constraint σ-ALIGN now higher-ranking than IDENT, we incorrectly choose the ablauting candidate as optimal in all cases, even though that should result in prefixation:

(16) magutif 'to cut with scissors, active'

Candidate (a) can be ruled out immediately because of its fatal violation of PPN, a constraint which is never violated on the surface in this language. However, candidate (b), which we need to rule out, is incorrectly predicted by the ranking in this tableau, as indicated by the backward-pointing hand. What this shows is that some constraint must dominate σ-ALIGN in order to rule out candidate (b) as the optimal form in favor of the actual output, candidate (c). Determining the nature of this constraint is the next issue. I claim that the choice of constraint should reflect what we observe empirically: that in the case of a verbal stem with a full vowel in the initial syllable, we always find prefixing allomorphy. This is in contrast with ablauting allomorphy, which occurs in every verbal stem whose initial vowel is schwa. These observations point to an important insight: that the featural (specifically, the place-featural) specifications of stem vowels require high-ranking faithfulness. In particular, my proposal is that such data provide evidence for an output-output constraint that demands that every place feature in a verbal stem be preserved in a related form:
Fixed Prosodic Effects in Austronesian: An Optimality-Theoretic Account

(17) **OO-MAX-PL(ACE)**

A place feature in a verbal stem has a correspondent in a related form.

The force of this constraint is to crucially preserve a stem vowel if that vowel is a full vowel; the analysis rests on the critical assumption that the vowel schwa is unspecified for place features. This is why ablaut is found only in cases where a verbal stem has schwa in the initial syllable: deleting the schwa does not violate the constraint **OO-MAX-PL**. However, when the stem-initial vowel is a full vowel, replacing it, in effect, with the vowel of the prefix does violate this constraint. Note that this constraint is crucially of the output-output variety, following work of Benua (1995, 1997): featural specifications of vowels in related output forms is at issue. Thus we do not always find the prefixal vowel surfacing faithfully; as observed above, in prefixing allomorphy this vowel is always neutralized, and only shows up in cases of ablaut. This type of “**MAX-F(EATURE)**” constraint has precedence in earlier work, e.g., Lombardi 1995, 1998, Causely 1996, Walker 1997; cf. Lamontagne & Rice 1995 on coalescence and feature parsing.

Recall that the motivation behind ablaut is to conform to the fixed prosodic constraint **σ ALIGN**. It must be the case, as discussed above, that some constraint dominate σ-ALIGN in order to prevent ablaut from occurring when the stem-initial vowel is not schwa. **OO-MAX-PL** serves this function, as the following tableau illustrates:

(18) _mægutig_ ‘to cut with scissors, active’

<table>
<thead>
<tr>
<th>/ni-,gu-tin/</th>
<th>PPN</th>
<th>OO-MAX-PL</th>
<th>σ-ALIGN</th>
<th>IDENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ni-qi-tin</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. gi-tin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. na~:gu-tin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because the ablaut candidate (b) violates **OO-MAX-PL**, it may not surface. However, consider the case of a stem whose initial vowel is schwa:

(19) _kikut_ ‘to be excavated, passive’

<table>
<thead>
<tr>
<th>/nij,kə-kut/</th>
<th>OO-MAX-PL</th>
<th>σ-ALIGN</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. nij,kə-kut</td>
<td></td>
<td>*!</td>
<td>**</td>
</tr>
<tr>
<td>b. ki-kut</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The winning candidate here has no violations of **OO-MAX-PL**: this is because although the first vowel of the stem has no correspondent in the optimal output, the unparsed vowel is schwa, which lacks place features (Jakobson 1938, Anderson 1982, Browman & Goldstein 1992). Therefore **OO-MAX-PL** is vacuously satisfied by such a candidate. **OO-MAX-PL** plays no role in determining the outcome in such a case; as the tableau shows, the competition is therefore passed down to the constraint **σ ALIGN**, which favors the bisyllabic output.

We have so far successfully accounted for the main split in the allomorphy exhibited in the active and passive verbal paradigms of Mukah. As we have seen, fixed prosody is emergent; that is, it occurs only in case it does not violate higher-ranking faithfulness constraints. Although the language has a strong desire for words to conform to a maximally bisyllabic size, this is only possible if such a prosodic shape does not involve the deletion of vowel-place features from the verbal stem. In the case where the relevant stem vowel is schwa, there are no vowel-place features to preserve, in which case fixed prosody dictates that a bisyllabic output form is optimal.

265
5. Compound ablaut as coalescence

An interesting portion of the data remain to be captured under this analysis, however. This portion involves what is termed by Blust (1997) compound ablaut, which involves a further alternation in the active verbal paradigm of some forms. In addition to the expected u ablaut in these forms, they also unexpectedly contain an initial m. The relevant data are repeated here for convenience.

(20) Compound Ablaut in Active Verbal Paradigm

<table>
<thead>
<tr>
<th>Unaffixed</th>
<th>Active</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. babah</td>
<td>mubah</td>
<td>'split (stative)'</td>
</tr>
<tr>
<td>b. babad</td>
<td>mubad</td>
<td>'tie'</td>
</tr>
<tr>
<td>c. banu?</td>
<td>munu?</td>
<td>'kill'</td>
</tr>
<tr>
<td>d. papah</td>
<td>mupah</td>
<td>'hit, whip'</td>
</tr>
<tr>
<td>e. ppepek</td>
<td>muppek</td>
<td>'a whip'</td>
</tr>
</tbody>
</table>

A crucial observation here, due to Blust (1997), is that all of the forms to which compound ablaut applies contain a labial plosive in initial position. Given the account so far, these forms are predicted to surface as normal ablauting forms since their verbal stems contain schwa in the initial syllable, resulting in, for instance, *bubah, *bubad, *bunu?, etc. However, such surface forms are routinely avoided in favor of outputs which have initial m, rather than initial b or p.

My proposal is that such forms involve a coalescence of two segments; specifically, the prefix-initial m and the stem-initial labial plosive in each case. I will show below that this coalescence, which violates the faithfulness constraint UNIFORMITY arises in order to satisfy a higher-ranking faithfulness constraint. Before elaborating on this point, however, let us focus on the phonetic restriction involved in cases of compound ablaut. Clearly the sequence mu favored over the sequences pu or bu. As discussed by Blust (1997), in morphologically complex forms the sequence mu is widespread, arising from prefixation. However, in morphologically complex forms, one never finds pu or bu. This provides evidence for an OCP-type constraint against sequences of a labial obstruent followed by a round (labial) vowel. This constraint is called *BU.

