

*The history and typology of
western Austronesian voice systems*

Also in Pacific Linguistics

Nikolaus P. Himmelmann, (compiler), 2001, *Sourcebook on Tomini-Tolitoli languages: general information and word lists*.

David E. Mead, 1999, *The Bungku–Tolaki languages of south-eastern Sulawesi, Indonesia*.

Linda A. Barsel, 1994, *The verb morphology of Mori, Sulawesi*.

Naomi Ferreirinho, 1993, *Selected topics in the grammar of Limos Kalinga, the Philippines*.

Sherri Brainard, 1991, *Theme, result, and contrast: a study in expository discourse in Upper Tanudan Kalinga (Philippines)*.

Stephen H. Levinsohn, ed., 1991, *Thematic continuity and development in languages of Sabah*.

Charles E. Grimes and Barbara D. Grimes, 1987, *Languages of south Sulawesi*.

Pacific Linguistics is a publisher specialising in grammars and linguistic descriptions, dictionaries and other materials on languages of the Pacific, the Philippines, Indonesia, Southeast and South Asia, and Australia.

Pacific Linguistics, established in 1963 through an initial grant from the Hunter Douglas Fund, is associated with the Research School of Pacific and Asian Studies at The Australian National University. The Editorial Board of Pacific Linguistics is made up of the academic staff of the School's Department of Linguistics. The authors and editors of Pacific Linguistics publications are drawn from a wide range of institutions around the world. Publications are refereed by scholars with relevant expertise who are usually not members of the editorial board.

The 470 publications before Publication 501 were numbered in four series, A, B, C and D, but the series were merged from the beginning of the year 2000 because the rationale for them had gradually disappeared.

FOUNDING EDITOR: Stephen A. Wurm

EDITORIAL BOARD: Malcolm D. Ross and Darrell T. Tryon (Managing Editors),
I Wayan Arka, John Bowden, Andrew Pawley, Paul Sidwell

The history and typology of western Austronesian voice systems

edited by

Fay Wouk and Malcolm Ross



Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University

Published by Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University
Canberra ACT 0200
Australia

Copyright © The authors

First published 2002

National Library of Australia Cataloguing-in-Publication entry:

The history and typology of western Austronesian voice and voice-marking

Bibliography.

ISBN 0 85883 477 4

1. Australian languages. I. Wouk, Fay. II. Ross, Malcolm.
III. The Australian National University. Research School
of Pacific and Asian Studies. III. Title. (Pacific Linguistics 518).

499.2

Typeset by Jeanette Coombes
Cover design by Emily Brissenden
Printed and bound by Union Offset Printers

Table of contents

	Introduction Fay Wouk	1
Part I: <i>Overviews</i>		
1	Voice in western Austronesian: an update Nikolaus P. Himmelmann	7
2	The history and transitivity of western Austronesian voice and voice-marking Malcolm Ross	17
3	Notes on the history of 'focus' in Austronesian languages Robert Blust	63
Part II: <i>Languages of Sulawesi</i>		
4	Voice in Tukang Besi and the Austronesian focus system Mark Donohue	81
5	A sketch of the primary transitive verbs in Pendau Phil Quick	101
6	Voice in two northern Sulawesi languages Nikolaus P. Himmelmann	123
7	Proto Celebic focus revisited David Mead	143
Part III: <i>Languages of the rest of Indonesia and Malaysia</i>		
8	Voice and valency alternations in Karo Batak Clodagh Norwood	181
9	Nominative and genitive case alternations in Bonggi Michael E. Boutin	209

10	The prefixes <i>di-</i> and <i>N-</i> in Malay/Indonesian dialects David Gil	241
11	Voice in the languages of Nusa Tenggara Barat Fay Wouk	285
12	Changes in word order and noun phrase marking from Old to modern Javanese: implications for understanding developments in western Austronesian 'focus' systems Gloria R. Poedjosoedarmo	311

Part IV: *Languages outside Indonesia and Malaysia*

13	The morphology and syntax of Seediq focus Arthur Holmer	333
14	The Sinai derived transitive construction Jun Akamine	355
15	Some aspects of 'focus' in Sama Bangingi' JoAnn Gault	367
16	Voice and role in two Philippine languages Walter Spitz	379
17	The position of Chamorro and Palauan in the Austronesian family tree: evidence from verb morphosyntax Erik Zobel	405

Part V: *Discussion notes*

18	Final words: the development of the focus system John Wolff	437
19	Final words: research themes in the history and typology of western Austronesian languages Malcolm Ross	451

Map 1:	The Austronesian family and major Austronesian language groups	18
Map 2:	Languages discussed in Part II	80
Map 3:	Languages discussed in Part III	180
Map 4:	Languages discussed in Part IV	332

Introduction

FAY WOUK

In December 1997, a one-day workshop on focus in western Austronesian languages was held at the Eighth International Conference on Austronesian Linguistics, in Taipei. The original aim of the workshop was to obtain an overview of the manifestations of focus in as many languages as possible. It was hoped that such an effort, by providing information about a wider range of languages than has previously been available, would stimulate useful discussion concerning the historical development of focus, and help to set up an agenda for further research.

Each speaker was invited to address some or all of the following questions and issues:

1. How many foci/focuses/voices can be distinguished?
2. Are noun phrases marked in any way, such that the marking is (or historically was) related to the use of focus?
3. How many pronoun sets are there? How do they function?
4. How (if at all) are the verbs marked in relation to focus? Is aspect part of focus marking?
5. How is focus selection determined?
6. Does the language employ reflexes of the Proto Austronesian (PAN) focus markers? (*-um-, *-in-, *-en, *-an, etc.) Even if there is no longer a focus system, some of these may have been retained.
7. How are stative verbs marked? It seems that PAN *ma- sometimes/often disappears with the focus system.
8. Is constituent order important in the language? What is the typical constituent order?
9. The development of proclitic/pronominal marking on the verb.
10. The development of alternative strategies for lost voice(s).

The workshop was originally the brainchild of René van den Berg, who also developed the list of questions for authors. Since he did not plan to attend the conference in Taipei, I took over organisation of the workshop, and then the editing of the proceedings. This volume is the fruit of that workshop. Many of the papers were presented there; others were commissioned after the conference, to provide coverage of unrepresented geographical

regions, and two were presented in general sessions but published here because they fit the theme of the workshop.

The book has five parts. Part I contains three overview contributions. The first is by Nikolaus Himmelman and outlines some of the typological issues involved in describing ‘focus’ languages. The other two, by Malcolm Ross and Robert Blust, concern possible histories of voice in western Austronesian. Ross outlines a scenario for its development, and discusses an alternative to the received position. Blust offers notes on the history of voice systems and on their study.

Part II, ‘Languages of Sulawesi’, contains four papers. Two of them describe individual languages: *Tukang Besi* by Mark Donohue, and *Pendau* by Phil Quick. *Tukang Besi* has two voices, distinguished from each other by presence or absence of object enclitics on the verb, and by the order and case marking of the nominal elements. Donohue argues that, although an ergative analysis seems initially attractive, *Tukang Besi* is in fact neither ergative nor accusative, and is best analysed as having a ‘Philippine-type’ voice system, although the traditional voice morphology is now found only in restricted environments such as nominalisation and external relative clauses. The behaviour of floating quantifiers is cited as evidence for the Philippine-type analysis. According to Quick, *Pendau* has a basically Indonesian-type voice system: there are only two voices, both clearly transitive, as demonstrated by a study of topic continuity in discourse, and applicative suffixes occur with both of them. However, *Pendau* also has constructions similar to Philippine instrumental and locative focus which can occur only with one of the voices.

The third paper, by Nikolaus Himmelman, describes the verbal systems of two languages, *Ratahan* and *Lauje*, and compares them with Tagalog (as an exemplar of the Philippine-type system). All three share certain features, but the two Sulawesi languages each diverge in a number of features, including word order (both languages freely allow SVO). *Lauje* shows the greater divergence, having only three voices, and making use of prenominal clitics and applicative suffixes.

The fourth paper, by David Mead, discusses the status of the reconstruction of Proto South Sulawesi. Mead takes Van den Berg’s (1996) proposed South Celebic group comprising the *Kaili-Pomona*, *Bungku-Tolaki* and *Muna-Buton* language groups as his starting point. He describes voice in six languages, *Da’a*, *Uma* and *Pamona* from the *Kaili-Pomona* group, and *Tolaki*, *Padoe* and *Kulisusu* from the *Bungku-Tolaki* group, and gives reconstructions of the voice systems of Proto *Kaili-Pomona* (PKP) and Proto *Bungku-Tolaki* (PBT). While his reconstruction of PKP voice is similar to Van den Berg’s, and easily derived from Van den Berg’s proposed South Celebic voice system, the voice system in PBT is so highly divergent that it is difficult to see it as deriving from the reconstructed PKP voice system. This, for Mead, then makes the idea of a South Celebic group less compelling than it initially seemed.

Part III, ‘Languages of the rest of Indonesia and Malaysia’, contains five papers. Four are synchronic descriptions: *Karo Batak* (spoken in Sumatra, Indonesia) by Clodagh Norwood, *Bonggi* (spoken in Sabah, Malaysia) by Michael Boutin, *Riau Indonesian* (and more briefly *South Sulawesi* — or *Macassar* — Indonesian and *Kuala Lumpur Malay*) by David Gil, and *Sasak* (spoken in Lombok, Indonesia), *Sumbawa* (spoken in western Sumbawa, Indonesia) and *Bima* (spoken in eastern Sumbawa, Indonesia) by myself.

Karo Batak has a two voice system which Norwood shows to operate in terms of discourse transitivity. She suggests that certain details of the *Karo* system, such as a set of emphatic particles and a transitive suffix that are found only in constructions with the nasal prefix,

provide evidence for subgrouping Karo with languages of the Philippines. Bonggi has a Philippine-type voice system with an active and three passives, as well as a periphrastic passive. Interestingly, the undergoer in goal focus clauses may be marked for either nominative (as is usual for this type of system) or genitive case. Boutin provides a Role and Reference Grammar based analysis of the syntax of Bonggi to explain this unusual case marking. Gil argues that in colloquial Riau Indonesian, and other non-standard local varieties of Indonesian, the verb morphology does not function as in a traditional voice system, as there is no change in syntactic subject. Gil suggests a continuum of voice markers from classical voice systems (like English) where all subject properties are shifted to a different participant, through a typical Austro-Indonesian type of system where some subject properties are shifted, to a very generalised marker that indicates simply the existence of an argument, but does not influence its syntactic status. In my paper, I describe the voice system of Sasak in detail, and the verbal alternation (no longer voice) found in Sumbawa, and provide a very brief comparative sketch of verbs in Bima, a neighboring Central Malayo-Polynesian language. I suggest a pattern in the loss of the focus system whereby actor focus constructions become more restricted in function, and thus less frequent, until finally they are dropped from the language, remaining only as frozen forms in the lexicon.

The last paper in Part III, by Gloria Poedjosoedarmo, compares the voice systems of Modern and Old Javanese. Old Javanese was in some ways like a Philippine-type language: clauses were verb initial and passive was marked with *-in-*, rather than with conjugated proclitics. Different pre-nominal markers provided some indication of syntactic status of the NP. However, the applicative suffixes *-i* and *-aken* were already present and had spread to actor focus. Modern Javanese has introduced conjugated verb forms, lost the pre-nominal markers, and become SV(O). Poedjosoedarmo attempts to describe a path of change that would lead from a Philippine type system to Old Javanese, and then from Old to Modern Javanese. She suggests that the influence of Sanskrit may be the source of the changes in the voice system, and that the loss of pre-nominal markers could have led to the word order shift between Old and Modern Javanese, and by analogy to the actor proclitics.

Part IV, 'Languages outside Indonesia and Malaysia', contains five papers. Four are synchronic descriptions, one of a Formosan language, Seediq, by Arthur Holmer, and three on languages of the Philippines: Manuk Mangkaw Sinama by Akamine Jun, Sama Bangingi by Jo Ann Gault, and Hiligaynon and Yogad by Walter Spitz. Holmer describes Seediq's four-way voice system, and its interaction with tense and aspect. He then goes on to discuss voice and case marking interact with auxiliaries, causatives and ditransitives, and to outline the factors that determine focus choice. Jun's paper on Sinama, which has a typical Philippine-type five-voice system, explores a construction that combines actor-focus morphology with an oblique agent (marked as agents are in goal-focus constructions). After providing a structural analysis of this construction, he compares it with goal focus and concludes that this construction is higher in discourse transitivity. Gault provides an overview of voice in Sama Bangingi, arguing that the five-voice system is best seen as ergative, with one antipassive and four transitive clause types. Spitz presents a detailed study of the semantics of verbal affixes in Yogad and Hiligaynon, arguing that verbal event semantics provides the best explanation for the choice of individual verbal constructions, and the clearest insight into the functioning of the voice systems of Philippine-type languages, while nominal-oriented transitivity relations are only of secondary importance.

The final paper in Part IV, by Eric Zobel, based on an analysis of the verbal morphology of Chamorro and Palauan, proposes a new subgroup within the western Malayo-Polynesian

language region. This subgroup, Nuclear Malayo-Polynesian, has two branches, one including Chamorro, Palauan and the languages of western Indonesia, and the other consisting of all Central-Eastern Malayo-Polynesian languages. Nuclear Malayo-Polynesian excludes the languages of the Philippines, north Sulawesi, and northeast and interior Borneo.

Part V of the book consists of two discussion notes by John Wolf and Malcolm Ross. These were written in reaction to the papers in this volume, and outline their current thinking on the development of voice from PAN to the current time.

I did the standard editing, working with authors to revise their contributions. Malcolm Ross performed this task for my paper, did some final editing and saw the book through the Pacific Linguistics production process.

— PART I —

Overviews

Voice in western Austronesian: an update

NIKOLAUS P. HIMMELMANN

The analysis of voice marking and grammatical relations in western Austronesian languages¹ continues to be an exciting and controversial issue for linguists of all persuasions. So far, the debate has mainly focused on a few languages, in particular Tagalog.² The primary purpose of the present book is to broaden the empirical basis of the debate by presenting relevant data from other western Austronesian languages. As may be expected, broadening the empirical basis for debating a given issue almost inevitably raises a number of analytical issues which are usually pushed aside all too easily. The present introductory remarks outline some of these issues.

1 How many types of voice marking are found in western Austronesian languages?

To date, very little is known about the extent to which voice marking and grammatical relations vary among western Austronesian languages. In particular, it is not clear if there are indeed different types of voice and grammatical relation marking among these languages. Alternatively, we could be dealing with a single basic type and a vast array of micro-level variation.

-
- ¹ For the purposes of this paper, the term *western Austronesian* is defined as a cover term for all Austronesian languages spoken in Taiwan, the Philippines, mainland Southeast Asia, western Indonesia (Sulawesi and all islands to the west of it), Borneo and Madagascar, and also including Palauan and Chamorro. That is, *western Austronesian* is used in a strictly geographic sense (roughly: all Austronesian languages to the west of 130° east longitude, excluding in Indonesia the languages east of Sulawesi, but including Palauan and Chamorro). It is not to be confused with the genetically defined term *Western Malayo-Polynesian* (for which see Blust 1978).
- ² See Kroeger (1993), Schachter (1995) and Ross (this volume) for summary reviews of the debate surrounding Tagalog. A somewhat less intensive debate has been concerned with the analysis of Indonesian (see Wouk 1996 for references and discussion). Very recently, Balinese is also becoming the object of some controversy (see Clynes 1995; Artawa & Blake 1997 and Wechsler & Arka 1998, among others).

Fay Wouk and Malcolm Ross, eds, *The history and typology of western Austronesian voice systems*, 7-15.
Canberra: Pacific Linguistics, 2002.

© Nikolaus P. Himmelmann

The major task here is to uncover interesting correlations between different morphosyntactic features. For example, is it possible to claim that the occurrence of an applicative suffix *-i* correlates with the absence of noun phrase markers? Does extensive person marking on the verb correlate with placing the subject in immediate preverbal position? And so on. Before it will be possible to search for such correlations, however, it will be necessary to identify a set of features which promise interesting correlations and to obtain the relevant facts for a broad range of languages. Most contributions in this book address this preliminary task of reviewing correlation candidates and filling in the data for lesser-known languages.

Note that it is common in the literature to assume a fairly rough and hardly ever explicitly discussed division of western Austronesian languages into the following two types: Philippine-type languages and the rest (occasionally also called *Indonesian-type* languages). Wolff (1996) and Zobel (this volume) are laudable attempts to make this distinction a bit more precise. They propose a number of morphological features (for example, preposed person markers, affix combinations involving the applicative suffix *-i*) which are claimed to occur only in non-Philippine-type languages. This two-way distinction provides a useful start for investigating the (internal) typology of Austronesian languages. However, it should be clearly understood that it is a hypothesis which needs a lot of further empirical scrutiny (see Himmelmann, this volume, for more discussion).

2 What does the term *focus system* actually refer to?

The correlations of primary interest to this volume are, of course, those which may be linked to what has been called 'the focus system', a term which in the Austronesianist literature is all too often assumed to have unambiguous reference. I believe this to be a misconception. In fact, I think that further progress in the typology of western Austronesian languages depends on spelling out the range of phenomena one has in mind when using this term.

There are at least two levels/domains to which the term 'focus system' has been applied: morphology and syntax.³ And on each level different sets of phenomena may be held to be within the scope of this term, as will be briefly illustrated in the following paragraphs.

When applied to morphology, the term 'focus system' seems to be used primarily to refer to Philippine-type voice related morphology, most of which has been reconstructed for PAn. What is not always clear, however, is which affixes precisely are deemed to be part of this system. Major candidates are displayed in Table 1.

³ A third level to which this term has also been applied is the discourse level. Since this plays only a marginal role in the literature, it is not further discussed here.

Table 1: Possibly voice-related morphology in Philippine-type languages

Prefix	Infixes	Suffixes
<i>si-/i-</i>	<i>-um-/M-</i> ⁴	<i>-ən/-in</i>
	<i>-in-</i>	<i>-an</i>
		<i>-i</i>
		<i>-a</i>

There are three interrelated issues concerning the affixes displayed in Table 1. First, it is not clear whether they all are actually voice marking affixes. It is widely agreed that the prefix *si-/i-* and the suffixes *-ən/-in* and *-an*, all of which occur in a wide range of languages, are always clearly voice-related (marking voices in which the undergoer is the syntactic pivot). The same appears to be true for *-um-/M-*, at least on a synchronic level.⁵ The infix *-in-*, on the other hand, appears to be primarily a mood (or tense) marker, as it occurs in many languages in both actor and undergoer voices (see Reid 1992 for examples and discussion). The suffix *-i* is widely attested as an applicative marker (for example in Indonesian, *Tukang Besi*, *Tomini-Tolitoli* languages, etc.), but it also alternates with *-an* in, for example, *Bisayan* languages (see Zorc 1977 for examples).

Second, it is not obvious in what sense these forms actually form a system. One of the most conspicuous features of voice-related morphology in many western Austronesian languages is its formal heterogeneity, usually involving prefixation, infixation, and suffixation as well as stem-alternations of various kinds. Furthermore, there is substantial variation across the western Austronesian languages as to the number and shape of affixes which can be claimed to have some bearing on voice-marking in a given language. This raises questions such as the following: Which elements could be missing from the ones listed in Table 1 above while still allowing the remaining inventory of forms to be called a 'focus system'? When other morphological formatives are in paradigmatic contrast with the formatives listed in Table 1 (e.g. Indonesian *di-* and *-kan*), are they part of the 'focus system'? To my knowledge, there are, at present, no straightforward and widely agreed on answers to questions of this kind.

Thirdly, it is unclear whether the idea of a 'focus system' on a morphological level includes morphological formatives beyond the formatives which appear to be directly involved in voice alternations. For example, many western Austronesian languages have elaborate morphology for marking involuntary actions and/or the mere ability to perform an action. Are these formatives part of the 'focus system'? What about the ubiquitous stative

⁴ The symbol *M-* is used here to refer to all kinds of stem-initial nasal alternations which signal 'actor-focus', often in combination with mood/tense such as Tagalog *pakiníg* 'hearing', *makiníg* 'listen', *nakiníg* 'listened'. Compare Balinese *daar - naar* 'eat', *baan - maan* 'obtain', etc.

⁵ There is some evidence for the view that *-um-* differs from the three undergoer voice affixes on a number of counts and should not automatically be assumed to be a voice-marking affix. For example, it is possible to give reasonably precise accounts of the semantics of the three undergoer voice affixes. The meaning and function of *-um-*, on the other hand, is much more elusive (the common characterisation of *-um-* as an actor voice marker is problematic in that *-um-* regularly also occurs in verbal predicates not involving actors such as predicates denoting meteorological events, e.g. Tagalog *umuulán* 'it's raining', and change-of-state predicates, e.g. Tagalog *gumandá* 'become beautiful').

morphology? Again, authors differ widely as to whether such morphology is considered part of the ‘focus system’, hardly ever providing explicit arguments for their analytical choices.

On the syntactic level, the term ‘focus system’ is also used in reference to ill-defined and differing sets of phenomena. In conservative usage it refers to some of the constructions in which the affixes displayed in Table 1 typically occur, i.e. simple declarative main clauses in Philippine languages. However, one of the more striking features of Philippine languages is the use of these forms in question word and cleft constructions and also in conjunction with existential quantifiers. The following Tagalog example shows a ‘focus’-marked word in construction with an existential quantifier:

- (1) *may ipàpakíta ako sa iyó*
 may i-RED-pa-kita akó sa iyó
 EXIST UG.T-RED-CAUS-visible 1SG LOC 2SG.DAT
 ‘(Come here,) I have something to show you.’

Is this construction part of the ‘focus system’ on the syntactic level? That is, is it a defining feature of ‘focus systems’ that ‘focus’-marked words can occur in direct construction with existential quantifiers?

There are several other widespread syntactic features in western Austronesian languages which so far have not been directly linked with the ‘focus system’ in its syntactic sense but which may well regularly co-occur with ‘focus’-morphology and in that sense actually be of central importance to clarifying what the notion of ‘focus system’ precisely means. For example, a semantically transitive ‘focus’-marked predicate usually forms a constituent with the immediately following NP, regardless of the specific ‘focus’ affix involved. This is shown by the fact that generally no other constituents (except clitics and possibly some adverbial expressions) may intervene between this NP and the predicate. Furthermore, their order is fixed while the order for other NPs is often somewhat more flexible. Here is a Totoli example:

- (2) *gaukan nogutu ponguman itu*
 gaukan N-po-gutu poN-uman itu
 king ACT.REAL-SF-make GER-story that
 ‘The king made the following announcement: ...’

- (3)a. **nogutu gaukan ponguman itu*
 b. **ponguman itu nogutu gaukan*

In this example, the post-verbal NP *ponguman itu* ‘that announcement’ expresses the undergoer of the verbal predicate *nogutu* ‘made’. It forms a constituent with the predicate since it has to follow immediately after the predicate (as shown by (3a)) and cannot be fronted (as shown by (3b)).

Again the typological question arises of whether this feature is to be considered essential to the syntax of the ‘focus system’ and if so, how this constituent compares to the well-known and widespread VP-constituents in other languages.

Another, somewhat more expansive use of ‘focus’ on the syntactic level makes reference to constructions which appear to be similar in some way or another to the Philippine voice constructions without actually involving ‘focus’-marked predicates. This is the case, for example, when the voice system of Indonesian is characterised as a ‘focus system’ (in Indonesian it is only the ‘actor-focus’ forms with *meN-* which can be considered clearly to

reflect 'focus' formatives). In this usage, the term *focus* is no longer linked primarily to a set of verbal affixations but instead refers to one or more syntactic characteristics of the voice system in a given language. Though rarely made explicit, the core feature that is of relevance here is the idea that the voice alternations in these languages are valency-neutral alternations, i.e. sets of constructions with equal transitivity values but different role-function alignments (see below end of §3).

This short review should make it clear that the reference of the term 'focus' in western Austronesian studies is far from clear. Given the various levels and differing sets of phenomena to which this term has been applied, it should not come as a surprise that considerable confusion exists in the literature as to what exactly is being claimed by a given author and how these claims are related to claims by other authors (for example, claims such as 'language X has a "focus system"' obviously depend on how a 'focus system' is defined). This confusion is aggravated by the fact that the term *focus* itself is prone to lead to misunderstandings, a topic to which I turn in the next section.

3 Why western Austronesian 'focus' is not focus

As is well known, there is a long tradition of claims, in particular for Philippine languages but also for other western Austronesian languages, that the voice system (or systems) found in these languages does not match any other system found in the world's languages. In this tradition, the term *focus* in its special Philippinist sense has been coined. This terminological choice is somewhat less than fortunate for two reasons. First, it obscures the fact that Philippine-type 'focus', though differing from the English active/passive alternation on a number of counts, is essentially a voice phenomenon. Second, it does not tally well with the concept of pragmatic focus which is widely used in general linguistics. Let me briefly elaborate on both of these problems.

The claim that Philippine-type 'focus' is essentially a voice phenomenon should not be confused with the claim that Philippine-type 'focus' alternations are essentially the same thing as the active/passive alternation in English. There is broad agreement today that the active/passive alternation should be distinguished from the Philippine-type alternations, as is obvious from the fact that almost all authors who use the term 'voice' in reference to the Philippine-type 'focus' alternations avoid the terms 'active' and 'passive'.⁶

Still, acknowledging that the Philippine-type 'focus' alternations are not the same thing as the active/passive alternation in English does not necessarily imply that these two kinds of alternation do not share any similarities. The essential point of similarity between the Philippine-type 'focus' and the English active/passive alternations is that in both kinds of alternations a different argument is put into pivot (or subject) function and that this change in the alignment between semantic role and syntactic function is marked morphologically on the verb. Compare the following two examples from Tagalog:

⁶ Even authors such as Wolff et al. (1991) who use the terms 'active' and 'passive' for pedagogical reasons make it clear that there are essential differences between the two types of alternations.

- (4) *humanap na ng bahay ang bata'*
 um-hanap na ng bahay ang bata'
 ACT-search now GEN house SPEC child
 'The child looked for houses/a house.'
- (5) *hinanap na ng bata' ang bahay*
 in-hanap na ng bata' ang bahay
 REAL(UG)-search now GEN child SPEC house
 'The child looked for the house(s).'

In (4) the verbal predicate *hanap* 'search' is marked by the infix *-um-* and the actor of the search, i.e. *bata'* 'child', is marked by the phrase marker *ang*, while the undergoer of the search, i.e. *bahay* 'house', is marked by the phrase marker *ng* (which though conventionally written as <ng> is actually /nang/). In (5), the distribution of actor and undergoer have been exchanged: it is now the undergoer which appears in the *ang*-phrase, while the actor appears in the *ng*-phrase. At the same time, the verbal morphology has changed, the verbal predicate now being marked by the infix *-in-*.⁷

In order for this alternation to count as a voice alternation, it is important to show that the *ang*-phrase is the syntactic pivot (or subject) of the two constructions illustrated in (4) and (5). And while there has been considerable controversy about the nature of subjecthood in Tagalog,⁸ it is widely agreed that the *ang*-phrases in the two examples above have more (and more important) subject-like properties than any other kind of noun phrase in Tagalog and therefore qualify as syntactic pivots, if not as subjects. Inasmuch as one accepts the claim that there are syntactic pivots or subjects in western Austronesian languages, it is clear that the change of 'focus' morphology on the verb regularly corresponds to a change in the alignment between semantic role and syntactic function, and that in this sense 'focus' alternations are voice alternations.

Of course, there are different ways of defining voice. Dixon and Aikhenvald, for example, in their recent typology of verbal alternations define voice as 'removing an argument from the (inner) core, and placing it in the periphery (valency reducing)' (1997:72). This definition, which essentially covers the passive and antipassive alternations, is somewhat narrower than the one employed here. And on this definition, it is dubitable whether the Philippine-type 'focus' alternation is a voice alternation because it is controversial whether 'focus' alternations involve a reduction in valency. Still, it seems important to me to make it clear that there is one essential point of similarity between Philippine-type 'focus' alternations and the voice alternations as defined by Dixon and Aikhenvald: all of these alternations involve a realignment between syntactic pivots and semantic roles.

Whether one captures this similarity by extending the use of the term 'voice' to also cover Philippine-type focus alternations or whether one uses another term as a cover term for the

⁷ Note that the translations for the two preceding examples are identical, except that in (4) 'house' is indefinite but in (5) it is definite. These translations reflect the most typical and unmarked reading of the Tagalog clauses. The difference in definiteness, however, is not a categorical one. Given an appropriate context and possibly also a somewhat different word order, example (4) could mean 'the children looked for the house(s)' and (5) 'the children looked for houses/a house'.

⁸ See Kroeger (1993) and Schachter (1995) for further references and discussion.

passive, antipassive and Philippine-type 'focus' alternations,⁹ is a secondary terminological matter. In line with most of the recent literature, all contributions to this volume have opted for the former option and use *voice* in reference to Philippine-type 'focus' alternations.

Turning now to the second point, it is also widely agreed today that Philippine-type 'focus' does not have anything to do with what is commonly understood as pragmatic focus. That is, Philippine-type 'focus' does not pertain to the marking of new information (so-called information focus) nor does it mark contrastive emphasis on one of the arguments (so-called contrastive focus). Thus, the NP which is said to be 'in focus', for example *ang bata* 'the child' in (4) above, is not new information, nor is it contrasted with another entity in a presupposed set of possible agents for the search of a house (i.e. (4) does not mean 'it was the child (and not its father) who looked for a house'). It is also not the case that the 'focus' morphology which appears on the predicate (such as the infixes *-um-* and *-in-* in the two examples above) marks any kind of pragmatic focus. Rather, it is clear, and widely agreed, that the so-called 'focus' affixes mark a combination of semantic roles and tense, mood and/or aspect (opinions differ quite widely whether it is tense or mood or aspect or a combination of two of these categories). Thus, for example, the infix *-um-* in (4) makes it clear that the referent of the *ang*-phrase is the actor of the search and not its undergoer. It does not highlight or emphasise the referent of the *ang*-phrase in any particular way.

The idea of highlighting or emphasising a particular constituent should be clearly distinguished from the idea that there is a special relation between the *ang*-phrase and the predicate. Undeniably, such a special relationship exists simply because it is only the *ang*-phrase which gets its semantic role directly marked on the predicate (by the 'focus' affix). The semantic role of any further argument of a verbal predicate is not directly marked anywhere in the clause. Instead, it has to be inferred from its noun phrase marker and the fact that it cannot be the semantic role which has been explicitly assigned to the *ang*-phrase. Thus in (4), the fact that *-um-* assigns the agent role to the *ang*-phrase makes it clear that the argument in the *ng*-phrase (*ng bahay*) cannot have this role. It must be an undergoer of some sort. Contrast this with (5) where the 'focus' morphology makes it clear that the argument in the *ang*-phrase bears an undergoer role. Hence, it is most likely that the argument in the *ng*-phrase is an actor.¹⁰

As already mentioned above, the special relationship existing between the predicate and the *ang*-phrase is most adequately characterised as the relationship between a predicate and its syntactic pivot. That is, what is involved here is a syntactic relationship and not some kind of pragmatic highlighting or emphasis. The syntactic nature of the relationship is clearly shown by the fact that the *ang*-phrase has a substantial number of subject properties, such as being the only argument that can launch floating quantifiers, control secondary predicates, be

⁹ One possibility that comes to mind is to use *diathesis*, a term which is used by many (in particular European scholars) as a synonym of *voice*, in a broader sense to cover both valency-changing and valency-neutral alternations concerning syntactic pivots.

¹⁰ The details of determining the semantic roles of non-pivot arguments warrant a much more elaborate discussion than is possible here. Essentially there are four different factors involved: the semantic frame (or lexical-conceptual structure) associated with the predicate, the semantic role explicitly assigned to the pivot, the marker used for the non-pivot argument(s) (in Tagalog either *ng* or *sa*), and last but not least the semantics of the non-pivot argument (a *ng*-phrase referring to a hammer or a knife is, all other things being equal, more likely to be interpreted as an instrument than one which makes reference to an animate being, which will be most naturally interpreted as an agent).

relativised and be omitted in conjunction reduction. None of these properties has anything to do with pragmatic focus.

In conclusion of this section, it may be noted that, in principle, it does no harm to call Philippine voice morphology ‘focus’ as long as it is clearly understood that ‘focus’ is used here in a very special sense which is not in any way directly related to the notion of pragmatic focus. However, as so often, while terminology *per se* is ‘harmless’ when handled with care and consideration, the discussion of Philippine-type ‘focus’ in the last fifty years has shown that this terminology adds unnecessary confusion and complexity to an issue which is already complex and confusing. This even holds true for some of the specialist literature (some of which is documented in Matsuda French 1988), not to mention the confusions in the typological literature which often appear to be misguided by the ‘focus’ terminology (a recent example is Dixon and Aikhenvald’s discussion (1997:89-91) of Philippine languages in their typology of verbal alternations).

Using ‘voice’ instead of ‘focus’ may not only be useful in that it avoids misunderstandings related to the pragmatic meanings of the term *focus*; it may also be of help when delimiting the set of phenomena which are deemed to make up the ‘focus system’. However, changing terminology alone does not solve any of the analytical problems associated with these phenomena. Whatever the thing is called in the end, the most important task is, of course, to provide an explicit analysis of voice phenomena in western Austronesian languages which allows a productive comparison with related phenomena in other language areas and families.

To my mind, there are currently two major approaches to the analysis of voice phenomena in western Austronesian languages. On the one hand, there are various proposals for an ergative analysis of Philippine languages as well as a substantial number of other western Austronesian languages (Indonesian, South Sulawesi, Uma, Balinese, etc.). In these approaches, one of the voices is analysed as the basic unmarked construction for transitive clauses while another voice (usually the so-called actor-‘focus’) is analysed as an antipassive. On the other hand, there is a fairly broad and heterogeneous set of approaches which analyse voice-related phenomena in western Austronesian languages as *valency-neutral alternations* (another term is *symmetrical voice systems*). The basic tenet of these approaches is that the different voices found in these languages do not change the overall transitivity of the constructions in which they occur. Instead, a change of voice signals a change in the alignment of semantic roles and syntactic positions (i.e. in actor voice the actor is the syntactic pivot, in undergoer voice the undergoer is the syntactic pivot, but both constructions share the same transitivity value). Among these approaches are the traditional multiple passive analyses¹¹ as well as the ‘focus’ analysis, although neither of these has been very explicit about the precise nature of the presumed realignment processes.

As mentioned at the beginning of these introductory remarks, it is far from clear that western Austronesian languages all belong to a single basic type with regard to their voice and grammatical relation marking. It is thus possible that an ergative analysis is appropriate for some of these languages while an analysis in terms of valency-neutral alternations may

¹¹ One should not let oneself be misled by the terminology. Most authors who use *passive* in reference to the undergoer voices — including the much scolded Bloomfield — are well aware of the fact that these voices do not affect the transitivity of the overall construction (see for example, Bloomfield 1917:153ff., 1933:173).

be more appropriate for others. In appraising the respective merits of these two analyses it would certainly be useful to stick to specific claims for particular languages rather than making sweeping generalisations for the whole language area — claims which are generally unfounded simply because very little is known about the details of grammatical relation marking in the majority of the languages included in such claims.

In line with this view, the contributions to this volume are concerned with voice-related phenomena in a single language or a small number of closely related languages. Furthermore, they do not focus narrowly on the two or three main clause constructions which may be deemed to constitute the core of the voice marking system of a given language. Instead, they generally include a number of related constructions and morphological items, thus contributing to the slowly growing database of potentially correlating features on which a future typology of these languages can safely be founded.

References

- Artawa, Ketut and Barry J. Blake, 1997, Patient primacy in Balinese. *Studies in Language* 21:483-508.
- Bloomfield, Leonard, 1917, *Tagalog texts with grammatical analysis*, 3 vols. Urbana, Ill.: University of Illinois.
- Bloomfield, Leonard, 1933, *Language*. London: Allen & Unwin.
- Blust, Robert A., 1978, Eastern Malayo-Polynesian: a subgrouping argument. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, 181-234. Canberra: Pacific Linguistics.
- Bybee, Joan, John Haiman and Sandra A. Thompson, eds, 1997, *Essays on language function and language type*. Amsterdam: Benjamins.
- Clynes, Adrian, 1995, Topics in the phonology and morphosyntax of Balinese. PhD thesis, The Australian National University.
- Dixon, R.M.W. and Alexandra Y. Aikhenvald, 1997, A typology of argument-determined constructions. In Bybee et al. eds, 1997:71-113.
- Kroeger, Paul R., 1993, *Phrase structure and grammatical relations in Tagalog*. Stanford: CSLI Publications.
- Matsuda French, Koleen, 1988, The focus system in Philippine languages: a historical overview. *Philippine Journal of Linguistics* 18/19:1-29.
- Reid, Lawrence A., 1992, On the development of the aspect system in some Philippine languages. *Oceanic Linguistics* 31:65-91.
- Schachter, Paul, 1995, *The subject in Tagalog: still none of the above*. UCLA Occasional Papers in Linguistics vol.15, Los Angeles: UCLA/Department of Linguistics.
- Wechsler, Stephen and I Wayan Arka, 1998, Syntactic ergativity in Balinese: an argument structure based theory. *Natural Language and Linguistic Theory* 16:387-442.
- Wolff, John U., 1996, The development of the passive verb with pronominal prefix in Western Austronesian languages. In Bernd Nothofer, ed. *Reconstruction, classification, description — festschrift in honor of Isidore Dyen*, 15-40. Hamburg: Abera-Verlag.
- Wolff, John U. with Maria Theresa C. Centeno and Der-Hwa V. Rau, 1991, *Pilipino through self-instruction*, 4 vols. Ithaca: Cornell Southeast Asia Program.

Wouk, Fay, 1996, Voice in Indonesian discourse and its implications for theories of the development of ergativity. *Studies in Language* 20:361-410.

Zorc, R. David, 1977, *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. Canberra: Pacific Linguistics.

The history and transitivity of western Austronesian voice and voice-marking

MALCOLM ROSS

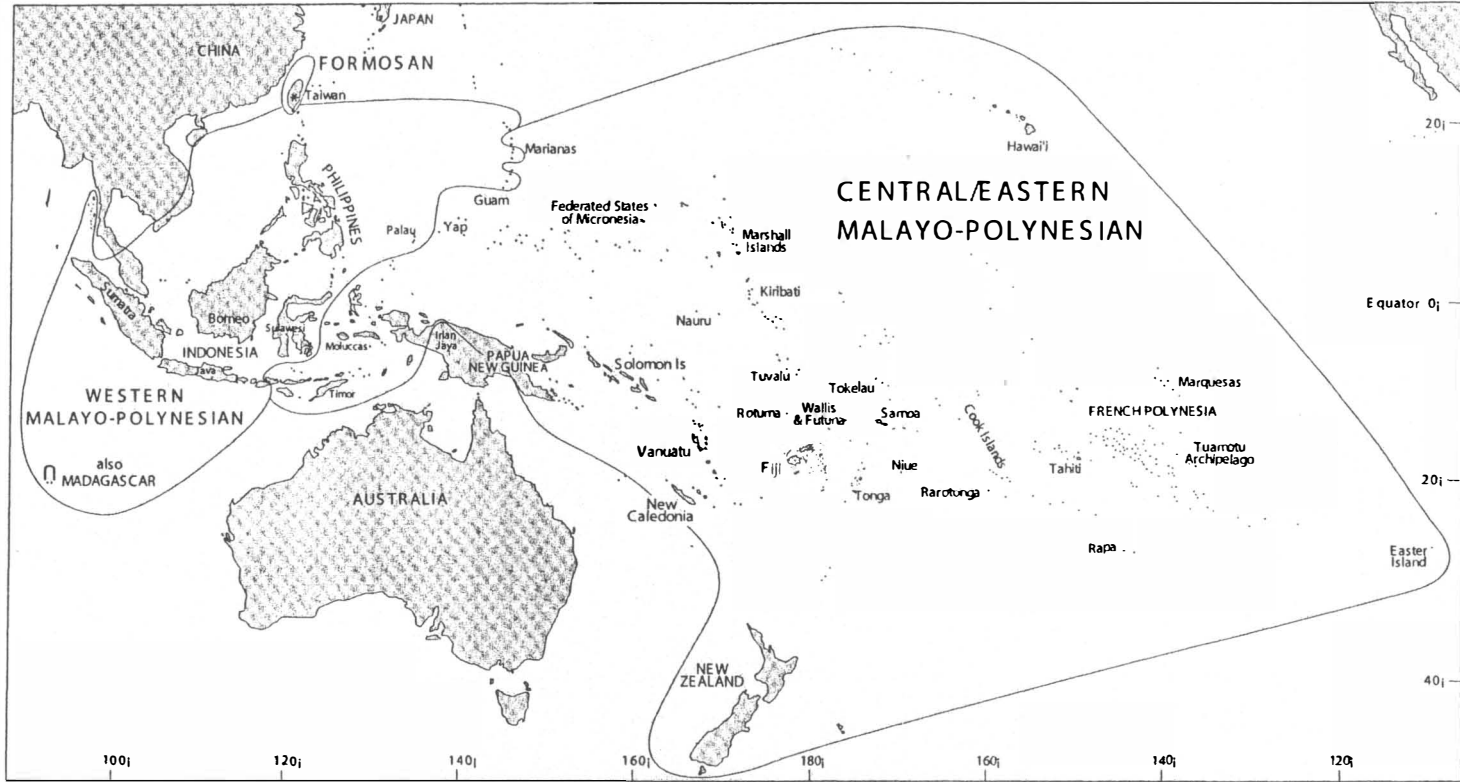
1 Introduction

One of the purposes of the present book is to publish analysed data from western Austronesian languages which will facilitate the reconstruction of the history of voice marking and grammatical relations in the Austronesian language family.¹ This is an area in which reconstruction has not progressed very far since Wolff's landmark reconstruction of Proto Austronesian (PAN) voice, mood and aspect morphemes (Wolff 1973). Arguably the most important development since then was presented in a 1981 paper by Starosta, Pawley and Reid (henceforth SPR), showing how some of this morphology had perhaps developed from nominalising morphemes which are still reflected in many present-day Austronesian languages.² Ross (1995a) summarises these and other contributions and examines evidence from the languages of Taiwan to produce a revised reconstruction of PAN verbal morphology which is not very different from Wolff's original version.³

The subgrouping hypothesis that has gained widest acceptance among Austronesianists is one whose highest nodes are as shown in Figure 1.

The italicised labels *Formosan subgroups* and *Western Malayo-Polynesian subgroups* in Figure 1 refer to sets of languages which each contain more than one subgroup but which do not themselves form a single subgroup. That is, there was — as far as we can tell — no “Proto Formosan”: the only ancestor which all Formosan languages have in common is PAN. And there was — again, as far as we can tell — no “Proto Western Malayo-Polynesian”: the common ancestor of the western Malayo-Polynesian languages, which occupy the large area

-
- 1 We adopt the convention of writing *western Austronesian* with lower-case *w-* because the languages thus labelled do not form a genealogical subgroup, despite their similarities.
 - 2 This paper was never published in its entirety: an abbreviated version appeared as Starosta, Pawley and Reid (1982).
 - 3 I am grateful to Wayan Arka, Robert Blust, John Bowden, Nikolaus Himmelmann and Andrew Pawley for their comments on earlier drafts of this essay, although, of course, the responsibility for its contents is mine.



Map 1: The Austronesian family and major Austronesian language groups

shown in Map 1, is Proto Malayo-Polynesian (PMP).⁴ Subgrouping among western Malayo-Polynesian languages in particular is controversial. The reasons for this are of two kinds. One is simply that much of the research which would be needed to determine well-founded subgroups has not been done. The other is that contact over millennia between neighbouring languages, together with the use of Malay as a lingua franca among speakers of many western Malayo-Polynesian languages, has altered much of the evidence that might otherwise have been used to determine subgroups.

Under the hypothesis represented in Figure 1, the Formosan languages represent a number of primary Austronesian subgroups (Blust 1999b:53-55), but all Austronesian languages outside Taiwan belong to a single subgroup, dubbed Malayo-Polynesian by Blust (1977).⁵ Since the reconstruction of a proto language should be based on evidence from more than one primary subgroup, this gives the Formosan languages considerable significance in the reconstruction of PAn. Section 3 is thus a potted version of Ross' (1995a) Formosan-based reconstruction of PAn verbal morphology with some revisions and additions, including an alternative explanation of the data (§3.2.2).

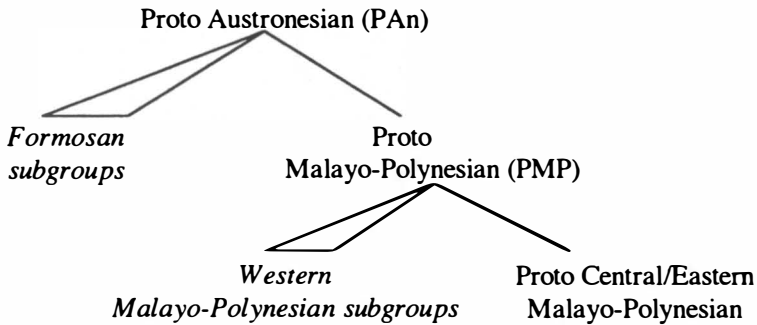


Figure 1: The uppermost nodes of the Austronesian genealogical tree (after Blust 1977)

Recently, Starosta (1995) has revised his view of the PAn system of grammatical relations, proposing that certain Formosan languages separated from the rest of the early Austronesian family before a system like that reconstructed by Wolff (1973), SPR and Ross (1995a) came into being.⁶ If Starosta is correct, then this means that a system of the kind reconstructed by Wolff, SPR and Ross arose not in PAn but in an interstage language a node or two below it in the Austronesian genealogical tree. This interstage would come between PAn and PMP in the tree in Figure 1 and would still be the ancestor of all Austronesian languages except perhaps four or five of those in Taiwan — as well as the ancestor of all the languages considered in this book. The conventional alternative to Starosta's revised

4 See Pawley and Ross (1993) and Ross (1995b) for summary reviews of Austronesian subgrouping. Blust (1999:68) also stresses that there is no Western Malayo-Polynesian subgroup of Austronesian. One largish subgroup within western Malayo-Polynesian has recently received stronger support, however: this is Malayo-Chamic (Thurgood 1999).

5 It is usually assumed that Malayo-Polynesian is a primary subgroup of Austronesian. However, Reid (1982) suggests that PMP may subgroup with one or more Formosan languages, and this is at least circumstantially likely.

6 There is an important difference in methodology between Starosta's reconstruction and the others mentioned here, since, as Blust (1999:62-67) points out, Starosta compares and reconstructs morphosyntactic types but not forms.

hypothesis is that the PAN system was indeed as reconstructed by Wolff, SPR and Ross, and that the Formosan languages which display other systems have undergone substantial innovations. This analytic disagreement arises largely because PAN is at the top of the tree. That is, we must reconstruct it entirely on the basis of its daughter-languages, whose primary subgrouping we are uncertain about. If we were debating the reconstruction of PMP instead, we could also draw on data from external witnesses (i.e. Formosan languages) for evidence about the kind of system that PMP inherited. We lack this corroborating external evidence when we reconstruct the language at the top of the tree.

Of the languages whose verbal morphology and grammatical relations are described in this book, Seediq is a Formosan language, and the others are western Malayo-Polynesian. Because of this bias, §4 offers a sketch of probable changes which had occurred in Proto Malayo-Polynesian and of subsequent developments among western Malayo-Polynesian languages. No account is taken here of the contributions in this book, as this is the task of the discussion notes by Wolff and Ross.

2 Transitivity and ergativity

2.1 Morphosyntax

The terminology employed here follows that used by Himmelmann in his introductory contribution to this book (henceforth ‘Himmelmann’ without further specification). PAN was a Philippine-type language in Himmelmann’s terminology. Note, though, that “Philippine-type languages” include not only Philippine languages but also some of the languages of northern and central Borneo, northern Sulawesi and Madagascar, as well as most of the Formosan languages. In a Philippine-type voice system, the semantic role of the syntactic pivot (the Philippinists’ ‘topic’) is marked by verbal affixes. The (made up) examples in (1) illustrate the major affixes for the four voices in Paiwan, a Formosan language (see Map 4). The four examples contain respectively the suffix *-ən* ‘patient voice’, the suffix *-an* ‘location voice’, the prefix *si-* ‘circumstantial voice’⁷ and the infix *<əm>* ‘actor voice’. In each case the syntactic pivot introduced by the specific phrase marker *a* assumes the role indicated by the verbal affix:

- (1) Paiwan⁸
- a. *təkəl-ən a vaua*
 drink-PV SPEC wine
 ‘the wine will be drunk’ (‘s/he/they will drink the wine’)

7 The circumstantial voice is commonly known in the literature as the ‘instrumental voice’, but its uses are usually wider. Keenan (1976:256) writes with regard to Malagasy: ‘subjects of circumstantial sentences can express the instrument, benefactee, location, time, purpose, manner...of that action.’ A common feature of its uses is that the syntactic pivot refers to something that is moved or is the goal of movement but is not affected by the event.

8 Abbreviations used in interlinear glosses: 1,2,3 first, second, third person; 1EP first person plural exclusive; 1IP first person plural inclusive; AT atemporal; AV actor voice; CJ conjunction; CV circumstantial voice; D disjunctive pronoun; GEN genitive (phrase marker or pronoun); IMPF imperfective; INVOL involuntary; IRR irrealis; LIG ligature; LOC location (phrase marker); LV location voice; NEG negative auxiliary; NPV non-pivot (=neither pivot nor agent); P plural (phrase marker or pronoun); PAN Proto Austronesian; PERS personal (phrase marker); PF perfective; PIV pivot; PMP Proto Malayo-Polynesian; PN personal (phrase marker); PV patient voice; R reduplication; RECIP reciprocal; S singular pronoun; SPEC specific (phrase marker); UV undergoer voice.

- b. *təkəl-an a kakəsan*
 drink-LV SPEC kitchen
 'the kitchen will be drunk in' ('s/he/they will drink it/them in the kitchen')
- c. *si-təkəl a kupu*
 CV-drink SPEC cup
 'the cup will be drunk with' ('s/he/they will drink it/them from a cup')
- d. *təməkəl a qata*
 <AV>drink SPEC stranger
 'the stranger will drink (something)'

It is appropriate to refer to the patient, location and circumstantial voices collectively as 'undergoer voices', as they have certain features in common (see below).

The voice-marked verb forms in (1) are worthy of comment. First, the fact that two voices are marked by suffixes, one by a prefix, and one by an infix is unusual crosslinguistically. I return to this in §3.2.1. Second, although 'drink' happens to have four voice forms in Paiwan, in Philippine-type languages generally neither the morphological shape nor even the occurrence of a particular voice form of a given verb is completely predictable. This means that voice forms must be listed in the lexicon, i.e. they are derived, not inflected, forms,⁹ and are more similar to the applicative verb-forms of, for example, Oceanic Austronesian languages than to the fully productive, largely predictable passive of a language like English. Recognising this, Starosta (1986) proposes the term 'recentralisation' instead of 'voice' in Philippine-type languages. The main reason we retain 'voice' here is that it is already well entrenched and is a decidedly better term than the Philippinists' 'focus' (see Himmelmann).

Starosta's account is important in another respect. The effect of applicative verb-forms is generally to allow a referent with a semantic role other than patient (e.g. location, instrument, beneficiary) to become the undergoer. This is the effect of *-an* and *si-* in (1b–c). We might therefore regard the patient voice in (1a) as the basic undergoer voice and (1b–c) as undergoer voice applicatives. I have decided against this analysis here because there is no morphological evidence that the patient voice with *-ən* in (1a) is more basic than those with *-an* and *si-*, and there is therefore no pressing argument for moving away from more conventional terminology.

A crucial feature of Philippine-type voice systems is that some of them seem to entail no reduction in valency (see Himmelmann). However, this is a matter of controversy to which I return below (§2.3). The Paiwan sentences in (2) each contain two noun phrases, one the syntactic pivot introduced by *a*, the other either the agent, marked with the genitive phrase marker *nua* or the patient, marked with the non-pivot phrase marker *tua*:

⁹ For a succinct statement of the claim that 'voice' affixes in Philippine languages are derivational, not inflectional, see Reid (1992:67–68). For similar views, see Starosta (1986) and Himmelmann (1991). De Guzman (1997) argues the opposing case, but her survey (318–322) suggests rather that voice forms are derived, but form a larger part of the language and show more regularities than derived forms in many languages.

(2) *Paiwan*

- a. *təkəl-ən nua qata a vaua*
 drink-PV GEN stranger SPEC wine
 'the wine will be drunk by a/the stranger'
 ('a/the stranger will drink the wine')
- b. *təkəl-an nua qata a kakəsan*
 drink-LV GEN stranger SPEC kitchen
 'the kitchen will be drunk in by a/the stranger'
 ('a/the stranger will drink it/them in the kitchen')
- c. *təkəl-an a kakəsan tua vaua*
 drink-LV SPEC kitchen NPIV wine
 'the kitchen will have wine drunk in it'
 ('someone will drink wine in the kitchen')
- d. *təməkəl a qata tua vaua*
 <AV>drink SPEC stranger NPIV wine
 'the stranger will drink wine'

The most-cited Philippine-type language is Tagalog (see Map 4). The phrase markers of Paiwan and Tagalog are shown in (3).