(21) *BU

The sequence of a labial obstruent followed by a labial vowel is prohibited.³

³ It might be advantageous to view this constraint as a local self-conjunction of the markedness constraint *LAB, with the local domain specified as either the syllable or the prosodic word. As pointed out by Kazutaka Kurisu (p.c.), this move could provide an explanation for the relative ranking between the conjoined constraints *LAB and *COR, following Spaelti's (1997) Universal Conjoined Constraint Ranking Hypothesis (UCCRH). According to Spaelti, if the ranking Cₐ ≻ Cₖ holds, then so must the ranking Cₐ ≻ Cₖ. Given the universal markedness harmony scale *LAB ≻ *COR, the ranking *LAB ≻ *COR follows by the UCCRH, and need not be stipulated. For an account of OCP effects analyzed as self-conjunction of featural markedness constraints, see Aldrete (1997). Ito & Mester (1998) provide a conjunction-based analysis in their account of Japanese sequential voicing effects. Crucially, for the Mukah case at hand, other constraints may be ranked in between these two self-conjoined constraints. However, this matter is not as simple as generalizing the *BU constraint to a self-conjunction barring adjacent labial segments. Doing so turns out to make the incorrect prediction that compound ablaut should occur when the passive morpheme ni- is prefixed to a coronal-initial base, a situation which is never found. Thus I adopt the more specific *BU constraint.
As an aside, such cases provide further evidence for the ranking between *σ*-ALIGN and MAX, since the fixed prosodic-conforming output surfaces. Relevant correspondence relations are indicated with subscripted numerals.

(22)  **Fixed prosody prevails**

<table>
<thead>
<tr>
<th></th>
<th>σ-ALIGN</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>/m₁u₂-b₃a₁b₄ah/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. m₁:Eb₃a₁b₄ah</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>b. m₁₁₃u₂b₄ah</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Returning now to the issue of coalescence, we note that the fixed prosodic constraint must dominate UNIFORMITY.⁴

(23)  **UNIFORMITY** (McCarthy & Prince 1995)

No element in the output has multiple correspondents in the input.

With UNIFORMITY dominated by *σ*-ALIGN, coalescence takes place in order to meet fixed prosodic requirements.

(24)  **Fixed prosody prevails yet again**

<table>
<thead>
<tr>
<th></th>
<th><em>σ</em>-ALIGN</th>
<th>UNIFORMITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>/m₁u₂-b₃a₁b₄ah/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. m₁:Eb₃a₁b₄ah</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>b. m₁₁₃u₂b₄ah</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

In the optimal candidate (b), the stem-initial $b₃$ has coalesced with the prefix-initial $m₁$ to yield $m_{1,3}$. Notice that this merger of the two segments is contingent on their sharing place features: they are both labial. Thus, the constraint OO-MAX-PL is satisfied in the optimal candidate, since the labial place features of both the prefix-initial $m$ and the base-initial $b$ are preserved. This will prevent compound ablaut from taking place with no restrictions: it is limited to strictly those cases in which the verbal stem happens to begin with a consonant of the same place of articulation as the active voice prefix. There is, however, another important candidate to consider: one in which the prefix-initial $m$ and the base-initial $b$ coalesce into a $b$. This candidate satisfies OO-MAX-PL, yet such a candidate does not surface (*bubah). However, this candidate is ruled out by the markedness constraint *BU.

---

⁴ See Pater (1999) for an account of coalescence in Austronesian that involves the constraint LINEARITY:

**LINEARITY** (McCarthy & Prince 1995)

The input is consistent with the precedence structure of the output, and vice versa.

Although I do not address this issue further, it is not clear that coalescence violates LINEARITY, since under McCarthy & Prince’s definition LINEARITY is violated only when precedence relations are reversed. Whether coalescence involves a reversal of precedence relations seems unlikely. What seems more plausible is that coalescence results in a loss of precedence relations. For this reason, I adopt UNIFORMITY as the constraint violated by coalescence.
Non-coalescence vs. coalescence

<table>
<thead>
<tr>
<th>/m₁u·b·a·b·ah/</th>
<th>OO-MAX-PL</th>
<th>*BU</th>
<th>UNIFORMITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. b₁u·b·ah</td>
<td>*!</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b. b₁u·b·ah</td>
<td></td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>c. m₁u·b·ah</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. m₁u·b·ah</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Candidate (a) violates OO-MAX-PL, since the [labial] feature of the stem-initial b₁ has no correspondent. Candidate (b) satisfies OO-MAX-PL, yet violates *BU. Candidate (c) is like candidate (a) with respect to its fatal violation of OO-MAX-PL but is phonetically identical to the optimal candidate, which violates only relatively low-ranking UNIFORMITY. Compound ablaut is thus analyzed as a way to both avoid the sequence bu or pu by being faithful to the features of the stem-initial consonant.

6. Conclusion

The following ranking diagram summarizes the analysis presented here:

```
(26) Final Ranking

PPN
\[ \rightarrow \]
OO-MAX-PL
\[ \rightarrow \]
*BU
\[ \rightarrow \]
σALIGN
\[ \rightarrow \]
UNIFORMITY
\[ \rightarrow \]
IDENT MAX
```

We have examined a case of fixed prosody in the Austronesian language Mukah Melanau. As we have seen, this language exhibits an interesting allomorphy in its active and passive verbal affixation. The two main allomorphs we have been concerned with involved prefixation on the one hand, and ablaut on the other. A subset of ablauting forms presented an additional puzzle: labial-initial bases show compound ablaut in the active paradigm.

Our analysis made use of several well-motivated faithfulness constraints. In particular, high-ranking OO-MAX-PL forces preservation of featural specifications of verbal stems, which is ultimately responsible for restricting the effects of fixed prosody to cases of stems whose initial vowel is schwa. In such cases, the fixed prosodic constraint σALIGN takes effect, limiting words to two syllables.

The fixed prosodic effects we have observed in Mukah are widespread within a particular morphological domain: that of active and passive affixation in the verbal paradigm. However, such fixed prosody is not observed with other affixational material in the language, at least, not according to the available data. Rather than viewing this as a weakness of the analysis presented here, however, I claim that this scenario in a consequence of the underlying forms of affixes in general in the language. Given the data in Blust (1988), all affixes in this language are prefixes (and in some cases, infixes). However, this is not the sole generalization that appears to hold on affixation.

It is also the case in Mukah that affixes contain at most one syllable. The typology of affixal segmentism is represented below:

---

5 One example containing a bisyllabic prefix *tala*- is given, though no meaning is explicitly attributed to it.
(27) Affixational segmentism

\[ C(\alpha(C)) - \]

That is, all affixes are prefixes that consist of either a single consonant, or a consonant followed by the vowel \( \alpha \), or a sequence of \( CaC \). The important generalization regarding these affixes, which contrasts to the affixes discussed at length in this paper, is that they all contain the vowel schwa underlyingly.

By contrast, the underlying forms of the active and passive morphemes, as seen earlier, are \( mu- \) and \( ni- \), respectively. These differ in that they crucially have full vowels specified in their inputs. This essential difference explains why ablaut occurs in the cases of the active and passive morphemes but not with any other morphemes in the language. If any affix did have a vowel other than schwa then the account here predicts that an ablauting paradigm would result, under the proper phonological circumstances: namely, when the affix is attached to a stem with schwa in its first syllable. Thus the fact that ablaut is observed only in the active and passive paradigm can be explained as a consequence of underlying representation, and supports the distinction in underlying specification.

References


Adam Ussishkin
Department of Linguistics
Stevenson College
UC Santa Cruz
Santa Cruz, CA 95064
ussishki@ling.ucsc.edu
http://ling.ucsc.edu/~ussishki
One of the most remarkable differences that one notices as one thumbs through dictionaries of Western Austronesian languages like Tagalog and Indonesian and those of Western Oceanic languages of the north coast of New Guinea like Gedaged (Mager 1952) and Yabem (Streicher 1982) is the relative lack of precategorial roots in the latter. Given that precategoriality is a striking feature of other Oceanic languages, especially the conservative languages in the Central-Eastern Oceanic subgroup like Fijian (Dixon 1988) and argued convincingly for Polynesian languages like Tongan (Broschart 1997), it would appear that these Oceanic languages of the New Guinea north coast are innovative in this aspect. In this paper I will trace the developments that led to this change in typology, but first I wish to discuss what I mean by precategoriality in AN languages and revise the reconstruction of the clause structure of Proto-Austronesian in the light of this idea.

In Foley (1998) I presented an analysis of Tagalog as a precategorial symmetrical voice language. The gist of this analysis is as follows. Most Tagalog roots are precategorial, i.e. not assigned to lexicosyntactic classes of Noun or Verb, just unspecified, X. Whether a given root is to function referentially or
predicationally is determined by its functional head, D or I respectively, of which it is the complement within a DP or IP phrasal projection. D is drawn from a set that contrasts proper versus common and also case, much like German determiners:

(1)  
\[ \begin{array}{c}
\text{DP} \\
\text{D'} \\
\text{D} \quad \text{X} \\
\text{sa} \quad \text{jalake} \\
\text{COMMON} \quad \text{man} \\
\text{OBL} \\
\text{‘to the man’}
\end{array} \quad \begin{array}{c}
\text{DP} \\
\text{D'} \\
\text{D} \quad \text{X} \\
\text{ni} \quad \text{Maria} \\
\text{PROPER} \quad \text{PN} \\
\text{CORE} \\
\text{‘by/of Maria’}
\end{array} \]

The I head is more complex. First because the roots are not verbs, they do not take argument structures, so that first they must be derived as verbs with their respective array of argument structures. This is accomplished by a set of voice affixes, commonly called ‘the focus affixes’, which not only derive a verb and its associated argument structure, but also select one of these arguments as the clausal PIVot and assign grammatical functions appropriately:
The structure of a verb derived so and its associated DPs, arguments and not, in traditional terms, an S, are what are specified for I. The I takes the full S as its complement. The I may be realized by what is commonly called an AUXiliary, in which case the verb of the S appears in a non-finite form, i.e. without I inflection:

(3) huwag mo-ng tawag-in ang duktor
    don't 2SG CORE-LIG call-UV PIV doctor

'Don't call the doctor'

Note the pronoun mo 2SG CORE occurs as an enclitic to the AUXiliary huwag 'don't' and precedes the verb tawag-in 'call s. o.' that governs it. This is a typical feature of Tagalog: CORE and PIVot pronouns, as well as a number of particles are specified as enclitics to an I head:
However, verbs can be directly inflected for I, i.e. be finite, along a dimension of a mood contrast, realis versus irrealis, and within realis, for aspect, imperfective versus perfective:
When the verb is inflected for 1 there is no AUXiliary, and it must occupy the 1 head position of the IP in order to meet requirements of feature compatibility:

(6a)  p-in-atay-O ko ang abogado
      PERF-kill-UV ISG CORE PIV lawyer

'I killed the lawyer'
It is a well known fact about Tagalog that it allows fully inflected verbs to behave nominally, i.e. be referential, as in (7):

(7) \[ \text{p-in-atay-O ko ang nag-a-abogado} \]
\[ \text{PERF-kill-VC 1SG CORE PIVE IMPERF.VC-lawyer} \]

'I killed the one becoming a lawyer'

In the terms of this analysis, this is simply a DP with an IP complement and in fact is the general structure of relative clauses in the language. Relative clauses in Tagalog like those of Western Austronesian languages generally are formed by
deletion: the DP within the IP relative clause coreferent with the head is deleted. But of course the argument structure of the verb of the IP must be satisfied and this is assured by stipulating that only the PIVot DP of the relative clause may be the relativized constituent (note that the PIVot DP, being specified by the voice affix on the verb, is the only fully recoverable argument) and further that the D of the higher DP projection bind it (Jelinek and Demers 1994):

(8)

```
<table>
<thead>
<tr>
<th>DP</th>
</tr>
</thead>
<tbody>
<tr>
<td>D'</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Z</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>I'</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>DP</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Z</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>PIV</td>
</tr>
</tbody>
</table>
```

For the DP in (7) the structure is:
It is clear when investigating diverse languages of Taiwan drawn from varying first order subgroups of Austronesian like Atayal (Egerod 1965, 1966; Huang 1993,
1995), Paiwan (Egli 1990; Ferrell 1982), Pazeh (Blust 1999) and Tsou (Starosta 1985; Szakos 1994; Tung 1964; Zeitoun 1992, 1993), that a system much like that of Tagalog can be reconstructed for Proto-Austronesian (see Ross 1995; Wolff 1973), with a few notable amendments. First the AUXiliary system seems much more robust in Taiwan languages and was likely so in PAN. Indeed Ross (1995) reconstructs a set of atemporal voice affixes, i.e. nonfinite [-l] forms, which followed AUXiliary I heads: Actor PIV: O, Undergoer PIV: -u (~ a) Locative PIV: -i.