(3)		SPEC	GEN	NPIV	LOC
	Paiwan	<i>a</i>	<i>nua</i>	<i>tua</i>	<i>tua</i>
	Tagalog	<i>ang</i>	<i>ng</i> [naŋ]	<i>ng</i> [naŋ]	<i>sa</i>

It is common in the literature to refer to a marker with the functions of Paiwan *a* and Tagalog *ang* as the marker of the syntactic pivot (or whatever the corresponding term is in each writer's terminology), but Himmelmann (forthcoming a) points out for Tagalog that this is not strictly accurate. *Ang* also occurs in predicate noun phrases, and marks the noun phrase as specific. The same is true of phrase markers in other Philippine-type languages which correspond to *ang*, and so these markers are glossed here as specific.

2.2 Functions and 'discourse ergativity'

Despite the controversy about transitivity, however, it is clear that in many Philippine-type languages there is an important distinction between the undergoer voices and the actor voice. The undergoer voices are the unmarked choice in a number of respects, the actor voice the marked choice.¹⁰ (This leads to the curious situation that in a language like Paiwan with patient, location and circumstantial voices, there are in this sense three 'unmarked' choices and one marked). In many Philippine-type languages there is a general requirement that the syntactic pivot have a specific referent, and the actor voice is reserved for independent transitive clauses where the undergoer referent is not specific,¹¹ and for subordinate clauses where the syntax requires an actor pivot. For example, in a relative clause the (deleted) noun

¹⁰ I owe much of the correlation of the sources referred to in this paragraph and its footnotes to S. Huang (2000).

¹¹ Patient specificity has been noted as a major determinant of voice in Seediq (Holmer, this volume), Tsou (S. Huang 2000), Yami (Ho 1993), Kapampangan (Mirikitani 1972), Tagalog (Wouk 1986), Cebuano (Bell 1988) and Karao (Brainard 1994).

phrase coreferential with the head noun is the pivot, and the actor voice is used if this is the agent. The undergoer voices are the default choice for foregrounded (= story-line) events in discourse (where the active voice is the default in European languages), and the actor voice tends to be reserved for marked functions. These observations are related to one made by Cooreman, Fox and Givón (1988) about topicality, where ‘topicality’ refers to the discourse continuity of a referent. Across languages, agents are more topical in discourse than undergoers, and this is as true of Philippine-type languages as of others, but it is undergoer topicality that affects Philippine-type voice selection: the higher the topicality of the undergoer referent, the greater the probability that it will be selected as syntactic pivot.¹² However, these observations cannot be taken as definitive, as there has not yet been enough research on the uses of the different voice forms in the discourse of Philippine-type languages for us to be certain how widely these generalisations hold.¹³

The effects of default undergoer pivot choice can be seen in (4), drawn from a Paiwan text.¹⁴ In English, the action sequence is best translated with a sequence of active verbs (‘loosened...saw...crushed...ate’), but in Paiwan the normal choice is a sequence of undergoer voice verbs (in bold). The passage is semi-literally translatable into English as ‘That monkey, the stones were loosened (by him), the water became muddy, the crabs were seen (by him), and (they) were crushed (by him) and (they) were eaten (by him).’ The semi-literal translation reads poorly because a major function of the English passive is to suppress the actor, a function not shared by the Paiwan patient voice.

(4) Paiwan

<i>a</i>	<i>zu'</i>	<i>a</i>	<i>ti</i>	<i>sa</i>	<i>ɬaiɬail</i>	<i>cəkətən</i>	<i>a</i>	<i>zu'</i>	<i>a</i>	<i>qacitai</i> ,
a	zua	a	ti	sa	ɬaiɬail	cəkət-ən	a	zua	a	qacitai
SPEC	that	LIG	PN	RESPECT	monkey	loosen-PV	SPEC	that	LIG	stone
<i>matimək</i>	<i>a</i>	<i>zatum</i> ,	<i>pacunan</i>	<i>a</i>	<i>zu'</i>	<i>a</i>	<i>ganj</i> ,	<i>qucəqucən</i>		
ma-timək	a	zatum	pacun-an	a	zua	a	ganj	R-quc-ən		
PASSIVE-mud	SPEC	water	see-PV	SPEC	that	LIG	crab	DUR-crush-PV		
<i>sa</i>	<i>kani</i>	<i>aya</i> .								
sa	kan-i	aya								
and.then	eat-PV.AT	thus								

‘That Mr Monkey, he loosened some stones, the water became muddy, he saw the crabs, and crushed and ate them.’

In natural discourse, verbs in Philippine-type languages often have no noun phrase accompanying them, like the last two verbs in (4), or only one, like the other verbs in the example. Verbal clauses which have two full noun phrases like the constructed examples in (2) are rare.

Observations such as those summarised above have led to a labelling of Tagalog and various other western Austronesian languages as ‘discourse-ergative’. This is a rather confusing use of the term ‘ergative’, as Cumming and Wouk (1987) show in a critique of ‘discourse ergativity’ in Austronesian languages. If ‘ergativity’ refers to a system in which

¹² This has been shown for Tsou (S. Huang 2000), Chamorro (Cooreman 1983, 1987; Cooreman et al. 1984, 1988), Tagalog (Cooreman et al. 1984, 1988) and Cebuano (Payne 1994).

¹³ The undergoer specificity criterion evidently does not apply to Ilocano (Baker 1991). The foreground/background distinction does not apply to Tsou (S. Huang 2000) or to Cebuano (Bell 1988).

¹⁴ The text is from Egli (1990:326-343); the interlinear glosses and free translation are mine.

the single argument of an intransitive verb (the S) and the non-actor (undergoer or O) argument of a transitive verb are treated in the same way, but differently from the actor (the A), then 'morphological ergativity' refers to a system in which S and O are marked by the same morphology, which is different from the marking of A, and 'syntactic ergativity' to a system in which S and O are subject to the same syntactic processes (see below) (Dixon 1979). In the corresponding 'accusative' systems it is S and A that are treated in the same manner. 'Discourse ergativity' ought then to refer to some system which treats S and O in the same way with regard to some discourse process, in contrast to a discourse-accusative system. What it in fact seems to refer to in Cooreman, Fox and Givón's (1984) usage is a preference for foregrounded transitive clauses whose syntactic pivot is O rather than A.

Note that it makes no difference to the discourse ergativity of Tagalog whether the actor voice is transitive or not. Since discourse ergativity is a preference for foregrounded transitive clauses whose syntactic pivot is O, a discourse-ergative language must allow such clauses. It does not matter whether it also allows transitive clauses with an A pivot. The one exclusion is that a discourse-ergative language cannot be syntactically accusative.¹⁵

2.3 Syntactic ergativity and the transitivity of the actor voice

Philippine-type languages have long been sources of puzzlement and controversy among syntacticians, and some of these issues are touched upon by Himmelmann. A significant feature of Philippine-type languages for linguists is that they force us to deconstruct the categories that we use in morphosyntactic analysis. One question which has been raised again and again in more recent literature is: Are some or all Philippine-type languages syntactically ergative? The hypothesis put forward by those who answer in the affirmative (e.g. De Guzman 1988; Gibson & Starosta 1987; Starosta 1988, 1999) can be simply stated:

The ergative hypothesis: Undergoer-voice clauses are transitive, actor-voice are intransitive.

The converse claim, that Philippine-type languages are syntactically accusative, with intransitive undergoer-voice clauses and transitive actor-voice clauses (e.g. Bell 1976), has faded from discussion, and there seems to be a consensus that undergoer-voice clauses like (2a–c) are transitive. A third claim, still on the table, is implicit in Kroeger's (1993:40–48) work:

The symmetrical-voice hypothesis: Both undergoer-voice and actor-voice clauses are transitive.

Although the symmetrical-voice hypothesis seems to hold for a number of Indonesian-type languages (§4.2), it is less clear that it is true of Philippine-type languages. The difference between the two hypotheses boils down to a single question: are actor-voice clauses in some or all Philippine-type languages transitive or intransitive?

The problem with this question is that it presupposes a crosslinguistically valid definition of 'transitive' and 'intransitive'. Dryer (1997), writing about grammatical relations, suggests

¹⁵ If we label a language with both O- and A-pivot transitives as 'symmetrical', then the entailments are: a discourse-ergative language is syntactically ergative or symmetrical; a discourse-accusative language is syntactically accusative or symmetrical; a syntactically ergative language cannot be discourse-accusative; a syntactically accusative language cannot be discourse-ergative; a symmetrical language may be discourse-ergative or discourse-accusative.

that grammatical relations are like phonemes: it makes no sense to define them crosslinguistically. There are, he suggests, language-particular grammatical relations, as varied as those found in Dyirbal, Acehnese, Cree and Cebuano. There may be similarities between the grammatical relations of these languages and common explanations for these similarities, but crosslinguistic labels like 'subject' belong to the metalanguages of various theories and lack empirical substance (cf. Dryer 1999).

Of Dryer's four examples, Cebuano is a Philippine-type language similar in structure to Tagalog, and he points back to Schachter's (1976, 1977) famous deconstruction of the Tagalog subject into the pivot and the actor. The pivot has reference-related functions which include being (i) referential; (ii) uniquely capable of relativisation; (iii) modified by a floating quantifier or a depictive predicate; (iv) the controller of raising; and (v) the controllee in a raising construction (Schachter 1976; Kroeger 1993). The actor has role-related functions.¹⁶ It is (i) the antecedent of reflexives; (ii) the controllee in equi constructions; and (iii) the imperative addressee.¹⁷

In a language like English, the pivot and the actor of a transitive clause coincide as the subject: such a language is syntactically accusative. In other languages they remain separate, and the pivot coincides instead with the undergoer.¹⁸ such a language is syntactically ergative. This formulation is due to Manning (1996:16-20 and *passim*). His crucial insight is that role-related functions are carried by the actor — and sometimes also by the undergoer — regardless of whether the language is syntactically accusative or syntactically ergative, and so these functions can be discounted in making the accusative/ergative distinction.¹⁹ Dyirbal and Tagalog, according to Manning, are syntactically ergative by this criterion, and Tagalog is thus made to look more ordinary than it has looked from earlier perspectives. (I will modify this assessment below.)

In a syntactically accusative language, there is typically a passive: it is intransitive, and the pivot coincides with the undergoer. In a syntactically ergative language, there is typically an antipassive: it is intransitive, and the pivot coincides with the actor. This takes us back to the debate about the transitivity of the actor voice in Philippine-type languages. If the actor voice in, say, Tagalog is intransitive, then it is an antipassive, and the language is

16 Not all languages have a syntactic pivot in this sense. In so-called 'split-S' languages like Acehnese (Austronesian), grammatical relations are based on actor and undergoer (Durie 1987). In Yimas (Papuan) privileged arguments differ from construction to construction (Foley 1993). In neither case is there an single pivot.

Actor and undergoer (Foley & Van Valin 1984) are quasi-semantic relations, in the sense that they are 'macro-roles': an actor is sometimes an agent, sometimes a force, sometimes an experiencer, and so on, and an undergoer is variously a patient, a theme, a beneficiary etc. They are semantic abstractions which receive *grammatical* expression in various ways in various languages.

17 For examples, see Schachter (1976, 1977), Kroeger (1993) and Manning (1996). Schachter (1984) describes a similar distribution of functions in Toba Batak.

18 For these reasons Kroeger's (1993) Lexical-Functional-Grammar-motivated use of 'subject' for the Tagalog pivot may sit uncomfortably with some linguists. Differences between Tagalog and English subjects are not limited to the fact that the English subject has role-related properties as well as reference-related. Unlike a Tagalog pivot, an English subject is not the only relation that can be relativised, nor is it the only controller of equi deletion and raising. These differences serve to underline Dryer's claim that a grammatical relation like 'subject' is not universal.

19 This represents a refinement of Dixon's (1979, 1994) account of syntactic ergativity, in which role- and reference-related functions are not distinguished. Manning's observations apply, incidentally, to only a subset of the world's languages: they do not apply to languages which lack a pivot.

syntactically ergative. But if it is transitive, like the undergoer voices, then there is no transitive/intransitive voice contrast and Tagalog has a symmetrical voice system.

But here a further deconstruction is needed. What does it mean to say that a clause is transitive? Unfortunately, in the last twenty years, ‘transitive’ has come to be used in at least two different senses, one semantic, the other morphosyntactic. We will see below that the semantic transitivity of the actor voice in Tagalog is ambiguous, or, more accurately, that its intransitive interpretation apparently depends on pragmatic inference. Morphosyntactic transitivity, as conventionally defined, depends on being able to determine whether a clause has a minimum of two core arguments. Tagalog, however, forces us to deconstruct the notion of ‘core’ into criteria which match in many languages — but not in Tagalog. I will deal with semantic and morphosyntactic transitivity separately in the next two sections.

2.3.1 *Semantic transitivity*

‘Semantic transitivity’ (perhaps one should call it ‘functional transitivity’) derives from the work of Hopper and Thompson (1980) and consists of features of the clause which include agentivity, perfective aspect, and individuation of the undergoer. ‘Individuation’ includes, among other things, specificity, and it is often pointed out that the undergoer of an actor-voice clause in a Philippine-type language is non-specific. We can begin to get a handle on this by looking at the (apparently elicited) Tagalog sentences in (5).

(5) Tagalog (Schachter 1976:494-495)

- a. *Mag-alis ang babae ng bigas sa sako para sa bata.*
AV-take.out SPEC woman NPIV rice LOC sack for LOC child
‘The woman will take some rice out of a/the sack for a/the child.’
- b. *A-alis-in ng babae ang bigas sa sako para sa bata.*
DUR-take.out-PV GEN woman SPEC rice LOC sack for LOC child
‘A/the woman will take the rice out of a/the sack for a/the child.’
- c. *A-alis-an ng babae ng bigas ang sako para sa bata.*
DUR-take.out-LV GEN woman NPIV rice SPEC sack for LOC child
‘A/the woman will take some rice out of the sack for a/the child.’
- d. *Ipag-alis ng babae ng bigas sa sako ang bata.*
take.out-CV GEN woman NPIV rice LOC sack SPEC child
‘A/the woman will take some rice out of a/the sack for the child.’

(6) Tagalog (De Guzman 2000:227)

- Nag-tanong ang bata sa/*ng kapitbahay*
PF.AV-ask SPEC child LOC/*NPIV neighbour
‘The child asked the neighbour.’

As the free translations indicate, *ng bigas* ‘some rice’ is interpreted in (5a), (5c) and (5d) as non-specific. Hence some scholars consider the actor voice in (5a) not to be transitive. Though rare, a definite patient in an actor voice clause may be expressed with *sa* LOC, as in (6). However, in (7a) and (8), where the actor voice verb occurs in a relative clause, the *ng*-phrase (henceforth ‘*ng*-patient’) may have either a non-specific or a specific

interpretation.²⁰ To guarantee a definite interpretation in (7), *sa* LOC may be used instead of *ng*, but this option is not available with the trivalent verb in (8). When the patient is a personal noun phrase, it is always specific and always preceded by *kay* PERS.LOC (examples from Schachter and Otnes 1972:382-383).

(7) Tagalog

- a. *Siya ang naka-kita ng aksidente.*
 NOM:3S SPEC PF.INVOL.AV-see NPIV accident
 'He's the one who saw a/the accident.'
- b. *Siya ang naka-kita sa aksidente.*
 NOM:3S SPEC PF.INVOL.AV-see LOC accident
 'He's the one who saw the accident.'
- c. *Siya ang naka-kita kay Jose.*
 NOM:3S SPEC PF.INVOL.AV-see PERS.LOC Jose
 'He's the one who saw Jose.'

(8) Tagalog

Siya ang nag-bigay ng premyo kay Ben.
 NOM:3S SPEC PF.AV-give NPIV prize PERS.LOC Ben
 'He's the one who gave a/the prize to Ben.'

The crucial question is, what is the status of *ng*-patient in these examples? The provisional answer is that it is a grammatical relation which encodes the common noun phrase patient of a non-patient voice. On the basis of (5), it is tempting also to say that the *ng*-patient encodes a non-specific patient and that the actor voice is therefore inherently less transitive than the undergoer voices. The evidence of (7) and (8), however, suggests that this is an oversimplification. The relevant facts appear to be these:

- a. The pivot, marked with *ang*, must be specific.
- b. *Ng* encodes the common noun phrase patient (*ng*-patient) of a non-patient voice, as in (5a), (5c) and (5d).
- c. In an independent clause, a specific common noun phrase patient will almost always be the pivot, as in (5b). This means that the *ng*-patient of an independent clause will almost always be interpreted as non-specific, as in (5a), (5c) and (5d).
- d. In a relative clause, the relativised noun phrase must be the pivot. If this is the actor, it may block a specific common noun phrase patient from being the pivot, resulting in an actor voice relative clause with a specific *ng*-patient, as in (7a) and (8).
- e. To guarantee an interpretation of a specific common noun phrase patient as definite, *ng* NPIV may be replaced by *sa* LOC, as in (7b). But this option is blocked if the verb is trivalent and there is therefore another a LOC-marked core argument, as in (8).

One conclusion can be drawn straight away from these facts: although the *ng*-patient of an independent clause is *interpreted* as non-specific, *ng* does not *encode* non-specificity. Instead, the non-specificity of the *ng*-patient is a pragmatic inference based on the fact that a specific common noun phrase patient would normally be the pivot (c). When it is blocked

²⁰ 'Relative clauses' here also includes the cleft construction and the existential construction (Kroeger 1993:55; Himmelmann forthcoming b).

from being pivot, as in a relative clause, the inference is not necessarily made (d). And, as Himmelmann (forthcoming b) points out, there are rare cases when the *ng*-patient of an independent clause may be specific, e.g. when the *ng*-patient is owned by the actor.

This has a bearing on a claim made by Hopper and Thompson (1980:289) about Tagalog. They write that semantic features of high transitivity (and these include specificity of the undergoer) tend to be collectively grammaticised across languages in transitive clause constructions. The undergoer voices in (5b–d) represent this kind of grammaticisation. They observe in a discussion of Tagalog data, however, that the actor voice is further down their transitivity continuum, as the undergoer is non-specific, i.e. unindividuated. They seem to imply that the actor voice thus represents the grammaticisation of lower transitivity. On the basis of the facts listed above, however, although the actor voice may express lower transitivity in independent clauses, it does not represent its grammaticisation.

Note that this conclusion about the actor voice is not drawn on the basis of a simple opposition between it and the undergoer voices. Although discussion of the status of the actor voice has often been couched in terms of this opposition, the conclusion depends crucially on the status of the *ng*-patient, and this may occur in any non-patient voice, as (5) shows.

2.3.2 *Morphosyntactic transitivity*

Since a transitive clause is one with a pivot plus at least one more core argument, the issue of the morphosyntactic transitivity of the actor voice boils down to the question, is the *ng*-patient in Tagalog core or oblique? This entails being able to define ‘core’, however, and it seems that in conventional definitions, there are three conditions for an argument being ‘core’:

- (a) The argument has a morphosyntactic relationship to the verb. This relationship may be marked by coding on the verb (e.g. agreement affixes), by coding on the arguments (e.g. case-marking), or by position in the clause. At the same time, the argument is not oblique: an argument is oblique if an argument with the same structure may also occur as a peripheral argument (one not required by verbal valency), as in *I was working on the floor*.
- (b) The argument is required by the valency of the verb (or, ‘subcategorised for by the verb’). This is a necessary, but not a sufficient condition, as verbal valency may also require an oblique argument, as in *I gave the apple to the man* or *I put the apple on the floor*.
- (c) The argument has reference-related functions. If the argument is not the pivot, then it will have fewer reference-related functions. This again is a necessary, but not a sufficient condition, as in some languages an oblique argument may also have reference-related functions.

Since the only sufficient condition is (a), this is the one which will ultimately determine whether the *ng*-patient is a core argument. However, it is hard to distinguish between core and oblique arguments in Tagalog. Across languages, an oblique is typically coded by a special structure, usually an adpositional phrase. But Tagalog usually does not use a special structure in peripheral phrases. Instead, a peripheral phrase, like a core argument, is coded only with a phrase marker.

Out of the phrase markers *ng* GEN, *ng* NPIV and *sa* LOC, only *sa* unambiguously introduces an oblique. It introduces phrases required by the valence of the verb, like *sa sako* in (5a,b,d), as well as peripheral adjuncts of time and place (Schachter & Otanes 1972:440-441, 450-452).²¹

The situation with *ng* is less clear, and is compounded by the fact that *ng* has at least two functions. Although I have glossed *ng* as GEN when it marks the agent of a non-actor voice and as NPIV when it marks the patient of a non-patient voice, there is little doubt that these are two functions²² of the same morpheme (rather than two homophonous morphemes), since both can be replaced by the genitive form of the deictic pronoun (Schachter & Otanes 1972:382-383). Since *ng* GEN marks a core argument, one may infer that *ng* NPIV does so too, otherwise we would have the unlikely scenario of the same morpheme marking both core and oblique arguments.

However, things are not as simple as this. *Ng* also introduces an instrument phrase, as in (9).²³

(9) Tagalog (Foley & Van Valin 1984:135)

Binilh-an ng lalake ng isda ng pera ang tindahan.
 ⚭PF⚭buy-LV GEN man NPIV fish INSTRUMENT money SPEC store
 'The man bought fish in the store with money.'

The instrument phrase in (9) appears not to be required by verbal valency, so it is hard to argue that this is a core noun phrase, as it doesn't satisfy (b) above. Instead, it is an oblique. One could argue, incidentally, that since agents and instruments are marked in the same way in many languages, this is a 'subfunction' of the agent use.

A morpheme *nang* (homophonous with *ng*) introduces temporal peripheral phrases, contrasting with *sa* in contexts like the one in (10).

(10) Tagalog (Schachter & Otanes 1972:440)

- a. *Dumating kami roon sa umaga.*
 ⚭AV⚭arrive PIV:2EP there LOC morning
 'We arrived there in the morning.'
- b. *Dumating kami roon nang umaga.*
 ⚭AV⚭arrive PIV:2EP there ? morning
 'We arrived there of a morning.'

The fact that the contrast between *nang* and *sa* here is one of specificity, parallel to the one noted for *ng* NPIV and *sa* in (7a-b), implies that *nang* and *ng* NPIV are in a sense subfunctions of a single function, and that the arguments they mark are obliques.

If the formulation under (a) above is correct, then the only way to unite the functions of *ng* is to infer that it marks obliques (thereby overturning my previous assumption about *ng* GEN). This would mean that Tagalog had a symmetrical voice system in which all voices were morphosyntactically intransitive, i.e. a system that was the converse of the one proposed in the symmetrical-voice hypothesis above whereby all voices are morphosyntactically transitive.

21 As a peripheral adjunct, *sa* may be preceded by *para*, as in (5a), (5b) and (5c), which marks a beneficiary.

22 For a different view, namely that *ng* always marks attributes, see Naylor (1980).

23 I am grateful to Wayan Arka for discussion of this point.

When we turn to condition (b), we get a different perspective. Crosslinguistically it is not uncommon for a verb to have a valency of three. Three voices of the root *alis* 'take out' in (5) are trivalent: they require an actor, a patient, and a location, marked respectively with *ng* GEN, *ng* NPIV and *sa* LOC when they are not the pivot. This means that in (5a–b) we have a trivalent pattern of *ang*, *ng*, *sa* and in (5c) a trivalent pattern of *ang*, *ng*, *ng*. These patterns also occur with other trivalent verbs in Tagalog. The voices of the root *bigay* 'give' have an actor, a patient (the thing given), and a location (the recipient) (Schachter 1976:506, 1977:280–281). Those of the root *hiram* 'borrow' also have an actor, a patient (the thing borrowed), and a location (the source) (Schachter 1977:294). Those of causative verbs like *pa-luto* 'cause (someone) to cook' have an actor (the causer), a patient (the thing cooked), and a location (the causee = the cook) (Ramos 1971b:148). The pattern of phrase markers is the same in each case.

Across languages, trivalent verbs usually have three core arguments, as in *I gave the man the apple*, or two core and one oblique, as in *I gave the apple to the man*. I am not aware of languages that have trivalent verbs with one core and two oblique arguments. One may thus infer from the trivalent patterns of Tagalog verbs that *ng* GEN and *ng* NPIV code core arguments and *sa* an oblique argument. This inference is supported by (8), where core *ng* NPIV is blocked by the presence of another LOC-marked argument from replacement by oblique *sa* LOC, as this would result in one core and two oblique arguments, all of them required by the valency of the verb, and this would be a crosslinguistically unusual pattern. However, this evidence is circumstantial, as it is based on a crosslinguistic generalisation which is assumed to have no exceptions, and typology indicates that exceptionless generalisations are rare.²⁴

This discussion has an interesting consequence: the circumstantial voice in (5d) appears to have a valency of four. This is crosslinguistically uncommon, and is the consequence of 'promoting' a peripheral argument, the beneficiary, to pivot, without the loss of any of the arguments required by the trivalent voices.

Condition (c) above requires that a core argument have reference-related functions, albeit fewer than the pivot. This is the mainstay of Kroeger's (1993:40–48) claims that arguments introduced by both *ng* GEN and *ng* NPIV are core. I will not repeat the evidence here, but it is clear that *ng* GEN and *ng* NPIV do have a few reference-related functions, although *ng* NPIV has very few.

What are we to make of this? Since (b) and (c) are not sufficient conditions for core status, but (a) is, should we accept the conclusion from (a) that *ng* GEN and *ng* NPIV mark oblique arguments? This would be the result of applying strict logic, but it would make trivalent verbs display a very odd pattern and, more generally, commit us to the position that there are no transitive clauses in Tagalog.

There is an alternative. This is to extend Dryer's (1997) position a little and to say not only (i) that there are no crosslinguistic grammatical relations, only similarities among language-particular grammatical relations, but also (ii) that there are no crosslinguistic categories of core and oblique, only similarities among language-particular encodings of arguments. On this understanding, Tagalog happens to be different from the majority of (non-Philippine) languages in lacking a morphosyntactic distinction between core and oblique arguments other than the pivot. but similar to them in the application of (b) and (c) to certain (morphosyntactically undistinguished) arguments.

²⁴ In principle, Tagalog might be the exceptional language in which a trivalent verb has one core and two oblique arguments.

2.3.3 Summary

To summarise, the *ng*-patient is a Tagalog-specific grammatical relation which encodes the common noun phrase patient of a non-patient voice. There is no unambiguous way to say that either the *ng*-agent or *ng*-patient is core or oblique, and therefore no unambiguous way of talking about the transitivity of the clauses. All we can say is that the system is morphosyntactically symmetrical.

In an independent clause, the *ng*-patient is interpreted as non-specific, but this is a matter of pragmatic inference, not of grammaticisation.

The antipassive-like character of the actor voice is — it seems — an epiphenomenon rather than something encoded by the grammar. But it is now clear why some linguists have been unhappy describing actor voice clauses as ‘transitive’: its application to a clause whose patient is interpreted as non-specific seems anomalous. Conversely, others have been worried by the thought of calling an actor voice clause with a core patient ‘intransitive’ (Gault 1999:399-400). As a result, clauses of this kind have occasionally been dubbed ‘semi-transitive’.

2.3.4 The actor voice in Philippine-type languages other than Tagalog

The examples in §2.3.1–2 are from Tagalog. This is fortuitous: Tagalog happens to be the only Philippine-type language with enough finegrained published analysis to make an investigation of the ergativity and transitivity questions remotely worthwhile. My purpose is to show that the analysis of just a single language raises complex issues and that each language should be carefully analysed on its own merits.²⁵ I do not want to suggest that PAN was similar in detail to Tagalog: on the contrary, the complexities revealed in attempting to analyse Tagalog — and they would be no less for any other Philippine-type language, one suspects — should make us aware of how crude any reconstruction of PAN must be.

In one respect Paiwan is more typical than Tagalog of Philippine-type languages. In Tagalog the actor of a non-actor voice (GEN) and the patient of a non-patient voice (NPIV) are marked in the same way, by *ng* (cf. (3)). In Paiwan NPIV is marked in the same way as LOC, by *tua*, as in (11). The Paiwan patterning of phrase markers seems to be more common across Philippine-type languages than the Tagalog pattern (De Guzman 2000:229).

(11) Paiwan (Egli 1990:287)

Na qə̀m̩ci a caucau tua vatu tua paŋul.

PF <AV>kill SPEC person NPIV dog LOC cudgel

‘The man killed a dog with a cudgel.’

In (11) *tua paŋul* ‘NPIV cudgel’ is peripheral and therefore oblique. This means that the patient *tua vatu* ‘NPIV dog’ is also oblique (Paiwan *tua* does not share ambiguity of Tagalog *ng* NPIV), and the clause is intransitive and behaves like an antipassive. That is, Paiwan, and other Philippine-type languages which pattern similarly,²⁶ is syntactically ergative under Manning’s definition.

²⁵ The practice of using data and analysis from one Philippine-type language to make a point about another one is common but dangerously flawed.

²⁶ Ho (1993) and Huang (1994) respectively present the cases for Yami and Atayal being syntactically ergative. In the literature the actor voice of Formosan languages has sometimes been treated as transitive.

A Paiwan-like pattern need not differ from the Tagalog pattern with regard to semantic transitivity. The languages of the Batan Islands, which include Yami and Ivatan, have a pattern of phrase markers similar to Paiwan, except that GEN, NPIV and LOC are all marked differently. However, Ho (1993:110) analyses *su* NPIV as an oblique. Reid shows that its Ivatan cognate *so* NPIV alternates with *do* LOC (1966:25) and forms peripheral phrases of manner (1966:69-70), so we can be reasonably certain that NPIV and LOC both mark obliques. Ho (1993:94) notes that Yami *su* NPIV is interpreted as non-specific in independent clauses but may be specific in dependent clauses, i.e. it has the same inference pattern as Tagalog (§2.3.1).²⁷ In how many Philippine-type languages this inference of non-specificity is made, we do not know. It is often claimed of a language that its equivalent of the *ng*-patient is non-specific, but it is almost always unclear whether this is a matter of inference or of grammaticisation, and there is often no mention of what happens in dependent clauses. But it seems likely that the specificity pattern of Tagalog and Yami extends to many Philippine-type languages (and Indonesian-type languages, as Wouk's 1984 analysis of Toba Batak shows).

There is some evidence, incidentally, that PMP (Table 10), and perhaps PAn (Table 2) had different phrase markers for GEN, NPIV and LOC. One can only speculate that, as in Yami, this made no difference to patient specificity patterns.

3 Proto Austronesian verbal morphology

3.1 Reconstruction

The voice system which can be reconstructed for PAn is rather similar to the Paiwan system. It seems likely that its usage was similar to what I have described, particularly in §2.3.2, but we cannot be sure of this. The reconstructed voice, mood and aspect morphemes of PAn are set out in Table 1 in schematic form, together with their applications to two PAn roots, **káRaw* 'scratch' and **kaRáC* 'bite'. This reconstruction is based on material from fifteen Formosan languages and various Philippine-type languages of the Philippines and northern Borneo (see Wolff 1973 and Ross 1995a for supporting data). Table 1 differs from the corresponding table in Ross (1995a), however, as it shows the forms for four voices. In Ross (1995a), the circumstantial voice was not reconstructed, as the Formosan data barely justify it. Its reconstruction remains very tentative, for reasons given in §3.2.1.²⁸

The root-and-morpheme combinations in Table 1 are intended only to illustrate the structure of PAn verbal forms: there is no guarantee that these forms all occurred, as verbs in Philippine-type languages often have defective paradigms. These roots represent the two PAn stress types.²⁹ PAn **káRaw* is a paroxytone root, i.e. a root with penultimate stress, **kaRáC* an oxytone, i.e. a root with final stress. Infixes do not cause stress-shift, but the suffixes were probably all what Zorc (1978: 92) calls "same-accent" suffixes, that is, stress

For Paiwan this would give *tua* two functions, marking *tua vatu* as an accusative (core) noun phrase and *tua papul* as an oblique. This analysis is usually given without justification.

²⁷ Unfortunately, the corresponding information for Paiwan is not available.

²⁸ The data on which the reconstructions in Table 1, including the circumstantial voice forms, are based are drawn from the appendix to Ross (1995a).

²⁹ Blust (1997) has shown that the Budai Rukai data used to reconstruct PAn stress in Ross (1992) do not reflect PAn stress as reconstructed. I retain the reconstruction of PAn stress here, but recognise that the evidence for it is not conclusive.

shifts one syllable to the right, so that after suffix-addition a paroxytone remains a paroxytone, and an oxytone remains an oxytone. On Zorc's Philippine evidence, *-ən, *-an, *-a and *-i are all same-accent suffixes (Zorc 1977: 64), and Tsou confirms this for *-a and *-i (Ross 1992), Thao for *-ən, *-an and *-i (Blust In press). To date no reflexes of *-aw, *-ay or *-u have been found in languages which are criterial for reconstructing stress, and it is simply assumed that the pattern covers all monosyllabic suffixes in the paradigm. Tsou *-[n]éni* and Aklanon *-án*, both reflecting *-áni, have their own stress, so this is reconstructed for *-áni and, by analogy, for *-ánay. The aspect and mood categories used in Table 1 are explained below.

Table 1: Proto Austronesian voice, mood and aspect morphemes

Key:

- √ verb root
- ⟨X⟩ X is infix, normally after the root-initial consonant
- X X is suffixed to the root
- R CV- or Ca- reduplication. C is a consonant identical to the root-initial consonant and V a vowel identical to the first vowel of the root. The latter is sometimes replaced by -a.³⁰
- (XXX) XXX is possibly reconstructable only for a post-PAn interstage.

	Actor	Patient	Location	Circumstantial
INDICATIVE				
Neutral	⟨um⟩√ *k <u>um</u> áRaw *k <u>um</u> aRáC	√-ən *kaRáw-ən *kaRaC-ən	√-an *kaRáw-an *kaRaC-án	Si-√ *Si-káRaw *Si-kaRáC
Perfective	⟨umin⟩√ *k <u>um</u> in <u>á</u> Raw *k <u>um</u> in <u>a</u> RáC	⟨in⟩√ *k <u>in</u> áRaw *k <u>in</u> aRáC	⟨in⟩√-an *k <u>in</u> aRáw-an *k <u>in</u> aRaC-án	Si-⟨in⟩√ *Si-k <u>in</u> áRaw *Si-k <u>in</u> aRáC
Durative	⟨um⟩R-√ *k <u>um</u> a-káRaw *k <u>um</u> a-kaRáC	R-√-ən *ka-kaRáw-ən *ka-kaRaC-ən	R-√-an *ka-kaRáw-an *ka-kaRaC-án	Si-R-√ *Si-ka-káRaw *Si-ka-kaRáC
NON-INDICATIVE				
Atemporal	√ *káRaw *kaRáC	√-u, √-a *kaRáw-u, -a *kaRaC-ú, -á	√-i *kaRáw-i *kaRaC-i	án-i + √, (√-áni) *án-i káRaw (*kaRaw-áni) *án-i kaRáC (*kaRaC-áni)
Projective	⟨um⟩√-a *k <u>um</u> aRáw-a *k <u>um</u> aRaC-á	√-aw *kaRáw-aw *kaRaC-áw	√-ay *kaRáw-ay *kaRaC-áy	án-ay + √, (√-ánay) *án-ay káRaw (*kaRaw-ánay) *án-ay kaRáC (*kaRaC-ánay)

³⁰ As Robert Blust (pers. comm.) points out, both *CV- and *Ca- reduplication are reconstructable for the PAn durative and the PMP imperfective. I have no explanation for this alternation.

There appear to have been four major formal classes of verb in PAN:³¹

- A. Verbs like those in Table 1, which took actor voice infixation of *⟨um⟩ into the root;
- B. A small class of verbs whose actor (and sometimes other) voice forms had no affixes;
- C. Verbs whose root began with *pa- and whose actor voice forms began with *ma-, derived historically from *⟨um⟩ + *pa-, e.g. actor voice neutral *maCáy 'die', actor voice atemporal *paCáy 'die'. Many of these verbs are complex roots formed with the causative prefix *pa-.
- D. Verbs similar to those in (c), but whose root began with *ka- and whose actor voice forms began with *ma-, derived historically from *⟨um⟩ + *ka-. Many of these verbs are complex roots formed with the prefix *ka-, and Zeitoun and L. Huang (2000) show that these were stative (or perhaps inchoative) verbs.³²

The Formosan data indicate that intransitive verbs had the same morphology as actor voice transitives and that they belonged to these same four major formal classes.³³ There is very little analysis of intransitives in descriptions of Formosan languages, but it can be inferred that verbs in *⟨um⟩ usually denoted processes with an actor pivot (like 'walk', 'weep', 'sing'), verbs in *ma- denoted involuntary processes (like 'sleep', 'fall') or temporary states (like 'be afraid', 'be alive', 'be drunk') with an undergoer pivot, and unaffixed verbs were a small class which included both actor-controlled processes and permanent states (like 'be good', 'be big').³⁴

The location voice also seems to have served as a beneficiary voice in PAN, as it does in a number of daughter languages. In other words, with semantically appropriate verbs, a human location was interpreted as beneficiary, as in these examples:

- (12) Paiwan (Egli 1990:296)
uri ku=su=pavay-an tua kakudan
 FUTURE GEN:1S=PIV:2S=give-LV NPIV power
 'I will give you power'
- (13) Seediq (Asai 1953:46)
skat-an-i=ku qəhuni
 cut-LV-AT=PIV:1S tree
 'Please cut the tree for me!'

Alongside the circumstantial voice prefix *Si-, a functionally similar prefix *Sa- is also reconstructable (Ross 1995a; Blust 1999a). What the division of labour was between *Si- and *Sa- is unclear, and *Sa- is not further discussed here.

31 The four classes are also supported by L. Huang's (2000) detailed analysis of Mayrinax Atayal verb classes.

32 Zeitoun (2000) provides further evidence for the reconstruction of verbs in *ka-.

33 The situation with regard to intransitives in Tagalog and other Philippine languages is different. Here, some intransitives carry AV morphology, others PVe, and so on. Tagalog examples in ⟨um⟩ AV are *bum-agyo* 'be stormy', *dum-ating* 'arrive'. Intransitives in -in PV are *antok-in* 'feel sleepy', *langgam-in* 'be infested with ants'. Intransitives in -an LV are *kilabut-an* 'feel terrified', *pawis-an* 'sweat'. Intransitives in -i CV are *i-kaway* 'wave (a hand)', *i-kasal* 'get married' (Schachter & Otanes 1972:306-310).

34 The reconstruction of a contrast between unaffixed state verbs and state verbs formed with *ma- is addressed by Evans and Ross (2001).

As in the examples thus far, PAN noun phrases evidently followed the verb, except where one was topicalised to clause-initial position. SPR argue that a genitive-marked agent noun phrase normally followed its verb, 'since otherwise it could be interpreted as Genitive attribute of the noun preceding it.' As in Paiwan, PAN noun phrases were evidently introduced by a phrase marker. Reconstructing these phrase markers is not easy: they were monosyllabic, and conflicting evidence about their forms suggests quite a complex paradigm which has been subject to various simplifications and/or analogical reorganisations in daughter-languages. Their reconstruction is also subject to the top-of-the-tree effect. However, a well enough distributed set of languages shares the three-way distinction made in Paiwan between specific, genitive and non-pivot phrase markers — and agrees on the forms of these markers — for us to reconstruct them for PAN. The data point unambiguously to a Paiwan-style distinction between GEN and NPIV, as in (3). It is less clear whether there was a distinction between NPIV and LOC, although there is some evidence for this distinction in PMP (see Table 10). There is also Formosan evidence of a contrast between common **a/*u* and **ka/*ku*, the former used in topicalised (fronted) noun phrases, the latter elsewhere, but there is no evidence of this contrast in Malayo-Polynesian languages. There is a well reflected distinction between markers of common and personal noun phrases, and a probable contrast among common noun phrase markers between present and absent (or perhaps proximal and distal) referents. The resulting partial system is shown in Table 2.

Table 2: Some Proto Austronesian phrase markers³⁵

	TOPIC	SPEC	GEN	NPIV
common (present)	<i>*a</i>	<i>*ka</i>	<i>*na</i>	<i>*Ca, *sa</i>
common (absent)	<i>*u</i>	<i>*ku</i>	<i>*nu</i>	<i>*Cu, *su</i>
personal	—	(<i>*i, *ti, *si</i>)	<i>*ni</i>	—

No marker is reconstructable for non-pivot personal noun phrases, but this is expected: NPIV noun phrases were non-specific, whereas personal noun phrases are always specific and definite. Evidence from Formosan languages implies that a personal noun phrase which occurred as undergoer of an actor voice verb was marked as a locative oblique. The three forms of the specific personal marker reflect formal problems which hamper reconstruction.

Philippine-type languages commonly have cliticised pronouns (see Starosta 1988). The Paiwan clitics *ku=su=* in (12) precede the verb, whilst the Seediq clitic *=ku* in (13) follows it. The PAN clitic pronouns, like their reflexes in a number of Philippine-type languages, were apparently second-position clitics. If the verb was preceded by an auxiliary-like element, the clitics followed that element; if the verb was the first constituent of the clause, the clitics followed the verb. Auxiliary-like elements seem to have occurred very frequently in PAN, with the consequence in some languages that — as auxiliary use has declined and some auxiliaries have disappeared — some clitic pronouns, and especially agent genitives, have remained stranded in front of the verb (see SPR). This has happened in the Formosan languages Paiwan and Puyuma, and also in the Indonesian-type languages described below (§4.2).

³⁵ This table is based on analysis reported in Ross (2001).

The PAn personal pronouns (Table 3) are at least as hard to reconstruct as the phrase markers, but it is clear that there were both free and clitic sets, as Dahl (1973), Blust (1977) and Harvey (1982) observed. It seems that there were two free sets, the members of one containing the politeness morpheme **k-* or **ka-*. The polite morphemes became the default in PMP and those without the politeness morpheme vanished except in certain relic forms.³⁶

Only one clitic set is reconstructable (as SPR note), serving both as syntactic pivot and as genitive. If an undergoer-voice verb took both a genitive and a pivot clitic, then these occurred in the order genitive–pivot.³⁷ We find evidence that as early as PAn there were incipient tendencies to express pronominal pivots and genitives differently. One tendency, reflected in Kavalan, Atayal, Seediq, Pazeh, Saaroa, Rukai, Paiwan and PMP (see Table 11), was to replace the pivot clitic with a free form which over time became a new pivot clitic. Second, two additional ways of expressing the genitive are reflected in daughter-languages. A number of Formosan languages reflect non-third person PAn genitive clitics with initial **=m-*, shown here as GEN2.³⁸ Of these, only **=mami* 1EP is reflected in Malayo-Polynesian languages. The disappearance of the others was probably due to the rise of a third genitive set, GEN3, whose members had by PMP times also become clitics. They consisted of the genitive personal phrase marker **ni* (Table 2) and the free (non-polite) pronoun.³⁹

Table 3: Proto Austronesian personal pronouns⁴⁰

	Free	Free polite	PIV, GEN1	GEN2	GEN3
1S	<i>*[i-]aku</i>	—	<i>*=ku</i>	<i>*maku</i>	<i>*n-aku</i>
2S	<i>*[i-]Su</i>	<i>*[i-]ka-Su</i>	<i>*=Su</i>	<i>*miSu</i>	<i>*ni-Su</i>
3S	<i>*s(i)-ia</i>	—	<i>*=(ia)⁴¹</i>	—	<i>*n(i)-ia</i>
1EP	<i>*i-ami</i>	<i>*[i-]k-ami</i>	<i>*=mi</i>	<i>*mami</i>	<i>*n(i)-ami</i>
1IP	<i>*([i-])ita</i>	<i>*[i-]k-ita</i>	<i>*=ta</i>	<i>*mita</i>	<i>*n-ita</i>
2P	<i>*i-amu</i>	<i>*[i-]k-amu</i>	<i>*=mu</i>	<i>*mamu</i>	<i>*n(i)-amu</i>
3P	<i>*si-da</i>	—	<i>*=(da)</i>	—	<i>*ni-da</i>

36 The reconstructions in Table 3 are based largely on an examination of Formosan and Philippine data, but they also owe much to Blust (1977) and Harvey (1982). In Blust's (1977) reconstruction, the only PAn free form reconstructed as a pair with and without **k-* or **ka-* was 2S **i-Su* and **i-ka-Su*. Harvey (1982) points out that other pairs with and without **k-* are reconstructable.

37 Among Formosan languages, Kavalan (Li 1978:590) and Paiwan (Egali 1990:156-157, 296-297) have the sequence genitive–pivot. (Egali, p.296, seems to miss the fact that the second clitic marks the pivot, referring to it as 'Obliquus', his term for the non-pivot noun phrase.)

Atayal (Starosta 1988:12) and Seediq (Pecoraro 1979:67-68) generally have the clitic order pivot–genitive, but portmanteau double clitics have the (fossilised) order genitive–pivot, indicating that this was the earlier order.

38 Evidence for these is found in Saisiyat, the Atayal dialects, Thao, Amis, Kakanavu and Siraya.

39 Blust (1977) reconstructed alternants with **i-* as well as **ni-*, but Harvey (1982) points out that the evidence does not support these.

40 Parentheses () indicate that one cannot be sure whether their contents should be reconstructed. Square brackets [] indicate that there are two versions of the reconstruction, one with and one without the contents of the brackets.

41 The main evidence for the reconstruction of PAn third-person clitics is paradigmatic, i.e. they occurred as partials in the free and genitive sets. Both **=ia* and **=da* are reflected in Malayo-Polynesian languages, but the only known Formosan reflex is the Saaroa clitic *-isa*, reflecting **=ia*.

Just as no non-pivot personal noun phrase marker is reconstructable, so too there are no non-pivot personal pronouns. Instead, we find Formosan reflexes of the free pronouns with a location suffix **-an* or **-nan* (noted by SPR) and both Formosan and Malayo-Polynesian reflexes of the free pronouns with a patient suffix **-ən* or **-n*.⁴² A possible history of this suffixation is touched on in §3.2.2.

The aspect and mood categories used in Table 1 require some explanation. These categories are divisible on both formal and semantic grounds into two higher-order classes, indicative and non-indicative. The formal division is discussed in §3.2.1. Within the indicative class, “neutral” refers to a finite verb form not marked for tense or aspect. It was apparently used for realis events and states which were neither perfective nor durative. The perfective and durative were the finite forms used respectively for completed events and for events viewed as ongoing at some point of time. Within the non-indicative class, the projective was evidently the finite form used for irrealis events and states, i.e. intention, possibility and exhortation.

Atemporal forms have three basic functions in daughter languages (and often have all three functions in the same language): (a) as imperatives; (b) as verbs subordinate to some auxiliaries; and (c) expressing non-initial sequential events in narrative.

The second use is illustrated in the Atayal examples in (14). As Starosta (SPR, Starosta 1985, 1988) has shown, Formosan (and some extra-Formosan) languages make considerable use of sentence-initial auxiliaries, called “pre-verbs” by some scholars, which carry information on aspect, time, negation, manner, location and so on. As noted above, the auxiliary hosts enclitic pronouns.

(14) Sqliq Atayal

- a. *Ini[?]=saku[?] hju[?] qsia[?] lukus.*
 NEG=PIV:1S AV.AT.soak water clothes
 ‘I have not soaked the clothes in water.’ (Egerod 1966: 273)
- b. *Ini[?]=sami kac-i na[?] mqu[?].*
 NEG=PIV:1EP bite-LV.AT GEN snake
 ‘We have not been bitten by snakes.’ (Egerod 1966:354)
- c. *Laxi zŋ-i snon-an=maku[?] isu[?].*
 PROHIB forget-LV.AT message-OBLIQUE=GEN:1S D:2S
 ‘You must not forget my message.’ (Egerod 1966:358)
- d. *Si=nha[?] sr[?]ag-i ma ai.*
 ACTUAL=GEN:3P go.along-LV.AT it.is.said INTERJECTION
 ‘They were following (the river).’ (Egerod 1969)

The first morpheme in each example in (14) is an auxiliary, and in these cases (but not in all Atayal sentences beginning with an auxiliary) the subordinate verb is atemporal. In (14a) *hju[?]* is the actor voice atemporal form (cf *həm[?]hju[?]* actor voice neutral). In (14b,c,d) the subordinate verb is a location voice atemporal marked with *-i*.

The third use of atemporals is to express non-initial sequential events in narrative. This is illustrated in the Paiwan examples in (15). The first verb has the neutral form, and verbs

⁴² Some Malayo-Polynesian reflexes are possessive pronouns (‘mine’, ‘yours’ etc).

following it have (apparently optionally) the atemporal form. Another example is the verb *kani* in (4) above.

(15) *Paiwan*

- a. *Ribu-in sa pa-d'yulu-i.*
defeat-PV CJ CS-be.simple-PV.AT
'He defeated and pacified it [i.e. the village].' (Egli 1990:226)
- b. *Kiqənəc-an sa pa-pa-piriq-i.*
look.at-PV CJ RECIP-R-divide-PV.AT
'He looked at and divided it.' (Egli 1990:242)
- c. *Vuʔuq-ən sa ka-d'yaməq.*⁴³
spear-PV CJ AT.PASSIVE-hit
'He speared it and it was hit.' (Egli 1990:226)

Atemporal verbs in narrative sequences are also common in the Dusunic languages of Sabah (Kroeger 1991).

3.2 The rise of the Proto Austronesian system

3.2.1 *The voice-from-nominalisation hypothesis*

There are four observations that can be made about the reconstructed PAN morphology shown in Table 1:

- (16)a. The mixture of a prefix (**Si-*), an infix (**um*) and two suffixes (**-ən*, **-an*) marking voice in indicative forms makes for a paradigm with unusual asymmetries.
- b. Indicative verb forms also occur as (apparent) nominalisations, but non-indicative forms don't.
- c. Despite the asymmetries of the voice morphemes in indicative forms, the aspect morphemes which occur in these forms are paradigmatically regular: the perfective is marked by **in*, the durative by reduplication. The one exception is that perfective **in* and patient voice **-ən* do not cooccur.
- d. In contrast with the indicative forms, the morphemes of non-indicative forms make up a fairly symmetrical paradigm, except for the presence of **um* in the actor voice projective form.

Observations (a) and (b) are not new, and (b) is illustrated by the following derivations from the verb root *kan* 'eat' (Ferrell 1982:17, 106):

(17)	<i>Paiwan</i>	verb form	nominalisation
	<i>kəmən</i>	actor voice neutral	'eater', 'someone who eats'
	<i>kan-ən</i>	patient voice neutral	'food', 'something to be eaten'
	<i>kinən</i>	patient voice perfective	'consumed food', 'something eaten'
	<i>kan-an</i>	location voice neutral	'place where one eats'
	<i>si-kan</i>	circumstantial voice neutral	'eating utensil', 'something to eat with'

⁴³ The verb *ka-d'yaməq* is the atemporal form of the Paiwan passive. The passive is unique to Paiwan, and is formally resembles an actor voice verb of the *ka-* class: its neutral form begins with *ma-* (cf. *malimək* in (4)), its atemporal with *ka-*.

Observations (a) and (b) are accounted for if we infer, with SPR, that the indicative verb forms are derived from the nominalisations, since there is no reason why a language's nominalising morphemes should form a symmetrical paradigm. Under this hypothesis, the original verb forms were those of the (symmetrical) non-indicative paradigm, the atemporals having originally been the neutral forms. Ross (1995a) posits a series of diachronic steps whereby this derivation occurred. The most important of these is that a predicate nominalisation was used to put a non-agent noun phrase into the syntactic pivot slot. We can see how this might have happened by examining the Paiwan sentences in (1) and (2). (1a) is repeated below.

- (1a) Paiwan
təkəl-ən a vaua
 drink-PV SPEC wine
 'the wine will be drunk' ('s/he/they will drink the wine')

If *təkəl-ən* in (1a) is interpreted as a nominalisation, i.e. 'something to be drunk', then the example can be reglossed as:

- (18) Paiwan
təkəl-ən a vaua
 drink-NOM SPEC wine
 'the wine is something to be drunk'

We turn now to the expansion of (1a) given as (2a) above:

- (2a) Paiwan
təkəl-ən nua qata a vaua
 drink-PV GEN stranger SPEC wine
 'the wine will be drunk by a/the stranger' ('a/the stranger will drink the wine')

Again interpreting *təkəl-ən* as a nominalisation, the example is reglossed as:

- (19) Paiwan
təkəl-ən nua qata a vaua
 *drink-NOM GEN stranger SPEC wine
 *'the wine is something of a/the stranger's to be drunk'

The verbs in (1a) and (2a) are patient voice forms, but similar considerations apply at least to location voice forms. Example (1c), *təkəl-ən a kakəsan* 'the kitchen will be drunk in' ('s/he/they will drink it/them in the kitchen'), is derived from 'the kitchen is the place of drinking'.

Under the hypothesis, this highly marked strategy became decreasingly marked until the nominalisations were reinterpreted as verb forms and ousted the original neutral (realis) verb forms from main and perhaps relative clauses, leaving them as atemporals in imperatives and the other contexts mentioned above. The same morphemes continued to be used to form nominalisations, with the result that sentences like (1a) and (2a), at least when taken out of context, are vague as to their predicate structure in some modern Philippine-type languages.

There is formal support for the inference that the PAn non-indicative morphemes reconstructed in Table 1 originally formed a system in which the atemporals were the realis verb forms, the projectives irrealis. These morphemes form a pattern of two elements in actor, patient and location voices, shown in (20), the first element opposing atemporal

(< realis) zero to projective *-a* (< irrealis), and the second making a three-way contrast between AV zero, PV **-u* and LV **-i*:

(20)	Actor	Patient	Locative	Circumstantial
Atemporal	√ - <i>θ</i> - <i>θ</i>	√ - <i>θ</i> - <i>u</i>	√ - <i>θ</i> - <i>i</i>	√ - <i>án</i> - <i>θ</i> - <i>i</i>
Projective	√ - <i>a</i> - <i>θ</i>	√ - <i>a</i> - <i>u</i>	√ - <i>a</i> - <i>i</i>	√ - <i>án</i> - <i>a</i> - <i>i</i>

As Table 1 indicates, **-a* is reconstructable as an alternant of atemporal patient voice **-u*. Clearly **-a* does not fit this pattern (see Ross 1995a for discussion). It is possible that **-a*, **-u* and **-i* all represent captured phrase markers and/or prepositions, as suggested by Starosta (1995) for **-a* and **-i*, and that **-a* and **-u* reflect captured phrase markers that contrasted on a proximate/distal axis (cf Table 2).

The circumstantial morphemes in (20) show a different patterning:⁴⁴ they consist of the morpheme **-án-* plus the locative voice suffixes. The fact that **-án-* took the locative suffixes suggests that it was itself once a verb, and that its suffixation to the root reflects grammaticisation. Two further pieces of evidence speak in favour of this suggestion. First, unlike all the other suffixes in Table 1, **-án-* was stressed, indicating phonologically incomplete grammaticisation. Second, in Sqliq Atayal, *an* and *anai* are auxiliaries which precede the verb (which is itself prefixed with *s-* CV), as illustrated in (21):

- (21) Sqliq Atayal (Egerod 1965:282)
Anai-ta? *s-blaq* *km>ai*l.
 CV.IRR-GEN:1IP CV-good <AP>tell
 'Let's talk it over on that [basis]'

Given that grammaticisation processes tend to be irreversible, it is likely that Sqliq Atayal reflects the PAN situation, i.e. there was still a verb or an auxiliary **án-*, and the grammaticisation process was also syntactically incomplete in PAN. Hence the non-indicative circumstantial forms with auxiliaries, **án-i* + √ and **án-ay* + √, in Table 1 are reconstructable for PAN, but the suffixed forms **√-áni* and **√-ánay* may only have arisen later.

The neatness of Table 1 suggests a more orderly set of developments than probably occurred. If the voice-from-nominalisation hypothesis for indicative verb forms were completely correct then we would expect these morphemes

- (i) to function in each language both as voice morphemes and as nominalisers;
- (ii) not to occur in non-indicative verb forms.

In fact, neither expectation is met. Table 4 shows the distribution of Formosan reflexes of PAN indicative voice morphemes. AV, PV, LV, CV and UV indicate that the relevant morpheme is used to mark that voice in that language, whilst 'nom' indicates that it is used to form a nominalisation.⁴⁵ The morpheme **<in>* is also included: PF indicates that it marks perfective aspect across voices.

44 These forms were not reconstructed by Ross (1995a). Reflexes of **-áni* are Sqliq Atayal *an*, Mayrinax Atayal and Saisiyat *-ani* and Puyuma *-an*, all CV atemporal, Tsou *-[n]əni* CV neutral, Paiwan *-an* CV atemporal/imperative, Aklanon and Samar Leyte *-án* CV dependent, Javanese *-ʔan* CV imperative/optative. Reflexes of **-anay* are Sqliq Atayal *anai* CV projective, Puyuma *-anay* CV indicative/imperative, Siraya *-anei* LV projective.

45 Table 4 is based on Table 7 of Ross (1995a), with information for Pazeh and Thao drawn from Blust (1999b) and Blust (In press) respectively.

We see from Table 4 that the distributions of the morphemes vary considerably. At one extreme is *⟨um⟩, which marks actor voice in every language except Rukai but serves as a nominaliser only in Paiwan and Puyuma.⁴⁶ This suggests that its story is different from those of the other voice morphemes: it was probably a verbal morpheme which became a nominaliser only in daughter-languages and by analogy with the other voice morphemes. This inference is supported by the fact that *⟨um⟩ also defies our second expectation, by appearing in projective actor voice forms in Atayal, Bunun, Kanakanavu, Saaroa, Siraya, and Puyuma. Furthermore, although Table 1 shows the atemporal actor voice form as consisting of the root alone, Puyuma makes a contrast between a root-only imperative and a dependent reflecting *⟨um⟩. There are also dependent actor voice forms in Seediq, Puyuma and Bonggi which reflect *⟨um⟩.⁴⁷ It is difficult to know whether these forms are inherited or whether dependents with *⟨um⟩ result from an analogical extension of the latter's use. Either way, however, *⟨um⟩ was not limited to indicative forms in PAn, and was therefore probably a pre-PAn verbal morpheme rather than a nominaliser.

If *⟨um⟩ was indeed a verbal morpheme, then what was its function in PAn? As noted in §3.1, it also occurred in intransitives denoting actor-controlled processes, and the best generalisation we can make is that it marked its verb as having an actor pivot and denoting a process, usually one which was under the actor's control.

Table 4: Formosan reflexes of PAn indicative voice morphemes

	*⟨um⟩		*-ən		*-an		*Si-		*⟨in⟩	
Saisiyat	AV	—	PV	nom	—	nom	CV	—	PF	nom
Atayal	AV	—	PV	nom	LV	nom	CV	nom	PF	nom
Seediq	AV	—	PV	nom	LV	nom	CV	—	PF	nom
Kavalan	AV	—	—	—	UV	nom	—	nom	PF	—
Amis	AV	—	PV	—	—	nom	—	—	—	nom
Tsou	AV	—	—	—	—	nom	—	—	—	—
Kanakanavu	AV	—	PV	—	LV	nom	—	nom	PF	—
Saaroa	AV	—	—	—	LV	nom	—	—	PF	—
Rukai	—	—	—	—	—	nom	—	—	—	—
Thao	AV	—	PV	nom	LV	nom	—	—	PF	nom
Pazeh	AV	—	PV	nom	LV	nom	CV	nom	PF	nom
Puyuma	AV	nom	—	nom	— ⁴⁸	nom	—	nom	—	nom
Paiwan	AV	nom	PV	nom	LV	nom	CV	nom	PF	nom

At the opposite distributional extreme in Table 4 from *⟨um⟩ is *-an, whose reflexes serve as a locative nominaliser in all the languages and as a location (or undergoer) voice marker in most languages which reflect any of *-ən, *-an and *Si- as voice morphemes

⁴⁶ Ironically, the other morphemes in Table 4 are not reflected as voice markers in Puyuma. This means that the analogy hypothesis is not directly valid for Puyuma. However, as Blust (1999) shows, Puyuma shows ample signs of borrowing from Paiwan, and this is a likely source of Puyuma nominalisations reflecting *⟨um⟩.

⁴⁷ I am indebted to Nikolaus Himmelmann for drawing my attention to the Seediq and Bonggi data.

⁴⁸ Ross (1995a) took Puyuma CV forms in -an to be reflexes of LV *-an, but this is probably incorrect. It is more likely that they reflect CV *-áni.

(Tsou, Rukai and Puyuma reflect none of them as voice morphemes⁴⁹). Since Philippine-type languages commonly have nominals in the predicate slot, the same was almost certainly true of PAN and earlier stages still. Nominalised forms in **-ən*, **-an* and **Si-* would have occurred in the predicate slot as a matter of course, with the possibility of being reinterpreted as verbs. Since the converse process — a verbal form frequently occurring in an argument slot — is far less probable, one can reasonably infer that **-an* was originally a nominalising morpheme. The same argumentation applies to **-ən* and to **Si-*, although the distributions of their reflexes in Table 4 are less decisive.

The foregoing discussion suggests a basic pre-PAN system that looked something like this:

(22)	Actor	Patient	Location	Circumstantial
verb	√, <um>√	√-u	√-i	—
nominalisation	—	√-ən	√-an	Si-√

If the actor, patient and location voice morphemes displayed in (22) do represent the basic system as it was before nominalisations were reinterpreted as indicative verbal forms, then there was already a three-way voice system which provided the template for this reinterpretation. If, as is implicit in the discussion above, circumstantial **-án-* represents a later development than the other non-indicative voice morphemes, then it would also not be surprising if, at the stage we reconstruct as PAN, circumstantial **Si-* had not yet been reinterpreted as an indicative voice morpheme. That is, its reconstruction in Table 1 remains very tentative.

The reader may notice that (22) leaves a significant chunk of morphology incompletely explained. Under the voice-from-nominalisation hypothesis, durative reduplication and the perfective infix **<in>* occurred in verb forms with **<um>* and in nominalisations (Table 1). This is an odd distribution and it provides a motivation for an alternative hypothesis presented in the next section. The hypothesis itself is new, but several of its features have been touched on in the literature.

3.2.2 *An alternative hypothesis*

Under the voice-from-nominalisation hypothesis, it is assumed that nominalisation and indicative voice were already discrete phenomena in PAN, that is, that derived nominals and indicative verbs were homophonous forms belonging to separate word classes, as, for example, in (17). The hypothesis also proposes that some of the morphemes in Table 4 originally formed nominalisations, but by PAN times had also been reinterpreted as indicative verbs. It follows that between these interstages predicate nominalisations were only gradually reinterpreted as homophonous indicative verb forms (cf examples (1a), (2a), (18) and (19)). In other words, there must have been an interstage when derived nominals and indicative verbs were not yet discrete. Most modern Philippine-type languages tend to be analysed as having homophonous forms belonging to separate word classes (indicative verbs and nominalisations), but one, Tagalog, has been analysed such that these forms comprise a single word class whose members occur in both predicate and argument slots. The question is, does Tagalog represent a direct continuation of the PAN situation? If it does, then the

⁴⁹ It is essentially the absence of these morphemes in these languages that causes Starosta (1995) to propose that Rukai and Tsou (he does not refer to Puyuma) separated from all other Austronesian lects before a system of the kind reconstructed in Table 1 came into being. The alternative explanation of this absence, discussed in Ross (1995a), is that these morphemes lost their verbal function in these languages.

voice-from-nominalisation hypothesis is wrong, and an alternative hypothesis is needed. This alternative will be formulated after a brief look at Himmelman's (forthcoming a) analysis of Tagalog.

Himmelman outlines an analysis of Tagalog as a language in which there are distinct morpholexical categories, but no distinct terminal syntactic categories. If pronominal clitics are ignored, all Tagalog phrasal categories except (most) clause-initial predicates consist of a phrase marker and a content word, as illustrated in (23):

- (23) Tagalog (Himmelman forthcoming a)
Iniabót ng manggagamot sa sundalo ang itlóg.
 i-⟨in⟩abót naŋ maŋgagamot sa sundalo aŋ itlóg
 CV-⟨PF⟩reach GEN doctor LOC soldier SPEC egg
 'The physician handed the egg to the soldier.'

Almost every content word may head a phrase in either the predicate slot or an argument slot. Unlike in European languages, there is no correlation between the class of the content word and the category of the syntactic slot in which it occurs (Sasse 1993:200). In (24) the voice- and aspect-marked content word *àalagaan* is used in an argument slot:

- (24) Tagalog (Himmelman forthcoming a)
Iuuwi=nya ang àalagaan=nya.
 i-REDUP-uwi?=niyá aŋ REDUP-alaga?-an=niyá
 CV-DUR-return=GEN:3S SPEC DUR-care.for-LV=GEN:3S
 'He would return the ones he was going to care for.'

In (25) *artista*, which cannot be marked for voice or aspect, is the predicate, whilst the voice-marked content word *yumaman* heads the phrase in the argument slot:

- (25) Tagalog (Schachter & Otnes 1972:62)
Artista ang yumaman
 actress SPEC ⟨AV⟩wealthy
 'The one who got rich is an actress.'

Tagalog content words fall into two major morpholexical classes: those which do not include voice- or aspect marking (in the examples above *manggagamot* 'doctor', *sundalo* 'soldier', *itlóg* 'egg', *bata* 'child', *artista* 'actress') and those which do (*iniabót* < *abót* 'within reach', *iuuwi* < *uwi* 'return', *àalagaan* < *alaga* 'pet, ward', *yumaman* < *yaman* 'wealth'). One might label the first class 'nouns' on the basis of the ontological category of THING/PERSON that its members usually denote. The second class (which falls into morpholexical subclasses) is less readily labelled, however, as its underived roots tend to denote items in the ontological categories of either THING/PERSON or STATE/PROPERTY, while forms derived from them denote items in both these categories and, crucially, in the category of ACTION/EVENT as well. The labelling difficulty reflects the fact that these word classes are only morphological, correlating neither with ontological nor syntactic categories. Here I will label the category of words which do not include voice- or aspect marking '-VM words' and those which do '+VM words'.⁵⁰

⁵⁰ Sasse (1993) makes a distinction in Cayuga between 'simplex words' and 'roots'. However, the first category in Tagalog includes derived words like *manggagamot* 'doctor' and so the term 'simplex' is inappropriate here. It is in any case appropriate to label the two classes of *words* rather than to refer to roots.

Was PAn like modern Tagalog? It certainly had both –VM and +VM words. For example, Blust (1998, 1999a) notes a PAn distinction between *Ca-* reduplication, which formed instrumental nouns, i.e. a subclass of –VM words, and **Si-/Sa-*, which formed instrumental words “categorially ambivalent between verbal and nominal uses” (1999a:359), i.e. a subclass of +VM words.

Unfortunately, most descriptions of Formosan languages rely heavily on elicited sentences in which predicates are marked for voice and aspect and arguments are not. Texts, however, sometimes show a Tagalog-like use of a voice-marked content word in an argument slot. In (26) the content word of the syntactic pivot is *c<in>abu* ‘(it was)wrapped’:

(26) Mayrinax Atayal (L. Huang 1995:259)

Si-he?e=nia? c-ku? ngaquwaq n-ku? nabakis ku?
 CV-pour=GEN:3S NPIV-SPEC mouth GEN-SPEC old.man SPEC
c<in>abu?=nia? c-ku? abag na? bakati?.
 <PF>wrap=GEN:3S NPIV-SPEC leaf GEN bakati

‘He poured the thing wrapped by him in the *bakati* leaf into the old man’s mouth.’

There are not enough good dictionaries of Philippine-type languages for us to determine the ontological categories denoted by underived roots in these languages, let alone to reconstruct them for PAn, but Formosan examples like (26) do display a Tagalog-like mismatch between word class and syntactic slot. In other words, it appears likely that a +VM word formed from a PAn root could occur either as a predicate or, preceded by a noun phrase marker (Table 2), as an argument.

If this inference is right, then the voice-from-nominalisation hypothesis is wrong, because it proposes that one word class — nominalisations — gave rise to two — nominalisations and indicative verbs. The evidence just reviewed indicates that this split had not taken place in PAn and still has not taken place in Tagalog and probably some other Philippine-type languages. The alternative hypothesis proposes that +VM words were a single class in PAn, and that their members were used in both predicate and argument slots.

The Tagalog analysis, however, requires us to go further. The alternative hypothesis does not simply say that the voice-from-nominalisation process was incomplete in PAn: it questions whether, at any reconstructable interstage, PAn +VM words ever were nominalisations. It is to this question that we now turn.

As Himmelmann (forthcoming a) points out, on his analysis Tagalog bears a resemblance to certain indigenous North American languages, namely Straits Salish in the northwest and Cayuga, an Iroquoian language, as analysed respectively by Jelinek and Demers (1994) and Sasse (1993). If PAn resembled Tagalog, it must also have been typologically like these languages. In Straits Salish all content words take markers of transitivity, voice, tense, mood and argument coreference; in Cayuga, roots take markers of tense, aspect and argument coreference; in PAn, +VM words, formed with **-ən*, **-an*, **Si-* and **<in>*, included voice- and aspect-markers. That is, in the three languages there is a major class of morphologically complex content words. These words are marked for categories which are associated in many languages with verbs and which occur in both predicate and argument slots. In all three languages, their use in an argument slot is/was indicated by a preceding noun phrase marker. However, there are differences. In Tagalog, PAn and Cayuga there is/was a distinction between +VM and –VM words. In Straits Salish the roots of content words are not divisible into major morpholexical classes and all can apparently be marked for transitivity, voice and tense.

In Straits Salish and Cayuga, content words are analysed as predications, and the phrase marker in an argument slot is analysed as marking an embedded predication.⁵¹ There is evidence that a similar analysis is appropriate for at least some modern Philippine-type languages, and therefore for PAN. In Tagalog the item which characterises the phrase semantically and syntactically — the traditional head — may be either preceded or followed by attributes, and head and attributes are linked by a ligature, regardless of the order in which they occur. The ligature has two phonologically determined allomorphs =*ŋ* and *na*, and the language allows pairs of phrases like those in (27), (28) and (29), where the attributes are an adjective, a ‘verb’, and a prepositional phrase respectively.

(27) Tagalog (Schachter & Otnes 1973:122–123)

- a. *aŋ bantog na siyudad*
SPEC famous LIG city
- b. *aŋ siyudad na bantog*
SPEC city LIG famous
‘the famous city’

(28) Tagalog

- a. *aŋ ni-luto=mo=ŋ pagkain*
SPEC PV-cook=GEN:2S=LIG food
- b. *aŋ pagkai=ŋ ni-luto=mo*
SPEC food=LIG PV-cook=GEN:2S
‘the food cooked by you’

(29) Tagalog

- a. *aŋ nasa mesa=ŋ libro*
SPEC on table=LIG book
- b. *aŋ libro=ŋ nasa mesa*
SPEC book=LIG on table
‘the book on the table’

The inference to be made from these examples is that Tagalog lacks a noun phrase construction with a head noun and that arguments are expressed by strings of embedded predicates, the first marked by *aŋ*, and any others by the ligature. Thus (27a) and (27b) can be roughly glossed respectively as ‘the [one that is] famous [that is a] city’ and ‘the [one that is a] city [that is] famous’. In this respect Tagalog seems to be typical of Philippine-type languages. For example, Ferrell (1980:13) analyses the Paiwan string in (30) as being interpretable as either ‘the female child’ or ‘the young female’, i.e. neither word is the head:

⁵¹ Only during the final stage of preparation of this paper did I come across Mithun’s (2000) analysis of word classes in Iroquoian languages, including Cayuga. She sets out to refute Sasse’s analysis, arguing that Cayuga has nouns and verbs, distinguishable on the grounds of morphological structure and of syntactic function: a verb may serve as either a predicate or an argument, but a noun never serves as a predicate. She thus rejects the analysis of arguments as embedded predications. Her arguments appear to be well grounded, and if they are correct, Cayuga is less similar typologically to PAN than suggested in this section.

(30) *Paiwan*

a atak a vavaian
 SPEC child LIG woman

Significantly, in *Paiwan* and in some other Philippine-type languages, the ligature is identical in form to the phrase marker, supporting — at least diachronically — the analysis of an argument as a string of predicates.

Since the major class of roots in Straits Salish and Cayuga is the language's main source of content words, the class embraces a wide range of ontological categories. The same is true of Philippine-type languages, and presumably of PAN. Not only did predications denote items in the categories of THING/PERSON, STATE/PROPERTY, and ACTION/EVENT: they also denoted the MANNER of an action or event and certain pronominal categories.

In (31) and (32), the words denoting 'slowly' are the main predications:

(31) *Kavalan* (Lee 1997:86)

M-ŋasan q<man tu ?may ya sunis=su.
 AV-slow <AV>eat NPIV rice SPEC child=GEN:2S
 'Your child ate the meal slowly.'

(32) *Tagalog* (Schachter & Otnes 1972:306)

Bagal-an=mo ang lakad=mo.
 slow-LV=GEN:2S SPEC walk=GEN:2S
 'Walk slowly.'

The most striking piece of evidence that most content words were predications in PAN, or perhaps at some pre-PAN stage, was mentioned in §3.1: personal pronouns took the voice markers **-ən* (or **-n*) and **-an* (or **-nan*). This suggests that, like the corresponding root in Straits Salish (Jelinek and Demers 1994:715), second person singular **Su*, for example, was a content word whose meaning might be translated as 'be you', **i-Su* a phrase meaning 'the one who is you' (**i-* being a determiner), **Su-n* a content word meaning '[the one that] is you-ed', i.e. '[the one that] is yours', and **Su-(n)an* a content word meaning '[the place that] you are at'. Forms in **-ən* with possessive meaning are reflected in the Philippines. Forms in **-an* retain their locative meaning in the Formosan language *Pazeh* (*yami?an* 'at our.EXC place', *imu?an* 'at your.PL place'; Ferrell 1968). In other Formosan languages they are the personal pronouns used where a common noun phrase would be marked with a non-pivot phrase marker, as in:

(33) *Wulai Atayal* (L. Huang 1995:129)

M-ihiy k-nan Tali?
 AV-beat 1S-LV Tali
 'Tali beat me.'

Despite the similarities between PAN and Tagalog that I have adduced here, PAN differed from Tagalog in a significant respect. In Tagalog there is no correlation between word class and syntactic slot: any word, whether -VM or +VM, may occur in either the predicate slot or an argument slot. But in Philippine-type languages which retain non-indicative verb forms (and Tagalog doesn't), a non-indicative form derived from a root may only occur in the predicate slot, and the same must have been true of PAN. That is, in PAN, a -VM or +VM word occurred in both predicate and argument slots, but non-indicative forms in zero, **-a* and **-i* occurred only in the predicate slot. On the basis of their correlation of morphology and

syntactic distribution, we may legitimately call words formed with zero, **-a* and **-i* ‘verbs’. This distribution is depicted in Table 5.

Table 5: Distribution of PAn word classes by syntactic slots

	-VM words	+ VM words formed with <i>*-ən, *-an, *Si-</i> and <i>*in</i>	Verbs formed with zero, <i>*-a</i> and <i>*-i</i>
As predicate?	yes	yes	yes
As argument?	yes	yes	no

The agentive process infix **⟨um⟩* is omitted from Table 5 because the alternative hypothesis proposed here raises afresh the question, did **⟨um⟩* behave like **-ən, *-an* and **Si-* or like zero, **-a* and **-i*? In other words, did it form words which occurred in both predicate and argument slots, or did it only form verbs? On the basis of Table 4, it was excluded above from the ‘nominaliser’ affix set **-ən, *-an* and **Si-*. However, under the present hypothesis, the distinction made in Table 4 between the verbal voice-marking and the nominalising functions of these affixes does not exist,⁵² and there is no logical ground for its exclusion from this set. Instead, the fact that **⟨um⟩* forms are reconstructed with the same durative and perfective morphology as **-ən, *-an* and **Si-* (Table 1) implies that, like them, it formed words which occurred in both predicate and argument slots. This leads to a reformulation of the basic PAn system in (22) as follows:

(34)

	Actor	Patient	Location	Circumstantial
verbs	√	√-u	√-i	—
+VM words	⟨um⟩√	√-ən	√-an	Si-√

The lack of correlation between word class and syntactic slot in Tagalog is the result of Tagalog’s loss of (non-indicative) verb forms, i.e. of the rightmost column of Table 5, and cannot be projected back onto PAn. The presence of these verbs in PAn, however, makes it different not only from Tagalog but also from Straits Salish and Cayuga, as shown in Table 6.

Table 6: Word classes in Straits Salish, Cayuga, Tagalog and PAn

	-VM words	+VM words	Verbs
Straits Salish	no	yes	no
Cayuga	yes	yes	no
Tagalog	yes	yes	no
PAn	yes	yes	yes

This distribution leaves us with a question: what was the functional distinction between the verb and +VM word categories in (34)? The most obvious answer is that it was the same as in Philippine-type languages which reflect this morphological distinction:

⁵² The differences in the distributions of the different morphemes in Table 4 are then attributable either to their different PAn distributions between predicate and argument uses, or faulty descriptions of the modern languages, as we do not know if some Formosan languages could be better analysed along the same lines as Tagalog.

- (i) the verb forms $*\sqrt{\quad}$, $*\sqrt{-u}$ and $*\sqrt{-i}$ were used (a) as imperatives; (b) as verbs subordinate to some auxiliaries; and (c) expressing non-initial sequential events in narrative;
- (ii) the verb forms $*\sqrt{-a}$, $*\sqrt{-a-u}$ and $*\sqrt{-a-i}$ were used in irrealis predicates;
- (iii) +VM words were used in realis predicates and (as embedded predicates) in arguments.

The AV form of (i) is a plain stem, and this is crosslinguistically consistent with imperative use and with uses where no marking of aspectual categories is required. The acquisition of suffixes in the PV and LV forms is consistent with earlier preposition-capture, as noted above.

This alternative hypothesis is put forth here in order to account for features of the data which do not fit too well under the voice-from-nominalisation hypothesis. However, I do not want to argue that the voice-from-nominalisation hypothesis is wrong and the alternative hypothesis right. Indeed, the alternative hypothesis seems typologically rather odd. My concern is rather to suggest that morphosyntactic reconstruction is fraught with pitfalls, not least in the case of PAn, and that the most obvious reconstruction is not necessarily the right one. If a choice between the two reconstructions ever becomes easy, it will probably be when fine-grained descriptions of more Philippine-type languages — especially those of Taiwan — have been written.

Finally, the two hypotheses are not necessarily mutually exclusive. If the alternative hypothesis is correct, it may be that its content words were nominalisations at some substantially earlier interstage. But this is speculation.

4 Reconstructing Proto Malayo-Polynesian and subsequent interstages

As most of the languages described in this volume are Malayo-Polynesian, it is appropriate to review the reconstruction of PMP and what may have happened subsequently.

4.1 Proto Malayo-Polynesian

The structure of the reconstructed PMP clause was basically the same as that of the reconstructed PAn clause, and reconstructed PMP verb forms are set out in Table 7. They are similar to the PAn forms in Table 1. The alternative categorisations of the indicative forms as homophonous verbs and nouns (§3.2.1) or as an undivided category of content words (§3.2.2) also apply to the forms in Table 7.

A large majority of Malayo-Polynesian languages outside the Philippines are what Himmelmann (this volume) labels ‘Indonesian-type’ languages — languages which, for example, have preposed clitic pronouns and affix combinations which include reflexes of the applicative markers $*-i$ and $*-án/*-[a]kən$. Languages of the Philippine-type are limited geographically to Taiwan, the Philippines, north and central Borneo,⁵³ Madagascar and northern Sulawesi. Because the Philippine type is geographically constrained, it has occasionally been suggested that the latter is simply an areal phenomenon and that PAn and/or PMP are more likely to have been Indonesian-type languages than Philippine-type. However, Figure 1 shows why this cannot be so: Indonesian-type languages occur only

⁵³ Clayre (1996) provides a survey of voice systems in the languages of northern and central Borneo.

within Malayo-Polynesian, i.e. within one subgroup of Austronesian, whereas Philippine-type languages occur within more than one Formosan group and within Malayo-Polynesian. It follows, therefore, that both PAn and PMP must have been Philippine-type languages.

Table 7: Proto Malayo-Polynesian voice, mood and aspect morphemes

See key to Table 1.

	Actor	Patient	Location	Circumstantial
INDICATIVE				
Neutral	⟨ um ⟩√ *k⟨um⟩áRaw *k⟨um⟩aRát	√-ə n *kaRáw-ə n *kaRat-ə n	√-a n *kaRáw-a n *kaRat-a n	i-√ *i-káRaw *i-kaRát
Perfective	⟨ umin ⟩√ *k⟨um⟩in>áRaw *k⟨um⟩in>aRát	⟨ in ⟩√ *k⟨in>áRaw *k⟨in>aRát	⟨ in ⟩√-a n *k⟨in>aRáw-a n *k⟨in>aRat-a n	i-⟨ in ⟩√ *i-k⟨in>áRaw *i-k⟨in>aRát
Imperfective	⟨ um ⟩ R -√ *k⟨um⟩a-káRaw *k⟨um⟩a-kaRát	R -√-ə n *ka-kaRáw-ə n *ka-kaRat-ə n	R -√-a n *ka-kaRáw-a n *ka-kaRat-a n	i- R -√ *i-ka-káRaw *i-ka-kaRát
NON-INDICATIVE				
Atemporal	√ *káRaw *kaRát	√-a *kaRáw-a *kaRat-á	√-i *kaRáw-i *kaRat-í	√-á n *káRaw-á n *kaRát-á n
Projective	√-a *kaRáw-a *kaRat-á	(√-a w) (*kaRáw-a w) (*kaRat-á w)	√-a y *kaRáw-a y *kaRat-á y	—

The similarity between Table 1 and Table 7 is somewhat deceptive. Whereas a majority of PAn verbs seem to have adhered to the paradigm in Table 1, PMP evidently had a much richer derivational morphology which interacted with the morphemes in Table 7 to produce a bewildering variety of forms. This is an area which needs much more research, but two prefixes can be singled out which formed secondary roots from primary ones: *paN- ‘distributive’ and *paR- ‘durative’.⁵⁴ The semantic labels are very tentative. Distributive verbs apparently denoted plural actions, actions done by one or more agents to several things or by several agents to one thing. Durative verbs apparently denoted events regarded as ongoing or repetitive, as opposed to events regarded as punctual or viewed in their entirety (this distinction cut across the perfective/imperfective distinction of Table 7 which divided events into complete and incomplete).⁵⁵

⁵⁴ This quick-and-dirty attempt to reconstruct PMP affixes was limited to an examination of Ilokano (Rubino 2000), Tagalog (Ramos 1971a), the Bisayan dialects (Wolff 1972; Zorc 1977) and Binukid (Post 1992). It is clear that there are many more forms which should be reconstructed, but a much wider collection of data will be needed to do this with a hope of success.

⁵⁵ It is not easy to sort out the semantics accurately here, since, for example, Ilokano and Tagalog differ in their treatment of the perfective/imperfective distinction (Reid 1992).

Table 8 shows part of the reconstructed paradigm of the secondary distributive root **panakaw* ‘steal’, formed from **paN-* and the primary root **takaw*. The **-N-* of **paN-* combined with root-initial **p*, **t*, **k* and **c/*s* respectively to give **-m-*, **-n-*, **-ŋ-* and **-ñ-*, disappeared before a root-initial nasal, and otherwise became a nasal homorganic with the root-initial consonant. Accent is not reconstructed here, for lack of evidence. The data assembled so far are insufficient to reconstruct imperfective and non-indicative forms solidly, but non-indicative forms can be inferred by analogy with Table 7, e.g. AV atemporal **panakaw*:

Table 8: Proto Malayo-Polynesian **paN-* with voice, mood and aspect morphemes

See key to Table 1.

	Actor	Patient	Location	Circumstantial
INDICATIVE				
Neutral	maN-√ <i>*manakaw</i>	paN-√-ən <i>*panakaw-ən</i>	paN-√-an <i>*panakaw-an</i>	i-paN-√ <i>*i-panakaw</i>
Perfective	naN-√ <i>*nanakaw</i>	⟨in⟩paN-√ <i>*p⟨in⟩anakaw</i>	⟨in⟩paN-√-an <i>*p⟨in⟩anakaw-an</i>	i-⟨in⟩paN-√ <i>*i-p⟨in⟩anakaw</i>

Table 9 shows the corresponding paradigm for the secondary durative root **paR-kaRat* ‘bite’. From available Philippine data, it seems probable that no PV or LV forms incorporating **paR-* occurred. Instead, the primary root was used.

Table 9: Proto Malayo-Polynesian **paR-* with voice, mood and aspect morphemes

See key to Table 1.

	Actor	Patient	Location	Circumstantial
INDICATIVE				
Neutral	maR-√ <i>*maR-kaRat</i>	√-ən <i>*kaRat-ən</i>	√-an <i>*kaRat-an</i>	i-paR-√ <i>*i-paR-kaRat</i>
Perfective	naR-√ <i>*naR-kaRat</i>	⟨in⟩√ <i>*k⟨in⟩aRat</i>	⟨in⟩√-an <i>*k⟨in⟩aRat-an</i>	i-⟨in⟩paR-√ <i>*i-p⟨in⟩aR-kaRat</i>

It is clear that there were also many other derivational prefixes, e.g. **paka-* ‘abilitative’, and that several affixes often combined to give morphologically complex forms, as modern Philippine languages attest.

Noun phrases in PMP were marked in the same basic way as in PAN (Table 2), and PMP noun phrase markers are shown in Table 10. The main differences between the PMP and PAN systems are (i) that there was no separate phrase marker for topics in PMP; (ii) there is evidence of three sets of common phrase markers in PMP rather than two.⁵⁶ There is also some evidence that GEN, NPIV and LOC were all distinctly marked in PMP, but we should be cautious about this, as the three-way distinction is made only in Yami, Ivatan and the other languages of the Batan Islands (between Taiwan and the Philippines).

⁵⁶ This table is based on analysis reported in Ross (2001) and based partly on Reid (1978, 1979). It is possible that three sets of common phrase markers also occurred in PAN, but the evidence is less clear.

Table 10: Some Proto Malayo-Polynesian phrase markers

	SPEC	GEN	NPIV	LOC
common (default)	*i	*ni	*si	*di, *i
common (present)	*a, (*sa)	*na	*ta, *sa	*da, *ka, *sa
common (absent)	*u, (*su)	*nu	*tu, *su	*du (?)
personal	*si	*ni	—	*ka [n]i

The (partial) PAN pronominal system shown in Table 3 evolved into the (partial) PMP system in Table 11. Important changes include what Blust (1977) calls the second politeness shift, a set of innovations that defines the Malayo-Polynesian subgroup. Its elements are:

- (i) the PAN plain free and polite free sets became a single PMP free set: the PMP free forms *ikahu 2S, *[i]kami 1EP, and *[i]kamu 2P reflect the polite PAN free forms *i-ka-Su, *[i-]k-ami and *[i-]k-amu, and the plain PAN free forms *[i-]Su, *i-ami and *i-amu are lost;⁵⁷
- (ii) PMP *=mu GEN:2S reflects the PAN 2P clitic *=mu, and the PAN 2S clitic *=Su is lost;⁵⁸
- (iii) PMP has 2P forms, free *[i]ka-ihu and *kamu-ihu and GEN *=ihu, *=nihu, *=mu-ihu which incorporate *-ihu, apparently reflecting the PAN free 2S *i-Su.⁵⁹

Where only one set of short clitic pronouns for PIV and GEN is reconstructable for PAN, separate sets are reflected in the singular in PMP: new pivot clitics have been created by cliticising free forms, leaving the old short clitic set to serve only as short genitives in PMP. PAN GEN2 clitics have disappeared, except for *=mami 1EP, and have otherwise been replaced by PAN GEN3, now cliticised.

Table 11: Proto Malayo-Polynesian personal pronouns

	Free ⁶⁰	PIV	GEN (short)	GEN (long)
1S	*[i]jaku	*=aku	*=ku	*=n(a)ku
2S	*ikahu	*=kaw	*=mu	*=nihu
3S	*[s]iya	*=ya	*=(y)a, *-ña	*=niya
1EP	*[i]kami	—	—	*=mami
1IP	*[i]kita, ita	*=ta	*=ta	—
2P	*[i]kamu	—	—	—
	*[i]ka-ihu	—	*=ihu	*=nihu
	*kamu-ihu	—	—	*=mu-ihu
3P	*sida	*=da	*=da	*=nida

⁵⁷ Blust (1977) notes this change only with regard to the 2S form. Note, incidentally, that the change does not apply to the 1IP forms, where PAN polite and plain forms are both retained in the PMP 1IP and 1ID.

⁵⁸ Bungku-Tolaki languages have GEN:2S -u alternating with -mu and nominative 2S u- (this series is also historically derived from the PMP genitives; Mead 1998:122-125, 130-131). It is just possible that these reflect PAN *-Su and that the latter had not been lost in PMP.

⁵⁹ Whilst changes (i) and (ii) are typical politeness shifts (the polite form becomes the default, the 2P becomes the 2S), change (iii) isn't. In all probability this is not a politeness shift but a reinforcement of the plural form: we find *kamu alongside *kamu-ihu. The form *ka-ihu may have been a dual, from *kahu-ihu 'you (and) you'.

4.2 The genesis of Indonesian-type languages

The term ‘Indonesian-type language’ is, as Himmelmann notes, a vague one. It refers to western Malayo-Polynesian languages with (usually) two-voice verbal systems in which there are (i) preposed clitic pronouns and (ii) affix combinations which include reflexes of the applicative markers **-i* and **-án/*-[a]kən*. These systems vary as to the number of members in the preposed clitic paradigm, but as a rule, if there is just one clitic, it is 1S; if two, then 1S and 2S; if three, then the singular persons (Himmelmann 1996). Indonesian-type languages also vary in the forms of the ‘passive’ affix/proclitic (**·in/*ni/*di=*) and the circumstantial applicative (**-án/*-[a]kən*), and in numerous details. A proto language ancestral to all Indonesian-type languages is not reconstructed here, as it is not clear that they form a subgroup within Malayo-Polynesian. Indeed, it seems very probable that they don’t, and that their similarities are at least in part the results of independent parallel developments and of language contact. Very little is known about any other than the lowest-order subgroups within the region occupied by Indonesian-type languages (southern Borneo, peninsular Malaysia, Sumatra except Aceh, Java, Bali, Lombok, western Sumbawa, and central and south Sulawesi and its southern offshore islands), and even some of the accepted groups are open to question (Ross 1995b). Published reconstructions of verbal morphology of interstage languages within the region are Adelaar’s (1992) of Proto Malayic and van den Berg’s (1996) of Proto Celebic.⁶¹

Wolff (1996) takes up insights from SPR and from Himmelmann (1996) to explain how Indonesian-type languages developed from Philippine-type, and my account here largely summarises his. As a prototype of the Indonesian-type system he takes the Standard Indonesian system shown in Table 12.

Table 12: Standard Indonesian voice and applicative morphemes

(Italicised forms are pronominal clitics)

	Active	Passive			
		1S actor	2S actor	3S actor	no actor
Patient object	meN-√	<i>ku=√</i>	<i>kau=√</i>	<i>di-√=ñā</i>	<i>di-√</i>
Location object	meN-√-i	<i>ku=√-i</i>	<i>kau=√-i</i>	<i>di-√-i=ñā</i>	<i>di-√-i</i>
Circumstantial object	meN-√-kan	<i>ku=√-kan</i>	<i>kau=√-kan</i>	<i>di-√-kan=ñā</i>	<i>di-√-kan</i>

Wolff identifies three fundamental changes which have occurred to produce the Standard Indonesian system from a Philippine-type system: (1) the formation of a paradigm of passive proclitics — person proclitics for 1S and 2S actors and a general proclitic *di=* otherwise; (2) combinations of voice prefixes and suffixes which do not occur in Philippine-type languages, particularly *meN-* (< **maN-*) and *-i* (< **-i*); and (3) loss of the neutral/perfective distinction. He observes that languages which have made innovation 1 have also made innovation 2 and *vice versa*, but that there are languages that have made innovations 1 and 2 but not 3.

⁶⁰ Forms with initial **i-* may have been clause-initial topic pronouns.

⁶¹ These two reconstructions differ in status. Adelaar defines the Malayic subgroup by its shared innovations. Van den Berg assumes the integrity of the Celebic subgroup on the basis of shared similarities and of the fact that he can integrate these languages into a common story: research on shared innovations remains to be done. Van den Berg’s reconstruction does not deal with applicative suffixes, although these occur in Celebic languages. Mead’s (1998) thesis provides a well argued, well founded reconstruction of aspects of Proto Bungku-Tolaki (southeast Sulawesi).

Languages of the latter type, all in central Sulawesi, represent an early stage of the transition to the Indonesian type, and Table 13 is a hypothetical picture of what such a language might have looked like at an earlier phase of its history.⁶² The 1S clitics *ku=* and *=ku* in the table are stand-ins for what would in some languages have been a defective paradigm with perhaps only (some) singular members. Table 13 is only an aid to presentation, not a reconstruction. It depicts what seem to be the essential features of an early Indonesian-type language, in order to facilitate comparison with Tables 7–9.

Table 13: Voice and applicative morphemes in a hypothetical early Indonesian-type language

See key to Table 1.

	Active	Passive
Patient undergoer		
neutral	maN-√, <um>√	√[= <i>ku</i>]
perfective	naN-√	<i>ku</i> =√, <in>√[= <i>ku</i>]
Location undergoer		
neutral	maN-√-i, <um>√-i	√-i[= <i>ku</i>]
perfective	naN-√-i	<i>ku</i> =√-i, <in>√-i[= <i>ku</i>]
Circumstantial undergoer		
neutral	maN-√-an	√-an[= <i>ku</i>]
perfective	naN-√-an	<i>ku</i> =√-an, <in>√-an[= <i>ku</i>]

Wolff does not deal with innovation 3, loss of the neutral/perfective distinction, but it is worth noting that Standard Indonesian (Table 12) and other languages which have lost this distinction seem to preserve neutral forms in the active voice, but perfective forms in the passive. It is also noteworthy that Indonesian-type languages have tended to abandon the *<um>-infixation (Table 7) in favour of **maN*-prefixing (Table 8).

Wolff illustrates the first step in innovation 1, the formation of actor proclitics to passives, with examples from the Philippine-type language Cebuano Bisayan. Cebuano has pre-verbal auxiliaries which are followed by an atemporal verb, as described in §3.1. With undergoer-voice verbs, the genitive clitic marking the actor follows the auxiliary, if there is one; otherwise it follows the verb:

(35) Cebuano Bisayan (Wolff 1996:26)

a. *Gi-hugás-an=ku ang manga plátu.*
 PF-wash-LV=GEN:1S SPEC P plate
 'I washed the plates.'

b. *Walaq=ku hugás-i ang manga plátu.*
 NEG=GEN:1S wash-LV.AT SPEC P plate
 'I didn't wash the plates.'

⁶² Wolff (1996:20-21) uses Totoli to illustrate what an early Indonesian-type language would look like, but Himmelmann (1996:123-124) analyses (apparently) the same set of data as a Philippine-type language, so I have preferred not to use it here.

SPR argue that by auxiliary deletion the genitive clitic became stranded in front of the verb. Wolff deletes the auxiliary by an argument based on analogy. Either way, one finishes up with a sentence like the pseudo-Cebuano **ku=hugás-i ang manga plátu* 'I washed the plates', with the genitive clitic in front of the passive verb, as in Tables 12 and 13. In such a sentence the (main) verb has the form of an atemporal from Table 7, because the verb in (35b) is subordinate to an auxiliary. Predictably, all three passive forms in Tables 12 and 13 reflect the PMP atemporals. However, we would expect the passive patient-undergoer form to be suffixed with **-a* (< PMP PV atemporal): instead, it is unsuffixed, perhaps reflecting a conflation of AV and PV atemporals.

Wolff then uses examples from Totoli to illustrate how innovation 2 occurred, whereby combinations of voice prefixes and suffixes arose. In Totoli, only singular pronouns distinguish pivot and genitive forms (see Table 11). Other pronouns and noun phrases are not marked for case. The pair of sentences below reflects the distinction between, in (36a), one of the newly created forms of the previous paragraph, the suffixless PV atemporal and, in (36b), the AV neutral form reflecting **maN-*:

- (36) Totoli (Wolff 1996:27)
- a. *Ku=kaan* *sagin.*
 GEN:1S=eat.PV.NEUTRAL banana
 'I eat the banana.'
- b. *Aku mangaan* *sagin.*
 PIV:1S AV.NEUTRAL.eat banana
 'I am eating a banana.'

This set the scene for the creation of new verb forms by analogy. The relation in (36) is shown as (37a). Clitic-stranding had created the forms on the left in (37), and the forms **maN-√-i* and **maN-√-an* on the right of these relations were created by analogy:

- (37)a. *ku=√* : *maN-√*
 b. *ku=√-i* : *maN-√-i*
 c. *ku=√-an* : *maN-√-an*

The outcome was a reorientation of the PMP system in Table 7 to give systems like the one hypothesised in Table 13. Where PMP had only one set of actor voice forms, an Indonesian-type system has three sets of active forms, with patient, location and circumstantial undergoers respectively. Their corresponding passives are descended from the earlier atemporal patient, location and circumstantial voice forms.

Despite the morphosyntactic changes that separate Indonesian-type languages from Philippine-type, the functions of the voice system have in many languages remained virtually unchanged. The default forms in narrative discourse are passive, whilst actives are reserved for special uses, including when the patient is non-specific and when the syntax requires an actor voice. Wouk (1984, 1986) reports a set of facts regarding Toba Batak pivot choice and the interpretation of undergoer specificity in actor-voice clauses which are parallel to those listed for Tagalog following (8). The condition that a specific patient must be pivot in an independent clause held for early modern Malay (Hopper 1988). Topicality has been shown to be a determinant of pivot choice in Balinese (Pastika 1999) and Sasak (Wouk 1999).

Some languages have undergone a further syntactic innovation. The noun phrase immediately following the verb has become strongly bound to it so that verb + noun phrase form a single constituent. The postverbal noun phrase is the patient with actor voice and the

actor with patient voice, i.e. the voice system is symmetrical. Similar observations have been made about Balinese (Artawa 1994; Arka 1998).

For Toba Batak the bonding of verb + noun phrase is attested by pitch-accent behaviour (Emmorey 1984), by the fact that an adverb cannot intervene between verb and noun phrase, by the fact that such 'verb phrases' can be co-ordinated, whether they are both AV or OV, and by the fact that post-verbal noun phrase cannot be fronted, whereas the pivot noun phrase can (Schachter 1984).

- (36) Toba Batak (Schachter 1984:123)
- a. *Mang-ida si Ria si Torus.*
 AV-see PERS Ria PERS Torus
 'Torus sees/saw Ria.'
- b. *Di-ida si Torus si Ria.*
 PV-see PERS Torus PERS Ria
 'Torus sees/saw Ria.'

We do not have direct evidence about how this innovation occurred, but it seems to represent the grammaticisation of frequently occurring (but not rule governed) constituent sequences resulting from the Philippine-type tendency to place the pivot noun phrase at the end of the clause. It was apparently motivated by the loss of phrase markers to indicate case.

A comparison of Tables 11 and 12 shows three other innovations in the Indonesian system which require comment. They are: (i) the form of the general passive proclitic *di=*; (ii) the extent of the paradigm of passive actor proclitics, namely *ku=* and *kau=*; and (iii) the form **-kan*.

As Table 13 shows, Indonesian-type languages may also have a general passive affix reflecting **in>*, often as *ni-*. If an actor pronoun cooccurs with this affix, it remains in its inherited position, as an enclitic to the verb. Indonesian and a number of other Indonesian-type languages have replaced this with *di=*. There is good reason to believe that **di=* was a Proto Malayic innovation whose reflexes have spread by contact into non-Malayic languages, replacing the inherited affix reflecting **in>* (this is directly attested for Javanese).⁶³

Standard Indonesian has two actor proclitics on passive verbs, *ku=* 1S and *kau=* 2S. Other Indonesian-type languages have only one, e.g. Totoli *ku=* 1S. Yet others, e.g. Kulawi (Wolff 1996:29, citing Adriani & Esser 1939), have a full set. More research is needed to understand fully what has happened here. Himmelmann (1996) interprets the Sulawesi data as indicating that pronominal proclitic sets have grown in membership over time. Van den Berg (1996) reconstructs a full set of proclitics for Proto Celebic, inferring that languages with smaller sets have lost members over time. In the case of Indonesian, however, it is unlikely that the language has ever had a full set of actor proclitics, as the enclitic actor pronoun *=ña* in a form like *di=makan=ña* PASS-eat-GEN:3S 'be eaten by him/her' reflects the state of affairs in PMP, i.e. before the rise of Indonesian-type languages.

Finally, Table 13 shows a hypothetical Indonesian-type language with the circumstantial undergoer suffix **-an*, reflecting PMP CV atemporal *-án*, and indeed many Indonesian-type

⁶³ The question of the origin of **di=* is beyond the scope of this paper. A short summary of the relevant literature and an evaluation of the alternatives is given by Ross (forthcoming).

languages, like Totoli, do reflect **-an*.⁶⁴ Others, however, have replaced it with a reflex of **-[a]kən*, like Standard Indonesian *-kan*. This form appears to have been a captured preposition, as Indonesian also has the preposition *akan*, but the origins of the suffix **-[a]kən* are not well understood. Adelaar (1992) presents a strong argument that it should not be reconstructed for Proto Malayic. Yet its reflexes have also replaced reflexes of **-an* in non-Malayic languages from Sumatra to Oceania. It is possible that this suffix has arisen sometime during the history of Malay, and that it has been borrowed into Malayic and non-Malayic languages alike as a result of bilingualism in those languages and Malay. But this is an *ad hoc* solution without direct support, and it does not explain the presence of apparent reflexes of **-[a]kən* in Oceanic languages.

The history of Indonesian-type languages outlined in the foregoing paragraphs fits some languages better than others. For example, the Bungku-Tolaki languages of southeast Sulawesi fit our assumed definition of an Indonesian-type language, except that they have lost **-i* 'location undergoer'. However, the alternative forms shown in Table 13 have undergone an interesting functional split. Reflexes of **maN-√* and **ku=√* are what Mead (1998) calls respectively 'antipassive' and 'active'. Despite the (justifiable) shift in terminology, the antipassive is clearly the functional descendant of the PMP actor voice and corresponds to the Indonesian-type active in being used only when the undergoer is non-specific. The active is the functional descendant of the PMP patient voice (!) and corresponds to the Indonesian-type passive as it is the default main-clause transitive form. Meantime, reflexes of **um√* and **in√ [=ku]* continue respectively as active and passive in various contexts other than canonic main clauses.

A quite different aberrant Indonesian-type language is Balinese—aberrant because it lacks passive proclitics altogether. Instead the passive has a plain stem and reflects **√ [=ku]*, whilst the active stem displays nasal assimilation, i.e. has the form *N-√* (Artawa 1994). Inscriptional Old Balinese, however, reflected passive **in√ [=ku]* and both active **um√* and active **maN-√*. Beratha (1992) suggests that modern *N-√* represents a conflation of the two Old Balinese forms. The applicative suffixes corresponding to **-i* and **-an/*-akən* are *-in* and *-an*, both unexpected forms.

As the discussion in this section implies, the history of Indonesian-type languages is not well understood. Their sheer typological variety requires more research, and should at the same time be a warning to us against jumping to historical conclusions.

References

- Adelaar, K. Alexander, 1992, *Proto Malayic: the reconstruction of its phonology and parts of its lexicon and morphology*. Canberra: Pacific Linguistics.
- Adriani, Nicolaus and S.J. Esser, 1939, *Koelawische taalstudiën*, 3 vols. Bandoeng: Nix & Co.
- Arka, Wayan, 1998, From morphosyntax to pragmatics in Balinese. PhD thesis, University of Sydney.
- Artawa, Ketut, 1994, Ergativity and Balinese syntax. PhD thesis, La Trobe University, Melbourne.

⁶⁴ It is possible that in some Indonesian-type languages *-an* may reflect PMP LV indicative **-an* or represent a conflation of this with CV atemporal **-án*. This needs investigation.

- Asai, Erin, 1953, *The Sedik language of Formosa*. Kanazawa: Cercle Linguistique de Kanazawa.
- Baker, John W., 1991, Semantic interactions of nominal determiners in Ilokano. Paper presented to the Sixth International Conference on Austronesian Linguistics, Honolulu.
- Bell, Sarah J., 1976, Cebuano subjects in two frameworks. PhD dissertation, University of California, San Diego. Reprinted 1979 by the Indiana University Linguistics Club.
- Beratha, Ni Luh Sutjiati, 1992, Evolution of verbal morphology in Balinese. PhD thesis, The Australian National University.
- Berg, René van den, 1996, The demise of focus and the spread of conjugated verbs in Sulawesi. In Steinhauer, ed., 1996:89-114.
- Blust, Robert A., 1977, The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *University of Hawaii Working Papers in Linguistics*, 9/2:1-15.
- 1997, Rukai stress revisited. *Oceanic Linguistics*, 36:398-403.
- 1998, Ca- reduplication and Proto-Austronesian grammar. *Oceanic Linguistics*, 37:29-64.
- 1999a, Notes on Pazehe phonology and morphology. *Oceanic Linguistics*, 38:321-365.
- 1999b, Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In Zeitoun and Li, eds, 1999:31-94.
- In press, *Thao dictionary*. Canberra: Pacific Linguistics.
- Bowden, John, 2001, A revised typology for three place predicates: direct objects, primary objects, and 'split' objects. Unpublished ms. Research School of Pacific and Asian Studies, The Australian National University, Canberra.
- Brainard, Sherri L., 1994, Voice and ergativity in Karao. PhD dissertation, University of Oregon.
- Clayre, Beatrice M., 1996, The changing face of focus in the languages of Borneo. In Steinhauer, ed., 1996:51-88.
- Cooreman, Ann M., 1983, Topic continuity and the voicing system of an ergative language. In Talmy Givón, ed. *Topic continuity in discourse: a quantitative cross-language study*, 425-489. Amsterdam: John Benjamins.
- 1987, *Transitivity and discourse continuity in Chamorro narratives*. Berlin: Mouton de Gruyter.
- Cooreman, Ann M., Barbara Fox and Talmy Givón, 1984, The discourse definition of ergativity. *Studies in Language*, 8:1-34.
- 1988, The discourse definition of ergativity: a study of Chamorro and Tagalog texts. In McGinn, ed., 1988:387-425.
- Cumming, Susanna and Fay Wouk, 1987, Is there "discourse ergativity" in Austronesian languages? *Lingua*, 91:271-296.
- Dahl, Otto Christian, 1973, *Proto-Austronesian*. Oslo: Studentlitteratur.
- De Guzman, Videia P., 1997, Verbal affixes in Tagalog: inflection or derivation? In Cecilia Odé and Wim Stokhof, ed. *Proceedings of the Seventh International Conference on Austronesian Linguistics*, 303-325. Amsterdam: Rodopi.
- 2000, Some remarks on the grammatical functions of the nonabsolute agent in Tagalog. In Videia P. De Guzman and Byron Bender, *Grammatical analysis: morphology, syntax and semantics: studies in honor of Stanley Starosta*. Canberra: Pacific Linguistics.
- Dixon, R.M.W., 1979, Ergativity. *Language*, 55:59-138.