Tsou has taken the robust AUXiliary system to its logical conclusion: all IPs are headed by an overt AUXiliary (Zeitoun, Huang, Yeh, Chang and Wu 1996):

\[
\begin{align*}
(10) (a) & \quad \text{moso} \quad \text{bonö} \quad \text{to} \quad \text{tacömö} \quad /o \quad \text{amo} \\
& \quad \text{PAST.AV} \quad \text{AV.eat} \quad \text{CORE} \quad \text{banana} \quad \text{PIV} \quad \text{father}
\end{align*}
\]

‘Father ate a banana’

\[
\begin{align*}
(10) (b) & \quad \text{i-}ta \quad \text{ana} \quad /e \quad \text{tacömö} \\
& \quad \text{PROG.nAV-3SG} \quad \text{CORE} \quad \text{act.UV} \quad \text{PIV} \quad \text{banana}
\end{align*}
\]

‘He has been eating a banana’
It should be noted that Tsou’s n-AV voice affixes descend from the atemporal series. This is to be expected if the basic IP type that survived into this language was that headed by AUXiliaries and the IPs headed by I inflected verbs were lost. Interestingly, however, as the paradigm collapsed, the AV affix was drawn from the neutral verb series (see below) a development which has gone to completion in Tagalog, where the neutral forms have taken over the role of the atemporal non-finite set entirely. Tsou seems a local development; PAN was much more
probably like Tagalog with AUXiliary-headed and AUXiliary-less IPs illustrated by Atayal (Huang 1993; Zeitoun et al 1996):

(11) (a) m-ihiy-saku/ tali/ hiva/ AV-beat-1SG PIV PN yesterday
   'I beat Tali/ yesterday'

   (b) AUX:PAST-2SG PIV wan-su/ m-ihiy sayun
       AV-beat PN
   'You beat Sayun'

In addition to the non-finite forms which followed AUXiliary I heads (with secondary uses as plain imperatives and clause chaining dependent verbs (Egli 1990; Ross 1995), to be discussed below), PAN appears to have had a verbal system inflected for I in a manner not too different from Tagalog. There was a neutral verb form derived from the root simply by the voice affix, AV: *-um-, UV:*-uv: ·*.-x-0, and LV: *-an. It is not clear what the actual semantics of these forms was, having no overt realization of I, but they could be inflected for aspect, in a way already familiar from Tagalog.

PERF: *-in-
neutral
IMPERF: REDuplication

Finally, there was a set of forms labelled by Ross (1995) as projective, derived from the atemporal, non-finite set by the addition of a suffix -a preceding the non-finite voice affixes, -O, -u and -i. It is not clear whether these forms were truly finite (i.e. inflected for I) or a second set of non-finite forms derived from the atemporal forms by a process of insubordination (Evans 1993). Their semantics
is clearly in the irrealis range, intention, possibility and exhortation, so that latter proposal seems likely: modal or irrealis forms often arise from non-finite embedded forms that have lost their embedded status. In any case both the projective and atemporal series are lost in Tagalog; the former neutral series takes over the function of being the non-finite form. Table 1 drawn from Ross (1995) presents his full reconstruction of PAN verbal inflection:

Table 1: PAN Verbal Inflection

<table>
<thead>
<tr>
<th></th>
<th>AV</th>
<th>UV</th>
<th>LV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-finite atemporal</td>
<td>O</td>
<td>-u(-a)</td>
<td>-i</td>
</tr>
<tr>
<td>[-I]</td>
<td>?projective</td>
<td>(-um-)+a+O</td>
<td>aw&lt;-a+u</td>
</tr>
<tr>
<td>Finite neutral</td>
<td>-um-</td>
<td>-an</td>
<td></td>
</tr>
<tr>
<td>[+I]</td>
<td>PERFective</td>
<td>-um- + -in-</td>
<td>-in- + O</td>
</tr>
<tr>
<td></td>
<td>IMPERFective</td>
<td>-um- + RED</td>
<td>RED + -an</td>
</tr>
</tbody>
</table>

Another difference concerns the behavior of the enclitic pronouns. In PAN, their encliticization to the I head seems to have been optional, although it is obligatory in many daughter languages. Their optional encliticization in PAN is witnessed by Rukai in examples like the following (Ross 1995):

(12) wa-kan«aku
    PAST-eat-1SG PIV
    ‘I ate’

The prefix wa- PAST is derived from the PAN AUXiliary *ua PAST (cf Atayal wan PAST in (11b) above) which has become procliticized to the verb. Note that the pronominal DP aku 1SG PIV follows the verb rather than occurring between the
AUXiliary and the verb where we would expect it if encliticization to the AUXiliary was obligatory. This argues that encliticization was optional in PAN and either remained optional or was lost in Rukai, but became obligatory in many daughter languages.

Another important difference between Tagalog and probably PAN concerns the headless relative clause structures or nominalizations illustrated in (9) and analyzed here as DP over IP structures. In Tagalog and most Western Austronesian languages including many Taiwan languages like Paiwan these are restricted to IP complements headed by inflected verbs. PAN appears to have allowed such structures to be formed on IPs headed by AUXiliaries. In Tsou these are the only DP over IP constructions, as would be expected given its constraint requiring IPs headed by AUXiliaries (Starosta 1985):

(13) (a) ic/o na oh-to ei/m-l
    that D AUX-1PL CORE live-LV

'That’s where we live'
Remarkably, however, in addition to the usual structures, such as (13b), DP over IPs headed by inflected verbs, are also found in Atayal, a language of a different highest order subgroup in Austronesian, strongly suggesting that structures like these must be reconstructed for PAN as well.