- 1994, *Ergativity*. Cambridge: Cambridge University Press.
- Dryer, Matthew S., 1997, Are grammatical relations universal? In Joan L. Bybee, John Haiman, and Sandra A. Thompson, ed. *Essays on language function and language type, dedicated to T. Givón*, 115-143. Amsterdam: John Benjamins.
- 1999, Functionalism and the metalanguage-theory confusion. Unpublished ms, State University of New York at Buffalo.
- Durie, Mark, 1987, Grammatical relations in Acehnese. *Studies in Language*, 11:365-399.
- Egerod, Søren, 1965, Verb inflexion in Atayal. *Lingua*, 15:251-282.
- 1966, Word order and word classes in Atayal. *Language*, 42:346-369.
- 1969, The origin of headhunting: an Atayal text with vocabulary. *Bulletin of the Institute of History and Philology, Academia Sinica*, 39:291-325.
- Egli, Hans, 1990, *Paiwangrammatik*. Wiesbaden: Otto Harrassowitz.
- Emmorey, Karen, 1984, The intonation system of Toba Batak. In Schachter, ed., 1984: 37-58.
- Evans, Bethwyn and Malcolm Ross, 2001, The history of Proto Oceanic *ma-. *Oceanic Linguistics* 40.
- Ferrell, Raleigh, 1968, The Pazeh-Kahabu language. *Bulletin of the Department of Archaeology and Anthropology, National Taiwan University*, 31-32:73-97.
- 1980, Phonological subgrouping of Formosan languages. In Naylor, ed., 1980:241-254.
- 1982, *Paiwan dictionary*. Canberra: Pacific Linguistics.
- Foley, William A., 1993, The conceptual basis of grammatical relations. In William A. Foley, ed. *The role of theory in language description*, 131-174. Berlin: Mouton de Gruyter.
- Foley, William A. and Robert D. Van Valin jr, 1984, *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Gault, JoAnn, 1999, Thoughts concerning ergativity, subject and focus in Sama Bangingi'. In Zeitoun and Li, eds, 1999:393-422.
- Gibson, Jeanne D. and Stanley Starosta, 1987, Ergativity east and west. In Philip Baldi, ed. *Linguistic change and reconstruction methodology*, 195-210. Berlin: Mouton de Gruyter.
- Halim, Amran, Lois Carrington and S.A. Wurm, eds, 1982, *Papers from the Third International Conference on Austronesian Linguistics 1: Tracking the travellers*. Canberra: Pacific Linguistics.
- Harvey, Mark, 1982, Subgroups in Austronesian. In Halim, Carrington and Wurm, eds, 1982:47-99.
- Himmelman, Nikolaus P., 1991, The Philippine challenge to universal grammar. *Institut für Sprachwissenschaft, Universität zu Köln, Arbeitspapier*, 15 (Neue Folge).
- 1996, Person marking and grammatical relations in Sulawesi. In Steinhauer, ed., 1996:115-136.
- forthcoming a, Lexical categories and voice in Tagalog. In Peter Austin and Simon Musgrave eds *Voice and grammatical relations in Austronesian languages*. Stanford: CSLI Publications.
- forthcoming b, Tagalog. In Nikolaus P. Himmelman and K. Alexander Adelaar, eds *The Austronesian languages of south-east Asia and Madagascar*. Richmond: Curzon Press.

- Ho, Arlene Y.L., 1993, Transitivity, focus, case and the auxiliary verb systems in Yami. *Bulletin of the Institute of History and Philology, Academia Sinica*, 62:83-147.
- Hopper, Paul J., 1979, Some discourse sources of ergativity. *University of Hawaii Working Papers in Linguistics*, 11:137-153.
- 1988, How ergative is Malay? In McGinn, ed., 1988:441-454.
- Hopper, Paul J. and Sandra Thompson, 1980, Transitivity in grammar and discourse. *Language*, 56:251-299.
- Huang, Lillian M., 1994, Ergativity in Atayal. *Oceanic Linguistics*, 33:129-143.
- 1995, *A study of Mayrinax syntax*. Taipei: The Crane.
- 2000, Verb classification in Mayrinax Atayal. *Oceanic Linguistics*, 39:364-390.
- Huang, Shuanfang, 2000, The pragmatics of focus in Tsou and Seediq. *National Taiwan University Working Papers in Linguistics*, 3:111-127.
- Jelinek, Eloise and Richard A. Demers, 1994, Predicates and pronominal arguments in Straits Salish. *Language*, 70:697-736.
- Kroeger, Paul R., 1991, The event line in Kimaragang narrative. In Stephen H. Levinsohn, ed. *Thematic continuity and development in languages of Sabah*, 93-104. Canberra: Pacific Linguistics.
- 1993, *Phrase structure and grammatical relations in Tagalog*. Stanford: CSLI Publications.
- Lee, Amy Pei-jung, 1997, The case-marking and focus systems in Kavalan. MA thesis, Tsing-hua University, Taiwan.
- Li, Paul Jen-kuei, 1978, The case-marking systems of the four less-known Formosan languages. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, 569-615. Canberra: Pacific Linguistics.
- Li, Paul Jen-kuei, Dah-an Ho, Ying-kuei Huang, Cheng-hwa Tsang and Chiu-yu Tseng, eds, 1995, *Austronesian studies relating to Taiwan*. Symposium Series of the Institute of History and Philology, Academia Sinica 4. Taipei: Institute of History and Philology, Academia Sinica.
- McGinn, Richard, ed., 1988, *Studies in Austronesian Linguistics*. Athens, Ohio: Center for Southeast Asian Studies, Ohio University Center for International Studies.
- Manning, Christopher D., 1996, *Ergativity: argument structure and grammatical relations*. Stanford: CSLI Publications.
- Mead, David, 1998, Proto-Bungku-Tolaki: reconstruction of its phonology and aspects of its syntax. PhD dissertation, Rice University.
- Mirikitani, Leatrice T., 1972, *Kapampangan syntax*. *Oceanic Linguistics Special Publication* 10. Honolulu: The University Press of Hawaii.
- Mithun, Marianne, 2000, Noun and verb in Iroquoian languages: multicategorisation from multiple criteria. In Petra M. Vogel and Bernard Comrie, ed. *Approaches to the typology of word classes*, 397-420. Berlin: Mouton de Gruyter.
- Naylor, Paz Buenaventura, 1980, Linking, relation-marking, and Tagalog syntax. In Naylor, ed., 1980:33-50.
- Naylor, Paz Buenaventura, ed., 1980, *Austronesian studies: papers from the Second Eastern Conference on Austronesian languages*. Ann Arbor: University of Michigan, Center for South and Southeast Asian Studies.

- Pastika, I Wayan, 1999, *Voice selection in Balinese narrative discourse*. PhD thesis, The Australian National University.
- Pawley, Andrew K. and Malcolm D. Ross, 1993, Austronesian historical linguistics and culture history. *Annual Review of Anthropology*, 22:425-459.
- Payne, Thomas, 1994, The pragmatics of voice in a Philippine language: actor-focus and goal-focus in Cebuano narrative. In Talmy Givón, ed. *Voice and inversion*. Amsterdam: John Benjamins.
- Pecoraro, Ferdinando, 1979, *Eléments de grammaire taroko, précédés de la présentation de la vie et de la culture des taroko*. Paris: Association Archipel.
- Post, Ursula, 1992, *Binukid dictionary*. *Studies in Philippine Linguistics* 9/2. Manila: Linguistic Society of the Philippines and Summer Institute of Linguistics.
- Ramos, Teresita V., 1971a, *Tagalog dictionary*. Honolulu: University of Hawai'i Press.
1971b, *Tagalog structures*. Honolulu: University of Hawai'i Press.
- Reid, Lawrence A., 1966, *An Ivatan syntax*. Oceanic Linguistics Special Publication 2. Honolulu: Pacific and Asian Linguistics Institute, University of Hawaii.
1978, Problems in the reconstruction of Proto-Philippine construction markers. In Wurm and Carrington, eds, 1978:33-66.
1979, Evidence for Proto-Philippine nominative marking. *Philippine Journal of Linguistics*, 10:1-20.
1982, The demise of Proto-Philippines. In Halim, Carrington and Wurm, eds, 1982:201-216.
1992, On the development of the aspect system in some Philippine languages. *Oceanic Linguistics*, 31:65-91.
- Ross, Malcolm, 1992, The sound of Proto-Austronesian: an outsider's view of the Formosan evidence. *Oceanic Linguistics*, 31:23-64.
1995a, Reconstructing Proto Austronesian verbal morphology: evidence from Taiwan. In Li et al., eds, 727-791.
1995b, Some current issues in Austronesian linguistics. In D.T. Tryon, ed. *Comparative Austronesian Dictionary* 1, 45-120. Berlin: Mouton de Gruyter.
2001, The reconstruction of Proto Malayo-Polynesian construction markers. Unpublished ms. Research School of Pacific and Asian Studies, The Australian National University, Canberra.
forthcoming, Notes on the prehistory and internal subgrouping of Malayic. In John Bowden and Nikolaus P. Himmelmann, eds *Papers in Austronesian Linguistics*. Canberra: Pacific Linguistics.
- Rubino, Carl R.G., 2000, *Ilocano grammar and dictionary*. Honolulu: University of Hawai'i Press.
- Sasse, Hans-Jürgen, 1993, Das Nomen — eine universale Kategorie? *STUF — Sprachtypologie und Universalienforschung*, 46:187-221.
- Schachter, Paul, 1976, The subject in Philippine languages: topic, actor, actor-topic or none of the above. In Charles Li, ed. *Subject and topic*, 491-518. New York: Academic Press.
1977, Reference-related and role-related properties of subjects. In Peter Cole and Jerrold M. Sadock, ed. *Syntax and semantics* 8: *Grammatical relations*, 279-306. New York: Academic Press.

- 1984, Semantic-role-based syntax in Toba Batak. In Schachter, ed., 1984:122-149.
- Schachter, Paul, ed., 1984, *Studies in the structure of Toba Batak*. UCLA Occasional Papers in Linguistics 5.
- Schachter, Paul and Fe T. Otones, 1972, *Tagalog reference grammar*. Berkeley: University of California Press.
- Starosta, Stanley, 1985, Verbal inflection versus deverbal nominalisation in PAN: the evidence from Tsou. In Andrew K. Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 281–312. Canberra: Pacific Linguistics.
- 1986, Focus as recentralisation. In Paul Geraghty, Lois Carrington and S.A. Wurm, ed. *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*, 73-95. Canberra: Pacific Linguistics
- 1988, A grammatical typology of Formosan languages. *Bulletin of the Institute of History and Philology, Academia Sinica*, 59:541-576.
- 1995, A grammatical subgrouping of Formosan languages. In Li et al., eds, 1995:683-726.
- 1999, Transitivity, ergativity and the best analysis of Atayal case marking. In Zeitoun and Li, eds, 1999:371-392.
- Starosta, Stanley, Andrew K. Pawley and Lawrence A. Reid, 1981, The evolution of focus in Austronesian. Paper presented to the Third International Conference on Austronesian Linguistics, Bali.
- 1982, The evolution of focus in Austronesian. In Halim, Carrington and Wurm, eds, 1982:145-170.
- Steinhauer, Hein, ed., 1996, *Papers in Austronesian linguistics No.3*. Canberra: Pacific Linguistics.
- Thurgood, Graham, 1999, *From ancient Cham to modern dialects: two thousand years of language contact and change*. Oceanic Linguistics Special Publication 28. Honolulu: University of Hawai'i Press.
- Van Valin, Robert D., jr, and Randy J. LaPolla, 1997, *Syntax: structure, meaning and function*. Cambridge: Cambridge University Press.
- Wolff, John U., 1972, *A dictionary of Cebuano Visayan*, 2. Ithaca N.Y. and Manila: Cornell University Southeast Asia Program and Linguistic Society of the Philippines.
- 1973, Verbal inflection in Proto-Austronesian. In Andrew Gonzalez, ed. *Parangal kay Cecilio Lopez*, 71-91. Quezon City: Linguistic Society of the Philippines.
- 1996, The development of the passive verb with pronominal prefix in western Austronesian languages. In Bernd Nothofer, ed. *Reconstruction, classification, description: festschrift in honor of Isidore Dyen*, 15-40. Hamburg: Abera.
- Wouk, Fay, 1984, Scalar transitivity and trigger choice in Toba Batak. In Schachter, ed., 1984:195-219.
- 1986, Transitivity in Batak and Tagalog. *Studies in Language*, 10:391-424.
- 1999, Sasak is different: a discourse perspective on voice. *Oceanic Linguistics*, 38:91-114.
- Wurm, S.A. and Lois Carrington, eds, 1978, *Second International Conference on Austronesian Linguistics: proceedings* Fascicle 1: *Western Austronesian*. Canberra: Pacific Linguistics.

- Zeitoun, Elizabeth, 2000, Dynamic vs. stative verbs in Mantaoran (Rukai). *Oceanic Linguistics*, 39:415-427.
- Zeitoun, Elizabeth and Lillian M. Huang, 2000, Concerning *ka-*, an overlooked marker of verbal derivation in Formosan languages. *Oceanic Linguistics*, 39:391-414.
- Zeitoun, Elizabeth and Paul Jen-kuei Li, eds, 1999, *Selected papers from the Eighth International Conference on Austronesian Linguistics*. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica 1. Taipei: Academia Sinica.
- Zorc, R. David, 1977, *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. Canberra: Pacific Linguistics.
- 1978, Proto-Philippine word accent: innovation or Proto-Hesperonesian retention? In Wurm and Carrington, eds, 1978:67-119.

Notes on the history of 'focus' in Austronesian languages

ROBERT BLUST

1 Identifying the phenomenon¹

The grammatical category of focus in Austronesian (An) languages has been a source of descriptive and theoretical confusion for the better part of a century. McKaughan (1962), like many others working in Philippine linguistics, considered focus to be the grammatical relationship between the main verb and the topic of a sentence. Speaking of Maranao in the southern Philippines he observed that:

when a substantive phrase is not the topic, its grammatical relationship to the verb is indicated either by particles, or by certain pronoun forms...Cutting across the grammatical relations marked by these particles is another, indicating that a certain substantive is the topic of the sentence, or has a primary relation to the verb...The relation of special emphasis, amplification, or topic may be called primary, since nontopic relations and verb vary depending upon the 'voice' of the topic. (pp.47, 48)

Elsewhere (1971:158ff.), speaking of Philippine-type languages in general he suggested that:

The most prominent grammatical feature, at least to the western observer, is that which indicates grammatical relations between the verb and the topic of the sentence. To mark these relations, the verb contains 'case-marking' morphemes indicating that the topic is the actor, goal, indirect referent, or instrument of the action denoted.

Two things are noteworthy about these remarks. The first is the quotative qualification of the terms 'voice' and 'case-marking', a practice which suggests that McKaughan did not feel completely comfortable with the use of either term alone as a description of the function of the verbal affixes in question. The second is the use of both terms to refer to the same phenomenon. From an Indo-European standpoint the notions of voice and case are fundamentally distinct, but in Philippine-type focus systems the two are fused in an unfamiliar union which has caused much vacillation and uncertainty.

¹ Elizabeth Zeitoun supplied many of the references to unpublished theses on the Formosan aboriginal languages, together with information on the terminology used to describe focus in these works. Although I am much indebted to her, she bears no responsibility for any of my conclusions.

It is for just this reason that the term 'focus' was adopted by members of the Summer Institute of Linguistics as a neutral alternative to the more familiar terms from Indo-European languages such as English. In her discussion of the verb system of Agta in the northern Philippines, Phyllis Healey (1960:103, fn.11) credits Alan Healey (1958) with coining the term 'focus':

Elsewhere in Philippine linguistics this phenomenon has usually been termed "voice", as, for example, in Howard McKaughan's "The inflection and syntax of Maranao verbs" (1958). Because there is no automatic translation equivalent between the two English voices and the three Agta ones, the term "focus" was adopted to emphasise the non-English nature of Agta grammar.

McKaughan (1970) points out that the usage appears at about the same time in Dean (1958).

Healey's remarks highlight another mismatch between the grammatical category of focus in Philippine-type languages and the more familiar category of voice. Most Indo-European languages have just two voices, an active and a passive. If focus is treated as voice the situation in Philippine-type languages is disconcertingly different from that in English, since most of these languages have an active and *three* passives. Moreover, where statistical data are available it is clear that the active voice occurs with a lower text frequency than the passives as a group, and may be less frequent than the direct passive ('Patient Focus') alone. It was this embarrassment of riches and apparent downgrading of the actor as much as the haziness of the voice/case distinction that drove many researchers on Philippine languages to seek a terminological innovation which might avoid the problems associated with use of the term 'voice'.

Although descriptions of Philippine languages began to appear as early as the sixteenth century through the efforts of the Spanish friars, these adhered closely to the framework of Latin grammar. The first description of a Philippine-type grammatical system in a non-traditional framework is evidently Adriani's (1893) grammar of Sangir, a member of the Philippine subgroup of Austronesian languages spoken in northern Sulawesi. Adriani's approach was clearly inspired by van der Tuuk's (1864-1867) classic grammar of Toba Batak. Although the structure of the Batak languages of Sumatra is somewhat different from that of Philippine languages, van der Tuuk spoke of three types of 'circumstantial passive' constructions. Despite the verbal associations of the term 'passive' van der Tuuk classified the passives of Toba Batak as substantives. Very similar ideas are reflected by Adriani (1893:188):

The active and passive forms in Sangir are sharply distinguished. The active, as has already often been noted, is actually the only verbal form, the passive is a noun, and the imperative likewise; the distinction between active and passive in the latter is thus observed only in speech (translation mine).

Adriani goes on to state that in Sangir there are three sorts of passive: simple passive, local passive, and subjective passive.

Shortly after the beginning of the twentieth century Blake (1906) pointed out that in Tagalog the same types of relationships which are marked by prenominal particles in non-topic arguments are marked by verbal affixes in topic arguments. It would seem to follow, then, that for topics at least case relations are marked by verbal affixes. In what is often regarded as a classic work, Bloomfield (1917) described Tagalog as having four voices: an active marked by *mag-*, *mang-* or *-um-*, a direct passive marked by *-in*, an instrumental passive marked by *i-*, and a local passive marked by *-an*.

Although the connection is generally overlooked, it seems clear that Bloomfield's description of Tagalog focus as voice was simply a continuation of the Dutch tradition

initiated by van der Tuuk and first applied to Philippine-type languages by Adriani in 1893. Perhaps in deference to Bloomfield's more prominent position in the field of linguistics Blake (1925) modified his earlier analysis in the direction that Bloomfield took, but spoke of three passive 'themes' rather than voices, evidently feeling some discomfort at the notion of multiple passives. Nonetheless, five years later Blake (1930) returned to his original position that the focus affixes of Tagalog verbs are case markers. Blake's vacillation provides perhaps the first indication in the literature that, despite its convenience, the term 'voice' is problematic in relation to Philippine-type languages.²

Although the languages of Sulawesi south of the Gorontalic group become progressively less Philippine-like, some of these can be described in general terms as having a Philippine-type syntax. For Mori of southeast Sulawesi Barsel (1994:65) uses the term 'case' to describe something highly reminiscent of focus:

Case is defined here as the different relations that the NPs of a clause, including the NPs of locative phrases, have to the predicator. It is analyzed on two levels. Firstly, case refers to the grammatical functions of the NPs, which are language-specific to Mori and marked by inflection on the predicate and by syntax. Second, it refers to their semantic functions, which are a set of situational roles.

One reason that the voice/case distinction appears to be so difficult to make in Philippine-type languages is that pronominal arguments are formally different depending upon their case relationship to the verb, whereas in many languages nominal arguments are not. Thus, in Tagalog *ang* marks a focussed nominal argument (for common nouns) regardless of its relationship to the verb (actor, patient, location, instrument, etc.). If these semantic roles are seen as case relationships then it is clear that they are marked in the verb, not in the pronominal particle *ang*. But traditional notions of case derived from the structure of Indo-European languages see case-marking as a property of noun phrases, not of verbs. Ramos (1971:16) states the matter succinctly:

These relationships tying noun phrases in the Tagalog sentence to verb centers are indicated by affixes in the verb rather than on the noun and by particles introducing noun phrases. Affixes in Tagalog verbs mark case relationships between the verb and the subject of the sentence, while the particles introducing noun phrases mark the same kinds of relationships, but do so between verbs and non-subject noun phrases...At least three terms have been used to describe the verb affixes referred to above: focus, voice and case.

All three of these terms continue to be used up to the present for what is essentially the same grammatical category. Appendix 1 (which does not pretend to be complete) provides a summary of terms used for this phenomenon by various researchers since Adriani. An inspection of Appendix 1 shows that the term 'focus' is used in 25 sources, 'voice' in 28, 'case' in eight, 'topicalisation' in three, and 'theme', 'verb class', 'recentralisation' and

² Schachter (1990:940ff.), who suggests the terminological innovation 'trigger' for the focussed nominal argument and 'trigger system' for the general phenomenon of focus in Tagalog, makes this point with particular clarity. Among other things he notes that 'the arguments that can be chosen as trigger show a much wider range of semantic roles than the arguments that can be chosen as subject in typical voice systems and consequently there are many more distinct verb forms than a voice system's typical two... Secondly, in voice systems the active can generally be regarded as the unmarked voice by virtue of its frequency, unrestricted distribution and the like. But Tagalog actor-trigger clauses are not unmarked in relation to their non-actor-trigger counterparts. Indeed, the latter turn out to be generally more frequent in texts and to have fewer distributional restrictions. For these reasons, it seems best not to describe the Tagalog trigger system as a voice system.'

'trigger' in one each. Since terminological usages tend to be transmitted from teacher to student, it is impossible to treat each of the above choices of terminology as independent. Nonetheless it does appear that 'focus' and 'voice' are about equally common, with 'case' a distant third. It is noteworthy that several writers (Blake, McKaughan, Schachter, Prentice) changed their terminology over time.

A number of more recent analyses see Philippine-type languages as ergative, and as a result have a fundamentally different view of the nature of focus. Starosta (1988:185), for example, whose views can only be fully understood within the context of his Lexicase theory, suggests that:

In ergative Austronesian languages of the Philippines, the 'focus' system is just a relatively elaborate verbal derivation mechanism for recentralization, that is, for reinterpreting an actant as Patient.

These definitions say nothing about how we distinguish Austronesian languages which have a focus system from those that do not. Throughout the literature there is an essentially implicit assumption that a language must have at least two morphologically distinguished 'passives' to qualify for inclusion in the typological category of 'Philippine-type focus language'. For convenience we can call these 'three-term systems'. A language such as Tagalog, which distinguishes actor, Patient, Locative and Instrumental/Benefactive focuses by distinct verbal affixes, thus exemplifies a 'four-term system'. While three-term systems may be necessary for meeting the definition 'Philippine-type language', they are not sufficient, since some languages, such as Malay, have two morphologically distinguished passives but are not considered to have a Philippine-type syntax. The essential difference in this case is that the two passives of Malay apply to arguments having identical case relationships to the verb while the multiple passives of Philippine-type languages normally do not.³

2 How old is focus in Austronesian?

Wolff (1973) reconstructed a system of four voice contrasts for PAn. Omitting some details, and correcting his Instrumental passive from PAn **i-* to PAn **Si-*, PMP **i-*, the core system of verbal affixes included the following:

- *-um-* active
- *-en* direct passive
- *-an* local passive
- *Si-* instrumental passive

Apart from **-en* each of these voice-marking affixes (in Wolff's terms) could co-occur with the past marker **-in-*: **C-in-um-VCVC* (or **C-um-in-VCVC*), **C-in-VCVC-an*, **Si-C-in-VCVC*. By contrast, **-en* had a zero allomorph in past constructions: **C-in-VCVC*. In effect, then, the tense/aspect marker **-in-* assumed a portmanteau function in the past of the direct passive, simultaneously marking voice and tense/aspect.

³ Various writers, as McFarland (1976:v) have pointed out that the semantic roles marked by verbal affixes in typical focus languages such as Tagalog must be understood as canonical types which allow some variation and, indeed, crossover. Much the same appears to have been the case in PAn, at least for the Locative focus (Blust 1996).

Wolff's reconstruction is static in the sense that he is concerned only with inferring a system at the PAN stage, without reference to the types of changes that affected it during its subsequent evolution. It is important to recognise that in many Malayo-Polynesian (MP) languages — but not in Formosan languages — the actor focus may be morphologically marked in more than one way. While a reflex of **-um-* marks actor focus in most attested focus languages, many languages that have been called 'Western Malayo-Polynesian' (Philippines, western Indonesia–Malaysia, Malagasy, Chamorro, Palauan) use reflexes of the prefixes **maR-* and/or **maŋ-* in a similar function. In Malagasy **-um-* is reflected only as a fossilised or nearly fossilised element in a handful of verbs, the marking of actor focus being handled almost entirely by reflexes of **maŋ-*.⁴ PMP thus evidently had at least two and perhaps three patterns of affixation for marking the actor focus. In some languages, such as Kelabit, the reflex of **-um-* is associated with intransitive verbs, and the reflex of **maŋ-* with transitive/causative verbs, as in *ebhen* 'sinking into the earth, subsiding': *m-ebhen* 'to subside, sink of its own accord': *ng-ebhen* 'to push a stick into the ground, a needle into the skin, etc.', *layuh* 'wilting, withering': *l-em-ayuh* 'to wilt, wither, as a plant in the sun': *nge-layuh* 'to make something wither, as by placing it close to the fire', or *turun* 'descending': *nurun* 'to lower something, as a ladder from a house', *t-em-urun* 'to descend, as a ladder'. As noted by Mead (this volume), in the Bungku-Tolaki and Muna-Buton languages of southeast Sulawesi, reflexes of PMP **-um-* and **maŋ-* are distinguished along very different parameters: both mark transitive verbs, but the former co-occurs with indefinite objects and the latter with definite objects. Insofar as he is referring to the simple contrast between two verbal affixes which can potentially occur on the same stem Mead is certainly correct in observing that these languages have 'preserved a living distinction between Proto Malayo-Polynesian **maN-* and **um-*.' With regard to the functional distinction associated with them, however, this Sulawesi usage seems almost certainly to be innovative. It is an odd and still poorly understood fact that PAN **-um-* can only be securely reconstructed in intransitive verbs such as **q-um-uzaN* 'to rain', **k-um-aen* 'to eat', or **N-um-anuy* 'to swim'. This leaves the question of how transitive verbs were marked in PAN obscure, unless transitivity was restricted to non-actor focus constructions.

One other point deserves some notice. In Austronesian focus languages generally agency and possession are marked in the same way. In other words, the agent of non-actor focus verbs co-occurs with the genitive marker, usually a reflex of PAN **ni* 'genitive of human nouns; agent of non-actor focus verbs'. As a consequence of this system of marking grammatical relationships the noun/verb distinction becomes blurred. As noted above, Adriani considered the 'passives' of Sangir to be nouns, a view that he may well have taken over from van der Tuuk's earlier analysis of Toba Batak. To accommodate this observation Starosta, Pawley and Reid (1982) argued — contrary to Wolff (1973) — that the focus system of Philippine-type languages evolved from a PAN system of nominalisation in which certain derived nouns were analogically reinterpreted as verbs. As noted by Blust (1998a) there are serious problems with this analysis, and it is almost certain that Wolff's reconstruction provides a more accurate account of the focus morphology of PAN. Nonetheless, the alternative offered by Starosta, Pawley and Reid has the merit of recognising that bases carrying 'passive' affixation may be nouns in given syntactic contexts. The key expression here is 'given syntactic contexts', since in virtually all Austronesian focus

⁴ Dahl (1951:163) notes that in the Sakalava dialect of Malagasy the infix *-um-* is found in several intransitive verbs. While it is rare, it still must be considered a functional morpheme. In the standard dialect of Merina, by contrast *-om-* apparently is best regarded as defunct.

languages non-actor focus forms may be it *either* verbal or nominal, sensitive to the syntactic environment in which they are found, a point made forcefully for Tagalog by Himmelmann (1987:78ff.).

This system, diverging in various ways, has survived in many of the Formosan aboriginal languages, in virtually all languages of the Philippines, in the languages of Sabah and in northern Sulawesi at least as far south as the Gorontalic group, in Malagasy, and in Chamorro. A glance at the map will show that the geographical distribution of focus languages is strikingly skewed: with a few notable exceptions, languages closer to the probable Austronesian homeland in Taiwan have preserved more of the original focus system than languages at a greater distance from it. It is an intriguing question as to why this should be the case.

If the typological boundaries in Austronesian corresponded more or less exactly to major subgroup boundaries we might feel confident in assuming that the transformation of focus languages was a product of changes in a small number of ancestral communities which were transmitted to their descendants. This may be true in some cases, but does not appear to be true in others. Thus, there is no basis for reconstructing a Philippine-type focus system for Proto Central-Eastern Malayo-Polynesian (PCEMP), the hypothetical ancestor of most of the languages of the Lesser Sunda and Moluccan islands of Indonesia, and of the Oceanic group in the Pacific. For these languages we can reasonably assume that a fairly radical restructuring of the morphosyntactic system took place in PCEMP, and that this restructured system then underwent various changes in the attested CEMP languages.

The case of Malagasy, however, is quite different. We know with some certainty that Malagasy is most closely related to the Barito languages of southeastern Kalimantan (Dahl 1951, 1977). To the extent that we have data on the verb systems of the Barito languages it appears that they lack the multiple voices/focus categories of Malagasy and are structurally more similar to such western Indonesian languages as Malay (Hardeland 1858; Sundermann 1914; Dahl 1978:383; Kawi et al. 1979-1980; Gudai 1988). Since the Malagasy evidently departed from Borneo within the past 1,300 years (Adelaar 1989; Dahl 1991) Philippine-type focus systems must have been found in Borneo south of Sabah as recently as AD 700, although apart from the borderline case of Lun Dayeh (Clayre 1991) no such systems are known today. Malagasy is thus instructive in shedding light on the relatively recent syntactic history of portions of western Indonesia. Given the close genetic relationship of the Barito languages it is improbable that Malagasy was the only Philippine-type focus language in the Barito group at the time of the migration to Madagascar. A more plausible scenario is that Malagasy was one of a number of Philippine-type focus languages which were still found in Borneo south of Sabah around AD 700, and that extensive morphosyntactic changes swept over the area after the Malagasy departed. As Dahl (1978) has noted, similar extensive changes over a period of less than two millennia are familiar from Indo-European linguistics, as with the loss of the Latin case-marking system and the rise of prepositional phrases in all of the modern Romance languages. What is different about the Malagasy case is that one language 'escaped' in time to preserve some features of the earlier system which was thoroughly transformed in those languages that remained behind.

A second puzzle concerning the distribution of focus systems is that they almost invariably are associated with verb-initial (or predicate-initial) languages. It seems reasonably certain that this was also the case in PAN, and that many attested focus languages have simply preserved the PAN word order and system of focus marking in the verb. Word-order change from VSO or VOS to SVO in much of western Indonesia, and in eastern Indonesia and the Pacific thus may have played a part in the reduction and eventual loss of focus-marking

morphology. In this connection it should be noted that the verb-initial languages of Fiji and Polynesia represent a secondary development from a Proto Oceanic stage which probably was SVO.

3 How has the PAN/PMP focus system evolved?

The preceding discussion raises the question of how the PAN/PMP focus system has evolved both in those languages that preserve some type of focus system and in those which have changed it in some fundamental way.

Wolff's reconstruction recognises four morphologically distinguished categories of 'voice': (1) active, marked by **-um-*, (2) direct passive, marked by **-en* in the non-past, and by **-in-* (as a portmanteau form) in the past, (3) local passive, marked by **-an*, and (4) instrumental passive, marked by **Si-* (Wolff writes **i-*). In some daughter languages, as Tagalog the same four morphological categories are maintained, but additional semantic roles are incorporated into the system. Tagalog *i-*, for example, marks both instrumental and benefactive, although Wolff assigns only the former semantic role to PAN **Si-*.⁵ Other languages have collapsed the four-term system of morphological marking into a three-term system. Clayre (n.d.) has suggested a typology of Bornean languages on the basis of their focus properties, with the following groups:

- (1) Philippine type. This is characterised by three or more 'focuses', verb affixation and nominal and pronominal marking. In Borneo today Type 1 is confined almost exclusively to Sabah. South of Sabah only Lun Dayeh of northern Sarawak/Kalimantan can be assigned to this category.
- (2) Reduced focus system with two 'focuses'. Languages of this type have no markers of NPs, reduced aspect marking, and SVO word order. Clayre further subdivides this into:
 - 2a Two 'focuses', two pronoun sets (Berawan, Melanau, Penan, Sebop)
 - 2b Two 'focuses', one pronoun set (Sa'ban, Lengilu')
 - 2c Two 'focuses' (Kayan)
- (3) No focus, or vestiges of focus only in subordinate clauses (many or all Kenyah dialects)

The classification of 'Type 2' as focus languages is atypical, as they are structurally much closer to western Indonesian languages such as Malay than they are to languages of the Philippines. I would suggest instead a typology which recognises the following distinctions:

- (1) Four-term languages. Alternatively, these might be called 'full focus' languages. This would include languages with at least four morphologically distinguished focuses, such as Tagalog.

⁵ Chang (1997:35ff.) notes that Kavalan marks both Instrumental and Beneficiary voices with the verbal prefix *te-*. Although *te-* does not reflect PAN **Si-*, the association of semantic roles in connection with this form is similar to that in non-Formosan languages such as Tagalog. If this pattern is inherited in both languages the mapping of five semantic roles onto four morphologically distinguished focuses in Tagalog is not an expansion of the original system. In addition, Dahl (1978) has raised the possibility that PAN may have had a fifth morphologically distinguished focus, marked by **Sa-*. Further evidence for this position appears in Blust (1999). Ross (1995:730), on the other hand, expresses doubt about whether an Instrumental voice can be posited at the PAN level.

- (2) Three-term languages. Alternatively, these might be called 'reduced focus' languages. This category would include languages such as Thao of central Taiwan, which distinguishes (a) actor focus, marked by *-um-*, (b) patient focus, marked by *-in* in the non-perfective and by *-in-* or *-in- + -in* in the perfective (Blust 1998b), and (c) locative focus, marked by *-an*; Kavalan of eastern Taiwan, which distinguishes (a) actor voice, marked by *-um-*, (b) patient voice, marked by *-an*, and (c) beneficiary/instrumental voice, marked by *te-*; or Lun Dayeh of northern Sarawak, which in the imperfective aspect distinguishes (a) actor focus, marked by *N-* (homorganic nasal substitution), (b) patient focus (or undergoer focus), marked by *-en*, and (c) instrument focus, marked by *pi-N-*. These languages have reduced the original system in two different ways: Thao by eliminating the use of affixation to mark instrumental focus in the verb, and Kavalan and Lun Dayeh by eliminating the use of affixation to mark locative focus in the verb. Until a larger sample of three-term languages is tested it is impossible to generalise about dominant patterns in the reduction of four-term languages to three-term languages.
- (3) Two-term languages with portmanteau infix *-in-*. These languages have what is essentially a two-voice system, but the passive is obligatorily perfective, a remnant of the portmanteau function of PAN **-in-* in the patient focus. Examples of several such languages in northern Sarawak are given in Blust (1997).
- (4) Two-term languages without a portmanteau infix *-in-*. Malay can be used to illustrate this type, with the proviso that it has two passives which appear to be semantically much more similar to one another than is true of the passives in a voice analysis of Philippine-type languages.

Only Types 1 and 2 should properly be called 'focus' languages, although Type 3 has properties which are difficult to account for in a synchronic description without recognising that such languages once had more complex systems of verb morphology in relation to focus-marking. It might be added that few linguists would quibble with the use of 'voice' to describe the active/passive contrasts of Types 3 and 4. To the extent that the results of such morphosyntactic reduction reflect on the nature of the original system, then, they provide some support for the view that focus is fundamentally a system of voice marking.

This typology of focus/voice possibilities in Austronesian languages raises another intriguing question. Although it has received very little attention in the literature to date, there is a remarkable correlation between word order and the presence of focus systems in Austronesian languages: almost without exception four-term and three-term languages have a verb-initial syntax. This relationship can be stated as an implication such that the presence of three or more focus possibilities implies verb-initial constituent order with almost perfect accuracy. The reverse implication fails to hold, since verb-initial syntax has developed secondarily in some Oceanic languages, such as Fijian and the Polynesian languages. The few exceptions to this correlation include moribund Formosan languages such as Thao or Saisiyat which have been subjected to very heavy contact influence from SVO Taiwanese (Minnan), and as a result have begun to favour SVO order. However, even in these languages verb-initial constructions continue to be offered as a more native-like alternative to their historically recent calqued equivalents.

To date this implication remains theoretically unexplained. Perhaps the most useful way to view it is in negative terms. Verb-initial syntax and a focus system were part of a package of PAN typological features which were retained in many of the syntactically more conservative daughter languages. The correlation in these languages is thus simply a product of shared

history. But what of those languages which have (i) changed the verb-initial constituent order, or (ii) lost the original focus system? Why do these two types of change, one affecting constituent order and the other verb morphology, appear to be so highly correlated in languages which have independently undergone either one?

For now this question remains unanswered. There are, however, some indications that the transition from a Philippine-type focus language to a language with some type of active/passive voice distinction involved changes of constituent order that affected active constructions before affecting intransitive or passive constructions. Poedjosoedarmo (this volume), citing the work of Cumming (1991), notes that 'In Classical Malay, though word order was variable, the most common word order for intransitive and passive sentences was VS. For active transitive sentences, however, the most common pattern was SVO.' In modern Malay this pattern has been levelled to SVO for all constructions. She notes that the evolution of Old Javanese to Modern Javanese involved a similar transition, with intransitive and 'passive' constructions remaining bastions of the historically older verb-initial constituent order after active constructions had passed through a word-order change.

Since it is well known that Malay and Javanese have a long and intricate history of mutual contact influence this agreement in the details of constituent order change may appear to have no bearing on broader theoretical questions. But a strikingly parallel development can be seen in comparing Lun Dayeh and Bario Kelabit of northern Sarawak. Lun Dayeh and Kelabit are either divergent dialects of a single language or very closely related languages. Clayre (1991) has shown that Lun Dayeh is a Type 2 language in terms of the typology sketched above — that is, a language with a reduced focus system. At the same time it retains a verb-initial constituent order in intransitive constructions, and in both actor focus and goal focus transitive constructions: *l-em-anguy ieh neh* (swim-AF 3SG particle) 'He is swimming', *m-eru' bigan ieh ina* (AF-wash dish 3SG just-now) '(S)he was washing dishes just now', *k-in-an ku bua' di' peh* (eat-PF-Perf. 1SG fruit particle already) 'I have already eaten the fruit'. In contrast with Lun Dayeh, Bario Kelabit is a Type 3 language, with only an active/passive voice contrast. Most strikingly, although Bario still permits verb-initial constituent order in both intransitive and passive constructions, unlike Lun Dayeh it requires all active constructions to be verb-medial: *m-udur ieh* (AF-stand 3SG) 'He is standing', *itep uku' ineh uih* (bite-PF-Perf. dog that 1SG) 'That dog bit me', *nih uku' sinih ng-etep uih* (this dog here ACT-bite 1SG) 'this dog is biting me'. A similar situation is found in Bintulu, a Type 3 language of coastal northern Sarawak, except that intransitive constructions appear to have already followed active transitive constructions in the change from VS to SV order, leaving only the passive with preferred, but optional verb-initial constituent order: *isa lupek bajew ineh* (3SG fold-ACT shirt that) 'He is folding that shirt', *lipek ña bajew ineh* (fold-PASS 3SG shirt that) 'He folded that shirt' (but *akew de-bukut ña* (1SG PASS-punch 3SG) 'He punched me'), *isa me-lakaw* (3SG ACT-walk) 'He is walking', *isa taba taba lalu* (3SG smile intensive) 'She is always smiling'.

Although closely related to one another, Lun Dayeh and Kelabit have not shared a common ancestor with Malay or Javanese within approximately the past 3,500 years. They are, moreover, languages of interior Borneo, which until the second half of the twentieth century had very little contact with Malay. Similarly, although Kelabit–Lun Dayeh and Bintulu both belong to the North Sarawak group of languages (Blust 1974) they are members of different primary branches within it, and are geographically widely separated. These shared details of constituent order change in Kelabit–Lun Dayeh, Bintulu, Malay and Javanese thus must be regarded as historically independent. To the extent that such

parallelism cannot plausibly be attributed to limited possibilities it suggests a linguistic motivation for the historically repeated sequence of changes:

1. loss of focus system;
2. VS order changes to SV order in active constructions;
3. VS order changes to SV order in intransitive and passive constructions.

The Bintulu data further suggest that the transition from VS to SV order may occur earlier in intransitive constructions than in passives, but until independent evidence can be found to confirm these preliminary indications this more detailed ordering must be considered provisional.

With regard to the transition from VSO to SVO in active transitive sentences Poedjosoedarmo (this volume) has suggested that 'It was the necessity of showing a close link between the verb and two noun phrases associated with it which initially prompted the word order change.' On the other hand, 'In Javanese, the adoption of the new order for passive sentences was probably related to loss of obligatory marking of the agent and associated with the option of having a non-agent (such as patient of a passive benefactive verb) follow the verb.' Poedjosoedarmo does not explain what she means by 'a close link' in this quotation. Since she cites 'the non-ubiquitousness of noun phrase marking' as one of the two most significant changes from a presumed Philippine-type ancestor to Old Javanese, what she appears to have in mind is that the change from VS to SV in active constructions was motivated by the loss of earlier prenominal particles which marked semantic role. But it is not at all clear why the loss of noun phrase markers would trigger a change in constituent order. In full focus languages with noun phrase markers such as Tagalog the nominal arguments enjoy a relatively free order in relation to one another, but always follow the verb. If the noun phrase markers were to disappear, as they have, for example, in Malagasy, fixed word order simply usurps their function, leaving the position of the verb intact. Moreover, no reason is given as to why the need to show a close link between the verb and its nominal arguments would be greater in active than in passive constructions, which is the most crucial observation in need of explanation. Although I am not proposing it as an alternative explanation, one could as easily argue that constituent order change in active constructions preceded the similar change in non-active constructions because active constructions had lower text frequency and were thus in a sense more marginal to the system than intransitive or passive constructions. In short, the reason that active constructions have undergone the change from VS to SV constituent order before intransitive and passive constructions in several genetically and geographically separated languages remains to be fully explained.

There are also problems in trying to generalise Poedjosoedarmo's explanation for the change from VS to SV order in passive sentences. Although 'the loss of obligatory marking of the agent' may have played a part in motivating this change in Javanese, we must keep in mind that similar changes have occurred more widely in western Indonesia. In Sa'ban, a phonologically and morphologically divergent dialect of Kelabit-Lun Dayeh, all traces of a Philippine-type focus system have been lost, transitive and intransitive constructions show SV order, and agents of passives appear to be obligatorily marked. There can hardly be any doubt that in reaching this state Sa'ban passed through an earlier stage similar to that of Bario Kelabit. The change of VS to SV constituent order in the earlier passive thus took place even though the conditions cannot have been the same as those in Javanese.

Finally, Van den Berg (1996) and Mead (this volume) have traced various features of the evolution of focus in the languages of Sulawesi, and Starosta, Pawley and Reid (1982) attempted to provide an overarching framework which unites the focus systems of

Philippine-type languages with the superficially very different verb systems of Oceanic languages. Virtually all of the languages of Borneo south of Sabah have the following morpheme-structure constraints: (1) vowels in prepenultimate syllables have merged as schwa (this began with *a, and later was extended to include *i and *u), (2) prepenultimate schwa is disallowed in initial position. Since the changes which produced these constraints would have eliminated the inherited reflex of *i-, it is possible that they initiated the reduction of focus systems in this area. In Sulawesi the mechanisms appear to have been very different, based on a change in the balance of forces within the verb system itself rather than introduced from without by phonological innovations with grammatically destructive potential. Much research remains to be done in understanding the interplay of forces that led to the simplification of focus systems in various parts of the Austronesian world and their preservation in others. This volume and others like it will undoubtedly contribute toward that goal.

Appendix I: Terms for the grammatical category of 'focus' in the Austronesian literature

No.	Language	Term	Source
1	Sangir	voice	Adriani (1893)
2	Tagalog	case	Blake (1906)
3	Malagasy	voice	Malzac (1908)
4	Tagalog	voice	Bloomfield (1917)
5	Tagalog	theme	Blake (1925)
6	Tagalog	case	Blake (1930)
7	Ilokano	voice	Bloomfield (1942)
8	Ilokano	voice	Vanoverbergh (1955)
9	Maranao	voice	McKaughan (1958)
10	Yogad	focus	Healey (1958)
11	Agta	focus	Healey (1960)
12	Atayal	voice	Egerod (1965)
13	Cotabato Manobo	case	Kerr (1965)
14	Ivatan	focus	Reid (1966)
15	Bikol	verb class	Mintz (1971)
16	Maranao	case/topicalisation	McKaughan (1970)
17	Kapampangan	case	Gonzalez (1971)
18	Timugon Murut	focus	Prentice (1971)
19	Tagalog	focus	Ramos (1971)
20	Tagalog	focus	Schachter and Otones (1972)
21	Chamorro	focus	Topping (1973)
22	Atayal	voice	Wolff (1973)
23	PAn	voice	Wolff (1973)
24	Tagalog	case	Ramos (1974)
25	Tondano	voice	Sneddon (1975)
26	Malagasy	voice	Keenan (1976)

27	Tagalog	focus	Llamzon (1976)
28	Tagalog	focus	McFarland (1976)
29	Tagalog	topicalisation	Schachter (1976)
30	Tsou	focus	Tsuchida (1976)
31	Bunun	case	Jeng (1977)
32	Tagalog	focus	Naylor (1980)
33	Timugon Murut	voice	Prentice (1980)
34	Amis	focus	Chen (1982)
35	Tagalog	voice	Foley and Van Valin (1984)
36	Malagasy	focus	Dahl (1986)
37	Tagalog	voice	De Guzman (1986)
38	Tagalog	recentralisation	Starosta (1986)
39	Malagasy	voice	Randriamasimanana (1986)
40	Tagalog	focus	Himmelman (1987)
41	Kadazan	focus	Hurlbut (1988)
42	Kimaragang Dusun	focus	Kroeger (1988)
43	Tagalog	voice	Shibatani (1988)
44	Ilokano	voice	Clausen (1990)
45	Tagalog	trigger	Schachter (1990)
46	Lun Dayeh	focus	Clayre (1991)
47	Saisiyat	focus	Yeh (1991)
48	Atayal	voice	Rau (1992)
49	Tsou	focus	Zeitoun (1992)
50	Atayal	voice	Huang (1993)
51	Tagalog	voice	Kroeger (1993)
52	Mori	case	Barsel (1994)
53	Tsou	focus	Szakos (1994)
54	Amis	focus	Wu (1995)
55	PAN	voice	Ross (1995)
56	Tagalog	topicalisation	Richards (1996)
57	Tagalog	voice	Voskuil (1996)
58	Amis	voice	Liu (1997)
59	Kavalan	voice	Chang (1997)
60	Kavalan	focus	Lee (1997)
61	Puyuma	focus	Tan (1997)
62	Seediq	voice	Chang (1997)
63	Tukang Besi	voice	Donohue (this volume)
64	Ratahan, Lauje	voice	Himmelman (this volume)
65	Seediq	focus	Holmer (this volume)
66	Proto-Kaili-Pamona	focus	Mead (this volume)
67	Sasak	voice	Wouk (this volume)

References

- Adelaar, K.A., 1989, Malay influence on Malagasy: linguistic and culture-historical implications. *Oceanic Linguistics* 28:1-46.
- Adriani, N., 1893, *Sangireesche spraakkunst*. Leiden: A.H. Adriani.
- Barsel, Linda A., 1994, *The verb morphology of Mori, Sulawesi*. Canberra: Pacific Linguistics.
- Berg, René van den, 1996, The demise of focus and the spread of conjugated verbs in Sulawesi. In H. Steinhauer, ed. *Papers in Austronesian Linguistics* No.3, 89-114. Canberra: Pacific Linguistics.
- Blake, Frank R., 1906, Expression of case by the verb in Tagalog. *Journal of the American Oriental Society* 27:183-189.
- 1925, *A grammar of the Tagalog language*. New Haven: American Oriental Society.
- 1930, A semantic analysis of case. In James T. Hatfield et al., eds *Curme volume of linguistic studies*, 34-49. Baltimore: Waverly Press.
- Bloomfield, Leonard, 1917, *Tagalog texts with grammatical analysis*. University of Illinois Studies in Language and Linguistics, Vol.3, Nos. 2-4. Urbana, Illinois: University of Illinois.
- 1942, Outline of Ilocano syntax. *Language* 18:193-200.
- Blust, Robert, 1974, The Proto-North Sarawak vowel deletion hypothesis. PhD dissertation, University of Hawaii.
- 1996, Notes on the semantics of Proto-Austronesian *-an 'locative'. In Marian Klamer ed., 1996:1-11.
- 1997, Ablaut in northwest Borneo. *Diachronica* 14:1-30.
- 1998a, Ca- reduplication and Proto-Austronesian grammar. *Oceanic Linguistics* 37:29-64.
- 1998b, A note on the Thao Patient focus perfective. *Oceanic Linguistics* 37:346-353.
- 1999, Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38:321-365.
- Chang, Yung-li, 1997, Voice, case and agreement in Seediq and Kavalan. PhD dissertation, Tsinghua University, Hsinchu, Taiwan.
- Chen, Teresa M., 1982, Verbal constructions and verbal classification in Nataoran-Amis. PhD dissertation, University of Hawaii.
- Clausen, Josie Paz, 1990, Ilokano verbs: argument structure. *Working Papers in Linguistics* 22.2:49-92. Department of Linguistics, University of Hawaii.
- Clayre, Beatrice, 1991, Focus in Lundayeh. *Sarawak Museum Journal* 42(63):413-434.
- n.d., *A preliminary typology of the languages of middle Borneo*. Typescript, 19pp.
- Cumming, Susanna, 1991, *Functional change: the case of Malay constituent order*. Berlin and New York: Mouton de Gruyter.
- Dahl, Otto Christian, 1951, *Malgache et Maanjan: une comparaison linguistique*. Oslo: Egede-Instituttet.
- 1977, La subdivision de la famille Barito et la place du Malgache. *Acta Orientalia* 38:77-134.
- 1978, The fourth focus. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: Proceedings*, Fascicle 1:383-393.
- 1986, Focus in Malagasy and Proto-Austronesian. In Geraghty, Carrington and Wurm eds, 1986:21-42.