Finally, in marked contrast to languages like Tagalog and Indonesian, in
which the same affixes can be used in different ways on roots to derive forms
which can function referentially or predicationally:

(14) -in: lagnat ‘fever’: lagnat-in ‘have a fever’
awit (um-awit ‘sing’): awit-in ‘song’
-tukod ‘cane: tukur-an ‘use as a cane’
hiram (h-um-iram) ‘borrow’; hiraman ‘place for borrowing’

it is a notable fact that Taiwan languages seem especially rich in derivational
affixes (Blust 1999; Ferrell 1982). In contrast to the Tagalog situation illustrated
above, some of these seem restricted to deriving either a verbal or a nominal
form. For example in Pazeh (Blust 1998, 1999), the prefix sa- derives IV voice verbs:

(15) sa-talek alaw ki bulayan
IV-cook fish PIV pan

‘The fish was cooked with a pan’

And through a nominalization by DP over IP structures, referential expressions
like: hiud ‘fish’, sa-hiud ‘fishing pole’; kuxus ‘scrape’ sa-kuxus ‘scraper’; tabuk
‘peck’, sa-tabuk ‘bird’s bill’. However, there is another affix consisting of
reduplication of the initial C of the root followed by the vowel a that also derives
referential expressions describing instruments, but in this case is never used as a
verbal voice affix:
As this affix Ca- is widespread in Taiwan languages and indeed found in other AN languages outside Taiwan, it must be reconstructed for PAN, and with it a contrast between referential expressions derived by a DP projection over IP complements and simple nominal derivation by affixation. This does not change the status of precategoriality in PAN: the roots must be derived to function referentially/nominally or predicationally/verbally, but it does indicate that D heads projecting a DP phrase were not the only way to derive referential expressions in PAN. Tagalog and Western Austronesian languages of like ilk seem to have moved from the flexibility of PAN into increasing reliance on functional heads and phrasal projections; hence the lack of obligatory D heads in languages like Atayal and Pazeh versus their pervasive presence in Tagalog.

To summarize our findings so far: PAN was a strongly left-headed language, dominated by DP and IP phrasal structures. Roots were precategorial and required derivation to function referentially or predicationally, and this was
commonly accomplished syntactically through being complements of the proper functional heads, D (referential) or I (predicational). I heads could be realized syntactically through the use of AUXiliaries or morphologically, by inflecting the verb for I, in which case the verb occupied the I head position. Pronominal DPs were always encliticized to the I head; if that was an AUXiliary, then they occurred between the AUXiliary and the verb. It is not clear if all CORE pronouns encliticized or only those denoting the Actor, especially when the PIVot was a non-Actor. Clearly the latter case was the strongest target for encliticization, as all languages allow or require encliticization, in this case. The form of the non-PIVot Actor pronominal DPs when encliticized was identical to that of the forms marking possession, as indeed generally non-PIVot Actor and possessive DPs are formally identical.

In the transition from PAN to Proto-Oceanic (POC) a number of major grammatical changes occurred. The basic sentence type with an AUX as I head and a non-finite atemporal verb survived into POC, but the sentence types with I inflected verbs serving as head were lost. The PAN AV atemporal form now became the intransitive clause type, while in the transitive clauses the UV atemporal verbs in *-a (~ *-u) merged with the form in *-i, leaving the PAN LV atemporal form as the basic transitive clause type. (17) exemplifies the normal clause types in Pre-POC:

(17) Pre-POC clause types

<table>
<thead>
<tr>
<th>ITR</th>
<th>AUX</th>
<th>PRO</th>
<th>V</th>
<th>(DP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR</td>
<td>AUX</td>
<td>PRO</td>
<td>V-*i</td>
<td>DP</td>
</tr>
</tbody>
</table>

287
The PRO forms could be either non-PIVot actor forms (morphologically identical to possessive pronominals) or the PIVot forms: in modern Oceanic languages the so-called subject proclitic pronominals can be descended from either of these sets, different languages exhibiting differing ancestry. However, given that there was no longer any productive voice system, the PRO always referenced the actor/subject DP. By POC, the AUX element is lost; hence the difficulty of reconstructing the POC tense-aspect-mood system noted by Ross (1988). This left the old enclitic pronominal now a proclitic to the following verb, and in addition in transitive clauses new pronominal enclitics referencing the object were innovated from older independent pronominal DPs:

(18) POC clause types

ITR: PRO·V (DP)

TR: PRO·V·*i·PRO (DP) (DP)

These changes had an especial impact in embedded clauses. With the loss of I inflected verb heads as an option in main clauses, these structures in embedded clauses in former DP over IP structures now became islands. As the transparency of the I inflection here was lost, it was no longer possible for speakers to identify them as IPs, i.e. relative clauses:
This had two effects. First new ways of forming relative clauses had to be innovated, and with the loss of a productive voice system, the restriction on PIVot only relativized DPs had to be dropped. Modern Oceanic languages mostly relativize DPs in situ leaving PROnominal copies. Secondly, because the XP was no longer categorically identifiable on its own, but the whole constituent was, as a DP, i.e. semantically referential, the formerly I inflected word was now reanalyzed as a normative complement of a D head, i.e. a referential word, in short a derived noun. What had formerly been an I inflected verb in PAN, now became a derived noun in POC (note this sequence is exactly the opposite of that proposed in Starosta, Pawley and Reid (1982)), so that by the time of POC a clear word level category of noun had emerged, and this process of nominal derivation through formerly verbal affixes remains productive in conservative Oceanic languages like Tolai: (Mosel 1984) and Nakanai (Johnston 1980):
(20) (a) Tolai

momo 'to drink' m-in-omo 'a drink'
mate 'to die' m-in-at 'death, corpse'
tata 'talk' t-in-ata 'word, language'
ian 'eat' ni-(i)an 'food'
kukul 'buy' k-un-ukul 'buying'

a k-un-ukul na kar ka-i ra tutana
D NOMLZ-buy LIG car POSS CLSF-POSS D man
'The buying of the car by the man'

(b) Nakanai

taga 'afraid' t-il-aga 'fear'
vago 'to pole a canoe' v-il-ago 'a pole'
pelho 'die' p-il-eho 'a corpse'
pou 'sit' p-ul-ou 'stay'

la p-ul-ou tamiteu kama taritigi
D NOMLZ-sit POSS 1PL EX NEG good
'Our stay wasn’t good'

However, in spite of this innovation of a word level lexicosyntactic category of noun, it is clear that very strong features of precategoriality remained in POC, as it does in most modern Oceanic languages, in that prototypically nominal roots can freely be used predicatively without derivation as exemplified by Tolai (Mosel 1984):

(21)

(a) a kaliku i ga boroil lua
D snake 3SG S REM pig before
'The snake “pigged” (i.e. became a pig) before'

(b) tuk i ga pui vanavana mulai
until 3SG S REM bush by and by again

'Until it by and by “bushed” again'

'Until the bush grew again'

The opposite, the use of underived prototypically verbal roots as referential DPs is rarer, due no doubt to the innovation of derived nouns by reanalysis of the older I inflected verbs. Yet it does appear: again Tolai (Franklin, Kerr and Beaumont 1974):

(22) (a) a tutana i ririvon
D man 3SG dream

'The man is dreaming'

(b) a ririvon i tar kaina
D dream 3SG PERF bad

'The dream was bad'

This contrast is buttressed by Mosel’s point (1984:136) that English nouns borrowed into Tolai are frequently used predicatively, but borrowed English verbs are not used referentially. However, as productivity of the processes deriving nouns from prototypically verbal roots declines, we might expect more robust precategoriality to reemerge, and this is indeed the case in Oceanic languages like Fijian and Tongan (Broschart 1997).

A few words on wider typological issues are in order here. PAN was and POC remained a strongly left headed language. Precategoriality seems particularly well adapted to left headed languages in that the functional head which instructs
how to take the following XP, predicationally, referentially (or even attributively, a topic beyond my scope here), always occurs first, a clear processing advantage, a claim buttressed by the fact that crosslinguistically strong precategoriality does seem particularly salient in language families strongly of a left headed structure, Salishan, Wakashan, Austronesian, as witnessed by the literature going back to Boas (1911) on this topic. In fact we can go further and contrast the typical left headed language with DPs to right headed languages (e.g. Latin) with NPs or KPs. The claim is that in right headed languages, processing constraints make D functional heads which follow precategorial XP complements dysfunctional:

(23) *
   \[
   \begin{array}{c}
   \text{DP} \\
   \downarrow \\
   \text{D'} \\
   \quad \begin{array}{c}
   \text{XP} \\
   \quad \text{D} \\
   \quad \text{X}
   \end{array}
   \end{array}
   \]

Or in optimality theoretical terms this is a very highly ranked constraint. In fact, a very widespread, but weaker constraint is that right headed languages don’t much like DPs at all and overwhelmingly across the globe, with high statistical reliability, there is an inverse correlation between right headed typology and DPs, as witnessed by strongly right headed families like Classical Indo-European, Uralic, Altaic, Japanese, Australian, Sino-Tibetan, Andean. Furthermore, as right headed languages have NPs rather than DPs, they typically have a robust contrast
between noun and verb roots, necessary so that the proper NP phrase can be projected from the syntactic N heads. The possibility of precategoriality in languages with KPs, i.e. the phrase is headed by a K category, a case suffix, enclitic or postposition which takes a precateorial XP complement has not yet been investigated and cannot be ruled out a priori, although in such cases it seems difficult to see how a principled distinction between K and D categories could be drawn.

When we turn to IP structures, somewhat the opposite picture emerges. Right headed languages from diverse families across the world are characterized by clause chaining, a pattern in which clauses headed by morphologically stripped down verbs precede ones headed by a fully inflected verb, from which the previous verbs take their specifications for these I features. Watam of the Papuan Lower Sepik-Ramu family illustrates:

(24) Watam (NAN: Lower Sepik-Ramu)

\[
\begin{array}{llllll}
\text{waut} & \text{nakan} & \text{i} & \text{mbo} & \text{ga-}r & \text{saNgga-}r \\
\text{stone} & \text{big} & \text{a LOC} & \text{climb-DEP} & \text{go-DEP} \\
\text{timoN} & \text{an} & \text{Ng-utki-}r & \text{ak-ri} & \\
\text{on top} & \text{this FOC-stand-DEP} & \text{call out-PAST} \\
\end{array}
\]

'(he) climbed up a big rock and stood on top of it and called out'

In this example only the final verb is specified for tense, i.e. I, by -ri PAST. The previous verbs simply mark their dependence through -r. A straightforward analysis here is that the tense is the I head with takes a complex complement, i.e. a number of conjoined S nodes:
(25) (a)  

(b)  

wa\textit{u}t nakan i mbo ga-r sa\textit{N}ga-r timoN an Ngutki-r ak-
stone big a LOC climb-DEP go-DEP on top this FOC-stand-DEP call out
The structure here is simplified; it is certainly necessary to posit layers of more deeply embedded I heads with circumscribed scope relations (Foley and Van Valin 1984; Cinque 1998), as well as lower level narrow scope I categories like aspect within the individual S constituents, but the above does give a clear idea of the overall structure. Now structures like (24) and (25) are strongly disfavored in left headed languages. While there are a few left headed AN languages that have analogs of clause chaining like Paiwan in Taiwan (Egli 1990) and some Oceanic languages in Southern Vanuatu (Crowley 1998; Lynch 1983), it is quite rare and is not reconstructable for PAN or POC. Overwhelmingly AN languages link clauses headed with fully inflected verbs or AUXiliaries through paratactic coordination or subordination. Clause chaining is simply not in their grain and again this is true of strongly left headed language families like Salishan, Wakashan, or Mayan. Again the following structure is highly dysfunctional or in OT terms, a very highly ranked constraint.