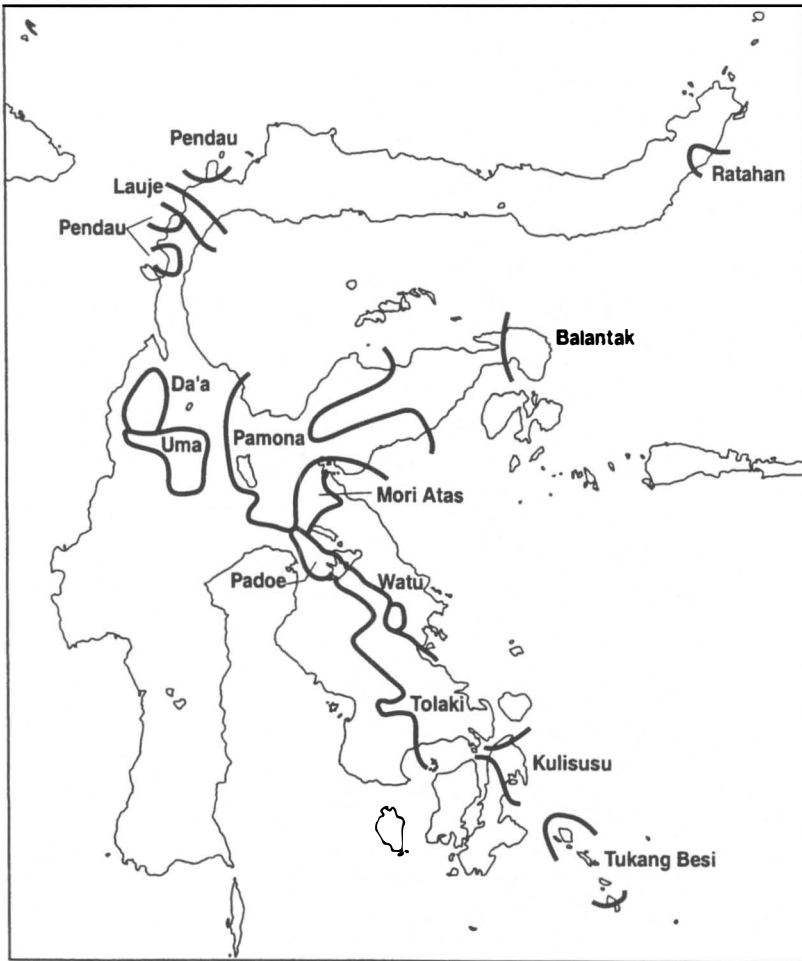
- 1991, *Voyage from Kalimantan to Madagascar*. The Institute for Comparative Research in Human Culture. Oslo: Norwegian University Press.
- Dean, James, 1958, Some principal grammatical relations of Bilaan. *Oceania Linguistic Monographs*, 3:59-64. Sydney: University of Sydney.
- De Guzman, Videa P., 1986, Some consequences of causative clause union in Tagalog. In Geraghty, Carrington and Wurm, eds 1986:59-72.
- Egerod, Soren, 1965, Verb inflexion in Atayal. *Lingua* 15:251-282.
- Foley, William A. and Robert D. Van Valin, jr., 1984, *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Geraghty, Paul, Lois Carrington and S.A. Wurm, eds, 1986, *FOCAL 1: papers from the Fourth International Conference on Austronesian Linguistics*. Canberra: Pacific Linguistics.
- Gonzalez, Andrew B., 1971, Outline of Generative Semantic description of Pampangan. Ph. dissertation, University of California.
- Gudai, Darmansyah H., 1988, A grammar of Maanyan: a language of Central Kalimantan. PhD dissertation, The Australian National University.
- Hardeland, August, 1858, *Versuch einer Grammatik der dajakschen Sprache*. Amsterdam: Netherlands Bible Society.
- Healey, Alan, 1958, Notes on Yogad. *Oceania Linguistic Monographs* 3:77-82.
- Healey, Phyllis M., 1960, *An Agta grammar*. Manila: Bureau of Printing.
- Himmelmann, Nikolaus jr., 1987, *Morphosyntax und Morphologie – Die Ausrichtungsauffixe im Tagalog*. Munich: Wilhelm Fink.
- Huang, Lillian M., 1993, *A study of Atayal syntax*. Taipei: Crane.
- Hurlbut, Hope M., 1988, *Verb morphology in Eastern Kadazan*. Canberra: Pacific Linguistics.
- Jeng, Heng-hsiung, 1977, *Topic and focus in Bunun*. Institute of History and Philology. Special Publication No.72. Taipei: Academia Sinica.
- Kawi, Djantera, Abdurrachman Ismail and Willem Ranrung, 1979-1980, Struktur Bahasa Maanyan. Typescript, 188pp. Pusat Pembinaan dan Pengembangan Bahasa, Jakarta.
- Keenan, Edward L., 1976, Remarkable subjects in Malagasy. In Charles N. Li, ed. *Subject and Topic*, 247-301. New York: Academic Press.
- Kerr, Harland B., 1965, The case-marking and classifying function of Cotabato Manobo voice affixes. *Oceanic Linguistics* 4:15-47.
- Klamer, Marian, ed., 1996, *Voice in Austronesian*. NUSA 39.
- Kroeger, Paul, 1988, Verbal focus in Kimaragang. In H. Steinhauer, ed. *Papers in Western Austronesian Linguistics*, No.3, 217-240. Canberra: Pacific Linguistics.
- 1993, *Phrase structure and grammatical relations in Tagalog*. Stanford: CSLI Publications.
- Lee, Amy Pei-jung, 1997, The case-marking and focus systems of Kavalan. MA thesis. Tsinghua University, Hsinchu, Taiwan.
- Liu, Tsai-hsiu, 1997, Cleft constructions in Amis. MA thesis, National Taiwan University.
- Llamzon, Teodoro A., 1976, *Modern Tagalog: a functional-structural description*. The Hague: Mouton.
- McFarland, Curtis D., 1976, *A provisional classification of Tagalog verbs*. Studies in Languages & Cultures of Asia & Africa Monograph Series, No.8. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

- McKaughan, Howard P., 1958, *The inflection and syntax of Maranao verbs*. Publications of the Institute of National Language, Manila: Bureau of Printing.
- 1962, Overt relation markers in Maranao. *Language* 38:47-51.
- 1970, Topicalization in Maranao — an addendum. In S.A. Wurm and D.C. Laycock, eds *Pacific linguistic studies in honour of Arthur Capell*, 291-300. Canberra: Pacific Linguistics.
- 1971, Minor languages of the Philippines. In Thomas A. Sebeok, ed. *Current Trends in Linguistics*, vol. 8: *Linguistics in Oceania*, 155-167. The Hague: Mouton.
- Malzac, R.P., 1908, *Grammaire malgache*. Tananarive: Mission Catholique.
- Mintz, Malcolm W., 1971, *Bikol grammar notes*. Honolulu: University of Hawaii Press.
- Naylor, Paz Buenaventura, 1980, Linking, relation-marking, and Tagalog syntax. In Paz Buenaventura Naylor, ed. *Austronesian Studies: Papers from the Second Eastern Conference on Austronesian Languages*, 33-50. Ann Arbor: Center for South and Southeast Asian Studies, The University of Michigan.
- Prentice, D.J., 1971, *The Murut languages of Sabah*. Canberra: Pacific Linguistics.
- 1980, "Beneficiary" and other semantic functions in Murut. Paper presented at the Colloquium on Linguistic Typology, Amsterdam, April, 1980.
- Ramos, Teresita V., 1974, *The case system of Tagalog verbs*. Canberra: Pacific Linguistics.
- 1981 [1971], *Tagalog structures*. 3rd printing. Honolulu: The University Press of Hawaii.
- Randriamasimanana, Charles, 1986, *The causatives of Malagasy*. Oceanic Linguistics Special Publication No.21. Honolulu: University of Hawaii Press.
- Rau, Victoria Der-Huo, 1992, *A grammar of Atayal*. Taipei: Crane.
- Reid, Lawrence Andrew, 1966, *An Ivatan syntax*. Oceanic Linguistics Special Publication No.2. Honolulu: Pacific and Asian Linguistics Institute, University of Hawaii.
- Richards, Norvin, 1996, Subjects in Tagalog and Icelandic. In Klammer, ed. 1906:31-49.
- Ross, Malcolm, 1995, Reconstructing Proto-Austronesian verbal morphology: Evidence from Taiwan. In Paul Jen-kuei Li, et al., eds *Austronesian studies relating to Taiwan*, 727-791. Symposium Series of the Institute of History and Philology, Academia Sinica, No.3. Taipei: Academia Sinica.
- Schachter, Paul, 1976, The subject in Philippine languages: topic, actor, actor-topic, or none of the above? In Charles N. Li, ed. *Subject and topic*, 491-518. New York: Academic Press.
- 1990, Tagalog. In Bernard Comrie, ed. *The world's major languages*, 936-958. New York: Oxford University Press.
- Schachter, Paul and Fe T. Otnes, 1972, *Tagalog reference grammar*. Berkeley: University of California Press.
- Shibatani, Masayoshi, 1988, Voice in Philippine languages. In Masayoshi Shibatani, ed. *Passive and voice*, 85-142. Amsterdam: John Benjamins.
- Sneddon, J.N. 1975, *Tondano phonology and grammar*. Canberra: Pacific Linguistics.
- Starosta, Stanley, 1986, Focus as recentralization. In Geraghty, Carrington and Wurm, eds, 1986:73-95.
- 1988, *The case for Lexicase: an outline of lexicase grammatical theory*. London: Pinter.
- Starosta, Stanley, Andrew K. Pawley and Lawrence A. Reid, 1982, The evolution of focus in Austronesian. In Wurm and Carrington, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2, 145-170. Canberra: Pacific Linguistics.

- Sundermann, Hermann, 1914, Der Dialekt der "Olon Maanjan" (Dajak) in Süd-Ost-Borneo. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië* 67:203-236.
- Szakos, József, 1994, Die Sprache der Cou: Untersuchungen zur Synchronie einer austronesischen Sprache auf Taiwan. Doctoral dissertation, Rheinische Friedrich-Wilhelms-Universität, Bonn.
- Tan, Cindy R., 1997, Simple sentences in Puyuma. MA thesis, National Taiwan Normal University.
- Topping, Donald M., 1973, *Chamorro reference grammar*. Honolulu: The University Press of Hawaii.
- Tsuchida, Shigeru, 1976, *Reconstruction of Proto-Tsouic phonology*. Study of Languages & Cultures of Asia & Africa monograph series No.5. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Tuuk, H.N. van der, 1971 [1864-1867], *A grammar of Toba Batak*. The Hague: Nijhoff.
- Vanoverbergh, Morice, 1955, *Iloko grammar*. Dagupan: Dagupan City Press.
- Voskuil, Jan E., 1996, Aspects of verbal morphology in Tagalog. In Klamer, ed. 1996: 51-67.
- Wolff, John U., 1973, Verbal inflection in Proto-Austronesian. In Andrew B. Gonzalez, ed. *Parangal kay Cecilio Lopez: Essays in honor of Cecilio Lopez on his seventy-fifth birthday*, 71-91. Philippine Journal of Linguistics Special Monograph Issue No.4. Quezon City: Linguistic Society of the Philippines.
- Wu, Joy J., 1995, Complex sentences in Amis. MA thesis, National Taiwan Normal University.
- Yeh, Meili, 1991, Saisiyat structure. MA thesis. Tsinghua University, Hsinchu, Taiwan.
- Zeitoun, Elizabeth, 1992, A syntactic and semantic study of the Tsou focus system. MA thesis. Tsinghua University, Hsinchu, Taiwan:

— PART II —

Languages of Sulawesi



Map 2: Languages discussed in Part II

Voice in Tukang Besi and the Austronesian focus system

MARK DONOHUE

1 Introduction

Tukang Besi is an Austronesian language based on the islands known as the *Kepulauan Tukang Besi* found off Southeast Sulawesi, in central Indonesia. In addition to this location there are also numerous trading communities (numbering in some cases above 10,000 people, such as in south-west Buru, southern Taliabu, and Fakfak in western Irian Jaya) scattered across eastern Indonesia (Donohue 1995, 1997). Tukang Besi is almost certainly a Sulawesi-area 'Western Malayo-Polynesian' language, but is difficult to subgroup with any confidence as a member of any of the recognised subgroups of the area. Due to its geographical position it is also the 'end of the line' as far as the western Malayo-Polynesian languages go to the south-east, before the Central Malayo-Polynesian languages begin.

Tukang Besi has two basic voices, and a full set of pronominal agreement markers on the verb: (obligatory) prefixes are used to mark the [S,A] of the clause, and also indicate the realis or irrealis state of the activity; enclitics are optionally used to mark an [O].¹ The presence or absence of the [O] enclitics controls the voice system, as independently witnessed by the nominal case marking, which distinguishes *na* 'nominative' (in the manner of Bell 1976, Kroeger 1993), *te* 'core' (but not nominative), *i* 'oblique' and *nu* 'genitive' (this last case may appear NP-internally only, and is not used to case-mark an argument in a main clause, as is found in many western Austronesian languages). These case-markers are used on all nominals, including pronouns: apart from the affixed forms, there is no difference in

¹ I shall use [S,A] and [O] following Dixon (1994), etc., to refer to the syntactic roles of the core arguments of a clause in an atheoretical manner; for definitions, see Andrews (1985:68). For what some people call 'grammatical subject' I shall use the term 'pivot', after Heath (1975), Foley and Van Valin (1984) and Dixon (1994).

The following abbreviations have been used; in the case of portmanteau morphemes, the individual components of the glosses have been separated out in the following list: 1,2,3 first, second, third person; ALL allative; CLASS classifier; COM comitative; CORE non-nominative; I irrealis; KP case phrase; NL nominaliser; NOM nominative; OBL oblique; OBJ object; OCC occupational; OP object prefix; REC reciprocal; PA paucal; PASS passive; PF 'perfective'; PL plural; POSS possessive; R realis; REC reciprocal; SI subject infix; SG singular; TOP topic.

pronoun sets. The nominative argument is also marked with *te* when preverbal, this word order being due to pragmatic focus (see §6).

Voice selection is controlled by factors of discourse continuity and prominence, with the most prominent argument being cast as the nominative one in a series of clauses, and subject to extensive zero anaphora. Some other constructions that refer to the nominative or non-nominative status of an argument include floating quantifiers, internal relative clauses and possessor ascension, all of which require nominative status.

The older PAn voice morphology is largely preserved in *Tukang Besi*, with both *-[um]-* and *i-/ni-* appearing in some subordinate clauses headed by an [S,A] or [O] (respectively), and *-[um]-* also appears as a pragmatic focus construction in main clauses. The locative * *-an* is preserved in the locative nominalising suffix *-'a* 'place of...', which is taking on more general nominalising properties, a property also found for *i-*. The function of this morphology in syntactic voice alternations has been taken over by the pronominal indexing system, as described earlier, and a series of passive(-like) verbal prefixes, *to-* 'passive', *te-* 'accidental passive' and *mo-* 'anticausative'.² The last of these, *mo-*, is identical in form to the fossilised verb marker which appears on many (but not all) adjectives (not verbs). Adjectives are also commonly marked with two other variants of the *mo-* marker: *me-*, which is also the productive frequentive prefix, and *ma-*, which has no further uses, and is the most infrequent of the three. All of these morphemes occur in the same position – following the [S,A] prefix and preceding the verb and any derivational morphology.

2 Basic verbal morphology

Verbs in *Tukang Besi* are indexed to indicate the person and number of their [S,A] and [O]. There are two sets of [S,A] prefixes, depending on the mood (realis or irrealis) of the verb, and one set of [O] enclitics. The forms of these pronominal agreement markers (and the other sets, the free pronouns and the possessive enclitics) are given in Table 1:³

2 An anticausative is a morpheme that removes the cause of a resultant state: 'She boiled the water', compared to 'The water boils(-anticausative)'. Compare the *Tukang Besi* sentences:

<i>No-rede na uwe.</i>	and	<i>No-mo-hengolo na uwe.</i>
3R-boil NOM water		3R-ANTCAUS-boil NOM water
'The water is boiling.'		'The water has been boiled.'

in which the choice of verbs is interesting – with *rede* being an intransitive verb, and *hengolo* a transitive one. The second sentence contains an anticausative, and implies the prior existence of an agent.

3 The phonemes found in *Tukang Besi* are the following: *p t kʔ b d dʒ g β s h m n ŋ mp nt ns ŋk mb nd ndʒ ŋg r l i ε a o u*. The following orthographic conventions are followed: *ʔ*, *'*; *β*, *b*; *d*, *d*; *β*, *w*; *ŋ*, *ng*; *dʒ*, *j*; *ʒ*, *l*; *ε*, *e*; *u*, *u*. See Donohue (1994) for details. A surface phonemic transcription has been used, with phonemes that are underlyingly present (as determined through paradigmatic alternations), but not surfacing in a particular sentence, shown in brackets ().

Table 1: *Tukang Besi* pronominal forms

	[S,A]		Free forms	Possessive	[O]
	Irrealis	Realis			
1SG	<i>ku-</i>	<i>ku-</i>	<i>iaku</i>	<i>=su</i>	<i>=aku</i>
2SG	<i>ko-</i>	<i>'u-/nu-</i>	<i>iko'o</i>	<i>=u</i>	<i>=ko</i>
3SG	<i>na-/a-</i>	<i>no-/o-</i>	<i>ia</i>	<i>=no</i>	<i>=e</i>
1PA	<i>ka-</i>	<i>ko-</i>	<i>ikami</i>	<i>=mami</i>	<i>=kami</i>
1PL	<i>ta-</i>	<i>to-</i>	<i>ikita</i>	<i>=nto</i>	<i>=kita</i>
2PL	<i>ki-</i>	<i>i-</i>	<i>ikomiu</i>	<i>=miu</i>	<i>=komiu</i>
3PL	<i>na-/a-</i>	<i>no-/o-</i>	<i>amai</i>	<i>=no</i>	<i>=e</i>

The [S,A] prefixes are used with all transitive and intransitive [S,A]s, regardless of the semantic roles of the arguments concerned. The [O] enclitics are similarly used for all primary [O]s of the verb, though there is a (semi-)archaic set of dative [O] enclitics that are occasionally encountered.⁴ Comparison with the free forms reveals that the [O] clitics are only minimally different from the free pronominal forms, and so probably represent a relatively recent development.⁵ The alternations for the 2SG realis and the third person [S,A] forms is insignificant, with no meaning or dialectal differences ascribable to them. The only difference that can be teased out is one of speech tempo, with the *'u-*, *o-* and *a-* forms tending to be used in faster speech, though this is not a hard and fast rule. Examples of the use of some of these affixes are given in (1)–(7) (only the realis set of [S,A] prefixes are illustrated to save space):

- (1) *Ku-gonti te kau.*
1SG-chop CORE wood
'I chopped the wood.'
- (2) *Ku-tinti=mo kua ito.*
1SG-run=PF ALL there:higher
'I ran away to the mountains.'
- (3) *Ku-mohoo.*
1SG-sick
'I'm sick.'
- (4) *"O-ha'a 'u-doito, La Kape'ingkape'i?"*
3R-why 2SG.R-cry La Fool
'Why are you crying, Fool?'

⁴ These have the forms (presented in the same order as the table above) *=naku*, *=nso*, *=ne*, *=nsami*, *=nggita* and *=ngkomiu*; there is no 3PL dative [O] enclitic.

⁵ The use of *s*, rather than *k*, in the 1SG.POSS form is unusual, and reflects the partial adoption of a **k > s* sound change that is prevalent in Southeast Sulawesi, but (apart from in this one morpheme) is not found in *Tukang Besi*.

- (5) *No-topa=aku ka'ano no-pa-muru.*
 3R-slap=1SG.OBJ because 3R-OCC-bald
 'She slapped me because she was angry.'⁶
- (6) *No-wila lego-lego.*
 3R-go arms.swinging
 'He was walking, swinging his arms.'
- (7) *No-buti='e.*
 3R-fall=3OBJ
 'He dropped them.'

As (1)–(7) show, there are no complications with the verbal indexing system found in *Tukang Besi*: all [S,A]s are prefixed, and [O] agreement is done by optional enclitics. There are no obvious morphological traces of split-intransitivity, ergativity, hierarchical systems or other exotica: it exhibits a simple nominative-accusative alignment.⁷

3 Basic clause order and case marking

The basic verbal clause in *Tukang Besi* is verb-initial, and, due to extensive head-marking, core arguments are optional if the identity of their referents has already been established, the information about their syntactic functions being carried by verb agreement. A nominal [O] of a transitive clause usually appears immediately following the verb, and the [S,A] follows this, but the order of these two constituents is not fixed. The basic clause can be modelled as follows:

Transitive:	s-V-o	na O	te A
Intransitive:	s-V	na S	

This is, on first inspection, an ergative-absolutive case marking system: an intransitive [S] or an [O] is marked with *na*, a transitive [A] is marked with *te*. Arguments against this analysis will be presented shortly. Sentences (8) and (9) illustrate these patterns:

Transitive:

- (8) *No-'ita='e na kene=no te ana.*
 3R-see=3OBJ NOM friend=3POSS CORE child
 'The child saw its friend.'

Intransitive:

- (9) *No-tinti na ana.*
 3R-run NOM child
 'The child ran off.'

⁶ When the stative verb *muru* 'bald' occurs with the occupational prefix ('one who habitually Vs'), it has the meaning 'angry', as in the example here.

⁷ There are in fact some exceptions to this, found in some serialised motion verbs (see Donohue 1998 for details), but these are not part of the basic alignment system. The syntactic correlates of split intransitivity are described in Donohue (1996a).

As illustrated above, NPs referring to a core argument are obligatorily preceded by a case marker, either *te* or *na*. If the argument is known, given information, and pragmatically prominent, it may be assigned nominative case, and is marked with the nominative case marker *na* (with variant *a*) (glossed as 'NOM'). Only one argument per clause may be nominative; other core arguments, not selected as filling the nominative position in the clause, are marked with the general non-nominative core case marker *te* (with variants '*e* and *e*') ('CORE').⁸ The unit that is made up of the case marker and the NP is referred to as the case phrase (KP) (see Lamontagne and Travis 1987 for discussion of this unit).

4 Transitive verbs without [O] enclitics

Constituent order and nominal marking strategy differ when a transitive verb appears without an [O] enclitic. When the enclitic is not used, the [S,A] prefixing on the verb does not change, but the basic constituent order of the arguments is [VO] A and, importantly, the use of the case markers is reversed in comparison with a clause with [O] enclitics. This is shown schematically as follows:

Normal transitive:	s-V-o	<i>na</i> O	<i>te</i> A
no [O] indexing:	s-V	<i>te</i> O	<i>na</i> A

In these transitive clauses without [O] enclitics, the marking at the KP level has changed, but the verbal indexing of the [A] argument has remained consistent with example (1). That is: the [A] argument is still prefixed onto the verb, but is no longer marked at the nominal level by the general core case marker *te*, but rather by the nominative case marker *na*. Comparing this transitive case-marking system with the intransitive system, we would now want to consider it to be nominative-accusative. Examples (10) and (11) contrast [O] agreement and non-agreement constructions:

Transitive verb with [O] enclitics:

- (10)a. *No-kiki'i=ko (na iko'o) te beka.*
 3R-bite=2SG.OBJ NOM you CORE cat
 'The cat bit you.'
- b. **No-kiki'i=ko te iko'o na beka.*
 3R-bite=2SG.OBJ CORE you NOM cat
 'The cat bit you.'

Transitive verb without [O] enclitics:

- (11)a. *No-kiki'i te iko'o na beka.*
 3R-bite CORE you NOM cat
 'The cat bit you.'

⁸ The term 'core' does not reflect all the functions of *te*. It is also used when any core argument is fronted (either clause-internally or topicalised), so it is perhaps better to call it the 'other' case marker; core argument other than the nominative, argument other than the post verbal ones. This makes a terrible label, however, is not very serious-sounding and is impossible to abbreviate.

- b. **No-kiki'i te beka na iko'o*
 3R-bite CORE cat NOM you
 'The cat bit you.'

In both (10a) and (11a) the agent *beka* is indexed on the verb by the third person realis [S,A] prefix *no-*, and additionally in (10) the second person singular [O] is indexed by means of the second person singular [O] enclitic =*ko*. When this enclitic is not used, as in (11), the agent *beka* must be marked by the nominative case marker *na*, and *iko'o* by the non-nominative case marker *te*. This variation in the presence or absence of [O] enclitics is the only way that the case assigned to nominals may be changed; note the ungrammatical (10b) and (11b). Clearly, the presence or absence of [O] agreement on a verb functions as a form of voice system. The analysis adopted here is that it is a Philippine-style voice system,⁹ with a small number of voice categories (two); various arguments that this is the most expedient analysis are given elsewhere in this paper, and in more detail in Donohue (1995). The choice of the voice used is dependent on the pragmatic status of the arguments in a clause, and their saliency in discourse, and also has interpretative consequences: clauses with [O] enclitics are more likely to encode a highly individuated [O], or a more highly affected one, in a perfective or punctiliar time setting.

5 Markedness and the type of voice system

Although the transitive verbs we have seen can appear either with or without [O] enclitics, there is evidence that the encliticised (i.e. morphologically more complex) versions are in some sense the 'basic' ones: they appear more frequently in texts (approximately 70% of transitive verbs in texts use [O] enclitics); they are the citation forms of most transitive verbs; and not only may all transitive verbs appear with [O] enclitics, there are some that cannot appear without them, such as the verb *molinga* 'remember'. The limited data available to me on child language acquisition suggests that children learning *Tukang Besi* acquire a command of the [O] enclitics earlier than they do of the [S,A] prefixes; children often substitute the near-frozen 'adjectival' prefix *mo-* in the place of [S,A] prefixes until they are about 5 or 6 years old, but seem to be able to manipulate the [O] suffixes on their verbs much earlier, indicating that [O] indexing is learnt earlier than is [S,A] indexing.

If we were to consider the forms in (10) as basic, and conclude that the language has an ergative-absolutive case marking paradigm, then logically the sentences in (11) are antipassive derivations of them. There are several problems with this analysis: the [A] in the 'antipassive' construction in (11) is still a core argument, and does not undergo demotion of any sort, a process taken to be normal treatment of *by*-phrases in passive and antipassive

⁹ By which I refer to the voice system that is found in the majority of the Austronesian languages of Taiwan, the Philippines, Madagascar and large parts of western Indonesia, characterised by a case marking system that monitors the degree of pragmatic saliency of an argument rather than its syntactic role. The symmetrical nature of the voices in this sort of system (none of the basic voices being unambiguously derived from the other) is the other main characteristic of this style of voice system.

constructions (Baker 1988:9; Dixon 1994:146, amongst others).¹⁰ Secondly, there is no derivational morphology involved in the derivation of the ‘antipassive’, rather the derivation is carried out by DROPPING verbal morphology. This is a surprising artefact of the analysis, more so given that *Tukang Besi* DOES have a morphological passive construction, one that (like all other reported examples of passives and antipassives) involves additional morphological material, in the form of a verbal prefix, not a suffix. That the antipassive construction would be so different, by both language-internal and cross-linguistic evidence, seems remarkable. An even more compelling reason to not regard this as an antipassive form emerges when we examine external relative clauses, later on.

Alternatively, the sentences in (11) can be considered basic, and those in (10) can be thought of as ‘passive’-like derivations of them. This would be consistent with the relative amount of morphology found on the verbs. We are then faced, however, with a peculiar passive morpheme that varies for person and number of its derived [S] (we’d presume), and in which the [S,A] prefixes of the verb do not agree with the (derived) [S]. Again, typologically very odd, and even less plausible when we remember that there are other, unambiguous passive morphemes in the language, including the prefix *to-*. Compare (10) and (11) above with (12), found with a *to-* passive form, in which no *by*-phrase may be mentioned, and the single argument¹¹ of the verb may be indexed on the verb by means of [S,A] prefixes:¹²

- (12) *‘U-to-kiki’i na iko’o.*
 2SG.R-PASS-bite NOM you
 ‘You were bitten.’

In (12) the patient nominal takes the nominative case marker just like the patient of an [O] encliticised verb form such as (10), but unlike that sentence, the patient of the passive verb in (12) is the [S], not [O], and is indexed by the prefixed set of pronominal affixes, as is an argument of an intransitive verb. Thus, while treating the patient of the verb alike, as far as its nominative marking goes, the indexing strategy on the verb is quite different. Notice also that in (10) the agent of the verb is present in the [S,A] prefixes on the verb; in (12) the agent may not be expressed in any way whatsoever.

These arguments show that the case marking and pronominal indexing system of *Tukang Besi* is best thought of as not representing either an ergative case-marking system with an antipassive, nor an accusative case-marking system with a passive. It is, however, similar to what appears to be found in *Kapampangan*, a well-described Philippine language (Mirikitani 1972 and others), and (perhaps) *Jarawara*, an Amazonian language. Although

¹⁰ Foley and Van Valin (1984:176-81) argue that there is a Jacalteco antipassive construction found in relative clauses that treats the [O] as a core argument, and that Sama also has a non-backgrounding antipassive construction. The analysis of Sama seems to me to be flawed, appearing to force the language into either an accusative or an ergative mould.

¹¹ Again, a slight oversimplification: if the [A] of the unpassivised sentence is an instrument, then it may be overtly present in the passive sentence; similarly, a ditransitive verb, or one with applicative morphology, allows more than one argument in its passive sentence forms.

¹² Though a third person prefix may always be substituted: *Notokiki’i na iko’o* is also grammatical as an alternative, even though the verb shows no agreement with the ‘derived-[S]’. This form shows that the apparent [S] of a passive construction is not in fact eligible for the grammatical salience that is attributable to other arguments of intransitive verbs. The passive construction here is, in Foley and Van Valin’s terms, a *demoting passive* that does not create a new pivot.

an 'ergative' analysis of Philippine-type languages (including Kapampangan; see Mithun 1994) does become fashionable every so often, I do not find this a convincing analysis for many reasons (see Donohue 1998 for a summary). Regardless of the 'rightness' of this analysis for the better-known Philippine-type languages, the *Tukang Besi* data is even less amenable to an ergative analysis. For these reasons I have chosen to analyse the voice alternation as being the result of a Philippine-style 'focus' system (further arguments that the ergative analysis is inappropriate for *Tukang Besi* can be found in Donohue 1995:160-66). In *Tukang Besi* the diachronic drift towards head-marking pronominal indexing, found throughout southeast Sulawesi and the islands to the south of that region, has proceeded to quite an extent,¹³ but at the same time the overt Philippine-style case system has been preserved, and its pronominal verbal cues have been reinterpreted as being those involving the presence versus absence of the [O] enclitics. This is strikingly similar to one recent analysis of voice systems in other, 'mainstream' Philippine languages as involving incorporated pronominal elements (see Sells 1995).

6 Variation in case marking: preverbal position

The basic order of constituents presented in §3 and §4 can be, and often is, modified through the appearance of an argument before the verb. There are two strategies by which a nominal can appear in a pre-verbal position: either fronting one of the core arguments to a position within the clause, which serves as a 'focussing' strategy, or fronting to a position outside the clause, topicalising the nominal. Only clause-internal fronting is discussed in detail in this article; what is here referred to as the "pre-verbal position" is very similar to the position that Durie (1987) called the CORE TOPIC. This term is not appropriate for *Tukang Besi*, however, since non-core time expressions may also occur in this position.

The topicalisation that can occur in *Tukang Besi* also creates a preverbal argument, but unlike the preverbal and clause internal position that is described here, the topic position is demonstrably outside the clause. Moreover, the topic is not restricted to being a particular argument: any argument, core or oblique, may appear in the topic position; and if non-core, it will retain its original case marker or preposition. Further discussion of the treatment of two pragmatically-determined preverbal positions (in Mayan languages) can be found in Aissen (1992).

Within the clause, the only argument nominals that may be fronted are those bearing the nominative pivot properties. The nominal is placed in a pre-verbal position, yet still within the clause, and the case marker of the nominal is not the nominative *na*, but rather the more general *te*. Thus we can say that arguments with nominative case are marked either by the case marker *na*, or by preverbal position (and the general case marker *te*). The pronominal marking on the verb is unaffected by this process. The constituent order and case marker use found in clauses with a preverbal argument can be summarised as follows:

¹³ Though this tendency towards head-marking has been carried even further in some of the Muna languages in the region, which allow more than one [O] agreement marker on the verb in trivalent clauses: see, for example, Muna (van den Berg 1989).

Transitive:	<i>te</i> O	s-V-o	<i>te</i> A
Intransitive:	<i>te</i> S	s-V	
Transitive, no [O] marking	<i>te</i> A	s-V	<i>te</i> O

Variations of sentences (8) and (9) showing fronting are presented below as (13) and (14), and a version of (8) without [O] agreement as (15):

- (13) *Te kene=no no-'ita='e te ana iso.*
 CORE friend=3POSS 3R-see=3OBJ CORE child yon
 'That child saw its friend.'
- (14) *Te ana iso no-tinti.*
 CORE child yon 3R-run
 'That child is running.'
- (15) *Te ana iso no-'ita te kene=no.*
 CORE child yon 3R-see CORE friend=3POSS
 'That child saw its friend.'

Note the difference between (13) and (15), in which the change in grammatical relations is signalled only by constituent order and the presence versus absence of the [O] clitic on the verb.

7 A short note on voice selection and word order

As has been mentioned in §3–5, one argument in a clause is selected, based on its pragmatic prominence, and assigned nominative case. This choice is motivated by the exigencies of discourse, since the nominative argument is the preferred controller and target of zero anaphora across coordinate clause boundaries. Since the nominative argument usually represents relatively older, known and more 'given' information, with newer participants appearing as non-nominative arguments, arguments are usually nominatively marked only after being introduced as a non-nominative argument.

This pattern is illustrated in (16) taken from the beginning of a story, in which there can be no assumed knowledge about the identity and relative prominence of the participants. The protagonist *Wa Sabusaburengki* is introduced as the object of an existential clause, and in the next clause becomes the predicate of an identificational clause, and then the nominative [A] of the transitive clause headed by *asumumbele*. Following the introduction of a new argument as the [O], *Wa Sabusaburengki* loses nominative status (but remains an [A]); the new character introduced as an [O] in the preceding clause, the chicken (*kadola*) becomes the new nominative argument and retains this status for the rest of the passage:

- (16) *Sapaira sapaira ana, ane kene wowine_i sa-mia, te ngaa=no_i*
 once.upon.a.time exist and woman 1-CLASS CORE name=3POSS
- te Wa Sabusaburengki.*
 CORE Wa Sabusaburengki
- Te Wa Sabusaburengki ana_i a_i-s[um]umbele te kadola_j.*
 CORE Wa Sabusaburengki this 3I-decapitate.SI CORE chicken

La'a=mo na_i-s[um]umbele-'e_j na kadola iso_j, no_j-pogau-mo
 just=PF 3I-decapitate.SI=3OBJ NOM chicken yon 3R-say=PF

na kadola iso_j kua...

NOM chicken yon :

'Once upon a time, there was a lady_i, and her name_i was Wa Sabusaburengki.

Wa Sabusaburengki_i was going to cut off a chicken_j's head. Just as she_i was about to cut off its_j head, that chicken_j said ".....". (WaSab: 1-3)

A short example from the middle of a text illustrates the mechanisms by which an argument is re-introduced as the main player in a text:

(17) *Ara ku_i-[m]o-busu na_j-t[um]alo=aku_i, kene te ia_j*
 if 1SG-REC.SI-forward.fist 3I-win.SI=1SG.OBJ and CORE 3SG

no_j-pande di lola-'a, jari labi ku_i-akala-'e_j.

3R-clever OBL fly-NL so better 1SG-trick=3OBJ

'If I_i want to fight he_j'll beat me_i, and he_j's good at flying, so it'd be better if I_i tricked him_j.' (RA: 24)

In these four clauses, 'I' begins as the nominative argument in an [S] role (with no [O] in the clause, the single argument must be the one with nominative case) in the conditional clause. The next clause sees a different argument ('he') introduced in [A] role, and 'I' continuing in an [O] role, still the nominative argument. The third clause uses fronting to highlight the change of grammatical relations; in this clause, the 'he' argument is continued, but placed preverbally as the single argument of an intransitive verb, making it necessarily nominative (although the overt marking is not nominative, because of its position). The final clause follows the same pattern as was seen in the second clause, the [S] argument now becoming an [O] but remaining nominative, and a new argument being (re-)introduced as an [A]. Notice also that in four clauses containing two transitive and two intransitive verbs, only once is a core argument expressed with a nominal as well as the pronominal affixes, and that occurred when there was a change in the identity of the nominative argument in the second clause. Since the referential information about the participants is already clear from the context of story, only the role information present on the verbs is needed, combined with occasional pragmatic marking of the nominals, to monitor which participant is being referred to at any time.

As would be expected, given the lack of previous information, the proportion of core nominals per clause is higher in the introductory fragment in (16), which serves to lexically expand the role information carried on the verbs. Of the five clauses in (16), two are transitive verbal clauses and one an intransitive verbal clause; these three clauses display a total of four KPs. The extract from the middle of a text presented in (17) has three clauses, and only one KP. This clearly reflects a preference for more fully (lexically) specified arguments at the beginning of a text than at some point in the middle (see work by Du Bois, e.g. 1987, on preferred argument structure for a discussion of the relevance of these facts to morphosyntax).

8 Evidence for the pivot status of the nominative argument: floated quantifiers

In *Tukang Besi*, *saba'ane* 'all' and some other quantifying expressions may occur in the NP or 'float' to a position outside the NP, and appear immediately pre- or post-verbally. In all cases that a quantifier appears outside an NP, it is launched by the NOMINATIVE argument of the clause, regardless of the syntactic or thematic role borne by that NP. Some examples are presented below. In the following example, the quantifier appears in its normal NP-internal position:

Launched by a nominative [A]:

- (18) *No-lemba te kaluku* [KPna amai [QUANT*saba'ane*]]
 3R-carry CORE coconut NOM 3PL all
 'All of them carried coconuts.'
 * 'They carried all of the coconuts.'

Alternative orders show the quantifier in either immediately preverbal, or immediately post-verbal, positions. In both cases the interpretation is unambiguously the same as the clause in (18).

Floated:

- (19) *Nolemba* [QUANT*saba'ane*] *te kaluku* [KPna amai]
 (20) [QUANT*Saba'ane*] *nolemba te kaluku* [KPna amai]

When the verb is found with [O] agreement, the [O] nominal is now the nominative one. An example of a non-floated quantifier appearing in the [O] NP is given below:

Launched by a nominative [O]:

- (21) *No-lemba=e* [KPna kaluku [QUANT*saba'ane*]] *te amai*.
 3R-carry=3OBJ NOM coconut all CORE 3PL
 'They carried all of the coconuts.'
 * 'All of them carried coconuts.'

As with the examples in (19)–(20), it is possible for the quantifier to float away from this position; examples of pre- and post-verbal quantifiers, still with the same reference as in (21), are given below.

Floated:

- (22) *Nolemba'e* [QUANT*saba'ane*] [KPna kaluku] *te amai*.
 (23) [QUANT*Saba'ane*] *nolemba'e* [KPna kaluku] *te amai*.

Finally, (24)–(26) show the use of a quantifier, floated and unfloated, with an intransitive verb. In these cases the reference is still unambiguous. Adding an oblique phrase such as *kua wunuano* 'to their house' to any of the sentences cannot force a possible 'They went home to all of their houses' reading, regardless of the position of the quantifier.¹⁴

¹⁴ The only way to get this reading with a floated quantifier is for the goal to be coded as a nominative applied [O]:

No-mbule=api=e=mo saba'ane na amai kua wunua=no
 3R-return=APPL=3OBJ=PF all NOM 3PL ALL house=3POSS
 'They returned to all of their houses.'

Launched by a nominative [S]:

- (24) *No-mbule=mo* [_{KPna} *amai* [_{QUANT} *saba'ane*]]
 3R-return=PF NOM 3PL all
 'They all went home.'

Floated:

- (25) *Nombulemo* [*saba'ane*]_{QUANT} [_{KPna} *amai*]
 (26) [_{QUANT} *Saba'ane*] *nombulemo* [_{KPna} *amai*]

In all the above sentences, regardless of the syntactic role borne by the quantified nominal, it is the nominative argument that is quantified by the floated quantifier *saba'ane*. Furthermore, this quantifier, when floated, may not refer to the non-nominative argument in a sentence. It is also worth noting that the ability to launch a quantifier is not dependant on the morphological case displayed. Recall from §6 that a preverbal argument appears with the *te* case-marker, but it is still able to launch a floated quantifier. Compare (19) with (27), which does not have an overt nominative case marker since it is preverbal, but has the same restrictions on interpretation,

- (27) [_{KP} *Te amai*] *nolemba* [_{QUANT} *saba'ane*] *te kaluku*.
 CORE 3PL 3R-carry all CORE coconut
 'All of them fetched coconuts.'

Other grammatical constructions that single out the (syntactically) nominative argument as the preferred pivot include conjunction reduction, internal relative clauses, external possession (Donohue 1999b) and temporal clauses. These all select an argument that must be nominative, and so provide evidence that the system of verbal marking and case alternations presented in §2–4 is indeed a voice system, with syntactic (and not just pragmatic) effects.

9 Relics of old focus morphology: external relative clauses and nominalisations

Despite the fact that the voice system in *Tukang Besi* makes no reference to the reconstructable Austronesian voice affixes, several of these affixes are reflected in *Tukang Besi*, in the form of the affixes *-[um]-*, *(n)i-* and *-'a*. The use of these affixes is described in the following sections, according to function.

9.1 External relative clauses

Tukang Besi uses the 'subject infix' (SI) infix *-[um]-* to mark a relative clause with an [S,A] as its head, and the object prefix (OP) *i-* to show that the head of the relative clause is an [O] (the [O] prefix surfaces as *di-* (in younger speakers, probably through the influence of Malay), and *ni-* or *i-*, idiolectally and dialectally). These two affixes clearly reflect the Austronesian actor voice and object voice affixes (terminology following Kroeger 1993), **-um-* and **-in-*, and are still functioning in roles easily relatable to these original meanings, though they are not found in main clauses. Indeed, they are not voice markers in the subordinate clauses; the alternation between the presence and absence of [O] marking on transitive verbs is found in relative clauses constructed with *-[um]-* as well as in main clauses (see example (37)).

Examples of these different relative clauses can be seen in (28)–(36) (more details on interesting aspects of relative clauses in *Tukang Besi* can be found in Donohue (1996b)). The first two examples simply illustrate the case marking patterns for the arguments of transitive verbs in main clauses.

Main clauses:

- (28) *No-balu te pandola na wowine.*
 3R-buy CORE eggplant NOM woman
 ‘The woman bought an eggplant.’
- (29) *No-balu=e na pandola te wowine.*
 3R-buy=3OBJ NOM eggplant CORE woman
 ‘The woman bought the eggplant.’

Relative clauses based on the sentences in (28) and (29) are given below. The basic use of the subject relative clause and object relative clause are given in (30) and (31), while (32) and (33) show that the different morphological markers may not be freely interchanged.

Relative clauses:

- (30) *Te ia iso te wowine [RCb[um]alu te pandola]*
 CORE s/he yon CORE woman buy.SI CORE eggplant
 ‘That’s the woman who bought the eggplant.’
- (31) *Te iso te pandola [RCi-bal(u) u wowine]*
 CORE yon CORE eggplant OP-buy GEN woman
 ‘That’s the eggplant that was bought by the woman.’
- (32) **Te iso te pandola [RCb[um]alu te wowine]*
 CORE yon CORE eggplant buy.SI CORE woman
 (Good for: ‘That’s the eggplant that bought the woman.’)
- (33) **Te ia iso te wowine [RCi-bal(u) u pandola]*
 CORE s/he yon CORE woman OP-buy GEN eggplant
 (Good for: ‘That’s that woman who was bought by the eggplant.’)

With intransitive verbs the only possible choice of relative clause is the subject relative clause; attempts to use the object prefix are not grammatical with intransitive verbs.

Intransitive clauses:

- (34) *No-kengku na uwe iso.*
 3R-cold NOM water yon
 ‘That water is cold.’
- (35) *Te iso te uwe [RCk[um]engku]*
 CORE yon CORE water cold.SI
 ‘That’s the cold water.’
- (36) **Te iso te uwe [RCi-kengku]*
 CORE yon CORE water OP-cold

Note that it is NOT simply the nominative argument of the relative clause that is the pivot for this construction. If there are [O] enclitics on the verb of the relative clause, then the [O] in the relative clause will receive nominative marking, just as in a main clause:

- (37) *Ku-ita te kalambe [RCk[um]ele='e na kaujawa]*
 1SG-see CORE girl carry.by.strap.SI=3OBJ NOM cassava
 'I can see the girl who's carrying the cassava.'

This fact is one of the strong arguments against considering the system of voice alternations presented in §3 and §4 as displaying either passive or antipassive characteristics – the choice of voice forms has no effect on the status of the argument, as [A] or [O], for the purposes of this construction. Other constructions that make explicit reference to the [S,A] vs [O] grouping of syntactic roles (the same split found in the verbal agreement) are control in complement clauses, [O] incorporation, and case marker adoption.

9.2 Nominaliser -'a

There are rather few affixes in *Tukang Besi* that can be used to derive nominals from otherwise verbal roots. There are, however, three strategies which are used commonly for the derivation of nominals, though they are not exclusively derivational constructions. One is the use of relative clauses without a head in the N position, as seen in (38), illustrating both a subject relative clause and an object relative clause:

- (38)a. *te [RCt[um]inti]*
 CORE run.SI
 'the running (person)'
- b. *te [RCmbeaka i-'ita]*
 CORE not OP-see
 'the spirit'¹⁵

Another strategy that exists is for a precatatorial root to simply be used in either a verbal or a nominal syntactic position, with no derivational morphology required. This pattern of alternations without any derivational morphology is seen in (39) and (40).

- | | Nominal/Referential | | Verbal/Predicative |
|--------|---|----|--|
| (39)a. | <i>te bose</i>
CORE paddle
'the paddle' | b. | <i>No-bose.</i>
3R-paddle
'S/he is paddling.' |
| (40)a. | <i>te tomba</i>
CORE mud
'the mud' | b. | <i>No-tomba=mo.</i>
3R-mud=PF
'It's become muddy (already).' |

Finally, the suffix -'a serves to derive a nominal concept from an explicitly (lexically specified as such) verbal concept. When applied to verbal roots, the result is usually an abstract noun, referring to the action of the verb. With 'verbal' concepts that are based on roots which are more precatatorial in nature, the derived nominal often refers to the place in which the action is conducted, though it can also refer to the conduct of the action itself. Here we see the connection with the Austronesian morpheme -an, which serves as a dative voice

¹⁵ The position of *mbeaka* 'not' is not unusual in (38b), and is found in headed relative clauses as well (e.g. *Te mia mbeaka i'ita* 'the person who was not seen').

marker in most more northerly Philippine-type languages. An example of each of these cases is seen in (41) and (42):

- (41)a. *No-wila.*
3R-go
'They are going.'
- b. *te wila-'a=no*
CORE go-NL=3POSS
'their going'
* 'the place that they go (to)'
- (42)a. *No-manga.*
3R-eat
'They are eating.'
- b. *te manga-'a=no*
CORE eat-NL=3POSS
'their eating'
'the place that they eat (at)'

With some verbs, the difference between these two senses (abstract nominalisation and locative nominalisation) has developed into a morphological distinction – the (rarely attested, and lexically determined) allomorphs *-ra* and *-ma* appear to have more specific semantic domains than does the more general *-a*. Compare (43b) and (43c) with (44b) and (44c):

- (43)a. *No-kede.*
3R-sit
'They are sitting.'
- b. *te kede-'a=no*
CORE sit-NL=3POSS
'their sitting'
* 'the place that they are sitting'
- c. *te kede-ma=no*
CORE sit-NL=3POSS
'the place that they are sitting'
* 'their sitting'
- (44)a. *no-'ita.*
3R-see
'they are looking'
- b. *te 'ita-'a=no*
CORE see-NL=3POSS
'(the fact of) their looking'
* 'the way that they look'
* 'the place that they look'
- c. *te 'ita-ra=no*
CORE see-NL=3POSS
'the way that they look'
* 'their looking'
* 'the place that they look'

Although the evidence in (43)–(44) suggests that several suffixes are developing, they are not yet productive enough to require special treatment. The suffix *-ma* has been observed on only one word, *kede* 'sit', and the *-ra* suffix on only two, *'ita* 'see' and *namisi* 'feel, taste' (with the irregularity that *namisi* + *-ra* yields not **namisira* but *namira*). A more regular (but still not completely predictable) alternation is the dissimilation that *-a* displays when following a

syllable with a glottal stop, appearing as *-ka*. This is not wholly regular, however, with some lexical items retaining the glottal stop in the nominalising suffix. For example, compare the forms in (45) and (46), in which the first consistently appears with [ka] as the form of the nominalising suffix, and the second always appears with [ʔa]:

- (45) *te motindo'u-ka='u di uwe*
 CORE thirsty-NL=2SG.POSS OBL water
 'your thirst for water'
- (46) *te helo'a-'a=(')u nu bae*
 CORE eat-NL=2SG.POSS GEN rice
 'your cooking of rice'

As explained, the function of *-a* and its alternants is to derive unambiguously nominal words from either precategoryal or verbal bases. It may not appear with an unambiguously nominal base:

- (47) **te komba-'a*
 CORE moon-NL
 'the mooniness' (?)

Once derived, the nominal displays all the properties associated with an N, and is otherwise unexceptional. It will be noted in the examples above that the [S,A] of the verb may be present in the derived nominal, expressed by a possessive enclitic. It may also be expressed by a genitive NP, rather than just its pronoun:

- (48) *te wila-'(a) u amai La Tonggi*
 CORE go-NL GEN 3PL La Tonggi
 'the going of La Tonggi and his group'

If the verb is transitive, then the [O] may appear as a genitive phrase. The normal interpretation is that the first genitive phrase refers to the [S,A] of the equivalent verbal expression, though this restriction is not an absolute one in nominalisations (object relative clauses are stricter in their requirement that the first genitive phrase refers to the *by*-phrase, and also more likely to include more than one genitively indexed argument):

- (49) *te 'ita-'(a) u Wa Ode Kiradati*
 CORE go-NL GEN Wa Ode Kiradati
 'the seeing of Wa Ode Kiradati'
- A: the act of seeing that Wa Ode Kiradati carried out, resulting in her seeing someone/something else
- B: the act of seeing that was carried out by someone/something, which resulted in Wa Ode Kiradati being seen.

With (most) ditransitive verbs, or verbs with applicative or other valency increasing morphology, all the core arguments may be present in this manner.¹⁶ An example of several genitive phrases on one nominalisation can be seen in (50):

¹⁶ Though more than one or two are unlikely to occur in natural speech, since the lack of strict rules on the position of postverbal arguments in different syntactic roles makes it difficult to interpret these sentences (such as the ambiguity of (49), in which *Wa Ode Kiradati* is not unambiguously identified as either the see-er or the seen).

- (50) 'E, te pa-manga-'a=(('u) u Aswi nu ik(a) atu,
 huh TOP CAUS-eat-NL=2SG.POSS GEN Aswin GEN fish that
 no-marasai na 'ita-ra-no la i.'
 3R-difficult NOM see-NL=3POSS ILL.FORCE FAMILIAR
 'Hey, the way you fed the fish to Aswin, it didn't look easy.'

Conceivably the nominalisation in (50) could be interpreted as 'Your feeding of Aswin to the fish', but this is pragmatically rather unlikely. With ditransitive verbs, whether of the ⟨[Agent], [Dative], [Theme]⟩ type or the ⟨[Agent], [Instrument], [Theme/Patient]⟩ type, all arguments may appear in the nominalisation with genitive case marking, though the instrument is unlikely to appear without the theme/patient appearing as well:

- (51) *No-'ita te tompa-'a=n(o) u Aswi nu watu.*
 3R-see CORE throw-NL=3POSS GEN Aswin GEN stone
 'They saw her throwing the rock at Aswin.'
- (52) # *No-'ita te tompa-'a=n(o) u watu.*
 3R-see CORE throw-NL=3POSS GEN stone
 'They saw her throwing the rock.'¹⁷
- (53) *No-'ita te tompa-'a=n(o) u Aswi.*
 3R-see CORE throw-NL=3POSS GEN Aswin
 'They saw her throwing at Aswin.'
- (54) *No-mele=ako te hu'u-'a=n(o) u ama=no nu doe.*
 3R-happy-APPL CORE give-NL=3POSS GEN father=3POSS GEN money
 'He's happy because of their giving his father some money.'

Nominalisations of this sort are clearly not restricted to being [S,A]s or [O]s, and are simply nominalisations of the verbal forms, preserving the argument structure of the verbal clause, as would be expected (see, for example, Hale & Keyser 1993).

10 Conclusions

The data above has shown that in *Tukang Besi* there is a clearly Philippine-type voice system, with the type of syntactic correlates reported for most Philippine-type languages (floated quantifiers (§8), conjunction reduction (mentioned in passing in §7; see also Donohue 1999a). In this respect, then, we have the same functional oppositions operating in *Tukang Besi* as in more 'typical' western Austronesian languages.

At the same time the superficial typology of the language is quite different from the more well described Philippine-type languages, with agreement on the verb and many indications that the language's typology has shifted to a more head-marking model. The lack of nominal-deriving morphology, and the proliferation of verbal morphology, supports this claim.

The complete loss of the classic Austronesian voice morphology in main clauses, and the loss of the use of that morphology in any position as a functioning voice system, has led to radical restructuring of the morphological appearance of *Tukang Besi*, with the addition of

¹⁷ More likely to be interpreted as 'They saw her throwing (something) at the rock.'

agreement markers to take on the role of monitors of pragmatic importance. The preservation of these old voice markers as markers of subordinate clauses only makes an interesting comment on the paths of grammaticalisation, with (presumably earlier) main clause morphology becoming restricted to subordinate clauses (a reversal of the usual trend observed in other languages in which subordinate clause morphology comes to be used in main clauses). In addition to being a reversal of the usually-observed trends cross-linguistically, it also goes against some proposed paths of grammaticalisation for earlier forms of Austronesian itself, with some writers claiming that the Austronesian main clause morphology arose from a grammaticalisation of earlier subordinate clause morphology.