(26)

```
* IP
  |
  I'
  |
I
  |
Tense
  |
MOOD
  |
IF
  |
V
  |
NP
  |
S
  |
S
  |
S
  |
S
  |
V
  |
NP
```
The only place in the domain of the AN languages where non-Austronesian languages survive is the region around New Guinea, where some 800 such languages are spoken. The north coast of New Guinea in particular is densely populated with hundreds of non-Austronesian (NAN) languages and has been so for some 40,000 years. While belonging to a number of language families, all languages of this region, with the exception of those of one family (the Torricelli family, not of much relevance to my topic here) are characterized by a strong right headed typology. All have a sharp noun/verb root distinction, lack DPs and exhibit extensive clause chaining—all prototypical features of right headed languages and all lacking in POC, a strongly and consistent left headed language. Oceanic speakers, however, have been in the region of the north coast of New Guinea for some three millennia (Spriggs 1997) and have long term ongoing and often intense trade and cultural relations with their NAN speaking neighbors, causing in many cases significant shifts in the typological profile of the Oceanic languages in the region.

Essentially there seem to be two outcomes, what I will call strip or convert. The strip solution I will dispose of rather quickly, as it is of less interest to my main concern. This is the case in which the language undergoes simplification during the process of converging with adjoining typologically divergent languages. Adzera (Holzknecht 1986) in the upper Markham river valley, in close contact with Trans New Guinea languages illustrates this well. While POC were already
morphologically rather simple, Adzera improves on this profile. All clitic pronouns are lost, as well as the transitive verb suffix *-i. The verbal structure, however, remains strongly left headed, with proclitics and prefixes to mark I, with a minimal obligatory contrast between REALis and IRRealis. The category of D is completely lost (this is general along the north coast, undoubtedly due to its lack in neighbouring NAN languages) and with it DPs. NPs now emerge as a basic syntactic category. What distinguishes whether a given root is to be taken referentially or predicationally, is the presence of I marking in the latter case, but, as expected, precategoriality is completely lost, in sharp contrast to Tolai, not to mention Tongan or Tagalog. What is especially notable is the extension of productive nominalization from the old PAN LV affix ·an (Adzera ·(d)a(n)) to uses of non-finite nominalization totally atypical of OC languages that preserve these, but widespread among NAN languages:

(27)

(a) non-finite relative clause (note preserved left head structure)

(i) garam man
[is-a catch-NOMLZ fish'

'a fisherman'

(ii) nam food
[nu/-an cook-NOMLZ

'cooked food'

(b) non-finite complement

(i) araNan 3SG i-ni REAL-say
[ba-dan come-NOMLZ

'He wants to come'
The OC languages along the coast of Madang Province exhibit the convert solution, taken to varying degrees of completion, in other words they have shifted in the direction of the right headed typology of their NAN neighbors. As expected, the functional head D (and consequently DPs) disappears in all these languages, given its dysfunction in right headed languages, but the real focus of interest are changes affecting the IP. Manam (Lichtenberk 1983) represents the more conservative development. Here the clausal constituent order has switched to the canonical SOV pattern of right headed languages:

(28)

```
tamóata bóro di-tao-táon-i
man pig 3PL S REAL-RED-hunt-3SG O
```

'The men were cahsing the pig'

And the language has postpositional PPs in contrast to POC’s prepositional phrases: bía /ana beer because ‘for a beer’. The nominalizer -(N)a derived from PAN *-an is also in evidence in Manam, and like Adzera is used to form non-finite complements, but it is not used in non-finite relative clauses. Beyond this, however, there is little shift. The IP structure remains strongly left headed (see (28) above) with the verbal structure inherited from POC little changed: I categories are prefixal and there are proclitic and enclitic PROnominals. In spite of the right headed clausal structure, clause changing structures are not innovated: it appears the left-headed verb puts pay to that. Manam conjoins
clauses together, each headed by a fully inflected I specified verb. There are, however, very incipient signs of things to come in that the conjunction be ‘and’ becomes encliticized to the fully inflected verb preceding it. Bound conjunctive suffixes in dependent verbs are a salient feature of clause chaining structures.

(29) tamóata bóro di-tao-taon-í-be man pig 3PLS REAL-RED-hunt-3SG O-CONJ

dáine réga di-/oto-/óto-O woman firewood 3PLS REAL-RED-break-3PL O

‘The men were chasing the pig and the women were gathering firewood’

Precategoriality, while rare from Lichetenberk’s description, as fully expected from the language’s semi-shift to a right headed typology, is attested. For example there is a root nanari which can be used as in intransitive verb ‘tell a story’ or the noun ‘story’ as in the possessive construction

(30) nanari ne-gu nanari-ta/-a-gu story CLSF-1SG POSS story-TR-NOMLZ-1SG POSS

‘my story’ (I told) ‘my story (about me)’

But this is rare; it appears even robustly categorially vague words like ‘rain’ are resolutely categorial in Manam:

(31) (a) úra i-púra rain 3SG S REAL-come

‘It started to rain’

(b) natu /ategisi i-púra (compare (21))

child teacher 3SG REAL-come

‘His son (literally ‘child’) became a teacher’

Examples like (30) should perhaps be analyzed then as lexical exceptions, rather
than representing precategoriality in the language.

Finally, the Bel subgroup of languages spoken around Madang has shifted most strongly in the direction of neighboring right headed NAN languages (Ross 1987). These share the innovations noted above for Manam, but have become canonical right headed languages, with structures that look for all the world like clause chaining (Ross 1994). The crucial innovation that seems to have made these possible is the shift from prefixal to suffixal I inflection, i.e. the IP projection is clearly, unambiguously right headed (the mood inflection, REALis versus IRRealis appears to be the highest I category in these AN languages, in contrast to many TNG languages it which this is illocutionary force. It is the position of the highest I category that determines the head typology).

(32)

(a) Gedaged (Dempwolff n.d.)

tamol i-l(a)-e
man 3SG S-go-REAL
'The man went'

(b) Takia (Ross 1987)

panu na N-au wa
village OBL 1SG S-go IRR
'I shall go home'

Interestingly, the contrast between IRRealis and REALis PRONominal prefixes, salient in Manam and so pervasive in the Oceanic languages of the New Guinea region, is lost in the Bel languages in favor of modal suffixes or enclitics, bringing the language wholly into line with a right headed typology. Clause
chaining appears to emerge, with the typical Madang area pattern of mood inflection for dependent verbs:

(33) Takia (Ross 1987, 1994)

(a) iN 3SG i-marsi-go fud 3SG S-eat REAL

3SG S-sit-REAL banana 3SG S-eat REAL

'He sat and ate a banana'

(b) iN 3SG i-marsi-pe fud i-ani wa

3SG S-sit-IRR banana 3SG S-eat IRR

'He will sit and eat a banana'

While these structures look NAN, the actual DEPendent verb markers, -go REAL and -pe IRR are clearly of AN vintage, -pe IRR being cognate with Manam -be ‘and’ in (29) above (Ross 1987). The system, however, is clearly calqued, for such a contrast in dependent verbs is pervasive in Madang area languages of diverse families:

(34) (a) Watam (NAN: Lower Sepik-Ramu)

(i) ma 3SG sit-REAL banana FOC-eat-PAST

birka-r naNas Ng-amb-ri

'He sat and ate a banana'

(ii) ma 3SG sit-IRR banana FOC-eat-FUT

birak-mbe naNas Ng-am-na

'He will sit and eat a banana.'