References

- Andrews, Avery D., 1985, The major functions of the noun phrase. In Timothy Shopen, ed. *Language typology and syntactic description, 1: Clause structure*, 62-154. Cambridge: Cambridge University Press.
- Aissen, Judith L., 1992, Topic and focus in Mayan. *Language* 68:43-80.
- Baker, Mark C., 1988, *Incorporation: a theory of grammatical function changing*. Chicago: The University of Chicago Press.
- Bell, Sarah, 1976, Cebuano subjects in two frameworks. PhD dissertation, MIT.
- 1983, Advancements and ascensions in Cebuano. In David M. Perlmutter, ed. *Studies in relational grammar 1*: 143-218. Chicago: The University of Chicago Press.
- Berg, René van den, 1989, *A grammar of the Muna language*. Dordrecht/Providence: Foris.
- Dixon, R.M.W., 1994, *Ergativity*. Cambridge: Cambridge University Press.
- Donohue, Mark, 1994, *Tukang Besi*. *Journal of the International Phonetic Association*: 24:39-41.
- 1995, *The Tukang Besi language of southeast Sulawesi, Indonesia*. PhD thesis, The Australian National University.
- 1996a, Split intransitivity in *Tukang Besi*. *Oceanic Linguistics* 35:294-305.
- 1996b, Relative clauses in *Tukang Besi*: grammatical functions and thematic roles. *Linguistic Analysis* 26:159-73.
- 1997, Some trade languages of insular South-East Asia and Irian Jaya (including *Map 77: Tukang Besi as lingua franca*). In S.A. Wurm, Peter Mühlhäusler and Darrell T. Tryon, eds *Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas*, 713-717. Berlin: Mouton de Gruyter.
- 1998, Transitivity in *Tukang Besi*. *Studies in Language* 22:83-111.
- 1999a, *A grammar of Tukang Besi*. Berlin: Mouton de Gruyter.
- 1999b, Syntactic roles vs. semantic roles: external possession in *Tukang Besi*. In Doris Payne and Immanuel Barshi, eds *External possession*. Amsterdam: John Benjamins.
- Du Bois, John W., 1987, The discourse basis of ergativity. *Language* 63:805-55.
- Durie, Mark, 1987, Grammatical relations in Acehnese. *Studies in Language* 11:365-399.
- Foley, William A. and Robert D. Van Valin, 1984, *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

- Hale, Kenneth L. and Samuel Jay Keyser, 1993, On argument structure and the lexical expression of syntactic relations. In Kenneth L. Hale and Samuel Jay Keyser, eds *The view from Building 20: essays in linguistics in honor of Sylvain Bromberger*, 53-109. Cambridge MA: MIT Press.
- Heath, Jeffrey, 1975, Some functional relationships in grammar. *Language* 51:89-104.
- Kroeger, Paul R., 1993, *Phrase structure and grammatical relations in Tagalog*. Stanford: CSLI Publications.
- Lamontagne, Greg and Lisa Travis, 1987, The syntax of adjacency. In Megan Crowhurst, ed. *Proceedings of the West Coast Conference on Formal Linguistics*, 173-186. Stanford, California.
- Mirikitani, Leatrice T., 1972, *Kapampangan syntax*. Honolulu: University of Hawaii Press.
- Mithun, Marianne, 1994, The implications of ergativity for a Philippine voice system. In Barbara Fox and Paul J. Hopper, eds *Voice: form and function*. Amsterdam: John Benjamins.
- Sells, Peter, 1995, The function of voice markers in the Philippine languages. Paper presented at the Conference on morphology and its relation to syntax and phonology, University of California, Davis, May 1995.

A sketch of the primary transitive verbs in Pendau

PHIL QUICK

1 Introduction

Pendau, a Tomini-Tolitoli language of Central Sulawesi, has two distinct transitive verb forms, which I will refer to as the *nong*-¹ and the *ni*- forms.² This paper will present an analysis of these forms in basic clauses. Examples (1) and (2) contrast the *nong*- and the *ni*-transitive verb forms. Note that the final consonant of *nong*- varies in place of articulation, conditioned by the first phoneme of the root. When that first phoneme is a voiceless stop, it is deleted. In transcription the deleted voiceless stop is shown in parentheses; thus in (1) *nonuju*, from the root *tuju*, is transcribed *non-(t)uju*.³

- (1) *Si=ama='u non-(t)uju si=ina='u.*
PNM:I=father=I:S:II:G *nong*-send PNM:I=mother=I:S:II:G
'My father sent my mother.'

-
- ¹ The velar nasal is written orthographically as <ng>. The glottal stop is written as an apostrophe, but is not written when it appears word initially.
- ² I want to thank the seminar participants of my thesis proposal seminar (27 September 1996) for providing clarifying criticism and comments which have been helpful in the analysis of Pendau. Those who have been involved with editing and/or contributing suggestions and constructive criticism are: Mark Donohue, Bethwyn Evans, Catharina van Klinken, Nikolaus Himmelmann, Rebecca Quick, Malcolm Ross, Darrell Tryon, Fay Wouk and Chikao Yoshimura. I especially want to thank Dr Hanafie Sulaiman, for sponsoring my field research in Indonesia under the Universitas Tadulako and to thank the Indonesian Institute of Sciences (LIPI) for granting me permission to conduct research in Indonesia (1997-1998), which has allowed me to include some of the newer data in this paper. Pendau is considered to be one distinct language in the Tomini-Tolitoli grouping in Central Sulawesi, Indonesia
- ³ Abbreviations and symbols are: = clitic boundary; 1, 2, 3 first, second and third person; I NP Set I; II NP Set II; A agent; BEN benefactive; CAUS causative; CNM common noun marker; CONT continuative aspect; COMP completive aspect; G genitive; INSTR instrument; IR irrealis; LOC locative; NP noun phrase; P patient; P plural; PE plural exclusive; PI plural inclusive; PNM proper noun marker; *pV(C)*-stem formant prefix type; RED reduplication; R realis; RD referential distance; S₀ intransitive subject = transitive object; s singular; ST stative intransitive verb; TAM tense, aspect, mode; TP topic persistence; V verb; VBSR verbaliser.

Fay Wouk and Malcolm Ross, eds, *The history and typology of western Austronesian voice systems*, 101-122. Canberra: Pacific Linguistics, 2002.

© Phil Quick

101

- (2) *Si=ama='u* *ni-tuju* *ni=ina='u.*
 PNM:I=father=IS:II:G *ni-send* PNM:II=mother=IS:II:G
 'My mother sent my father.'

The verbs in the sentences presented above can be interpreted as *primary transitive verbs* (Andrews 1985). Each of the sentences has an agent (A) and a patient (P) argument.⁴

In this paper I analyse Pendau argument structure. In §2 I present the basic paradigms of nominal and verbal morphology. In §3 and §4 I discuss the use of nominal expressions with *nong-* and *ni-* verbs. Syntactic transitivity and word order variation in transitive clauses is presented in §5 and §6. Section 7 shows that the *nong-* and *ni-* verb clause constructions can each have an instrument noun phrase as a non-applicative second object (an important prelude to understanding §9 on applicatives). In order to complete the sketch of these two primary transitive verb constructions I include data that indicates that Pendau has passive voice in addition to two transitive verb constructions (§8) and a brief discussion of applicatives (§9), and of tense, aspect, and mode (TAM) (§10). Section 11 addresses the question of voice selection through discourse analysis of a folktale using topic continuity methodology as developed in Givón (1983, 1994), and provides supporting data for the claim that the *nong-* and the *ni-* verb constructions are both syntactically transitive.

2 Nominal and verbal paradigms

Pendau has two pronoun sets and two corresponding sets of noun phrase markers as seen in Figure 1. Noun phrase marking distinguishes two classes of nouns, common and proper. I will refer to the two sets as NP Set I and NP Set II. NP Set I is used for most core argument functions, while NP Set II is largely restricted to the A argument of *ni-*verbs, as will be demonstrated below. Note that when the singular Set II pronouns are cliticised to the verb, they are incorporated as part of the phonological word.

		Set I	Set II ⁵
singular	1	<i>a'u</i>	<i>=u</i> (' <i>u-</i> , <i>no'u-</i>)
	2	<i>oo</i>	<i>=mu</i> (<i>mu-</i>)
	3	<i>io</i>	<i>=nyo</i>
plural	1 inclusive	<i>ito</i>	<i>=to</i>
	1 exclusive	<i>ami</i>	<i>mami</i>
	2	<i>emu</i>	<i>miu</i>
	3	<i>jimo</i>	<i>nijimo</i>
proper nouns		<i>si=</i>	<i>ni=</i>
common nouns		<i>∅/(u=)</i>	<i>nu=</i>

Figure 1: Pronouns and noun phrase markers in Pendau

⁴ Following Dixon (1979, 1994) and Andrews (1985), agent and patient can be represented as prototypical arguments that have been symbolised as A and O respectively. Others, such as Comrie (1989), have used the same idea with the slightly different labels of A and P respectively. In this paper I will use Comrie's labels A and P to refer to the basic argument positions of transitive clauses.

⁵ The Set II pronouns also include the fronted pronouns '*u-*' and *mu-*' for first and second person respectively, effectively becoming verbal prefixes. The Set II pronouns are a mixed set; some are enclitics, and some are free words (distinguishable by phonological criteria).

Pendau verbs are marked for mode, distinguishing realis and irrealis. Examples (1) and (2) illustrate the realis forms of the two transitive clause constructions. Stative verbs take a realis prefix *no-* and an irrealis prefix *mo-*. The basic paradigm for the Pendau transitive verb forms is presented in Figure 2 with the word base *oli* 'buy', and for statives with the word base *tou* 'finish'.⁶

	Mode:	irrealis	realis
<i>nong-</i>	Verb	<i>mong-oli</i>	<i>nong-oli</i>
<i>ni-</i>	Verb	<i>ro-oli</i>	<i>ni-oli</i>
<i>no-</i>	Verb	<i>mo-tou'</i>	<i>no-tou'</i>

Figure 2: Primary transitive verb forms and the stative verb

To simplify the discussion most of the examples in this paper will make use of the realis forms. Whatever is stated for the realis verbs with regard to the argument structure can also be said of the irrealis.

3 Core nominal expressions with *nong-* verbs⁷

With *nong-* verbs, the P may be indefinite (3) or definite (4). Both the A (3) and the P (5) may be pronouns, and both may be full nouns (3).

- (3) *A'u nong-oli vea.*
 1S:I *nong-buy* rice
 'I bought rice.'
- (4) *Si=papa nong-komung vea uo.*
 PNM:I=grandpa *nong-take* rice yonder
 'Grandpa took yonder rice.'
- (5) *Io neng-ebiling a'u.*
 3S *nong-leave* 1S:I
 'He left me.' (Not: *I left him.)

Both A and P will be Set I nominals, even when both are pronouns, as can be seen by comparing examples (5) and (3). In (3) *a'u* 'I' is the agent, while in (5) it is the patient. The positions of the two Set I pronouns in example (5) could be reversed and the sentence would still be grammatical (although the meaning would change of course). Thus, in this construction grammatical role of the arguments is indicated by word order, not by form of the pronoun.

⁶ The *o* in *nong-*, *mong-* and *ro-* is analysed as the underlying vowel. The *o* fronts to an *e* or *i* in predictable environments. Stative *no-* and *mo-* have the allomorphs *na-/ma-* and *ne-/me-*, which vary according to the same vowel harmony rules.

⁷ The reader should note that there is another prefix *mepe-/nepe-* which selects the A argument as the pivot and can on some words be substituted for *mong-/nong-* and not change the meaning. While some roots can take either prefix, many roots generally occur with only one of the two prefixes. For simplicity and clarity I have not used these as examples in this sketch.

Both A and P proper nouns are preceded by the proper noun phrase marker *si* regardless of the argument position ((6) and (7)). As with pronouns, who sees whom is dictated by the argument position together with the *nong-* verbal prefix.

- (6) *Si=kai neng-ita-i si=be'e.*
 PNM:I=grandfather *nong-see-LOC* PNM:I=grandmother
 'The grandfather saw the grandmother.'
- (7) *Si=be'e neng-ita-i si=kai.*
 PNM:I=grandmother *nong-see-LOC* PNM:I=grandfather
 'The grandmother saw the grandfather.'

Common nouns are typically not preceded by a noun phrase marker. Argument structure is indicated by word order, as with pronouns and proper nouns (8 and 9). Only the noun phrase preceding the verb can be the A argument. The post-verbal argument position has to be the P argument.

- (8) *Ø=Tagu='u non-[t]uju Ø=unga='u.*
 CNM:I=friend=1S::II:G *nong-send* CNM:I=child=1S:II:G
 'My friend sent my child.'
- (9) *Ø=Unga='u non-[t]uju Ø=tagu='u.*
 CNM:I=child=1S:II:G *nong-send* CNM=friend=1S:II:G
 'My child sent my friend'

There appears to be an optional, but very rare, common noun phrase marker *u=*. In one of the few texts where it appears, it is only used three times in 238 clauses. Elicitation shows that it may precede both A and P nouns, in the position indicated as zero above ($\emptyset=$). Himmelmann (pers. comm.) suggests that it may have some kind of highlighting effect.

While Set II noun phrase constituents are never used for the A argument of the *nong-*verb, the Set II singular pronouns are occasionally found in the P argument position, as shown in (10), which may be contrasted with (11). In cases such as these, although Set II pronouns are used, they clearly represent core arguments, and the clauses are definitely transitive.

- (10) *Jimo mong-komung=onyo.*
 3P:I *nong-carry=3S:II*
 'They will carry him/her.'
- (11) *Jimo mong-komung io.*
 3P:I *nong-carry* 3S:I
 'They will carry him/her.'

This use of Set II pronouns to mark P is consistent with the occurrence of singular Set II pronouns as recipients of benefactive applicative constructions, as in (12). Since in both cases the verbs are clearly semantically transitive, we assume that syntax follows semantics, and that Set II pronouns can be used to express core arguments in these cases. A full discussion of the use of Set II noun phrases in the P argument position is outside the scope of this sketch, and is not counter-evidence to the conclusions made here.

- (12) *A'u nong-oli-a'=onyo vea.*
 1S:I *nong-buy-BEN=3S:II* rice
 'I bought rice for him.'

4 Core nominal expressions with *ni-* verbs

In contrast to the *nong-* verb constructions, the *ni-* verb constructions use both NP Set I and NP Set II to mark the P and the A arguments respectively. Note that the *ni=* proclitic is used to mark A argument proper nouns in (13) and (14) instead of the *si=* proclitic found with *nong-* verbs, and that the common nouns (13) and pronouns (14) in the P argument position are from NP Set I.

- (13) *Vea uo ni-'omung ni=papa='u.*
 rice yonder *ni*-take PNM:II=grandpa=1S:II:G
 'My grandpa took yonder rice.'

- (14) *A'u moo ni-tuju ni=papa.*
 1S:I here *ni*-send CNM:II=grandpa
 'Grandpa sent me (I that am here).'

Example (13) uses the verb base *'omung* (cf. *nong-komung* in example (4) above) 'carry, take', but has the *ni-* verb prefix. Notice that while the pre-verbal nominal in the *nong-* prefix clause is interpreted as an A, in a clause with a *ni-* verb it is interpreted as a P. The A argument cannot precede a *ni-* verb form as a free pronoun or a noun phrase, as illustrated in example (15) below.

- (15) **Ni=papa='u ni-'omung vea uo.*
 PNM:II=grandpa=1S:II:G *ni*-take rice yonder
 'My grandpa took yonder rice.'

Example (16) looks similar to example (9). The noun phrase preceding the *ni-* verb form however must be interpreted as the P argument.⁸ This contrasts with the *nong-V* clause in example (9) above where preverbal *unga'u* 'my child' must be an A argument. Note that the common noun agent in (16) is preceded by *nu=*, an NP Set II marker.⁹

- (16) *Ø=Unga='u ni-tuju nu=tagu='u.*
 CNM:I=child=1S:II:G *ni*-send CNM=friend=1S:II:G
 'My friend sent my child.'

In (17) the first person enclitic, which is the A argument of the *ni-* verb, is again from NP Set II.

- (17) *Si=papa ni-tuju='u*
 PNM:I=grandpa *ni*-send=1S:II
 'I sent Grandpa.'

In the last two sections I demonstrated that the A argument in the *ni-* verb has a unique marking, and that all other core argument positions of the *ni-* and the *nong-* verb constructions are from NP Set I. The next section will compare the *nong-* and the *ni-* verb constructions.

⁸ Note that elicitation has recently demonstrated that this position can optionally take the *u=* clitic for common nouns.

⁹ The NP Set II constituents are also used to possess the noun directly preceding the pronoun or noun phrase marker (for example see (16)) i.e. they are genitives (G).

5 Syntactic transitivity in *nong-* and *ni-* verb clauses

Not only are the *ni-* and the *nong-* verb forms semantically transitive, but they are syntactically transitive as well, although core arguments need not be overt. Each verb prefix semantically requires two different noun phrase entities for its clausal predication, and when a core argument is overtly realised it is obligatorily a member of either NP Set I or NP Set II.

Examples (18) and (19) clearly demonstrate that neither clause type contains oblique markers. Although in *ni-* verb clauses it could be argued that the *nu=* and the *ni=* are oblique markers, an analysis as markers of direct arguments is preferable, based on an understanding of the wider paradigm, and given that discourse analysis shows no significant difference in topicality between the As of the two clause types (see §10).

- (18) *Ami non-(t)uju jimo.*
 IPE:I *nong*-send 3P:I
 'We sent them.'

- (19) *Ami ni-tuju nijimo.*
 IPE:I *ni*-send 3P:II
 'They sent us.'

The *ni-* prefix is often replaced by one of three pronominal prefixes (see Figure 1): '*u-*' (1:IR), illustrated in (20), '*no'u-*' (1S:R), illustrated in (21), and '*mu-*' (2S:IR). *Ni-* marked verbs require a second formative to indicate the pronoun, but pronominally prefixed verbs collapse the pronoun and the voice into one affix. This suggests that the A of a *ni-* verb is a core argument and therefore *ni-* verb constructions are transitive.

- (20) *Oo u-raga, u-lava-i, paey u-pate-i.*
 2S:I 1S:IR-chase 1S:IR-obstruct-LOC and.then 1S:IR-kill-LOC
 'I will chase you, corner you, and then I will kill you.'

- (21) *No'u-pate-i jimo*
 1S.IV:R-kill-LOC 3P:I
 'I killed them.'

6 Word order variation of core arguments in transitive clauses

In the previous sections, all examples illustrated basic word order, AVP for *nong-* verbs, and PVA for *ni-* verbs. However, both clause types allow word order variation.

In the non-basic word order for the *ni-* verb form the P argument moves to word final position as shown in examples (22) and (23) below. It is important to note that the position of the A argument remains the same or is *rigid* when contrasting the two word order possibilities for the *ni-* verb clause construction. In contrast, the position of the P argument is flexible. I will call this the *flex* position.

- (22) *Ni-ra'op=onyo odo moo.*
ni-catch=3S:II monkey this
 'He caught this monkey.'

- (23) *Paey ni-'ito nu=too a'u.*
 and.then ni-see CNM:II=person IS:I
 'And then people saw me.'

Example (24) shows that for the *nong-* verb clause construction the P argument has the *rigid* word order position, and the A argument has the *flex* position when comparing the two word order possibilities for the *nong-* verb clause.

- (24) *Nong-komung asu jimo ono mbengi=mo ri='uo.*
nong-take dog 3P:I if night=COMP at=yonder
 'They took dogs over yonder when it was already night.'

Figure 3 compares word order possibilities for transitive clauses in Pendau. Each verb type has a *rigid* argument position that is postverbal, and each verb type has a *flex* argument position that may be either in a pre-verbal position or in a post-verbal position which must follow the *rigid* argument position. The *flex* positions are marked in bold in figure 3 by bolding the arguments which have more than one word order position. Word order variation is pragmatically conditioned, but the details of this conditioning are not relevant to the discussion in this paper.¹⁰ However, what is relevant is that this pattern suggests that both the *nong-* verb clause and the *ni-* verb clause have the same underlying word order possibilities. The *flex* position is tentatively identified as the *pivot* and the *rigid* position as the *non-pivot*. For both clause types, non-pivots must immediately follow the verb, while pivots may be clause initial or clause final.

1.	A	<i>nong-V</i>	P	
2.		<i>nong-V</i>	P	A
3.	P	<i>ni-V</i>	A	
4.		<i>ni-V</i>	A	P

Figure 3: A and P argument positions in Pendau transitive clauses

7 Instrument NP as a non-applicative second object

Both *nong-* and *ni-* verb clause constructions can have a second object, an instrument (INSTR) used by the A argument on the P argument. This third core argument is marked with the Set II common noun phrase marker *nu=*. In example (25) notice that the last word in the second clause *nu=uram* 'medicine' has the proclitic *nu*, and that the P argument 'them' is inferred from the context.

- (25) *Ila mai uo ni-alap=onyo uram,*
 from come yonder ni-get=3S:II medicine

¹⁰ Although I have not yet done a statistical analysis of correlations between the two word order possibilities and various pragmatics factors, my impression is that the pre-verbal pivot word order signifies less continuous information and the post-verbal pivot word order signifies highly continuous information.

paey ni-rembas-i=nyo nu=uram.
 and.then *ni-hit-LOC=3S:II INSTR=medicine*
 'From there then he took the MEDICINE, and then he applied the medicine on THEM.' (lit. From there then he got the MEDICINE, and then he hit THEM with the medicine.)

The preferred word order for such constructions when all arguments are overt is PVAIO, as in (26).

- (26) ...*paey unga ni-rembas-i=nyo nu=uram*
 and.then child *ni-hit-LOC=3S:II INSTR=medicine*
 '...and then he hit the CHILD with the medicine.'

A clause may have two noun phrases marked with *nu=*, but one marks the A argument and the second one marks the instrument (27).

- (27) ...*paey asu ni-rembas-i nu=unga nu=uram.*
 and.then dog *ni-hit-LOC CNM:II=child INSTR=medicine*
 '...and then the child hit the DOG with the medicine.'

The order of the A argument and the instrument can be reversed as in (28), and instrument arguments may be added to clauses exhibiting VAP order as in (29). However, elicited examples like these are not accepted by all speakers.

- (28) ...*paey unga ni-rembas-i nu=uram ni=Yusup.*
 and.then child *ni-hit-LOC INSTR=medicine PNM:II=Joseph*
 '...and then Joseph hit the CHILD with the medicine.'
- (29) ...*paey ni-rembas-i=nyo unga uo nu=uram uo.*
 and.then *ni-hit-LOC=3S:II child yonder INSTR=medicine yonder*
 '...and then he hit the CHILD with the medicine.'

When an instrument noun phrase is used in a *nong*-verb construction, the order is typically AVPIO, as in (30).

- (30) *Si=Yusup mony-(s)ambale japing uo nu=piso.*
 PNM:I=Joseph *nong-butcher cow yonder INSTR=machete*
 'JOSEPH will slaughter the cow with a knife.'

The order AVIOP, illustrated in example (31) was accepted by my language assistant during an elicitation session, but does not seem to be the preferred word order, and has not been found in any natural texts.

- (31) *Si=Yusup mony-(s)ambale nu=piso japing uo.*
 PNM:I=Joseph *nong-butcher INSTR=machete cow yonder*
 'JOSEPH will slaughter the cow with a knife.'

Many of the examples in this section were elicited based on patterns found in texts. Elicitations were collected from one speaker and then the data were checked by another Pendau speaker. There seemed to be a preferred word order, but there was some disagreement on word order when the *instrument* NP was not used in the final position. The instrument noun phrase was always rejected in any preverbal position of the clause.

In many of the cases described above, the instrumental construction could be replaced by an oblique construction using comitative *sono* 'with' rather than *nu* 'via, by means of (INSTR)' with no change of meaning. However, the final examples below show clearly the difference between an instrumental usage and the comitative usage. In (32) my language assistant laughed and said that it was grammatical, but that it meant that the glass was drunk along with the water (the exclamation mark indicates that it is grammatical, but semantically strange). He said it would be understood to mean that someone used the glass to drink with, but that the actual meaning was to drink the water and the glass (normally an impossibility). Example (33) clearly gives the appropriate instrumental use of the glass.

(32) !*Ogo ni-inung=o'u sono galas.*
 water *ni*-drink=1S:II with glass
 'I drank the WATER along with the glass.'

(33) *Ogo ni-inung=o'u nu=galas.*
 water *ni*-drink=1S:II INSTR=glass
 'I drank the WATER using a glass.'

Finally, example (34) shows that both the comitative *sono* and the instrumental *nu* can appear in the sequence of *sono nu* 'with the means of'.

(34) *Ogo ni-inung=o'u sono nu=galas.*
 water *ni*-drink=1S:II with INSTR=glass
 'I drank the WATER with the use of a glass.'

8 Stative verbs

In Pendau the stative affix occurs mainly on adjectival-type words, as in (35) and (36). Since statives are intransitive the S equals the P argument (S_0 in Dixon's terminology), and it is expressed as a Set I noun. Stative verbs can be recognised by the realis prefix *no-/na-/ne-* or the irrealis prefix *mo-/ma-/me-*, with choice of allomorph being conditioned by vowel harmony rules.

(35) *Joo jalang ri='uo ndau no-mbosi'*,
 however road at=yonder no ST:R-good
ma-rate' jalang ri='uo.
 ST:R-wicked road at=yonder
 'However that road is not good, that road is a terrible road.'

(36) *Aniong ne-riri*
 rice ST:R-yellow
 'The rice is yellow.' (OR: Yellow rice.)

Stative verbs sometimes occur in constructions that are semantically transitive, including an A argument. Syntactically they are probably intransitive, and may be best analysed as a middle voice. When this construction occurs, the A and P arguments follow the same encoding pattern as the *ni-* verbal construction, that is the P is Set I and the A is Set II, as in (37).

- (37) *Aniong no-tou' nijimo.*
 rice ST:R-finish 3P:II
 'The rice was finished by them.'

The above mentioned construction notwithstanding, roots of stative verbs are inherently intransitive, and are generally made transitive through causative operations (in which case they can occur in either a *nong-* verb or a *ni-* verb construction). However, I have found a few roots which may take either the stative verb construction (with or without a semantic agent) or the *ni-* verb construction without intermediary derivations. These include *alap* 'get, take, find', and *gansing* 'damage'. When an agent is present, there is a consistent meaning difference between the sentences with *ni-* affixed verbs and those with *no-* affixed verbs, the former implying a higher degree of actor involvement, as with (37).

Examples (38) and (39) illustrate this contrast with *alap* 'get, take, find'. Note that the Indonesian translation my language assistant provided used two different lexical words to attempt to convey the difference in meaning, *dapat* 'find, get' for the stative verb construction, and *ambil* 'take, carry' for the *ni-* verb construction.

- (38) *Bau uo na-alap=onyo.*
 fish yonder ST:R-get=3S:II
 'That fish was found by him.'
 (Indonesian: 'Ikan itu dia dapat.' lit. fish that he got)
- (39) *Bau uo ni-alap=onyo.*
 fish yonder ni-get=3S:II
 'He took that fish.'
 (Indonesian: 'Ikan itu dia ambil.' lit. fish that he took)

Examples (40) and (41) illustrate another minimal sentence pair in which my language assistant suggested the stative verb construction was unintentional whereas the *ni-* verb construction was intentional. However, since there is a nonvolitional prefix *te-* available,¹¹ I believe the difference he was attempting to communicate to me was one of passive versus "active" transitivity in which the degree of involvement of an agent would be significantly different.

- (40) *Motor=o'u na-gansing=onyo.*
 motorcycle=1S:II:G ST:R-damage=3S:II
 'My motorcycle was damaged by him.'
- (41) *Motor=o'u ni-gansing=onyo.*
 motorcycle=1S:II:G ni-damage=3S:II
 'He damaged my motorcycle'

In addition to the exact minimal pairs presented above, there are some near-minimal pairs that suggest that the stative verb construction with semantic agent really is intransitive, since the root requires the applicative locative *-i* suffix in order to appear in a *ni-* verb construction. This is contrasted in examples (42) and (43).

¹¹ The prefix *te-* can have either a non-volitional or abilitative meaning, and can be used on stative verbs, *ni-* verbs, and *nong-* verb type constructions, for example *metemeas* 'whitest', *nite'omung* 'able to carry', *mete'omung* 'happen to carry'.

- (42) *Oto='u na-rampung nu=too.*
 car=1S:1:G ST:R-burn CNM:II=person
 'My car was burned by a person (or: by someone).'
- (43) *Oto='u ni-rampun(g)-i nu=too.*
 car=1S:1:G ni-burn-LOC CNM:II=person
 'Someone burned my car.'

The following section will present more details on how the applicatives are used in primary transitive verb constructions.

9 Applicatives¹²

Pendau has two applicative suffixes which can be added to intransitives and transitives, increasing their valency. These are the benefactive *-a'* and the locative *-i*. With these constructions, the pivot noun phrase is identified by the presence of *nong-* or *ni-*, as previously described. In addition to these straightforward applicatives, there are additional *ni-* verb constructions, involving an applicative suffix plus an additional prefix, which can be used to raise an instrumental or locative noun phrase to the pivot position. Structurally Pendau is quite similar to Indonesian in many aspects of its grammar. However, the constructions which raise instrument and locative noun phrases to pivot position are an important exception to this similarity. These constructions have more in common with the 'instrument focus' and 'locative focus' constructions found in Philippine languages. Pendau 'instrument focus' is discussed below. The 'locative focus' construction will not be discussed in this paper, as I have not yet completed analysis of it. The reader should note that in this section the pivot noun phrase is indicated in the free translation with capital letters

9.1 Benefactive applicative

In clause constructions that have an inherent transitive verbal root the applicative adds a second object (or "indirect object", (IO)) as in examples (44), (45) and (46). In these cases the A=Donor, the P=Recipient and the IO=Gift (adapted from Dixon and Aikhenvald 1997). For intransitive bases the applicative usually adds only the P argument. An overt Recipient is not required provided it is recoverable from the context, as in (45) and (46). Intransitive verbs receiving the benefactive *-a'* become transitive and the direct object is the recipient of the verbal action (47). The preferred word order is: *Anong-VPIO* (as in 44) and *Pni-VAIO* (as in 48).¹³

¹² Part of this section was originally written as part of a working paper titled 'Valency changing operations in Pendau' that was presented at the *Workshop on Grammatical Categories: Valency Changing Derivations* led by Bob Dixon (1997). I am grateful for his and other participants' input.

¹³ Word order is free only in the sense that the IO can float anywhere after the verb, but the P and A must always maintain the PA sequence for the *nong-* construction and the AP sequence for the *ni-* construction when both arguments occur postverbally even when the IO comes between, before or after these two arguments.

- (44) *A'u nong-oli-a' io vea.*
 1S:I nong-buy-BEN 3S:I rice
 'I bought him rice.'
- (45) *Jimo uo ni-sambale-a'=omo manu' sensiama.*
 3P:I yonder ni-butcher-BEN=COMP chicken male
 '(They) butchered a ROOSTER for them./ (They) butchered them a ROOSTER.'
- (46) *Ni-bagi-a' ni=gibang pepitu karung moje.*
 ni-give-BEN PNM :II=water.monitor seven sack more
 'The water monitor gave him SEVEN MORE SACKS.'
- (47) *A'u no-tagu-a'=omo bau.*
 1S:I VBZR:R-friend-BEN=COMP fish
 'I now school with fish/I have become a friend of fish.'
- (48) *...paey uram uo ni-rembas-a'=onyo unga.*
 and.then medicine yonder ni-hit-BEN=3S:II child
 '...and then he applied the MEDICINE on the child.'

9.2 Locative applicative

The locative suffix *-i* has the A move towards the P or become closer to the spatial area of the P, or to do something within or approaching the spatial confines of the P. The locative *-i* has its corollary in the locative (oblique) preposition *ri* 'at, in, on, etc.'. This verbal correlation allows the agent to participate in a locative or otherwise deictic sense of the verb with the P which would only be allowable in an oblique argument. Below are typical examples:¹⁴

- (49) *Ni-pene'-i=nyo taipang uo.*
 ni-go.up-LOC=3S:II manggo.tree yonder
 'He climbed up YONDER MANGGO TREE.'
- (50) *Si=rapi=nyo langkai moo ni-pate-i nu=too na-ate.*
 PNM :I=spouse=3S:II:G male this ni-kill-LOC CNM:II=person STR-die
 'A person killed THIS MAN'S WIFE.'
- (51) *Oo 'u-lava-i, paey 'u-pate-i.*
 2S:I ni(1S:IR :II)-corner-LOC then ni(1S:IR :II)-kill-LOC
 'I will corner you and then I will kill you.'

The main difference between *-i* and *-a'* is that the applicative *-i* does not add a second object in non-derived transitive clauses. The locative *-i* could be interpreted as allowing an oblique NP to become a core argument of a transitive clause.

¹⁴ The *-i* locative suffix appears to have derivational and inflectional uses, but neither of these seem to be predictable. Compare for instance: *mongkomung* 'carry' and *mongkomuni* 'touch'.

9.3 Instrumental applicative

The instrument noun phrase can become the selected argument of the clause via applicativisation (compare this to the Philippine ‘instrument focus’). When the instrument noun phrase becomes the pivot through applicativisation it is not usually indicated by a case marker, as it is when the instrument noun phrase is not the pivot (see §7). Its status as pivot is indicated by word order (i.e. it must be in the P argument position), and is morphologically marked on the verb by the combination of a *pV(C)-* prefix and the benefactive *-a’*. The form of the prefix varies according to which root is taking instrument focus, depending on the verb class of the root, as illustrated in (52) and (53).¹⁵

- (52) *Baliung=o’u mu-po-gutu-a’=omo piso.*
 axe=1S:G:II *ni*(2S:IR:II)-INSTR-make-BEN=COMP machete
 ‘You make machetes for me by using my AXE.’ (A blacksmith makes
 machetes from the axe by forging).’
- (53) *Doi’ mo-luar ro-pong-oli-a’ nijimo gulang o pita nilon.*
 money *si*:IR-want *ni*-INSTR-buy-BEN 3P:II rope and ribbon nylon
 ‘They wanted to buy rope and fishing line with (their) MONEY.’

This construction may be used even if the pivot instrument noun phrase is not overt, as long as it is recoverable from the context as in (54). The noun *doi’* ‘money’ is given in parentheses because, although it did not occur in the original text, it was implied by the context and elicitation showed that it could occur. It is placed in preverbal position because overt instrument NP pivots are pre-verbal, while the recipient of the instrument occurs in postverbal position after the A argument as in (55). This is similar to the preferred word order of benefactives in the *ni-* verb construction illustrated above (48).

- (54) *(Doi’) u-pong-ongkos-a’ unga=u*
 (money) *ni*(1S:IR:II)-INSTR-cost-BEN child=1S:II:G
 ‘I will pay for my child by using this MONEY.’
- (55) *Piso uo ni-pony-(s)ambale-a’ ni=Yusup japing uo.*
 machete yonder *ni*-INSTR-butcher-BEN PNM:II=Joseph cow yonder
 ‘Joseph used the MACHETE to butcher the cow.’

The instrument case marker *nu=* may appear optionally (at least in an elicitation session), but only when the instrument NP is postverbal, as in (56).¹⁶ However, my language assistant suggested it was better without the *nu* particle.

- (56) *Ni-pong-kolog-a’ ni=Desmon ayu uo (nu)=sensar uo*
ni-INSTR-cut-BEN PNM=Desmon wood yonder (INSTR)=chainsaw yonder
 ‘Desmon used a CHAINSAW to cut that wood.’

¹⁵ However, note that some verbs have a stem which already fills the prefix prerequisite, e.g. ‘*gabu*, *pogabu*’ ‘cook’, and therefore it is only the benefactive suffix that is added, which turns the verb construction into instrument focus in combination with the appropriate word order.

¹⁶ This suggests that when the non-pivot instrument noun phrase marker is used it is a core argument.

When the derivational causative prefix is present, as in (57), the Instrument focus prefix is absent. Instrument focus is indicated by a combination of the causative prefix, the benefactive suffix *-a'*, and word order, as in (57).

- (57) *Upang uo ni-pa-inan(g)-a'=oto bau uo.*
 bait yonder *ni-CAUS:INSTR-eat-BEN=2P:II* fish yonder
 'We used the BAIT to feed the fish.'

Reciprocal and instrumental affixes can combine in the same verb, as in (58).

- (58) *Ogo uo ni-posi-pon-(t)uan(g)-a' nijimo api uo.*
 water yonder *ni-REC-INSTR-pour-BEN 3P:II* fire yonder
 'Together they poured WATER on the fire.' (It is implied that water is taken from one place or container.)

10 Tense, aspect, and mode on primary transitive verbs

The use of tense, aspect, and mode (TAM) in Pendau is a very complicated subject since the formatives which mark them often interact with other formatives or are fused into one formative. At least one of these formatives is always present in primary transitive verbs, while others appear optionally.

Mode is indicated, as described above, by the form of the verbal affix. The alternation between the two modes has temporal implications; alternation between realis and irrealis modes can be roughly categorised as distinguishing past and non-past tense for transitive verbs, although it is better to view the distinction in non-tense terms as contrasting actual and non-actual events.

Aspect in Pendau can be marked by one of two clitics,¹⁷ the completive clitic *=mo* and the continuative/sequential clitic *=po*. Both verbal and non-verbal clauses can be marked for aspect, as *=mo* and *=po* can cliticise to non-verbal as well as verbal roots. Completive *=mo* can be roughly interpreted as non-future, while *=po* is roughly interpreted as future.

The temporal difference between the two clitics can be clearly seen in the following pairs of non-verbs: *ndau=po* 'not yet', *ndau=mo* 'not now', *oo=po* 'your turn' (lit. 'you again'), *oo=mo* 'your turn is over' (lit. 'you already'), *ruo-mbengi=po* 'the day after tomorrow' (lit. 'two nights more'), *ruo-mbengi=mo* 'the day before yesterday' (lit. 'two nights past'), *na-sae=mo* 'when, how long has it been?', *ma-sae=po* 'when, how much longer?'. In the last pair mode interacts with aspect to achieve the meaning; the first member of the pair being realis and the second irrealis.

Figure 4 shows how tense correlates¹⁸ with the aspectual markers *=po* and *=mo*, and shows how tense correlates¹⁹ with realis and irrealis modes (the 'X' on the time line in figure 4 indicates the present).²⁰ Notice that there is a mismatch where the 'X' (present tense) occurs

¹⁷ These are not the only aspectual markers, but these are the two that can intersect with irrealis/realis to specify events on the temporal continuum. Other aspects marked in Pendau include iterative/durative, telic/inchoative, and plural distributive aspect.

¹⁸ This correlation is only one component of the aspectual markers.

¹⁹ This correlation is only one component of the modality marking.

²⁰ The *mong-* and *nong-* affixes represent all irrealis and realis affixes in Figure 4, so include for example the *ro-/ni-* affixes respectively.

above and below the line. This temporal mismatch allows four different combinations of mode and aspect to intersect (see §1.0.2 for the summary of these combinations as well as their interpretations in Figure 5).

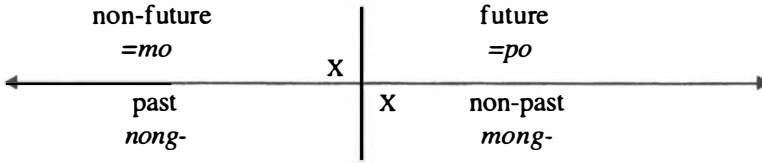


Figure 4: Correlation of tense with aspect and mode in Pendau

10.1 Interactions of mode and aspect

Either clitic can occur with a verb in either mode, allowing for four different shades of meaning. These are presented in §10.1.1–4 and summarised in §10.2.

10.1.1 Irrealis with =mo

When the completive enclitic =mo is combined with a verb in irrealis mode, the event described is non-future and non-past, thus forcing the clause to be understood as a present immediate act, as in (59).

- (59) *Oo moo ra-pate-i=mo mami.*
 2S:I this ni-kill-LOC=COMP 1PE:II
 'We (monkeys) are going to kill you (turtle) now!'

10.1.2 Realis with =mo

When =mo is combined with a realis mode verbal affix, the action is generally interpreted as a completive punctual past event, as in (60) and (61).

- (60) *Jimo uo ni-sambale-a'=omo manu' sensiama.*
 3P:I yonder ni-butcher-BEN=COMP chicken male
 '(They) butchered a rooster for them./ (They) butchered them a rooster.'
- (61) *Ila uo nerema mai nong-komung=omo asu jimo.*
 from yonder daylight come nong-carry=COMP dog 3P:I
 'When daylight came they took the dogs (to go hunting).'

However, sometimes the same combination can be used for a hypothetical situation in the future. In (62) it serves to emphasise the inevitability of the consequence, should the antecedent come to pass (i.e. it is functioning as relative tense where the locus of the 'present' tense is in the future).

- (62) *"Nyaa ro-bolilo asi a'u moo saba' ni-bolilo=mo*
 don't ni-club too.bad 1S:I here because ni-club=COMP

paey te'e=u neng-eng-ke-kee' moo."
 and.then back=1S:II:G *nong-RED?-RED-bumps* here
 "Please don't club me, because after I am clubbed then my back will be all
 bumpy!" (turtle responding to the monkeys)

10.1.3 *Irrealis with =po*

When *=po* is combined with an irrealis mode verb, as in (63) and (64), it indicates a future event which may require some elapsed time to reach its endpoint (present tense is specifically excluded).

- (63) *Ila mai moo re-pe-lampa-i=po lima-mbengi.*
 from come this ni(IRR)-CAUS-go-LOC=CONT five-night
 'From here it takes five days to get there.'
- (64) *Alea' be'e, ami meng-kan(g)-i=po loka lulu.*
 allow grandma 1PE:I nong(IRR)-eat-LOC=CONT banana first
 'Grandma, let us still eat bananas first.'

10.1.4 *Realis with =po*

It is also possible for *=po* to combine with a realis mode verb, as in (65), although this is not a frequent occurrence. It is difficult to capture the aspectual distinctions that are implied by this construction in the English translation. The events cited in example (65) below have already happened in the narrative, and are here reported in a way that emphasises their sequentiality.

- (65) *Panganganta moo na-ndasa=mo ni-rasa-i=nyo apa*
 man.eater this ST:R-critical=COMP ni-feel=LOC=3S:II because
no-'orop=omo, ni-tinjung=opo nu=tatambuang,
 ST:R-hungry=COMP ni-sting=CONT CNM:II=bumblebee
ni-ti'ap=opo nu=saa.
ni-bite=CONT CNM:II=python
 'The man-eater was already feeling critically injured (or: mortally wounded),
 because he was hungry, and then he was stung by bumblebees, and then bitten
 by a python.'

10.2 A summary of TAM in Pendau

Figure 5 shows the four possible TAM combinations that result from the interaction of the verbal prefixes and aspectual clitics.²¹

²¹ Kroeger (1993:15-17) shows a similar matrix for Tagalog with similar parameters, although the cells are interpreted differently.

	= <i>mo</i>	= <i>po</i>	
<i>nong-</i>	1	2	realis
<i>mong-</i>	3	4	irrealis
	completive	continuative	

Figure 5: Intersection of tense, aspect, and mode

The values of the four cells in Figure 5 are given in the list below (matrix numbers match the list numbers):

1. past completive (excludes present)
2. past sequential/continuative
3. present (excludes past and future)
4. future sequential/continuative (excludes past and present)

11 Selection criteria and transitivity evidence from discourse based quantification

In this section I will examine discourse data from one folktale ('The turtle and the monkey who were friends' by Josep Piri in Doge et al. 1992) and I will explore two problems. First, is the *ni-* verb construction really a syntactically transitive clause? Secondly, can the choice between a *ni-* verb and a *nong-* verb be predicted based on discourse level information, and if so what are the parameters? These two problems will be explored utilising Givón's topic continuity methodology (Givón 1994). Givón (1994:10) states:

These methods are based on the assumption that more topical, (thematically important) referents tend to be both more *anaphorically accessible* ('continuous') and more *cataphorically persistent* ('recurrent'). Neither measure assesses topicality directly. Rather, they measure the referential continuity properties of referents, in two — opposite — textual directions. It is assumed then that the two measures should correlate with the two respective cognitive dimensions of topicality.

Topic continuity analysis measures the frequency of occurrence of nominal arguments that are tracked in core argument positions. This is done by examining each core argument of each transitive clause in a text and counting "back" to find a match (called referential distance) and counting "forward" to find a match (called topic persistence). Whenever a match is made the distance of the span is noted and classed according to its span grouping or class. Givón (1994) has refined the topic continuity methodology so that counting is restricted to three clauses to the left and ten clauses to the right. If no matches are made for referential distance spans within three clauses to the left than the noun argument is automatically assigned the span "greater than three". If no matches are made within ten clauses to the right for topic persistence, then nothing is registered for any of those ten span classes.

Referential distance (RD) has three span classes used in the tally: (1) distance of one, (2) distance of two or three, and (3) greater than three. Topic persistence (TP) has ten span classes, one class for each of the first ten clauses following a clause. The tallying works differently depending on the direction of the count. For referential distance, the counting stops when the first match is encountered or after three clauses have been examined and no match is reached. For topic persistence, a tally is noted each time a referent is matched for each of the following ten clauses for the respective span distance (i.e. a maximum of one is added to a particular span class).

The composed folktale examined has a total of 108 transitive clauses with 66 *ni*- clauses and 42 *nong*- clauses.²² Table 1 shows the referential distance (RD) values for *ni*- and *nong*-verb constructions for the clause distances of 1-3 (topical) and greater than 3 clauses (not topical).

Table 1: Referential distance values

RD	<i>ni</i> - P	<i>ni</i> - A	<i>nong</i> - P	<i>nong</i> - A
1-3	29 (58%)	42 (82%)	9 (33%)	29 (69%)
>3	21 (42%)	9 (18%)	18 (67%)	13 (31%)
Total	50 (100%)	51 (100%)	27 (100%)	42 (100%)

Figure 6 shows a scatter-plot display for the RD for *ni*- and *nong*- verb constructions when the RD is equal to one for either A or P. This shows that for both the *ni*- and *nong*- verb constructions the A is of a significantly higher topicality than the P, and in fact the A of the *ni*- clause is more likely to have been mentioned in the immediately preceding three clauses (42 tokens or 82%) than any other argument type. If the A in the *ni*- clause was actually the oblique of a passive, then the A should be expected to be much lower in topicality than the P, as well as lower than the A of the *nong*- verb construction. What appears dramatically here is that the A of the *ni*- verb is actually 13% higher in topicality than the A of the *nong*- verb. Figure 6 also illustrates that the RD of the P in the *ni*- verb construction is significantly higher than the RD of the P in the *nong*- verb construction. This is what I would expect from a transitive verb construction which “focuses” on the P argument. Referential distance data in table 1 and figure 6 support the claim that the *ni*- verb construction is a syntactically transitive clause.

²² The statistics were calculated with a new computer program I have been developing called *Multilinear Discourse Analysis* (Beta 97.8.15). This software analyses specially tagged language databases such as *Shoebox*, in which additional fields for each clause are added (cf. Quick 1996). These fields contain succinct information about each participant for each clause. This information is then used to do discourse quantification among other things.

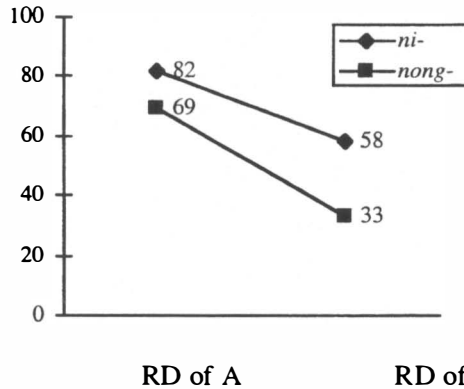


Figure 6: Percentages of *ni-* and *nong-* verb constructions with RD = 1-3 for P and A

Figure 7 shows that for gaps greater than three clauses (i.e. discontinuous topics as contrasted to uninterrupted topics in figure 6) the P is not very topical for both *ni-* (42%) and for *nong-* (67%) verb constructions in comparison to the A for each type. Comparing figure 6 and figure 7 suggests that A arguments generally occur in “runs”, i.e. the same participant is frequently a topical A for several continuous clauses, whereas the P in both verb constructions may or may not be continuous.

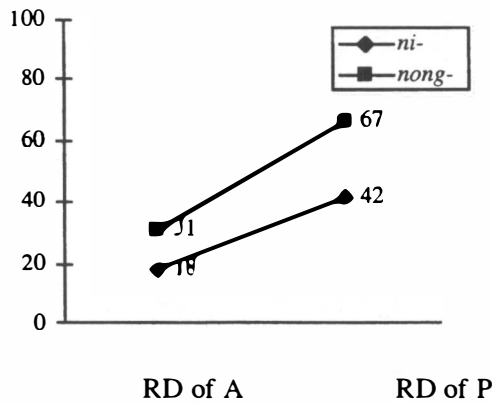


Figure 7: Percentages of *ni-* and *nong-* verb constructions with RD > 3 for P and A

Table 2 shows that the A of *ni-* verb constructions is highly persistent with 77% of these tokens persisting for more than three clauses (this is similar to the high RD in Table 1 of the A in *ni-* clauses). Topic persistence corroborates the conclusion from the referential distance above; the high topicality of the A in *ni-* clauses indicates that this is a syntactically transitive clause. Also notice that the A of the *nong-* verb constructions at 74% is nearly as highly persistent as the A of the *ni-* verb construction (see Table 2).

Table 2: Topic persistence values

TP	<i>ni</i> - P	<i>ni</i> - A	<i>nong</i> - P	<i>nong</i> - A
1:	10	10	4	3
2:	7	4	2	8
3:	3	6	4	7
4:	0	8	2	3
5:	0	10	0	1
6:	0	11	2	5
7:	5	5	1	6
8:	9	5	1	4
9:	3	2	1	5
10:	0	0	0	0
1-2:	17 (46%)	14 (23%)	6 (35%)	11 (26%)
3-10:	20 (54%)	47 (77%)	11 (65%)	31 (74%)
TOTAL:	37	61	17	42

Figure 8 below shows in a graph form the topic persistence of A and P arguments for both *ni*- and *nong*- verb clauses for topic persistence greater than two clauses. Notice that the lines both descend, but that there is a greater difference between the TP of the A and P of the *ni*- verb construction (about 23%) than for the TP of the A and P of the *nong*- verb construction (about 9%). Thus, by this measure also, Ps are somewhat less topical than As, but are nonetheless relatively topical, which is what I would expect from transitive clauses as opposed to say a passive verb construction. However, by this measure, the P of the *nong*- verbs is more topical than the P of the *ni*- verbs, the reverse of the results for RD. Thus when the two measures are considered together, it does not appear that they can shed any light on the choice between the two verb constructions.

The statistics for TP and RD do not seem to provide a clue to selection of choice between these two transitive clauses, but they do characterise a similarity between these two verb constructions, i.e. that they are both syntactically transitive clauses.

This section has demonstrated that Givón's method does not seem to be helpful in determining selection criteria between the *ni*- verb and the *nong*- verb constructions, but that it does show that there are strong similarities between the two verb constructions that would indicate they are both transitive verb constructions. A secondary methodology based on Givón's topic continuity methodology was used by Dryer (1994). This method was followed and the preliminary statistics from this one text have not provided clear results that can be presented here. Dryer's method may yet prove to be a valuable method that gets closer to the heart of the selection criteria (as he meant it to be), but more texts need to be analysed using this methodology to see if consistent results are obtained, and whether any clear selection parameters emerge that can be used to predict which type of verb will be chosen. Future research using discourse analysis methodology on a large number of texts should eventually yield the parameters that will provide a higher degree of predictability in the choice made between these two verb constructions.

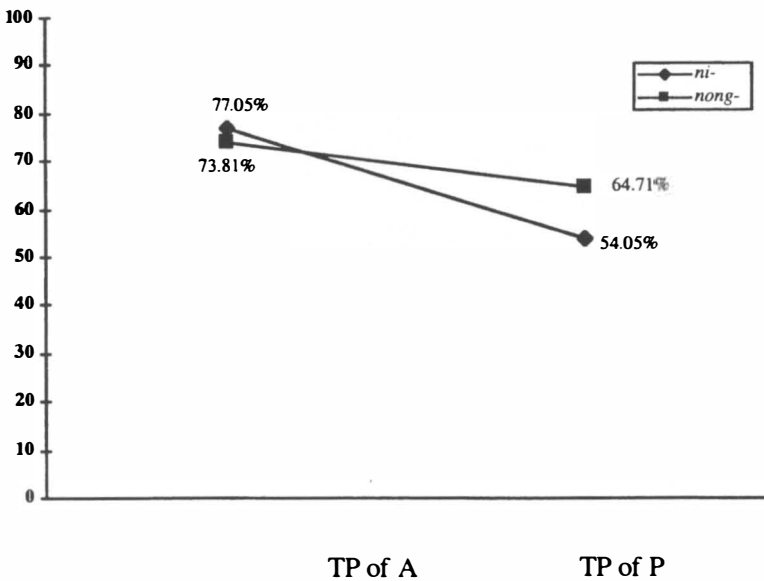


Figure 8: Topic Persistence of A > 2 and P > 2

12 Conclusion

The Pendau data and discussion can be summarised as below:

Pendau has two transitive verb forms distinguished by *nong-* and *ni-* prefixes. These verb constructions both have A and P arguments.

- nong-* AVP or VPA word order
 - NP Set I is used in A and P positions
 - NP Set II is rarely used in the P position
- ni-* PVA or VAP word order [also: PA-V and A-VP]
 - NP Set I is used in the P position
 - NP Set II is used in the A position

In the *ni-* verb construction the A is in the non-pivot position, and the P is in the pivot position. In the *nong-* verb construction the A is in the pivot position, and the P is in the non-pivot position. Initial research indicates that the Instrument NP, the Indirect Object, and Obliques may appear in several other positions, but the relative sequence of the A and P for each of the verb constructions must not be violated. Criteria for the choice between *ni-* and *nong-* verb cannot be determined as of yet.

All the topics touched on in this paper obviously need further research. A discourse study should show exactly what the pragmatic difference is between the different pivot positions (i.e. preverbal versus postverbal). Syntactic tests, including further study of relative clauses, to provide evidence for which position is the pivot or non-pivot will also be important. Further study employing discourse quantitative analytical techniques should yield further insights and/or evidence for establishing the identity and/or function of the NP Set II. Especially

crucial to furthering an understanding of the data presented in this paper will be studying eligibility criteria for choosing between the *ni-* and *nong-* transitive verbs (see Quick 1997 for a preliminary view), and whether one is derived from the other or whether they are both basic, as this sketch seems to indicate.

References

- Andrews, Avery, 1985, The major functions of the noun phrase. In Timothy Shopen, ed. *Language typology and syntactic description, 1: Clause structure*, 62-154. Cambridge: Cambridge University Press.
- Comrie, Bernard, 1989, *Language universals and linguistic typology*. (Second edition) Chicago: University of Chicago Press.
- Dixon, R.M.W., 1979, Ergativity. *Language*, 55:59-138.
- 1994, *Ergativity*. Cambridge: Cambridge University Press.
- Dixon, R.M.W. and Alexandra Y. Aikhenvald, 1997, A typology of argument-determined constructions. In John Haiman, Joan Bybee and Sandra Thompson, eds *Essays on language function and language type, dedicated to T. Givón*, 71-113. Amsterdam: John Benjamins.
- Doge, Mesak, Josep Piri, Philip Quick and Rebecca Quick, eds, 1992, *Sompulu Uma-umanong*. (*Ten Folktales* in Pendau). Manuscript, Palu, Indonesia.
- Dryer, Matthew, 1994, The discourse function of the Kutenai inverse. In Talmy Givón, ed. 1994:65-99.
- Givón, Talmy, ed., 1983, *Topic continuity in discourse: a quantitative cross-language study*. Amsterdam: John Benjamins.
- 1994, *Voice and inversion*. Amsterdam: John Benjamins.
- Kroeger, Paul, 1993, *Phrase structure and grammatical relations in Tagalog*. Stanford: CSLI Publications.
- Quick, Philip, 1996, Multilinear discourse analysis software demonstration. In H. Andrew Black, Alan Buseman, David Payne and Gary F. Simons, eds *Proceedings of the 1996 general CARLA conference, November 14-15, 1996*, 291-309. Waxhaw, NC/Dallas: JAARS and Summer Institute of Linguistics.
- 1997, Active and inverse voice selection criteria in Pendau, a western Austronesian language. In *Proceedings of the Seventh International Conference on Austronesian Linguistics*, 461-482. Amsterdam/Atlanta: Editions Rodopi B.V.

Voice in two northern Sulawesi languages

NIKOLAUS P. HIMMELMANN

1 Introduction

In this paper, the basic clause structure and some aspects of the verbal morphology of two northern Sulawesi languages, Ratahan and Lauje, are compared with each other and with Tagalog, the language chosen to represent a prototypical Philippine language (which, in fact, it is only with some qualifications). The paper is intended to contribute to the typology of voice-related phenomena in western Austronesian languages by providing primarily descriptive information on two languages for which such information has so far not been available. It should be noted, however, that the grammatical analysis of the two languages is still in its initial stages. Hence, it is almost certain that the information provided here is incomplete, if not erroneous in some respects.

Apart from providing descriptive data, the paper also aims to make some headway in developing a typological grid for western Austronesian languages. Thus, the comparison between the three languages is not limited to the features which come readily to mind when comparing voice-related phenomena in these languages, in particular the basic clause structure (§2), the voice morphology (§3), the paradigms for noun phrase markers (§4) and pronouns (§5), and pronominal prefixes (§6). Instead, it also comprises phenomena which to date have received little attention but may prove to be of major importance for a more elaborate and fine-grained grammatical typology of western Austronesian languages. Among such phenomena are applicative suffixes (§7), the use of voice-marked forms in non-predicative functions, in particular in construction with quantifiers (§8), and voice in stative derivations (§9).

All of these phenomena are further defined and exemplified in the following sections. Their potential typological relevance is discussed in §10, which also provides a summary of the comparison between Tagalog, Ratahan and Lauje.

Ratahan belongs to the Sangiric group of languages and is spoken southwest of the Minahasan area in the province of North Sulawesi. The discussion is based on fieldwork data collected in collaboration with John Wolff in 1996. Lauje belongs to the Tomini-Tolitoli group of languages and is spoken in northern Central Sulawesi. Lauje examples are from my fieldwork data collected in 1988 and 1993. See Himmelmann and Wolff (in prep.) and

Himmelmann (2001) for more background information and full acknowledgments. Tagalog examples are taken from Bloomfield (1917) or my own recordings.

A conscious (and at times quite time consuming) effort has been made to illustrate all grammatical phenomena with examples drawn from realistic spontaneous speech (conversations or narratives). These examples are presented in intonation units, each unit being delimited by a pair of curly brackets (numbers in curly brackets show pause length in seconds, though pause length has not been measured in all instances). In the few instances in which elicited data are given, the available spontaneous data would have required lengthy comments in order to serve as evidence for a given observation.

2 Basic clause structure

In Ratahan, in both transitive as well as intransitive clauses, there is always one core argument (the single core argument in the case of intransitive verbs) which is characterised by three features not shared by any other core argument: (a) its position is variable, i.e. it may occur in pre- or post-verbal position; (b) its semantic role is marked by the voice affix on the verb; (c) it is never marked by any kind of phrase marker.¹ This argument is called the *subject* of a verbal clause. The following two intransitive clauses exemplify these characteristics:²

- (1) Ratahan uc_tw 121
hairen araq kumukuk e manuk {}
 hairen araq um-kukuk Ce manuk
 later if ACT-cry_out COMPL chicken
 'later when the chicken cried out'
- (2) Ratahan uf2_tw 127
araq isé kumukuk e {1.2}
 araq isé um-kukuk Ce
 if 3SG ACT-cry_out COMPL
 'when he would cry out'

The following two examples, in which the verb is marked for undergoer voice, illustrate the same characteristics for transitive clauses:

- (3) Ratahan uc_ew 217
naq sawu ntoo Sepus winuno mangasé {}
 naq sawu N-to Sepus in-wuno mangasé
 ?? time LK-NR Joseph REAL(UG)-kill 3PL
 'at the time they killed Joseph'

¹ Proper nouns occur with the proper noun article in this position (see §4).

² Abbreviations used: ACT actor; AND andative; APPL applicative; CAUS causative; COMPL completive; DAT dative; DET determiner; DIST distal; EX exclusive; GEN genitive; GER gerund; IMP imperative; IRR irrealis; LK linker; LOC locative; NR nominaliser; PL plural; PM predicate marker; PN proper noun; PROX proximal; REAL realis; RED reduplication; RPRT reportative; SG singular; SPEC specific (article); STAT stative; UG undergoer; UG.L undergoer.locative; UG.T undergoer.theme.

- (4) Ratahan uc_tw 583
kinambei ne yaq { }
 in-kambei ne yaq
 REAL(UG)-embrace 3SG.GEN 1SG
 'he embraced me'

In transitive clauses, the non-subject core argument always follows immediately after the verb. If it is a common noun, it may optionally be marked with the genitive marker *nu* (for proper nouns the marker is *ni*). In examples (3) and (4) the non-subject argument is a pronoun (*mangasé* and *ne*, respectively). Examples with full lexical NPs for both actor and undergoer voices:

- (5) Ratahan uf2_tw 113
mamake nu babuq mangasé {0.9}
 maN-pake nu babuq mangasé
 ACT-use GEN slave 3PL
 'they had (female) slaves'
- (6) Ratahan et_up 013
niraes e nu waa wusak.
 ni-raes Ce nu waa wusak
 REAL(UG)-finish COMPL GEN monkey banana
 'The monkey had eaten all the bananas.'

For Lauje essentially the same observations hold. There is, however, one major distinction with regard to the use of NP-markers with core arguments, as discussed in §4.

In both Ratahan and Lauje it is very common for core arguments to follow the verb, which is also the most common order in Tagalog. But there is a crucial difference among the three languages with respect to examples in which the subject precedes the verb. Compare the Ratahan examples (2) and (3) above and the following example from Lauje:

- (7) Lauje UF_LM 150
bangkola' ngintugome galaa. {0.4}
 bangkola' N-intug-O-me galaa
 monkey REAL.ACT-look_for-O-COMPL turtle
 'The monkey looked for the turtle.'

There is no element corresponding to the Tagalog inversion marker *ay* in these examples. There is also no prosodic break between the clause-initial subject and the following verb. Thus these examples show that SV(X) is indeed an alternative basic word order in Ratahan and Lauje.

The examples of SV(X) order should be strictly distinguished from topicalised constituents, i.e. constituents which are separated from the following segment by an intonational boundary, as in:³

³ Capital letters in Lauje examples (primarily <E> but also <A>) represent paragogic vowels. These vowels occur fairly regularly at the end of phonological words (if consonant-final) but may also have some morphosyntactic function. See Himmelmann (1997) for further discussion.

- (8) Lauje UF_LD 029f
too ntoilapi injéine {}
 too N-toilapi injéine
 person LK-married_couple PROX

na'agaad E unga {}
 no-'o-gaad E unga
 STAT.REAL-??-take E child
 'this couple, (they) had a child.'

Such topicalised constituents are probably possible in all western Austronesian languages, with or without special markers such as Tagalog *ay*. Note that even in Tagalog the use of *ay* is not obligatory:

- (9) Tagalog donat 288
yung Eta namán {}
 iyón:LK Ayta namán
 DIST:LK Ayta truly

takot na takot sa iyó {}
 takot na takot sa iyo
 fear LK fear LOC 2SG.DAT
 'those Ayta, (they) were very afraid of you'

A prosodic break is sufficient to separate a topicalised constituent from the remainder of the clause. 'True' SV(X) order, on the other hand, is impossible in Tagalog and many other Philippine-type languages.

3 Basic voice paradigms

In Ratahan, as in Tagalog, the basic voice paradigm in declarative main clauses consists of four voices and two moods. There is, however, a conspicuous difference with regard to the morphological formatives used to express the voice/mood distinctions. While Tagalog has special formatives for each of the three undergoer voices (i.e. the suffixes *-in* and *-an* and the prefix *i-*), Ratahan has only one formative which exclusively occurs in the undergoer voices, i.e. the suffix *-an*. Compare Tables 1 and 2.

Table 1: Basic voice-mood paradigm in Ratahan

	+REALIS	-REALIS
ACTOR	<i>-im-/N-</i>	<i>-um-/M-</i>
UNDERGOER	<i>-in-</i>	<i>-an</i>
UNDERGOER.LOCATIVE	<i>-in--an</i>	<i>-an</i>
UNDERGOER.THEME	<i>-in-</i>	\emptyset

Table 2: Basic voice-mood paradigm in Tagalog

	+REALIS	-REALIS
ACTOR	-um-/N-	-um-/M-
UNDERGOER	-in-	-in
UNDERGOER.LOCATIVE	-in--an	-an
UNDERGOER.THEME	i--in-	i-

The three undergoer voices in Ratahan can only be distinguished based on the paradigmatic interaction with the realis-marking formative *-in-*. Thus, in undergoer voice there is no suffix in realis contexts (see (3) above) but the suffix *-an* is used in non-realistic contexts:

- (10) Ratahan uc_tw 762
kau wunón masé { }
 kau wuno-an masé
 2.SG kill-UG 3PL
 'they will kill you, no doubt about it'

In locative undergoer voice, on the other hand, the suffix *-an* occurs in both realistic and non-realistic contexts:

- (11) Ratahan et_ww G16a
laliwuqan le yaq mangewong buk tiqí!
 laliwuq-an le yaq maN-ewong buk tiqí
 help-UG.L IMP 1SG ACT-carry book PROX
 'Help me to carry these books!'

- (12) Ratahan et_ww G16b
isé nilaliwuqan ku.
 isé ni-laliwuq-an ku
 3SG REAL(UG)-help-UG.L 1SG.GEN
 'I helped him.'

In thematic undergoer voice, there is no formative for voice. In realistic contexts, the verb is marked for realistic mood by the infix *-in-*:

- (13) Ratahan et_ww G2
sabel nu winei ku si tuari nu
 sable nu in-wei ku si tuari nu
 machete 2SG.GEN REAL(UG)-give 1SG.GEN PN.LOC younger_sibling 2SG.GEN
 'I gave your machete to your younger sibling'

In non-realistic contexts, the verb remains unmarked, i.e. the non-realistic form of the thematic undergoer voice is identical with the basic form of the verb:

- (14) Ratahan et_up 067
wias té wei sisé taawi.
 wias teqé wei si =isé taawi
 rice DIST give PN.LOC=3SG tomorrow
 'That rice will be given to him tomorrow.'

In Ratahan, then, voice and mood marking are even more intricately intertwined than in Tagalog and other Philippine-type languages. The formal differences between Tagalog and Ratahan are paralleled by semantic ones in that in Ratahan many more verbal roots occur with thematic undergoer voice than in Tagalog. In fact, the thematic undergoer voice in Ratahan appears to be well on its way to becoming not only the formally but also the semantically unmarked undergoer voice (see Himmelmann & Wolff in prep. for more discussion).

The voice-mood paradigm found in Lauje is very different from the one found in Ratahan and Tagalog. To begin with, there are only three basic voices in the Lauje paradigm, as shown by Table 3.

Table 3: Basic voice-mood paradigm in Lauje

	+REALIS	-REALIS
ACTOR	<i>N-/(-um-)</i>	<i>M-/(-um-)</i>
UNDERGOER	<i>-in-</i>	<i>no-</i>
UNDERGOER.LOCATIVE	<i>-in--ang</i>	<i>no--i</i>

Furthermore, locative undergoer voice is marked by two different suffixes, i.e. *-ang* in realis mood and *-i* in non-realistic mood:

- (15) Lauje UF_LM 139
tinambunangonyome *nu* *petu* *batanganonye* {0.9}
 in-tambung-ang-O-nye-me *nu* *petu* *batangan-O-nye*
 REAL(UG)-pile-UG.L-O-3SG.GEN-COMPL GEN soil body-O-3SG.GEN
 'she covered her body with soil'
- (16) Lauje UF_LD 264
inyaa nrape'i *a'e* { }
inyaa no-rape'-i *a'e*
 don't IRR(UG)-close_by-UGL 1SG
 'don't get closer to me'

Finally, possibly the most remarkable feature of the Lauje voice paradigm is the prefix *no-* which occurs in the non-realistic forms of the undergoer voices (with the exception of the first person singular, see §6). This feature is remarkable on two counts. First, in typical Philippine-type languages such as Tagalog non-realistic mood is always morphologically unmarked. Second, while a prefix marking non-realistic mood in undergoer voice occurs in several Sulawesi languages (see van den Berg 1996), Lauje and its closest Tomini relatives (Dondo, Tialo and Ampibabo) are, to my knowledge, unique with regard to the segmental shape of this prefix. In southern Tomini languages (Tajio, Taje, Dampelas and Pendau) the functionally equivalent prefix has the shape *ro-* or *ho-* (which in some of the languages undergoes vowel harmonic alternations) and in Kaili-Pamona languages it generally has the shape *ra-*. Apart from its uniqueness, the Lauje form *no-* is also highly conspicuous and somewhat confusing in that *n-*initial formatives in western Austronesian languages generally signal realistic mood (see for example, Tagalog *nag-*, *na-*, *naka-*, *naki-* etc.).

4 NP markers

As in Tagalog, noun phrase markers in Ratahan and Lauje come in two sets, one for common nouns and one for proper nouns. The following discussion pertains primarily to common noun phrase markers. Table 4 lists the forms. Optional markers appear in parentheses.

Table 4: NP-markers

	TAGALOG		RATAHAN		LAUJE	
	common noun	proper noun	common noun	proper noun	common noun	proper noun
ARTICLE	<i>ang</i>	<i>si</i>	–	<i>i</i>	(<i>nu</i> ₁)	<i>si</i>
GEN	<i>ng</i>	<i>ni</i>	(<i>nu</i>)	<i>ni</i>	<i>nu</i> ₂	<i>ni</i>
LOC	<i>sa</i>	<i>kay</i>	<i>su</i>	<i>si</i>	<i>li</i>	–

The most obvious difference between Tagalog and the two Sulawesi languages pertains to the fact that in the Sulawesi languages there is no element corresponding to the specific article *ang* for common nouns. Consequently, common noun phrase subjects in both Ratahan and Lauje may occur without a noun phrase marker, as amply illustrated by the examples in the previous and following sections.

In Ratahan, furthermore, the non-subject argument in both actor- and undergoer-oriented constructions may also remain unmarked. Compare the following two examples with examples (5) and (6) above:

- (17) Ratahan tw 009
ngo: yaq mangewong e sinapang { }
 ngo yaq maN-qewong Ce sinapang
 like 1SG ACT-carry COMPL weapon
 'I'll take the gun'
- (18) Ratahan uc_tw 022
nilaweqan burung yaq {1.3}
 ni-laweq-an burung yaq
 REAL(UG)-lie_in_wait_for-LOC bird 1SG
 'a bird lay in wait for me'

In this regard, Ratahan differs sharply from Tagalog where the non-subject arguments in these constructions must be marked as genitives.

However, the differences between Ratahan and Tagalog are to a certain degree 'superficial' in that they only pertain to the overt marking of core arguments. The basic distinctions made within the noun phrase marking systems are the same: subjects exhibit coding properties different from all other core arguments; actors in undergoer-oriented constructions are coded as genitives; undergoers in actor-oriented constructions are also coded as genitives (if the undergoer is a proper noun it is marked as oblique); recipients and addressees are coded as obliques in both actor- and undergoer-oriented constructions.

The use of the Lauje noun phrase markers, on the other hand, differs quite distinctly from the one found in Tagalog and Ratahan, despite the fact that essentially the same labels may be used in describing them and that the Lauje formatives are cognate with the Ratahan ones, with the exception of the oblique marker *li*. There are two major differences: first, undergoers

in actor-oriented constructions are not marked as genitives (by the genitive marker *nu*₂) but rather receive the same marking as subjects (i.e. they may optionally be marked by *nu*₁, an article-like element). The difference between the two markers of the shape *nu* is further discussed below. Second, recipients and addressees in ditransitive constructions are not marked as obliques but generally remain unmarked.

Evidence for the second difference is relatively straightforward. In Ratahan (and in Tagalog), the additional core argument in ditransitive constructions is marked by the general purpose preposition *su* (unless it is made the subject of the construction):

- (19) Ratahan et_ww G2
tali nu winei ku su anak nu.
 tali nu in-wei ku su anak nu
 rope 2SG.GEN REAL(UG)-give 1SG.GEN LOC child 2SG.GEN
 'I gave your rope to your child'

In Lauje, it remains unmarked:

- (20) Lauje et_mt_g 002
'alolongome binee'e unga'e.
 'alolong-O-Vme in-bee-'u unga-'u
 rope-O-2SG.GEN REAL(UG)-give-1SG.GEN child-1SG.GEN
 'I gave your rope to my child'

Whether other morphosyntactic differences correlate with this difference in noun phrase marking requires further investigation.

The first major difference in noun phrase marking between Ratahan and Lauje mentioned above is in need of a somewhat more elaborate discussion. To begin with, note that the two markers of the shape *nu* are not only distinguished with regard to their use but also differ on a purely morphological level.⁴ Thus, genitive *nu* is regularly reduced to *u* after consonant-final words, as in:

- (21) Lauje PF_DJ 022
li lalong u ogo
 li lalong nu ogo
 LOC inside GEN water
 'in the water'

No such regular alternation occurs with article-like *nu*. Furthermore, there is a special emphatic form '*onu* for article-like *nu* which is found most commonly in clause-initial position:

- (22) Lauje UF_LM 020
'onu bangkola' pogombo'onya nu (0.4) tatambuang E
 'onu bangkola' po-gombo'-O-nye:A nu tatambuang E
 DET monkey GER-word-O-3SG.GEN:A DET bee E
 'as for the monkeys, they talked to the bees'

No such emphatic form exists for genitive *nu*.

⁴ The following account is simplified in that it ignores several variations which occur in spontaneous discourse and often render the two markers indistinguishable. See Yoshimura (in prep.) for a detailed account of essentially the same phenomenon in Tialo, Lauje's northern neighbour.

Article-like *nu* may be used to mark subjects (and topicalised constituents), as in:

- (23) Lauje UF_LM 122
'uminangma nu:: galaa {0.6}
 um-'inang-O-me:A nu galaa
 ACT-eat-O-COMPL:A DET turtle
 'the turtle already started eating'
- (24) Lauje UF_LM 059
ame goong A nu tatambuang E {0.3}
 ame goong A nu tatambuang E
 like gong A DET bee E
 'the bees were (arranged) like a gong.'

This use of article-like *nu* is optional,⁵ as shown by (7) above and:

- (25) Lauje PF_DJ 017
nomeelu A tuping: { }
 noN-peelu A tuping
 REAL.ACT-inform A eel
 'the eel said.'

Article-like *nu* is used not only with subjects but also with undergoers in actor-oriented constructions, as in:

- (26) Lauje et_kl_e 006
tuá'ie nonobong E nu kopi
 tuai-'u noN-tobong E nu kopi
 younger_sibling-1SG.GEN REAL.ACT-cut E DET coffee
 'My younger sibling chopped down the coffee tree (that I had planted).'

Again, this usage is optional and, in fact, rather infrequent in spontaneous discourse. Typically, undergoers in actor-oriented constructions remain unmarked, as in (7) above and:

- (27) Lauje UF_LM 078
li'ó mogutu manisanonya o'e
 li'oe M-po-gutu manisan-O-nye:A o'e
 you ACT-??-make honey-O-3.SG.GEN:A DIST
 'You make that honey'

Actors in undergoer-oriented constructions, on the other hand, are marked as genitives:

- (28) Lauje UF_LM 089
'ininang u pani'i goot E {0.5}
 in-'inang nu pani'i goot E
 REAL(UG)-eat GEN bat many E
 '(these bananas) had been eaten by many bats.'

⁵ The fact that article-like *nu* is optional is the major reason for calling it 'article-like' rather than an article. I take it as a defining characteristic of definite and specific articles that they have to be used whenever a nominal expression is definite or specific. The overall distribution of article-like *nu* is very patchy, with some speakers using it fairly consistently for marking subjects, but others hardly using it at all.

To sum up the presentation so far, there are two important morphosyntactic differences between Lauje and Ratahan which are related to article-like *nu*. First, common noun phrase subjects in Ratahan are never accompanied by a noun phrase marker while common noun phrase subjects in Lauje may optionally be preceded by article-like *nu*. Second, and more importantly, common noun phrase undergoers in actor-oriented constructions may optionally be marked as genitives in Ratahan, but in Lauje they may optionally receive the same kind of marking as subjects, i.e. article-like *nu*.

Note that this difference in noun phrase marking is not mirrored by differences with regard to other aspects of the morphosyntax of core arguments in the two languages. As mentioned in §2 above, subjects in both Lauje and Ratahan are characterised by the fact that their position with respect to the verb is variable and that their semantic role is marked by the voice affix on the verb. Non-subject core arguments in both languages, including undergoers in actor-oriented constructions, have to occur immediately after the verb.

Still, the different noun phrase marking in Lauje is highly significant in that it leads to a change in the markedness relation between actor- and undergoer-oriented constructions. In Ratahan (and in Tagalog) actor- and undergoer-oriented constructions are almost perfect mirror-images of each other, as indicated by the following structural patterns:

- (29) actor-oriented construction: NP_{ACT} ACT-Verb GEN-NP_{UG}
 undergoer-oriented construction: NP_{UG} UG-Verb GEN-NP_{ACT}

There is no direct evidence for deriving one construction from the other. (Note that this state of affairs only holds for clauses in which the non-subject argument is a common noun. If the non-subject argument in an actor-oriented construction is a pronoun or a proper noun, it is marked as oblique rather than as genitive. Pronouns are further discussed in the following section.)

For Lauje, on the other hand, it may be argued that the actor-oriented construction is slightly less marked than the undergoer-oriented construction (in terms of morphosyntactic marking). In any event, the two constructions are no longer perfect mirror-images of each other:

- (30) actor-oriented construction: (DET=)NP_{ACT} ACT-Verb (DET=)NP_{UG}
 undergoer-oriented construction: (DET=)NP_{UG} UG-Verb GEN-NP_{ACT}

That is, in Lauje the morphosyntactic properties of the two major types of non-subject core arguments (undergoers in actor-oriented constructions and actors in undergoer-oriented constructions) are not identical. They share the fixed post-verbal position, but they differ with regard to noun phrase marking (in a rather subtle way, since the marker in both cases has the basic shape *nu*).

5 Pronoun sets

Compared to Tagalog, there are fewer pronoun sets in Ratahan and Lauje, namely two instead of three (for Ratahan plural pronouns there is, in fact, only one set). The distribution of one of the two sets is adequately covered by the label *genitive* (it is used for actors in undergoer-oriented constructions and for possessors). The other set appears in a broad variety of syntactic functions. Since the use of any case label such as *absolute* or *nominative* as a label for this set is prone to lead to misconceptions, the rather nondescript label *free* is chosen here to refer to this set.

Table 5: Ratahan pronouns

	SG		PL
	FREE	GEN	FREE
1INCL	<i>yaq</i>	<i>ku</i>	(i) <i>kite</i>
EXCL			(i) <i>kami</i>
2	(i) <i>kau</i>	<i>nu</i>	(i) <i>kumú</i>
3	<i>isé</i>	<i>ne</i>	(i) <i>mangasé</i>

Table 6: Lauje pronouns

	SG		PL	
	FREE	GEN	FREE	GEN
1INCL	(li)a'e	-u	(li)'ite	-te
EXCL			(li)'ame	mame
2	(li)'oe	-Vme	(li)'eme	mie
3	(li)io	-nye	jimo	(nu jimo)

The differences between Tagalog and the two Sulawesi languages are not restricted to the inventory of forms. A much more important difference pertains to the fact that Ratahan and Lauje pronouns, unlike Tagalog subject and genitive pronouns, are *not* second-position clitics. The genitive pronouns in Ratahan and Lauje are either suffixes or enclitics, occurring always immediately after their heads (the verb when they are used to express actors in undergoer-oriented constructions, the possessed noun when they express possessors).

Free pronouns in Ratahan and Lauje occur in almost exactly the same positions as full lexical NPs. In particular, they may be used in preverbal position (see, for example, (2) and (49)) and they follow the non-subject argument in postverbal position:

- (31) Ratahan uc_tw 022
nilaweqan *burung yaq* {1.3}
 ni-laweq-an burung yaq
 REAL(UG)-intercept-UG.L bird 1SG
 'a bird lay in wait for me'

- (32) Lauje UF_LD 179
nongkoni *bisa io juniaa ine* {}
 noN-'oni bisa io juniaa ine
 REAL.ACT-carry poison 3SG world PROX
 'she poisoned this world (lit. she brought poison to this world)'

In Lauje, free pronouns may also be used as non-subject arguments in actor-oriented constructions:

- (33) Lauje PF_DJ 038
li'ó bela jeiang E {} *monyampuat a'e baoanomá'* {}
 li'oe bela jeiang E moN-sampuat a'e báoan-O-ma'e
 you buddy friend E ACT-save 1SG land-O-AND
 'you my friend, will rescue me by getting me ashore.'

In Ratahan, however, free pronouns cannot be used as non-subject arguments in actor-oriented constructions. Instead, free pronouns must be marked with the locative proper noun marker *si* when referring to the non-subject argument of an actor-oriented construction:

- (34) Ratahan et_ww N22
apa to namatakuq si kau?
apa to naN-pa-takuq si kau
 what NR REAL.ACT-CAUS-afraid PN.LOC 2SG
 'what frightened you?'

- (35) Ratahan uc_tw 077
nunongkoliwu si yaq {}
nu-nongko-liwu si yaq
 REAL.ACT-?-crowd_around PN.LOC 1SG
 '(they) crowded around me'

Note that this is the only way in which pronominal undergoers in actor-oriented constructions may be expressed in Ratahan. In this regard, pronominal undergoers differ from common noun phrase undergoers. As mentioned in the preceding section, if the undergoer in an actor-oriented construction is a common noun phrase, it may optionally be marked as genitive. But genitive pronouns in Ratahan can never be used to express the undergoer of an actor-oriented construction.

6 Pronominal prefix

In addition to the pronouns shown in Table 6, Lauje has a single pronominal prefix 'u- for first person singular actors in transitive events. This prefix only occurs in the non-realis undergoer voices:

- (36) Lauje PF_DJ 072
moopus E 'u'inang {}
mo-opus E 'u-'inang
 STAT-done E 1SG.ACT-eat
 'since I will devour it completely'
- (37) Lauje UF_LM 187
'udendenima ine {0.5}
'u-dendeng-i-me:A ine
 1SG.ACT-hit-UG.L-COMPL:A PROX
 'I will bang it.'

This is the standard form for non-realis undergoer-oriented constructions involving a 1SG actor. That is, there are no derivations involving a prefix *no-* and a suffix *-u*. Furthermore, 'u- is in complementary distribution with the prefix *no-* which occurs in all the other non-realis forms of the undergoer voices. Thus, strictly speaking, 'u- is also part of the basic voice-mood paradigm in Lauje and should be added as an alternative prefix to the two cells containing *no-* in Table 3.

7 Applicative suffixes

There is one more feature with regard to which Lauje verbal morphology sharply diverges from Tagalog and Ratahan. Lauje has at least one applicative suffix, i.e. the suffix *-a'e*. This suffix occurs in both actor- and undergoer-oriented constructions as illustrated by the following two examples:

- (38) Lauje UF_LD 209
setela pinogutuá' balung E e'e { }
setela in-po-gutu-a'e balung E o'e
 after REAL(UG)-GER-make-APPL provisions E DIST
 'when those provisions had been made (for him)'
- (39) Lauje et_ml_g 017
tuai'e mongkonía'e lia'e.
tuai-'u moN-'oni-a'e lia'e
 younger_sibling-1SG.GEN ACT-carry-APPL 1SG
 'my brother is going to carry (it) for me.'

This applicative suffix is used consistently with verbs denoting an event which involves a beneficiary (as in the two preceding examples). No alternative non-applicative construction exists for such events because there is no preposition for beneficiaries which would allow the beneficiary to be introduced in a grammatically oblique role (in elicitation, the Indonesian *untuk* is occasionally used in this function).

The applicative suffix is in complementary distribution with the locative undergoer voice suffixes *-ang* and *-i* (there are no formatives of the shape *-ana'e*, *-ia'e*, etc.). That is, in applicative constructions the basic three voice system is reduced to a two voice system.

It is possible that there is a second applicative suffix *-i* in Lauje, which would be homophonous with the non-realis locative undergoer voice suffix *-i* (see Table 3 above). So far, however, possible examples for an applicative *-i* have only appeared in elicitation and thus could be calques from Indonesian.⁶ If there is such a suffix, its distribution would be limited due to the homophony with the voice marking suffix *-i*. For example, the form *mo-linjo'-i* 'to leave, go away from' was given as a possible non-realis actor voice plus applicative suffix derivation from the root < *linjo'* 'run' (corresponding to Indonesian *meninggalkan*). However, there is no realis counterpart to this derivation. The form *no-linjo'-i* which would be the formally correct realis derivation is, in fact, the regular non-realis locative undergoer voice form (corresponding to realis *ilinjo'ang*). Note that *nolinjo'i* is not ambiguous. It is consistently translated by Indonesian *ditinggalkan* 'be left behind'.

8 Use of voice-marked forms in non-predicative functions

Voice-marked forms in Tagalog freely occur in non-predicative functions. For example, they occur after the NP-markers *ang*, *ng* and *sa*, as in:

⁶ Note that an applicative suffix *-i* has been found to occur in non-elicited material in other Tomini-Tolitoli languages, including Totoli, Pendau, and Tialo (the latter being the immediate neighbour of Lauje).

- (40) Tagalog Bloomfield (1917:30/14)
at ang pare at siyá ay nag-hinúy
 and SPEC priest and 3SG PM REAL.ACT-wait
ng sà-sabih-in ng sundalo.
 GEN RED1-statement-UG GEN soldier
 'and the priest and he waited for what the soldier would say.'
- (41) Tagalog Bloomfield (1917:72/6)
na-kà-tanaw siyá ng bahay na ma-ilaw
 REAL.STAT-??-in_sight 3SG GEN house LK STAT-light
sa p-in-a-tù-tunguh-an ng kalabáw
 LOC ??-REAL(UG)-RED1-direction-UG.L GEN water_buffalo
 'he saw a lighted house in the direction toward which the caribou was going.'

They may also be used in construction with the existential quantifier *may* (and other quantifiers):

- (42) Tagalog pep 094
ay mayroon palang nagàalaga {0.1}
 ay may-doón palá-ng nag-RED1-alaga'
 PM EXIST-DIST.LOC so!-LK REAL.ACT-RED1-cared_for
doón sa ibun {1.3}
 doón sa ibon
 DIST.LOC LOC bird
 'there was already someone looking after those birds.'
- (43) Tagalog ahas 096
may ipàpakíta ako sa iyo {0.6}
 may i-RED1-pa-kita akó sa iyo
 EXIST UG.T-RED1-CAUS-visible 1SG LOC 2SG.DAT
 'Come here, I have something to show you.'

In Ratahan and Lauje, voice-marked forms are used primarily in predicative function. They must not be used in construction with the existential quantifier (i.e. there are no examples parallel to the Tagalog examples (42) and (43)). They may be used in argument positions of verbal predicates but such use requires nominalisation by a particle which also functions as a relative clause marker (in Ratahan this is the particle *to*, in Lauje the particle *sau*):

- (44) Ratahan uc_tw 415
waktu itu siningkapen ku e to: (0.3)
 waktu itu in-singkap-an ku Ce to
 time DIST REAL(UG)-answer-UG.L ISG.GEN COMPL NR
pinakinak nge si yaq { }
 in-pakinak ne si yaq
 REAL(UG)-ask 3SG.GEN PN.LOC 1SG
 'at the time I answered what he asked me'

- (45) Ratahan uf_tw 088f
tapi to napók tee { } nangule taa { }
 tapi to N-pa-pok teqé naN-qule ta-ná
 but NR REAL.ACT-?-cut DIST REAL.ACT-return AND-DOWN
 'But those who split them (the stones) returned'
- (46) Lauje UF_LM 040
ma'o mongintug E sau no'inang E {0.5}
 ma'o moN-intug E sau no-'inang E
 go_out ACT-look_for E NR IRR(UG)-eat E
 '(they went to the river) to look for food.'

In Ratahan, a voice-marked form nominalised by *to* may then also co-occur with the NP-markers *su* or *nu*:

- (47) Ratahan uf_tw 049
nayuma na sto kunaq (0.2) pinangaren Pinatén {0.9}
 na-yuma na su-to kunaq in-paN-aren Pinaten
 REALSTAT-arrive ?? LOC-NR like? REAL(UG)-GER-name Pinaten
 'they arrived at the place called Pinaten'

9 Voice in statives

Philippine-type languages make a basic distinction between two event types, i.e. dynamic and stative. There is no special morphological marking for dynamic events, which are characterised by the fact that they involve actors who are in full control of the event. Statives are morphologically marked by a prefix (in Tagalog *ma-* and/or *ka-*) and comprise a broad range of event classes, ranging from events in which the actor is not fully in control but just manages or happens to do something, to events which do not involve actors.

Statives allow for up to four basic voice-mood derivations, involving at least in part the same morphology as the basic voice-mood derivations for dynamic events. The details of these derivations vary quite substantially, depending on the meaning of the root (for roots denoting perceptions there are different derivational possibilities than for roots denoting emotional states, etc.).

In Ratahan, statives may be marked for voice and mood in essentially the same way as in Tagalog (there are differences with regard to many details and in particular the overall productivity and generality of these formations). For example, with roots denoting emotional states the prefix *ma-* (realis *na-*) marks the theme, i.e. the entity that is or was undergoing a given emotion:

- (48) Ratahan uf_kera 034
tomponú na-awuq e rapa {1.3}
 tomponú na-awuq Ce rapa
 turtle REAL.STAT-annoyed COMPL RPRT
 'the turtle got angry'

When the same root is marked for stative locative undergoer voice (affix combination *ka--an*), the subject is the cause of the emotion (the place at, and because of which, the emotion is felt), while the theme occurs in the genitive slot:

- (49) Ratahan et_ww N23b
yaq kinaawuqan e
yaq in-ka-awuq-an ne
 1SG REAL(UG)-STAT-annoyed-UG.L 3SG.GEN
 'he got mad at me (I became the reason/object of his anger)'

A similar contrast is expressed by these derivations with roots denoting physical states. Here, however, the subject of the stative locative undergoer derivation is an experiencer:

- (50) Ratahan et_lw 24_4
nakawus e susu.
na-kawus Ce susu
 REAL.STAT-used_up COMPL milk
 'The milk is finished/used up.'
- (51) Ratahan et_lw 24_2
kinakawusan nu susu kami
in-ka-kawus-an nu susu kami
 REAL(UG)-STAT-used_up-UG.L GEN milk 1PLEX
 'We ran out of milk/the milk ran out on us.'

As far as I can currently ascertain, Lauje does not have voice-marked stative derivations. This assessment, however, needs further investigation and testing.

10 Summary and discussion

This section presents a brief summary of the (possibly) voice-related morphosyntactic features discussed in the preceding sections, focussing on the distribution of these features among the three languages under review and putting them into a wider typological perspective. To begin with, let us briefly note some voice-related features which are shared by Tagalog, Ratahan, and Lauje:

- the voice system consists of more than two basic voices.
- all voices involve some kind of morphological marking (with one exception in Ratahan, see Table 9).
- non-subject core arguments regularly occur in immediate postverbal position.
- actors in undergoer-oriented constructions are marked as genitives.

These features have not been discussed at length in the preceding sections since the focus there was on morphosyntactic features which differ among the three languages. Still, these shared features are of major typological import because they are among those features which render the voice system in (some) western Austronesian languages remarkable when compared to the voice systems found in other linguistic areas and language families. Note, however, that though they are widely shared among western Austronesian languages, it is far from clear whether they are shared by all of them.

As for the features which are not shared by Tagalog, Ratahan, and Lauje, there are altogether eleven features which have been reviewed in the preceding sections. For six of these, Tagalog groups with Ratahan, as shown in Table 7.

Table 7: Features shared by Tagalog and Ratahan

	Tagalog	Ratahan	Lauje
second core argument in ditransitive construction marked oblique	yes	yes	no
undergoer core argument in actor-oriented constructions marked as genitive (common noun phrases only)	yes	yes	no
voice derivations for statives	yes	yes	no
applicative suffix(es)	no	no	yes
pronominal prefixes in basic voice-mood paradigm	no	no	yes
non-realis prefix	no	no	yes

From a typological point of view, the interesting question is, of course, whether some of the features in Table 7 (and in the following tables) correlate with each other in such a way that the presence of one implies the presence of another. More than just three languages would have to be considered in order to make the search for correlations a typologically interesting one. Nevertheless, some very preliminary observations and suggestions can be derived from the data presented here and the (rather scarce) comparative literature on western Austronesian languages.

For two features, i.e. the presence of pronominal prefixes and applicative suffixes, it has been suggested repeatedly that they correlate with each other. More precisely, the presence of both pronominal prefixes and applicative suffixes is held to be the crucial characteristic that distinguishes Philippine-type languages from Indonesian-type languages.⁷ And while it appears to be true that none of the generally recognised Philippine-type languages⁸ exhibits both these features, it is not true that the languages of western Indonesia all exhibit both of them. Thus, for example, there are no pronominal prefixes or proclitics in Balinese, a language which otherwise appears to share many characteristics with neighbouring languages such as Javanese and Sasak, including applicative suffixes.⁹ Still, the co-occurrence of

⁷ To date, the distinction between Philippine-type and Indonesian-type languages (these are the terms used by Wolff (1996)) has been discussed exclusively in historical terms, i.e. as innovations found in Indonesian-type languages with respect to an older, possibly Proto Austronesian or Proto Malayo-Polynesian system which is hypothesised to have been inherited more or less intact by Philippine-type languages (Wolff 1996; Zobel). Note that the historical and the typological perspectives, though overlapping to a considerable degree, differ with regard to the inferences which may be drawn from the same set of data. A particular set of features may not qualify as an innovation from a historical point of view but may still provide a useful typological grid. Here, we are only concerned with typological groupings and generalisations.

⁸ That is, in addition to all the Austronesian languages spoken in the Philippines, the languages of Sabah and Sarawak and northern Sulawesi, and Yami.

⁹ The typological (and historical) relevance of pronominal prefixes is also somewhat questionable because of the fact that there appear to be pronominal prefixes in several Formosan languages, including Paiwan and Puyuma. But then, the typological position of Formosan languages, though often subsumed without much discussion under the Philippine type, is far from clear.

applicative suffixes with pronominal prefixes is so pervasive in western Indonesia that it surely constitutes a strong typological tendency.

A correlation of similar strength exists, I suspect, between the presence of applicative suffixes and the lack of voice derivations for statives. Rather than having two partially separate voice paradigms (one for dynamic predicates and one for stative predicates), it appears to be the case that in some languages of western Indonesia at least, stative morphology (such as Indonesian *ter-* and Balinese *ka-*) is in complementary distribution with voice marking morphology. However, this topic is hardly ever addressed in the literature dealing with the voice morphology in languages of western Indonesia, and thus certainly needs further study before any serious typological claims can be advanced.

The search for further correlations for two of the remaining features mentioned in Table 7 will be limited by the fact that the morphological category in question does not exist in many languages of western Indonesia. Most of these languages do not mark modal and/or aspectual distinctions on the verb. Thus, the question of whether or not there is a non-realis prefix in the basic voice paradigm does not arise. Similarly, there is no genitive marker in many of these languages. Thus, again, the question of whether or not the undergoer in actor-oriented construction is marked as a genitive does not arise. But in the latter instance it is still possible and useful to investigate the fundamental issue, i.e. whether actors in undergoer-oriented constructions exhibit essentially the same morphosyntactic features as undergoers in actor-oriented constructions (see (29) and (30) above).

Finally, it may be noteworthy that of the six features mentioned in Table 7 four pertain to verbal morphology. This is in contrast with the four features shown in Table 8, three of which pertain to the morphosyntax of nominal expressions. These features are shared between the two Sulawesi languages and distinguish them from Tagalog and other Philippine-type languages which resemble Tagalog.

Table 8: Features shared by Ratahan and Lauje

	Tagalog	Ratahan	Lauje
no (or optional) NP-marker for subject	no	yes	yes
subject may occur in immediate preverbal position	no	yes	yes
pronouns are second position clitics	yes	no	no
voice-mood marked derivations can be used with quantifiers	yes	no	no

If Ratahan is taken to be a Philippine-type language – and the features relating to verbal morphology in Table 7 strongly suggest that – then the distribution of the features in Table 8 shows that certain very typical features of nominal and pronominal expressions in Tagalog and most other languages spoken in the Philippines are not necessary correlates of Philippine-type voice morphology. Or, perhaps more productively, it suggests a division of Philippine-type languages into two subtypes: those which are similar to Tagalog in that all noun phrases are marked with some kind of marker, pronouns typically occur in clitic positions, and some kind of prosodic break or inversion marker has to be used when a subject precedes the predicate; and those which are similar to Ratahan in that subjects freely occur in pre-predicate position, pronouns are not second position clitics, and noun phrase marking is either

optional or non-existent. From Clayre's (1996) survey of the languages of Sabah and Sarawak it appears that these languages pattern with Ratahan with regard to these features.

The last feature mentioned in Table 8 — that voice-mood marked derivations can be used with quantifiers — is of interest in that it provides some indication of the degree to which a systematic distinction can be made between nouns and verbs in a given language. The noun-verb distinction in western Austronesian languages is, in general, less clearly developed than in the more familiar European languages. However, these languages certainly differ to the extent to which semantically nominal and semantically verbal expressions may occur in the same morphosyntactic contexts and hence can be more or less sharply distinguished from each other.

For Tagalog, it has been argued that there are actually no morphosyntactic contexts which provide a basis for distinguishing between nouns and verbs (understood as primary morphosyntactic form classes).¹⁰ The fact that in Tagalog putative verbs (i.e. words which express actions and are marked for voice, aspect and mood) may be used with quantifiers without undergoing further derivation (see examples (42) and (43) above) is perhaps the strongest argument for this view. Indonesian-type languages, on the other hand, appear to be much more restrictive with regard to the possibility of directly quantifying voice-marked derivations. Thus, in (standard) Indonesian it is not possible to say **ada menjaga anak-anak* for 'someone is looking after the children'. Instead, the phrase *menjaga anak-anak* has to be nominalised with *yang* when it is to be used as the argument of the existential operator *ada* (thus *ada yang menjaga anak-anak* is fine). Unfortunately, for most western Austronesian languages no reliable information is available for this potentially very important typological parameter.

To conclude this discussion of the morphosyntactic features not shared between Tagalog, Ratahan, and Lauje, we may note that there is one feature which is found only in Ratahan (Table 9).

Table 9: Feature unique to Ratahan

	Tagalog	Ratahan	Lauje
morphologically unmarked form as part of the basic voice paradigm	no	yes	no

For Philippine-type languages, the occurrence of a morphologically unmarked form as part of the basic voice paradigm may appear to be unusual. Ratahan, however, is not the only Philippine-type language in which this phenomenon is found. According to Table 14 in Clayre (1996:75), it is also widespread among the Philippine-type languages of Sarawak and Arms (1996) reports it for Sindangan Subanen, a language spoken in Western Mindanao. Note also that the thematic undergoer prefix *i-* is quite often dropped in Tagalog spontaneous discourse (and probably in many other Philippine-type languages as well):

¹⁰ See Himmelmann (1991) and Gil (1993) for further discussion and references.

- (52) Tagalog donat 273
 lalatag mo doón
 i-RED1-latag mo doón
 UG.T-RED1-spread_out 2SG.GEN DIST.LOC
 'you'd spread it (the sack) out there'

Thus, I would surmise that the occurrence of a morphologically unmarked form as part of the basic voice paradigm is not a feature of particular typological import. This does not preclude the possibility that it is of major historical import (the occurrence of a morphologically unmarked form may lead to a restructuring of the whole paradigm, possibly with further consequences for the overall system of grammatical relations).

References

- Arms, D.G., 1996, Categories of the Sindangan Subanen verb. In Steinhauer, ed. 1996: 89-114.
- Berg, René, van den, 1996, The demise of focus and the spread of conjugated verbs in Sulawesi. In Steinhauer, ed. 1996:89-114.
- Bloomfield, Leonard, 1917, *Tagalog texts with grammatical analysis*. Urbana: The University of Illinois.
- Clayre, Beatrice, 1996, The changing face of focus in the languages of Borneo. In Steinhauer, ed. 1996:51-88.
- Gil, David, 1993, Tagalog Semantics. *Berkeley Linguistics Society* 19:390-403.
- Himmelmann, Nikolaus P., 1991, *The Philippine challenge to universal grammar*. Arbeitspapier Nr. 15, Köln: Institut für Sprachwissenschaft.
- 1996, Person Marking and Grammatical Relations in Sulawesi. In Steinhauer, ed., 1996:115-136.
- 1997, The paragogic vowel in Lauje (Tomini-Tolitoli): phonology or morphosyntax? In Cecilia Odé and Wim Stokhof, eds *Proceedings of the Seventh International Conference on Austronesian Linguistics*, 81-103. Amsterdam: Rodopi.
- Himmelmann, Nikolaus P., compiler, 2001, *Sourcebook on Tomini-Tolitoli languages*. Canberra: Pacific Linguistics.
- Himmelmann, Nikolaus P. and John U. Wolff, in prep., *Ratahan morphosyntax*. München: Lincom Europa.
- Steinhauer, Hein, ed., 1996, *Papers in Austronesian linguistics No.3*. Canberra: Pacific Linguistics.
- Wolff, John U., 1996, The development of the passive verb with pronominal prefix in Western Austronesian languages. In Bernd Nothofer, ed. *Reconstruction, classification, description: festschrift in honor of Isidore Dyen*, 15-40. Hamburg: Abera-Verlag.
- Yoshimura, Chikao, in prep., A descriptive grammar of Tialo. Manuscript, The Australian National University, Canberra.

Proto Celebic focus revisited

DAVID MEAD

1 Introduction¹

The title of this paper is aptly appended with the word ‘revisited’, because throughout the following discussion I am indebted to René van den Berg and his pioneering investigation of focus in languages of central and south-eastern Sulawesi. In his paper, ‘The demise of focus and the spread of conjugated verbs in Sulawesi’ (originally presented 1991, published 1996) van den Berg both set forth arguments for recognising a Celebic macrogroup — comprising the Kaili-Pamona, Bungku-Tolaki and (the then) Muna-Buton language families — and within that framework discussed the nature of the Proto Celebic focus system and, by way of illustration, its development in seven daughter languages. As the title of his paper indicates, conjugated verbs — verbs with agent prefixes historically related to possessive enclitics — played a crucial role. They originally appeared as irrealis counterparts to realis goal focus verbs with **in*, and later spread to other (non-goal focus) verb forms.

The results of my own investigations, however, which have primarily been in the Bungku-Tolaki languages of southeastern Sulawesi, have led me to question some of van den Berg’s conclusions. There is, for example, little doubt in my mind that his Proto Celebic focus system accurately represents the source from which Kaili-Pamona languages sprang. However the basis of languages further to the south-east — namely Bungku-Tolaki and Muna-Buton — is significantly different. Not only have these languages preserved a living distinction between Proto Malayo-Polynesian **maN-* and **um*, but the role of conjugated verbs within the system of verb inflection is so disparate when compared to Kaili-Pamona — conjugated verbs are anything **but** goal focus forms in these languages — that one must question: what innovations exactly do all these languages share that they should be subgrouped together?

¹ My thanks go foremost to René van den Berg, without whose inspiration and feedback this paper would have been a much more formidable task. I am also indebted to Nikolaus Himmelmann for his comments on an earlier draft of this paper, as well as to Michael Martens for his participation in a running conversation concerning both the Uma language as well as Kaili-Pamona languages in general. My data on Tolaki was collected in 1989 under sponsorship of the then UNHAS-SIL Cooperative Program. Sponsorship by the Indonesian Institute of Sciences (LIPI) in 1996 and a grant from Rice University allowed me to collect the Kulisusu data presented herein.

I begin first by noting in §2 certain revisions to the notion of 'Celebic' which have resulted from improved understanding of sound change in this part of Sulawesi. In §3 I present my view of Proto Kaili-Pamona verbal inflection (only slightly modified from the Proto Celebic system of van den Berg 1996) and briefly outline its development in three daughter languages. In §4 I do the same for Proto Bungku-Tolaki, again tracing developments into three present-day languages. With an understanding of both protosystems in hand, in §5 I bring their differences to the fore. Although both may be derived from a system like Proto Malayo-Polynesian (PMP), the place of innovated conjugated verbs within the respective systems is very different. Implications and directions for further research are discussed in §6.

2 Changes to the notion 'Celebic' as a genetic group

When van den Berg first presented his article, he included under Celebic two established microgroups, Kaili-Pamona and Bungku-Tolaki, and one putative group, Muna-Buton. The most important revision since that time was made by Donohue (in press), who deduced from patterns of historical sound change that Wotu, Wolio and three other languages of the previously assumed Muna-Buton group constitute their own subgroup. Furthermore this deconstruction allows a close relationship to be recognised between the remaining languages Muna, Pancana, etc. (that is, the reformulated 'Muna-Buton' family, this term to be used henceforth only in this sense) and the Bungku-Tolaki languages. Both groups share in the lowering of PMP final **-iq > *e* (with **u* unaffected by a following **-q*); the raising of pretonic **a > *o*; the unconditioned split of **s > *s, *h*; and the subsequent merger of **Z* and remaining **s > *s* (van den Berg 1991b; Mead 1998). Of the four language groups under consideration, namely Kaili-Pamona, Wotu-Wolio, Bungku-Tolaki and Muna-Buton,² this constitutes at present the only incontrovertible subgrouping argument among them. In the same article mentioned above, Donohue also suggested (in a footnote) that it may be possible to relate Wotu-Wolio and Kaili-Pamona by virtue of the sporadic change of PMP **e > *a*. Some examples which I have gleaned include: Uma *mo'eta*, Pamona *maeta*, Wotu, Kamaru, Wolio *maeta* 'black' < PMP **qitem*; Ledo, Da'a, Pamona, Wolio *-aka* 'causative suffix' < **-ake(n)*, Uma *taja*, Pamona *mataja* 'sharp', Wolio *tadami* 'sharpen to a point' < PMP **taZem*. However, as there are notable exceptions (e.g. Ledo, Da'a, Pamona *aono*, Uma *ono*, Badaic *ini*, Wotu, Laiyolo, Wolio, Kamaru *ana* 'six' < PMP **enem*), this phenomenon requires further study.