(b) Bargam (Hepner 1995) (NAN: Trans New Guinea)

(i) leh-ad 1PL go-REAL yell-IMPERF-1PL S

ekton-y-auq

'As we were going, we were shouting'
'If you go, we will kill a pig for you'

In addition, as is typical of NAN languages, the dependent verbs may be marked for relative tense/aspect by suffixes which immediately follow the verb, but precede the DEPendent marker:

(35) Takia (Ross 1987)

(a) In fud ta i-ani-du-go Nai
   3SG banana a 3SG S-eat-SIM-REAL 1SG
   you Na-luk a water 1SG S-drink REAL

   'He was eating a banana while I was drinking water'

(b) Nai N-ani-gu-p panu na N-au
    1SG 1SG S-eat-SEQ-IRR village OBL 1SG S-go IRR

   'I shall eat and then go home’

(c) tamol di-ani-na-g i-sida ya
    man 3SG S-eat-DUR-REAL 3SG S-enough REAL

   'The men ate until they were full’

It is clear that the basic use of these suffixes is aspectual, as is transparent in their use with independent verbs (Ross 1987):
(36) Mait  nal  pem pem  iN  sa-n  wog  lo
Mait  day  every  3SG  CLSF-3SG POSS  canoe  OBL

i-sida-na   ya
3SG S-board-DUR   REAL

'Mait used to board his canoe every day'

While not obvious in Takia, the source of these aspetual suffixes is from earlier serial verb constructions in which the second verb was used as an aspectual modifier, a very common feature of Oceanic languages. Takia’s close relative Gedaged exhibits this transparently:

(37) Gedaged  (Dempwolff n.d.; Ross 1987)

(a)  i   mot   i-nau-la-g  me
3 snake  3SG S-do-SEQ-REAL  that

gasaN   lon   i-l(a)-e
forest  inside  3SG S-go-REAL

'After she gave birth to the snake, she went into the forest'

(b)  o   u-seg-me-g   u-nasi-lak
2SG  2SG S-come-SIM-REAL  2SG S-see-PERF

'While you were coming, you saw'

(c)  pain i   mot   oi   di-e
di-me  woman  3 snake  OBL  3PL S-fear  3PL S-IMPERF

Women fear snakes’
Note that the SEQuential form is -la and the SIMultaneous -me. These are undoubtedly identical to the aspectual markers with independent verbs -lak PERF and -me IMPERF, though, interestingly, -lak has become fully integrated as a suffix, while -me remains a verb in its own right in a serial verb construction, as witnessed by its own subject PRonominal marker. Both of these in turn descend from the POC verbs *lako ‘go’ and *mai ‘come’ (Ross 1987). But note that in contrast to the canonical Papuan pattern in which dependent verbs are quite stripped down and the morphological patterns of dependent and independent verbs very different, for example, the elaborate tense distinctions in Watam independent verbs are collapsed into a binary mood REALis-IRRealis contrast in dependent verbs, in Takia and Gedaged this is not the case. In these languages, the contrasts made are actually the same, the difference being the formal realization: Takia -p(e) IRR on dependent verbs, wa IRR on final verbs. This suggests that in contrast to the true clause chaining structure of (25) with nested conjoined S nodes within a higher IP projection from a final main I head, these AN structures are really conjoined IPs, each with their own I head:
(38) = (35b)

\[
\begin{tikzpicture}
  \node (ip) {IP};
  \node (ip1) [below left=of ip] {IP'};
  \node (ip2) [below right=of ip] {IP'};
  \node (s) [below=of ip1] {S};
  \node (i) [below=of s] {I};
  \node (p) [below=of ip2] {S};
  \node (i1) [below=of p] {I'};
  \node (np) [left=of i1] {NP};
  \node (v) [left=of np] {NP};
  \node (p1) [left=of v] {V};
  \node (pp) [right=of np] {PP};
  \node (v1) [right=of v] {V};
  \node (wa) [right=of v1] {wa};
  \node (irr) [left=of i1] {IRR};
  \node ( irr1) [right=of irr] {IRR};
  \node (nai) [above=of np] {Nai 1SG};
  \node (nani-gu) [above=of p] {N-ani-gu-1SG};
  \node (panu) [above=of pp] {panu 1SG};
  \node (na) [above=of irr] {na village OBL};
  \node (nau) [above=of irr1] {N-au 1SG};
  \node (s-go) [above=of irr1] {S-go};

\end{tikzpicture}
\]

'I shall eat and then go home'

or schematically:
In other words, these are virtual clause chaining structures only, in superficial appearance similar to NAN structures, but in nature different. These are really still structures of paratactically linked conjoined clauses, like those of Manam, with suppletive I heads for non-final versus final verbs (the shift from a conjunction to an I category is itself an interesting phenomenon). It is the shift to final I heads and the thoroughgoing REALis/IRRrealis distinction across clauses that is the crucial innovation here and one undoubtedly due to strong pressure from neighboring NAN languages like Bargam. While the clause linkage structures of Bel languages may look Papuan, they are not. They unquestionably betray their Oceanic heritage in their preference for fully I inflected clauses loosely
conjoined. It seems that this a pervasive property of AN typology that is very conservative and not easily given up, even in the face of strong NAN pressure as witnessed by the Bel languages. It would be valuable to compare other cases of language contact between language families of strongly contrastive basic typologies to see what grammatical properties are easily negotiable in these situations and what are not. This would be invaluable data in building informative theories of grammatical structure.