² Comprising the following languages:

Kaili-Pamona: *Northern*: Kaili (including Da'a, Ledo, Kulawi, Lindu), Pamona; *Southern*: Uma, Rampi, Badaic (Napu, Bada, Besoa) (Martens 1989b)

Wotu-Wolio: Wotu, Laiyolo, Kalao, Wolio, Kamaru (Donohue forthcoming)

Bungku-Tolaki: *Eastern*: Moronene, East Coast (Kulisusu, Koroni, Taloki, Wawonii, Bungku, Bahonsuai, Mori Bawah); *Western*: Interior (Mori Atas, Tomadino, Padoe); West Coast (Tolaki, Waru, Rahambuu, Kodeoha) (Mead 1998)

Muna-Buton: *Munan*: Busoa, Munic (Muna, Pancana, Liabuka, Kaimbulawa); *Buton*: West Buton (Cia-Cia, Masiri, Island Cia-Cia), East Buton (Lasalimu, Kumbewaha) (Donohue in press)

No subgrouping hypothesis within Wotu-Wolio has yet been advanced. Martens (1997, pers. comm.) has suggested to me that the Badaic languages (Napu, Bada and Besoa) may perhaps be better regarded as a sister group of Kaili-Pamona rather than being subsumed thereunder.

At the same time the case for Celebic as a genetic subgroup (see van den Berg 1996:94 for a list of subgrouping arguments) has been weakened. In particular, we now know that Proto Bungku-Tolaki retained **R* in medial and final position, as well as reflexes of almost all PMP consonants in final position (Mead 1996). Therefore **R* > \emptyset and final consonant loss can no longer be cited as instances of shared innovations. By my reckoning this leaves three possible sound changes by which Celebic may yet be established: (a) the merger of PMP **ay* and **ey* as **e*; (b) the monophthongisation of PMP **aw* as **o*; and (c) as Donohue has suggested, PMP **d* > **r*. However, as these changes are found with some frequency in other parts of Austronesia — these three changes also characterise Oceanic languages, for instance — they constitute a weak subgrouping argument.

3 Kaili-Pamona

The following is the focus system which is reconstructible for Proto Kaili-Pamona (PKP). It is in essence the same system which van den Berg proposed for his Proto Celebic, with one notable revision. Where he proposed three classes of verbs depending on which of three prefixes, **ma-*, **me-* and **mo-*, a verb was marked with — sometimes accompanied by prenasalisation, and possibly correlating with transitivity — I have found it imperative to extract out of this the prefix **maN-* (abstractly **um* + **paN-*) which was definitely accompanied by prenasalisation, and definitely ‘transitive’.³ As this is the sole transitive prefix — other prefixes being intransitive or stative — van den Berg’s failure to make this recognition must be regarded as his only oversight in an otherwise effective presentation of KP languages. I leave open the question of how many intransitive prefixes need to be reconstructed. The verbal inflection system of PKP is given in Chart 1.

	Realis	Irrealis
Stative and intransitive	<i>*na-V</i> <i>*ne-V</i> <i>*no-V</i> etc.	<i>*ma-V</i> <i>*me-V</i> <i>*mo-V</i> etc.
Transitive actor focus	<i>*naN-V</i>	<i>*maN-V</i>
Transitive goal focus	<i>*ni-V</i>	Set A + <i>*V</i>

Chart 1: Proto Kaili-Pamona verbal inflection

³ Comparative evidence suggests that in Proto Kaili-Pamona the nasal coda was realised as /ŋ/ before vowel initial stems, and as prenasalisation of a following **p*, **t*, **k* or **s* (preceding other consonants it was realised as zero or possibly as /ŋ/ followed by an epenthetic vowel /a/). In regard to the transitive nature of this form, compare Barr’s treatment of the Da’a prefixes *naN-/neN-/noN-* — “the forms with *N*, nasal, add the component of explicit transitivity, indicating the presence of an object” (1988a:19) — and Adriani’s discussion of the Pamona transitive prefix *ma-I* (1931:172ff.).

This system distinguished two moods, realis and irrealis, and two foci, actor focus and goal focus. An account of verb forms must also be paired with a description of pronoun sets, of which I reconstruct the following for Proto Kaili-Pamona:⁴

Table 1: Proto Kaili-Pamona pronoun sets

	Set A	Set P	Set F
1SG	* <i>ku-</i>	*- <i>ku</i>	* <i>i-aku</i>
2SG	* <i>mu-</i>	*- <i>mu</i>	* <i>i-ko</i>
3SG	* <i>na-</i>	*- <i>nya</i>	* <i>si-ia</i>
1PL.INC	* <i>ta-</i>	*- <i>ta</i>	* <i>i-kita</i>
1PL.EXC	* <i>ki-</i>	*- <i>mami/-kami</i>	* <i>i-kami</i>
2PL	* <i>mi-</i>	*- <i>miu</i>	* <i>i-komiu</i>
3PL	* <i>ra-</i>	*- <i>ra</i>	* <i>si-ira</i>

Set A clitic pronouns marked the agent of irrealis goal focus verbs, and were obligatory. Set P pronouns, in addition to their usual function of indicating the possessor in possession phrases, were also used to mark the agent of a realis goal focus verb. In other words, both Set A and Set P can be regarded as non-topic pronouns. Set F pronouns were used in other contexts, i.e. they were topic pronouns, used to mark focussed patients, focussed agents, and the subject of intransitive verbs.⁵ Set F pronouns were free forms which had alternate short forms (that is, without the **i-/si-* formative) in postverbal position. In many places a split along these very lines has continued with the short, postverbal pronouns undergoing further stages of grammaticalisation (phonological attrition, obligatorification, etc.). However as some languages do not show evidence of such a split, we must assume that in Proto Kaili-Pamona the stage was set for grammaticalisation, but where it has occurred it has been a post-PKP process.

This reconstruction requires little explanation, and is in fact very similar to the system found presently in Kulawi (van den Berg 1996:99–101; Wolff 1996:28–29). Set P pronouns were used with goal focus realis forms because **ni-* forms arose from what were originally nominalisations (see among others Pawley & Reid 1980; Starosta, Pawley & Reid 1982; Ross 1995). The development of Set A pronouns — from the fronting of Set P pronouns in goal-focus contexts — has been discussed elsewhere, particularly by Wolff (1996), and although confirmatory evidence has yet to be adduced for each step, there is nothing in these languages themselves to suggest Wolff's hypothesis is incorrect.

⁴ The labels Set A, Set P and Set F follow van den Berg (1996), who proposed these labels as a means of discussing form separately from function. These letters, of course, were not chosen randomly; the letter A can be thought of as mnemonic for 'agent', the letter P for 'possessive', and the letter F for 'free'. In the subsections devoted to individual present-day languages, however, I depart from van den Berg's convention, and employ instead the labels presented and used by the authors who have written about these languages.

⁵ In this paper I do not make a distinction between 'topic' and 'focus(ed constituent)'. Unlike in Philippine languages, in general Kaili-Pamona languages do not allow locatives, instruments, etc. to be focussed. Subjects of intransitive verbs could also be regarded as topics, albeit by default, as they are the only focusable constituent.

Certain developments which have occurred in the post-PKP period are illustrated briefly here with data from three present-day languages: Da'a, Uma, and Pamona (see also van den Berg 1996).

3.1 Da'a

Da'a is one of seven closely related languages or dialects collectively known as Kaili, spoken in the area of the Palu River in western Central Sulawesi. Pronoun sets in Da'a are, following Barr (1988a:39):

Table 2: Da'a pronoun sets

	Non-focussed actor proclitics	Possessive (= Non-focussed actor enclitics)	Focussed phrase
1SG	<i>ku-</i>	<i>-ku</i>	<i>aku</i>
2SG	<i>mu-</i>	<i>-mu</i>	<i>iko</i>
3SG	—	<i>-na</i>	<i>i'a</i>
1PL.INC	—	<i>-ta</i>	<i>kita</i>
1P.EXC	—	<i>-kami</i>	<i>kami</i>
2PL	—	<i>-mu</i>	<i>komi</i>
3PL	(<i>ra-</i>)	<i>ira</i>	<i>ira</i>

A primary difference between PKP and Da'a is that the latter no longer continues a full set of actor proclitics (Set A), a development which is coupled with the evolution of the third person plural pronoun **ra-* first into an indefinite actor marker and thence into a generalised marker of goal focus irrealis (van den Berg 1996; Wolff 1996). For example, although Barr adopts the glossing convention as shown in (1):

- (1) *Duria etu kana ra-pakanawu.*
 durian that must GF/IRR-drop
 'That durian must be dropped.' (Barr 1988a:24)

it is clear from a diachronic perspective that *ra-* continues the older pronoun and therefore a sentence like this must have earlier meant 'they (goal focus) must drop that durian', 'that durian must be dropped by them/by someone'. However other sentences with overt realisation of the non-topic agent as in (2) and (3) make it clear that *ra-* has lost its connection with third person plural:

- (2) *Pade ra-ala-ta kulimba nu bando...*
 then GF/IRR-get-1PL.INC hide of dwarf.buffalo
 'Then we get (GF/IRR) the hide of a dwarf buffalo...' (Barr 1988b:101)
- (3) *Da'a ma-mala ra-raga nu asu.*
 NEG INTR/IRR-able GF/IRR-chase by dog
 '(He) couldn't be chased by the dog.' (Barr & Barr 1988:149)

As further evidence that *ra-* continues the older pronoun, in the present language *ku-* '1SG' and *mu-* '2SG' are still found in paradigmatic relationship with it. However, constructions such as in (4) are now used only in highly intimate conversation:

DAA⁶

- (4) *Da'a ma-mala aku mu-raga.*
 NEG IRR-able 1SG 2SG-chase
 'You can't chase ME!' (Barr 1988a:40)

Da'a has otherwise maintained the PKP focus system intact, as illustrated in brief with examples (5) through (10). Where they occur, Da'a free pronouns may only reference a focussed constituent, i.e. subject of intransitive as in example (6), agent of transitive actor focus as in (8), or patient of transitive goal focus as in (10).

DAA

- (5) *No-rongo-mo tau etu.* (INTR)
 INTR/R-spouse-PERF person that
 'That person is already married.' (Barr 1988a:20)
- (6) *Na-lau-mo ira mpaka ri potomu.* (INTR)
 INTR/R-go-PERF 3PL to — market
 'They went to the market.' (Barr 1988a:19)
- (7) *Nang-goni-mo ira ntali matu'a.* (AF)
 AF/R-eat-PERF 3PL and in-laws
 'They and their in-laws ate.' (Barr 1988a:92)
- (8) *Kita mam-paresa tombi kalawata etu.* (AF)
 1SG AF/IRR-inspect room paddy that
 'We inspect the paddy.' (Barr 1988b:105)
- (9) *...ni-sambale-na bo ni-tompo-na.* (GF)
 GF/R-slaughter-3SG and GF/R-chop-3SG
 '...he slaughtered (him) and chopped (him) up.' (Barr 1988b:85)
- (10) *Ira ni-pakeni-ku mpaka ri Palu.* (GF)
 3PL GF/R-take-1SG to — market
 'They were taken by me to Palu.' (Barr 1988a:27)

3.2 Uma

Uma is spoken to the south of Kaili in the highlands of western Central Sulawesi in an area traversed by the Lariang River. The Uma language was studied by the linguist Esser in the first half of this century (Esser 1964), and more recently by Martens (1988a,b,c). Realis forms were lost in Uma. The Uma prefix *N-* is clearly a reduced form of PKP **maN-*. Although the optional *-po-* is synchronically analysable as a 'spacer' often used with *N-*, comparative evidence suggests that it arose from the overlay of **maN-* on top of intransitive morphology.⁷ The verbal affixes of Uma are given in Chart 2.

⁶ Examples are preceded by an abbreviation indicating their language: CIA Cia-Cia; DAA Da'a; KUL Kulisusu; MRA Mori Atas; MRB Mori Bawah; PAD Padoe; PAM Pamana; TOL tolaki; UMA Uma.

⁷ It appears that the *-po-* can be omitted only when the stem begins with a vowel (compare *m-po-'inca* ~ *ng-inca* 'know') or with *p*, *t*, *k* or *h* (e.g. *m-po-pali* ~ *m-pali* 'hunt'; *m-po-hilo* ~ *n-cilo* 'see'). According to

Intransitive	<i>ma-V</i> <i>me-V</i> <i>mo-V</i> etc.
Antipassive ----- Transitive actor focus	<i>N-(po)-V</i>
Transitive goal focus	ACT PRN + V

Chart 2: Uma verbal inflection

The difference between antipassive and transitive actor focus verbs — both of which make use of the same inherited verbal morphology — is significant, and is elucidated below. Uma also has a set of clitic pronouns which developed from — and is now distinct from — the original free pronouns. Pronoun sets here are from the Kantewu dialect (Martens 1988c: 169).

Table 3: Uma pronoun sets

	Non-topic actor	Possessive	Clitic	Independent
1SG	<i>ku-</i>	<i>-ku</i>	<i>-a</i>	<i>aku'</i>
2SG	<i>nu-</i>	<i>-nu</i>	<i>-ko</i>	<i>iko</i>
3SG	<i>na-</i>	<i>-na</i>	<i>-i</i>	<i>hi'a</i>
1PL.INC	<i>ta-</i>	<i>-ta</i>	<i>-ta</i>	<i>kita'</i>
1PL.EXC	<i>ki-</i>	<i>-kai</i>	<i>-kai</i>	<i>kai'</i>
2PL	<i>ni-</i>	<i>-ni</i>	<i>-koi</i>	<i>koi'</i>
3PL	<i>ra-</i>	<i>-ra</i>	<i>-ra</i>	<i>hira'</i>

Martens (1988b,c) has suggested that Uma be analysed as a morphologically ergative language, and indeed if clitic pronouns had continued in their original uses, we would expect Uma to have emerged with straight ergative marking, namely:

- (a) agents of transitives marked with non-topic actor (ergative) pronouns;
- (b) patients of transitives and subjects of intransitives marked with clitic (absolutive) pronouns;
- (c) agents of *N-(po)-* (formerly actor focus) verbs marked with clitic (absolutive) pronouns (= antipassive construction).

These patterns **are** found; compare examples (11) through (13).

Martens, a form such as *m-po-weba* 'hit' cannot be shortened to **N-weba*, since consonant clusters are impossible in Uma. It is possible that this particular *po-* prefix developed in Uma as a 'spacer' to separate the *N-* from the verb root and thus avoid consonant clusters' (1988c:173). Although this does seem in many respects to be its synchronic function, comparative data suggests that *-po-* was originally an 'embedded' intransitive prefix. For example, in Pamona the verb *mepali* is an intransitive form meaning 'seek'. There is also a corresponding transitive form which has been derived from it, namely *mam-pe-pali* (Adriani 1928:s.v.). It is prefixes such as this embedded *-pe-* which then become reinterpretable as 'transitivising' prefixes, or, if one will, 'spacers' used with transitive verbs. If this view is correct, then not only has this reinterpretation taken place in Uma, but the use of *-po-* has been greatly extended.

UMA

- (11) *Mo-dungka-ko.* (INTR)
INTR-fall.down-2SG
'You fall down.' (Martens 1988a:240)
- (12) *Ku-weba'-ko.* (TRANS GF)
1SG-hit-2SG
'I hit you.' (Martens 1988c:174)
- (13) *Ng-ala'-ko kiu.* (ANTIPASS)
N-take-2SG corpse
'You took a corpse.' (Martens 1988c:210)

Clitic pronouns, however, have an additional use: they may also be employed with *N-(po)-* verbs to cross-reference the **patient**. Compare example (14) where the clitic pronoun *-i* '3SG' marks the agent — this is also true of the clitic pronoun *-ko* of example (13) immediately above — versus example (15) where it marks the patient:

UMA

- (14) *M-po-hilo-i romeha' sakea.* (ANTIPASS)
N-po-see-3SG two boat
'He saw two boats.' (Martens 1988c:177)
- (15) *Tumai-a-ma m-po-pali'-i.* (TRANS AF)
come-1SG-PERF *N-po-search-3SG*
'I have come searching for him.' (Martens 1988c:173)

Although the **patient**-marking function of clitic pronouns following *N-(po)-* verbs is an innovation in Uma, it is now *N-(po)-* verbs with a clitic pronoun marking the **agent** which have a restricted distribution in discourse: 'In the story from which [example (14)] is taken, this clause is introducing the boats into the story. This is precisely the type of situation where antipassives usually occur in Uma, in the setting of a story to introduce some person or thing' (Martens 1988c:177). No such discourse restriction is found with *N-(po)-* verbs when the clitic refers to a patient, which Martens collectively labels transitive actor focus forms. Transitive actor focus forms are standard in relative clauses in which the agent is relativised, when the verb is preceded by an auxiliary or some other verb as in *come searching* of example (15), and in other contexts where 'the actor or the activity of the actor is the topic of the sentence or when the actor is being contrasted or highlighted' (Martens 1988c:173). Compare also examples (16) and (17) (note a clitic pronoun need not be present).

UMA

- (16) *Aku' to m-po-weba'-ko.*
1SG REL *N-po-hit-2SG*
'I'm the one who hit you.' (Martens 1988c:174)
- (17) *Hira' hante ana' boko' m-po-keni tujaa'.*
3PL and child behind *N-po-carry* seed
'They along with the children who came later CARRIED the seeds.'
(Martens 1988c:173)

3.3 Pamona

Pamona is a Kaili-Pamona language spoken in several dialects over a relatively large but thinly populated area in the heart of Sulawesi. It is also well known by the endonym Bare'e due in large measure to the writings of N. Adriani, whose posthumous dictionary (1928) and grammar (1931) are the sources of the following data. Pamona has three pronoun sets:

Table 4: Pamona pronoun sets

	Agent	Possessive	Free
1SG	<i>ku-</i>	<i>-ku</i>	<i>yaku</i>
2SG	<i>nu-</i>	<i>-mu</i>	<i>siko</i>
3SG	<i>na-</i>	<i>-nya</i>	<i>si'a</i>
1PL.INC	<i>ta-</i>	<i>-ta</i>	<i>kita</i>
1PL.EXC	<i>ka-</i>	<i>-mami</i>	<i>kami</i>
2PL	<i>ndi-</i>	<i>-mi</i>	<i>komi</i>
3PL	<i>na-nda-</i>	<i>-al-nya</i>	<i>si'a</i>

Adriani describes Pamona as distinguishing only intransitive and transitive verbs, the latter of which may either be conjugated (with agent prefix) or else a participle. The realis/irrealis distinction has been lost in standard Pamona, though according to Martens (1997, pers. comm.) it is still to be found in at least the Tojo dialect. I have given the verb forms in Chart 3, arranged to emphasise the continuity of Pamona with Proto Kaili-Pamona:

Stative and intransitive		<i>ma-V</i> <i>me-V</i> <i>mo-V</i> etc.
Transitive	participle conjugated	<i>maN-V</i> AG PRN + V

Chart 3: Pamona verbal inflection

It should not be supposed that the former agent/goal focus distinction has totally collapsed in Pamona. It is still evident, for example, in relativisation: when the patient is relativised the verb of the relative clause must be a conjugated (former goal focus) form, but when the agent is relativised a *maN-* participle (former agent focus) is required. Compare respectively examples (18) and (19).

PAM

- (18) *Nce'e-mo pau anu ku-pokono.*
 this-PERF talk REL 1SG-desire
 'Those are words which I gladly hear.' (Adriani 1931:357)
- (19) *Bare'e re'e anu ma-incani.*
 NEG exist REL TRANS-know
 'There is no one who knows it.' (Adriani 1931:452)

However, in main clauses the distinction is blurred, especially in regard to the treatment of the patient. As shown in (20) and (21), both transitive participles and conjugated verbs may be followed by a free pronoun, which cross-references the patient.⁸

PAM

(20) *Da ku-tinti siko.*

FUT 1SG-hit 2SG

'I shall hit you.' (Adriani 1928:845)

(21) *Isema mam-pokau siko da ma-pone kayuku-ku?*

who TRANS-order 2SG FUT INTR-climb coconut-1SG

'Who told you to climb on my coconut tree?' (Adriani 1931:353)

Taken together, Da'a, Uma and Pamona thus illustrate a progression in the use of PKP free pronouns. In Da'a a *maN-* (irrealis *naN-*) verb can occur with a pronoun which refers only to the agent. In Uma cognate *N-(po)-* verbs can be followed by a clitic pronoun referring either to the agent or the patient, being an innovation in the latter function. In Pamona this progression has advanced to the point where a free pronoun following a *maN-* verb refers exclusively to the patient. As others have noted (Martens 1988b; Himmelmann 1996; van den Berg 1996), this is a progression which has led away from a typical 'focus' system.

Pamona transitive participles and conjugated verbs are treated alike in other ways, for example both may have zero realisation for a patient known from context, illustrated respectively in (22) and (23).

PAM

(22) *Imbe'i labu-ku? — Bare'e ku-kita.*

where blade-1SG NEG 1SG-see

'Where is my knife?' 'I haven't seen it.' (Adriani 1931:452)

(23) *Mawela karama-mu, bara asu ma-mama.*

wounded finger-1SG perhaps dog TRANS-bite

'Your finger is wounded, did a dog bite it?' (Adriani 1931:175)

And as seen in (24) and (25), both transitive participles and conjugated verbs may have an indefinite patient.

PAM

(24) *Wawu pai lagiwa na-asuki.*⁹

pig and deer 3SG-hunt

'He hunts pigs and deer.' (Adriani 1931:453)

⁸ The agent, when emphasised, may be realised by a free pronoun **preceding** a transitive participle, i.e. *yaku mam-pokau* 'I sent him out' (transitive base *pokau* 'order, command') (Adriani 1931:458). Usually however the relativiser *anu* is used, in which case the sentence is more clearly a cleft. For example:

PAM *Kami anu mang-keni.*

1PL.EXC REL TRANS-carry

'We are the ones who brought it, we are the bringers.' (Adriani 1931:459)

⁹ This sentence is possibly a cleft, i.e. 'pigs and deer are what he hunts'.

- (25) *Tau mang-keni ju'a, ju'a mang-keni tau.*
 people TRANS-carry illness illness TRANS-carry people
 'The people bring disease, and the disease carries away the people.'
 (Adriani 1928:274)

It appears, therefore, that the choice of conjugated versus participle form is more dependent on the discourse status of the agent, but this needs to be confirmed by further study. Parallel to Da'a, the Pamona third plural pronoun *nda-* shows signs of developing into an indefinite agent marker. Compare (26); see further Adriani (1931:453–454) and van den Berg (1996:103).

PAM

- (26) *Se'i baula nda-roro.*
 this buffalo.meat 3PL-roast
 'This is roasted buffalo meat.' (Adriani 1931:341)

Another feature of Pamona grammar is what Adriani called 'conjunctive' forms and van den Berg 'consecutive' forms, which are found for example following the conjunction *pai* 'and'. Unlike in main clauses where agent pronouns are restricted to transitive verbs, the agent markers which characterise conjunctive verbs occur not only with transitives but also with intransitive and stative stems, as in respectively (27) through (29) (in these examples the relevant form is the one which follows *pai* 'and').¹⁰

PAM

- (27) *Na-wai-ka yaku kina'a pai ku-pang-koni.*
 3SG-give-BEN 1SG rice and 1SG-TRANS-eat
 'He gave me rice, therefore I ate.' (or: '...I had to eat.') (Adriani 1931:476)
- (28) *Ku-pedasi nyara-ku pai na-po-lonco-mo.*
 1SG-strike horse-1SG and 3SG-INTR-go.fast-PERF
 'I struck my horse, so that he went trotting.' (Adriani 1931:476)
- (29) *Ma-dago pang-koni-ta pai ta-ka-gasi se'i-se'i.*
 STAT-good TRANS-eat-1PL.INC and 1PL.INC-STAT-fit at.present
 'We ate well, and therefore we are now fit.' (lit: 'Our eating was good...')
 (Adriani 1931:479)

Adriani supposed that conjunctive forms were a recent innovation, having as their provenance gerunds with transposed possessive suffix. He supported this hypothesis with three pieces of internal evidence: (a) the form of the conjunctive base is identical to that of the gerund (note in Pamona transitive gerunds are formed with *paN-*, intransitives with *po-*, *pe-*, etc. and statives with *ka-*; compare for example the gerund *pangkoni-ta* 'our eating' of example (29) above); (b) this furthermore accounts for why conjugated transitives, e.g. *ku-koni* 'I eat' are not found in conjunctive contexts (because there are no gerunds such as

¹⁰ Contrary to the impression which examples (27) through (29) may give, it is also possible for the subject of both clauses to be the same. For example:

PAM *Kami ma-oro, setu pai ka-ka-dusu.*
 1PLEXC STAT-hungry that and 1PLEXC-STAT-scrawny
 'We hunger, because of that we are gaunt.' (Adriani 1931:478–479)

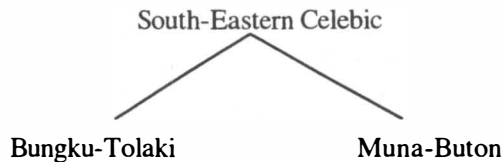
**koni-ku* 'my eating', etc. from which they could be derived); and (c) corresponding sentences with the gerund may in some cases still be formed, e.g. the sentence of (30) is possible but uncustomary.

PAM

- (30) *Na-wai-ka yaku kina'a pai pang-koni-ku.*
 3SG-give-BEN 1SG rice and TRANS-eat-1SG
 'He gave me rice, and (there followed) an eating of mine, my eating of it.'
 (Adriani 1931:476)

4 Proto Bungku-Tolaki

We now leave the Kaili-Pamona language family, and turn to languages of south-eastern Sulawesi. Most of the languages of this area belong to one basic stock, labelled here South-Eastern Celebic, which in turn comprises two subgroups:



Ideally Proto Muna-Buton should be compared to Proto South-Eastern Celebic (PSeCel). However, for practical reasons I have limited the comparison to the Bungku-Tolaki languages, for the reason that the reconstruction of Proto Bungku-Tolaki (PBT) is at a much further advanced stage. Fortunately for the purposes at hand Bungku-Tolaki languages have been conservative, therefore have transparently preserved many aspects of what must have been the PSeCel system. This is particularly true in regard to pronouns, where in most cases reconstructed PMP pronouns (Blust 1977 *inter alia*) are straightforwardly reflected in Bungku-Tolaki languages. Compare this to the situation in Muna-Buton where not only have agent proclitics tended to coalesce with following verbal prefixes, but also a number of person shifts have taken place (van den Berg 1991b:23–26). Hopefully, therefore, little has been lost in restricting our attention to Bungku-Tolaki languages.¹¹

Chart 4 gives the system of verbal inflection which I reconstruct for Proto Bungku-Tolaki. There is no evidence to my knowledge to suggest that a realis/irrealis verbal distinction was maintained either in PBT or in its ancestor, PSeCel.

¹¹ On most points mentioned in this paper regarding Proto Bungku-Tolaki verbal inflection, parallel forms are found to such a degree in Muna that I foresee few problems in reconstructing Proto South-eastern Celebic, at least to the extent that verbal morphology is discussed herein. To be sure there are some differences — for example the prefix which marked a transitive verb as having an indefinite object is clearly reconstructible in Proto Bungku-Tolaki as **poN-*, but van den Berg (1991b) proposes **me-* for Proto Muna (based on an internal reconstruction), and in Cia-Cia the corresponding form appears to be *pi* (van den Berg 1991c) — but hopefully further investigation, particularly of lesser known Muna-Buton languages, will lead to satisfying resolutions on such issues.

	M-form	C-form
Stative	* <i>mo</i> -V	Set A + * <i>mo</i> -V
Intransitive	* <i>me</i> -V * <i>mo</i> -V etc.	Set A + * <i>pe</i> -V Set A + * <i>po</i> -V etc.
Transitive indefinite object	* <i>moN</i> -V	Set A + * <i>poN</i> -v
Transitive definite object	*⟨ <i>um</i> ⟩-V	Set A + *V + Set B
Passives	*⟨ <i>in</i> ⟩-V	

Chart 4: Proto Bungku-Tolaki verbal inflection

Note that the transitive and intransitive prefixes **moN*-, **me*-, **mo*-, etc. at a certain level of abstraction can all be analysed as the morpheme *⟨*um*⟩ plus a *p*-initial prefix (that is, **moN*- is underlyingly *⟨*um*⟩ + **poN*- etc.), and it is this *⟨*um*⟩ which is absent in corresponding conjugated forms (from which derive the respective column labels, M-form and C-form, of Chart 4). The stative prefix **mo*- (from PMP **ma*-), however, had an invariant form whether or not preceded by a Set A pronoun. Passive verbs with *⟨*in*⟩, I will maintain, stood outside of this system especially in terms of pronominal inflection.

The difference between definite and indefinite object verb forms is most easily illustrated. Compare this pair of sentences from Tolaki:

TOL

(31)a. ...*lako mo-lolaha o ambo*.
go UM/TRI-search ART goodness
'...go look for goodness'

b. ...*lako l[um]olaha-i ina-no i Dapi*.
go UM/search-3SG mother-3SG PN David
'...go look for David's mother' (Scott Youngman 1990, pers. comm.)

Both sentences contain verbs derived from the transitive base *lolaha* 'search'. The form *mololaha* in (31a) is a transitive **indefinite** (TRI) object form, and is required in this case because the patient, 'goodness', is a non-specific quantity. The form *lumolaha* in (31b) on the other hand is a transitive **definite** object form, and is required because here the patient, 'David's mother', is a specific, known referent. As is typical for BT languages, however, even when indefinite the patient NP appears without oblique marking. Regarding the particulars of how 'definite' and 'indefinite' are defined in Bungku-Tolaki languages, see further Mead (1999). Transitive indefinite object forms could also be termed 'antipassives' in that they indicate lowered referentiality/topicality of the patient. In Bungku-Tolaki languages they have the same case marking potential as do ordinary intransitives.

I reconstruct four pronoun sets for Proto Bungku-Tolaki, shown in Table 5.

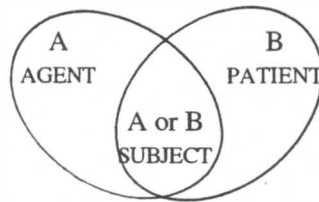
Table 5: Proto Bungku-Tolaki pronoun sets

	Set A	Set P	Set B	Set F
1SG	* <i>ku</i>	*-(<i>ng</i>)/ <i>ku</i>	* <i>aku</i>	* <i>i(n)aku</i>
2SG	* <i>u</i>	*- <i>mu</i> , - <i>u</i>	* <i>ko</i>	* <i>i(n)ko</i>
3SG	* <i>i</i>	*- <i>no</i>	* <i>io</i>	* <i>hia</i>
1PL.INC	* <i>to</i>	*-(<i>n</i>)/ <i>to</i>	* <i>kita</i>	* <i>i(n)kita</i>
1PL.EXC	* <i>ki</i>	*- <i>mami</i>	* <i>kami</i>	* <i>i(n)kami</i>
2PL	* <i>mi</i>	*- <i>miu</i>	* <i>komiu</i>	* <i>i(n)komiu</i>
3PL	* <i>ro</i>	*-(<i>n</i>)/ <i>do</i>	* <i>ira</i>	* <i>hi ira</i>

Set F pronouns were free pronouns, with a distribution similar to nouns.

Set A pronouns occurred in preverbal position and were clitics which either attached to the verb which followed (proclitic position) or to one of a certain number of particles preceding the verb (enclitic position). They were used to mark the subject/agent of intransitive and transitive verbs, and as already noted above were incompatible with the cooccurrence of **um*.

Set B pronouns occurred postverbally. In Proto Bungku-Tolaki, Set B pronouns were obligatory following transitive, definite object verbs in order to mark the patient, a situation which as far as I know has been maintained into all present-day daughter languages. Set B pronouns also occurred with other classes of verbs to mark the subject thereof. Together Set A and Set B pronouns constituted what could be regarded as a split-ergative system in main clauses (large capital letters indicate pronoun sets):



Set A pronouns occurred with intransitive verbs in a restricted environment — in Proto Bungku-Tolaki, primarily limited to consecutive clauses, conditional clauses, imperatives and following negative particles — while Set B pronouns can be regarded as the unmarked choice.

Set P pronouns were possessive pronouns which were also used on passive **in* verbs to mark the demoted agent. However, it appears that by the time of Proto Bungku-Tolaki, main clause passives had become strictly agent-deleting, with pronominal marking for demoted agents being restricted to relative clauses and content interrogatives.

Let us consider now three present-day Bungku-Tolaki languages. Tolaki is the most useful for illustrating Proto Bungku-Tolaki, as it has closely retained the proposed original system. Padoe and Kulisusu exhibit more extensive modifications, but in different directions.

4.1 Tolaki

With an estimated 280,000 speakers spread across the southern portion of mainland south-eastern Sulawesi, Tolaki is by far the largest of all Bungku-Tolaki languages, but

comparatively little information has yet been published about it. The following data, representing the prestigious Konawe dialect, have been taken from my own field notes or, as indicated, from various Indonesian sources.

Tolaki has four pronoun sets which are relevant to the present discussion,¹² shown in Table 6.

Table 6: Tolaki pronoun sets

	Nominative	Possessive	Absolutive	Independent
1SG	<i>ku</i>	<i>-nggu</i>	<i>-aku</i>	<i>inaku</i>
2SG	<i>u</i>	<i>-mu</i>	<i>-ko</i>	<i>inggo'o</i>
3SG	<i>no</i>	<i>-no</i>	<i>-i, (-o, -e)</i>	<i>ie'i, iee</i>
1PL.INC	<i>to</i>	<i>-ndo</i>	<i>-keito</i>	<i>inggito</i>
1PLEXC	<i>ki</i>	<i>-mami</i>	<i>-komami</i>	<i>inggami</i>
2PL	<i>i</i>	<i>-miu</i>	<i>-komiu</i>	<i>inggomiu</i>
3PL	<i>ro</i>	<i>-ro</i>	<i>-'iro</i>	<i>ihiro</i>

The split-ergative system described for the protolanguage has been fairly transparently maintained. I begin with illustrations of transitive, definite object verbs; as can be seen in the clauses of examples (32) through (34), a nominative (Set A) pronoun preceding the verb is used to mark the agent, while an absolutive (Set B) pronoun following the verb marks the patient.

TOL

(32) **Ku-wutiwuti-ko.**

1SG-deceive-2SG

'I deceived you.'

(33) ...*a-ro tiro-'i i Oheo laa*

and-3PL look.down.on-3SG PN Oheo be

me-reu-rehu ine nohu.

UM/INTR-REDP-sit at rice.mortar

'...and they spied Oheo below sitting by the rice mortar.'

(34) **No-wiso-iro o wula ana-ako-no i aa m-botingu.**

3SG-enter-3PL ART moon child-all-3SG at inside LKR-basket.cage

'The moon put all her children into a basket cage.'

Subjects, on the other hand, can be marked by either set. Nominative marking for subjects is in fact required in imperative contexts,¹³ after negatives, and following the concessive

¹² Not illustrated is a fifth set of 'dative' or 'indirect object' pronouns which arose historically from the fusion of the PBT instrumental/benefactive marker **ako* with a following Set B pronoun; see further Mead (1998:208ff.).

¹³ In cases where a second person singular addressee is understood from context, the imperative verb may appear without any agreement marker. For example:

TOL *Lako-to pong-kaa, Oheo.*
 go-PERF TRI-eat Oheo
 'Go eat, Oheo.'

marker *ke* 'if' or the sequential marker *a* 'and, so that' (although belonging with the verb, the nominative pronoun occurs in enclitic position with respect to these particles). Compare examples (35) through (38).

TOL

- (35) *I-pe-wiso-to ona i une baki-landaka!*
 2PL-INTR-enter-PERF EMPH at inside sago.filter.basket
 'You two get in the sago filter basket!'
- (36) *Ku-karu-karu-'i-kee woroko-no,*
 1SG-REDP-scratch-3SG-BEN:3SG throat-3SG
kioki no-po-lua ringgi wulaa.
 NEG 3SG-TRI-expel coin gold
 'I scratched its throat for it, (but) it never coughed up a gold coin.'
- (37) *Ke-ku po-wohiki o ta'i...*
 if-1SG TRI-clean ART faeces
 'If I clean up excrement...'
- (38) *...a-ro lako.*
 so.that-3PL go
 '...so that they go.' (Muthalib, Alimuddin, Pattiasina, et al. 1985:21)

Nominative subject agreement also occurs in a few other contexts, such as scene-setting at the beginning of narratives, in certain kinds of complement clauses, and in clauses which are specifically marked as durative (Mead 1998:322–324). However, as examples (39) through (42) illustrate, subjects can also be marked using absolutive pronouns.

TOL

- (39) *Tangga-ko-to wodo, wonggi, n[um]unu-o elo-mu...*
 endure-2SG-PERF truly snail UM/pull-3SG tongue-2SG
 'You must really endure a lot, Snail, pulling your tongue...'
- (40) *Me-rapu-aku-to.*
 UM/INTR-marry-1SG-PERF
 'I am already married.'
- (41) *Mom-behawe-'i akala.*
 UM/TRI-think-3SG tactic
 'He was thinking of a plan.'
- (42) *Ari-iro-to mong-gaa.*
 finish-3PL-PERF UM/TRI-eat
 'They've already eaten.'

In a significant move away from a strict split-ergative system, however, Tolaki has developed yet a **third** way to mark subjects in main clauses, namely with possessive pronouns, illustrated here in examples (43) through (45).

TOL

- (43) *Lako-nggu-to mo-lasu lako i Kolaka me-wuta.*
 go-1SG-PERF UM/INTR-flee go to Kolaka UM/INTR-land
 'Then I fled to Kolaka on foot.'

- (44) *Laulau-ro-to m-be-luuwako.*
do.immediately-3PL-PERF PL.SUB-INTR-take.off
'Immediately they flew off.'
- (45) *Sabutu-no hae nggo r[um]ako-'i,*
exact-3SG in.addition FUT UM/catch-3SG
pe-dapasako-no-to tudu tumoko i mumu inea.
INTR-glide-3SG-PERF arrive perch at peak areca
'Just as he was again about to catch her (the parrot goddess), off she glided,
landed and perched at the peak of an areca palm.'

The origin of possessive marking in main clauses appears to have been postposed¹⁴ subordinate temporal clauses. Compare in (46) through (48) the similar patterning which is to be found, among other places, in present-day Mori Bawah, Pamona and Uma:

MRB

- (46) *Ko-hawe-no men-toro a meda, pong-kaa-no-mo.*
just-arrive-3SG UM/INTR-sit at table TRI-eat-3SG-PERF
'No sooner had he come and sat at the table than in he delved.' (Esser 1933:190)

PAM

- (47) *Ku-pedasi nyara-ku, po-lonco-nya-mo.*
1SG-strike.with.object horse-1SG INTR-go.fast-3SG-PERF
'I struck my horse, off he went trotting.' (Adriani 1931:476)

UMA

- (48) *Na-tompoi', "Aku'-mi, mama."*
3SG-answer 1SG-PERF daddy
Po-me-hoko'-ra-mi hira' tau tolu.
INTR-RECIP-hug-3PL-PERF 3PL person three
'She answered, "It's me, daddy". (And so) the three of them hugged each other.'
(Martens 1988c:229)

In Tolaki, however, such verb forms have become ubiquitous in narrative texts — the verb stem most commonly found in this construction is *lako* 'go', which consequently has come to have a semantically bleached sense translatable as 'then' — where they reliably signal event-line happenings, that is, salient events which move the story forward in time. For example:

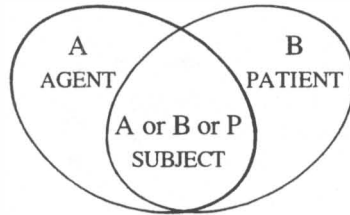
¹⁴ In *sabutu-no* of example (45) we also see the use of a possessive pronoun to mark the subject in a **preposed** subordinate temporal clause, a pattern which itself must be very old; compare for example Tagalog *pagalis niya* 'his leaving, when he leaves/left' (Schachter & Otanes 1972:160, 446) among many others. Compare also *ari-no* 'his finishing, his having finished' of example (49). Doubtless the use of P-forms of verbs and possessive marking for subjects in preposed and postposed temporal clauses is related to the fact that these constructions are in origin (if not in some languages still in fact) nominalisations.

As Hwang has noted for English, there is some tendency for postposed (but not preposed) temporal clauses to be used for dramatic surprise, as in *He was up in the tree, picking apples, when the wolf came along* (1990:68). This predisposition is probably also to be observed in these languages, and helps account for why postposed forms developed (in Tolaki) into main clause forms.

TOL

- (49)a. *Ingoni-ako-no-kaa nggiro'o, lako-no-to lako mo-salei.*
 near.present-INTS-3SG-just that go-3SG-PERF go UM/TRI-clear
 'Right after that, he went and cleared.'
- b. *Saa ari-no mo-salei, lako-no-to t[um]uehi-'i.*
 when finish-3SG UM/TRI-clear go-3SG-PERF UM/clearcut-3SG
 'After his having finished clearing, he clearcut it.'
- c. *Saa ari-no t[um]uehi-'i, lako-no-to h[um]unu-'i.*
 when finish-3SG UM/clearcut-3SG go-3SG-PERF UM/burn-3SG
 'After his having finished clearcutting it, he burned it.'

As illustrated by these sentences, there still exists some tendency for such verb forms to occur in the second clause of a sentence. However, both because of the discourse status of such clauses and the fact that they **can** occur independently — see examples (43) and (44) above — verbs with possessive subjects are arguably no longer dependent verb forms. If this is accepted, then Tolaki has an agreement system in main clauses which is no longer strictly split-ergative, but (in comparison with Proto Bungku-Tolaki) could be characterised as:



Passive verbs with *<in>* have two distinct uses in Tolaki. On the one hand, they are found in relative clauses and content interrogatives in which the patient is respectively relativised or questioned. As illustrated in (50) and (51), in this use the demoted agent may be expressed by a possessive pronoun.

TOL

- (50) *o gandu s[in]olongako-ro i tonga m-bada*
 ART corn PASS/pour.out-3PL at middle LKR-field
 'the corn which had been poured out by them in the middle of the field'
- (51) *O hapo laa k[in]aa-mu?*
 ART what be PASS/eat-2SG
 'What are you eating?' (Muthalib, Alimuddin, Chalik, et al. 1985:36)

As the predicate in main clauses, on the other hand, passive verbs with *<in>* are strictly agent-deleting, and thus can be viewed as ordinary intransitives with the potential for only one argument (the underlying patient/derived subject) to be marked on the verb. Example (52) illustrates a passive verb with a nominative pronoun marking its derived subject; example (53) illustrates the same with an absolutive pronoun.¹⁵

¹⁵ It is also possible for a main clause *<in>* verb to be unmarked for its (derived) subject. For example:

TOL *Koa oleo hopulo kiniku ni-gere.*
 every day ten carabao PASS-slaughter
 'Every day, ten carabao are slaughtered.' (Muthalib, Alimuddin, Chalik, et al. 1985:34)

TOL

- (52) *Ni'ino, iamo no-in-ala.*
 this NEG.IMPV 3SG-PASS-take
 'This must not be taken.' (Muthalib, Alimuddin, Pattiasina, et al. 1985:27)
- (53) *Inotu mo-lua nggiro'o t[in]amo-'ako-'i-to Kolo'imba*
 swamp STAT-broad that PASS/name-INSTR-3SG-PERF Kolo.Imba
 'That broad swamp became named (with the appellation) Kolo Imba.'

As far as I am aware, however, the use of possessive pronouns to mark subjects as described above has not been extended to include the derived subjects of passive verbs.

4.2 Padoe

Padoe is a Bungku-Tolaki language originally spoken in the interior south of Lake Towuti. Although Esser considered Padoe to be a dialect of Mori, it is now usually regarded as a language in its own right. Padoe data have been taken from Karhunen (1994) and Vuorinen (1995). Although van den Berg (1996) made extensive use of Vuorinen's data, his uncertainty about its accuracy (it consists of many elicited sentences) made him chary of giving it full weight. However, based upon my own reading of Esser (1927, 1933) and my experience with closely related languages, the constructions found in Vuorinen seem not only accurate, but valid representations of patterns still to be found in many Bungku-Tolaki languages.

Padoe is somewhat distinct, however, in that here the original Set B pronouns have had a curious history. In Padoe, reflexes of these pronouns have become differentiated into three sets — in order to capture this continuity, I depart from Vuorinen's terminology and label these respectively sets B1, B2 and B3 — depending on which of three functions they were used in: marking patients of transitive verbs (B1); marking subjects of intransitive and other verbs (B2); and, only in future contexts, marking both agents of transitives and subjects of intransitives (B3). In these first two uses Padoe continues functions of these pronouns which are well attested in other Bungku-Tolaki languages and assigned to the protolanguage; the third use, however, which represents an innovation, is discussed at some length below.

Table 7: Padoe pronoun sets

	Unmarked subject	Possessive	B1	B2	B3	Independent
1SG	<i>ku-</i>	<i>-nggu</i>	<i>-aku</i>	<i>-aku</i>	<i>aku</i>	<i>iaku</i>
2SG	<i>(a)u-</i>	<i>-mu</i>	<i>-ko</i>	<i>-iko</i> ¹⁶	<i>iko</i>	<i>iiko</i>
3SG	<i>no-</i>	<i>-no</i>	<i>-o</i>	<i>-o(to), -lo'o</i>	<i>o(no), lo'o</i>	<i>umono</i>
1PL.INC	<i>to-</i>	<i>-ndo</i>	<i>-kito</i>	<i>-kito</i>	<i>kito</i>	<i>ikito</i>
1PL.EXC	<i>ki-</i>	<i>-mami</i>	<i>-kami</i>	<i>-kami</i>	<i>kami</i>	<i>ikami</i>
2PL	<i>(a)i-</i>	<i>-miu</i>	<i>-komiu</i>	<i>-komiu</i>	<i>komiu</i>	<i>ikomiu</i>
3PL	<i>ro-</i>	<i>-ro</i>	<i>-iro</i>	<i>-iro, -lo'iro</i>	<i>iro, lo'iro</i>	<i>umboro</i>

¹⁶ Esser (1927:123) gives this form as *-ko* rather than *-iko*.

Set B1 pronouns (but not B2 or B3 pronouns) occur in four allomorphic sets (i.e. first singular *-aku*, *-'aku*, *-haku*, *-ngangu*), where allomorph selection is determined by the verb stem to which the suffix is attached. An analysis of the available data has established that these allomorphs reliably trace back to the stem final consonant of an earlier stage of the language (Mead 1998:76–79). I investigate the use of each Padoe pronoun set in turn. Pronoun sets are illustrated with all classes of verbs (intransitive, transitive, etc.) except for statives, for which unfortunately the crucial data is usually lacking in my sources.

Set B1 pronouns are used as patient markers, and as illustrated in (54) through (56) are exclusively used with transitive definite object verb stems. They are in fact obligatory with such stems, regardless of whether the patient is realised elsewhere in the clause:

PAD

- (54) *Iko kaa t[um]o'ori-aku kee?*
 2SG EMPH UM/know-1SG INTERROG
 'Do YOU know me?' (Vuorinen 1995:110)
- (55) *Ki-rodohiro.*
 1PLEX-hit-3PL
 'We (exclusive) hit them.' (Vuorinen 1995:110)
- (56) *Ki-angga'o io galu la'a.*
 1PLEX-work-3SG ART rice.field that
 'We are working that rice field.' (Vuorinen 1995:111)

Set B2 pronouns, on the other hand, are used for marking subjects, as shown in (57) and (58).¹⁷ Although Set B1 and B2 pronouns are both suffixes, they are in complementary distribution: Set B2 pronouns can occur with all manner of stems **except** transitive definite object verbs. Set B2 pronouns are usually followed by the completive marker *-to*, which thus provides an additional measure of separation between the two pronoun sets (and from a functional perspective, usually indicates that the pronoun marks a subject rather than a patient). Active verbs in such cases are always M-forms.

PAD

- (57) *Me-wuni-o-to umono ai te'olo.*
 UM/INTR-hide-3SG-PERF 3SG in woods
 'HE hid in the woods.' (Vuorinen 1995:105)
- (58) *Mo-nahu-aku-to inehu.*
 UM/TRI-cook-1SG-PERF vegetables
 'I cooked vegetables.' (Vuorinen 1995:105)

Because Set B2 pronouns do not occur in allomorphic sets, van den Berg supposed that patient marking must therefore represent an older situation, and subject marking an innovation (1996:109). However, when comparing the Bungku-Tolaki languages it is apparent that reduced pronouns must be reconstructed for Proto Bungku-Tolaki in **both** functions (Mead 1998:332 ff.). One possible explanation for this difference, then, is that

¹⁷ The predicate of (58) is notionally transitive. As noted in §4, however, verbs with indefinite objects (viz. prefixed by *poN-*) are — from the viewpoint of case marking potential — treated syntactically as intransitive.

Set B2 pronouns simply represent a later **capture** as verb suffix. It is also conceivable that this difference results from an analogical leveling whereby Set B2 pronouns came to more resemble Set B3 pronouns (see below).

So-called unmarked subject pronouns (from Set A) are not restricted in the type of stem which they may precede; however, <um> in any of its allomorphs must be absent.¹⁸ As illustrated in (59) through (61), these pronouns mark only subjects and agents.

PAD

- (59) *Aso n-ta'u-to ku-po-ia ai Poso.*
 one LKR-year-PERF 1SG-INTR-live at Poso
 'I have lived FOR A YEAR in Poso.' (Vuorinen 1995:108)
- (60) *Inehu mbio au-po-nahu.*
 vegetable what 2SG-TRI-cook
 'What vegetables are you cooking?' (Vuorinen 1995:107)
- (61) *Umari sie ku-nahu-o.*
 finish that 1SG-cook-3SG
 'After that I cooked it.' (Vuorinen 1995:107)

When preceded by one of a certain number of functor words, as in (62) through (64), the pronoun will cliticise to the functor rather than the following verb, but otherwise the patterning is the same.

PAD

- (62) *...ako-ku ma-haki.*
 because-1SG STAT-sick
 '...because I was sick.' (Vuorinen 1995:111)
- (63) *Amba-ambano ba-to po-nahu boka...*
 first if-1PL.INC TRI-cook oil
 'First, if we make coconut oil...' (Vuorinen 1995:109)
- (64) *La-ku to'ori-o nee-no.*
 NEG-1SG know-3SG name-3SG
 'I do not know his name.' (Vuorinen 1995:109)

In the part of the grammar we have investigated so far, the split-ergative system of Proto Bungku-Tolaki is still very much in evidence. The primary difference is a slight formal differentiation of original Set B pronouns into two sets, depending on whether they were used to mark patients or subjects.

In contrast to the above pronoun sets, compare now Set B3 pronouns, illustrated in examples (65) through (67).

¹⁸ Certain intransitive stems with frozen *-um-* constitute only **apparent** exceptions to this claim, for example:

PAD *Ambo no-umari.*
 not.yet 3SG-finish
 'It is not yet finished.' (Karhunen 1994:39)

PAD

- (65) **Lo'iro** [m]o-nahu inehu.
 3PL UM/TRI-cook vegetables
 'They will cook vegetables.' (Vuorinen 1995:103)
- (66) **Lo'o** kaa beehe kee me-'eka
 3SG EMPH want INTERROG UM/INTR-ascend
 ai lere saru ikito?
 to garden with 1PL.INC
 'Does he want to go to the garden with us?' (Vuorinen 1995:101)
- (67) **Kito-to** men-tarima-'o kee onie?
 3PL-PERF PL.SUB-recvie-3SG INTERROG this
 'Shall we receive this?' (Vuorinen 1995:115)

In addition to form, Set B3 pronouns distinguish themselves from Sets B1 and B2 pronouns in three other ways: (a) Set B3 pronouns precede the verb, while Set B1 and B2 pronouns always follow the verb; (b) while Sets B1 and B2 together reflect an older absolutive set of pronouns (i.e. used to mark subjects and patients), Set B3 pronouns constitute a set of nominative pronouns (i.e. they mark subjects and agents); and (c) Set B3 pronouns are only used in irrealis and/or future¹⁹ contexts. All these differences, however, have a unified account if we make the assumption that historically the formative *lo* in the third person pronouns *lo'o* and *lo'iro* was originally a future-oriented verb²⁰ which was then followed by the appropriate absolutive pronoun. In this view, an expression such as found in example (65) — *lo'iro monahu inehu* — would in origin actually have been a two-predicate structure 'want/must/come-they to cook vegetables'.²¹ A growing opacity — whereby the verbal nature of *lo* was no longer recognised — later gave rise to a new set of pronouns.

¹⁹ Vuorinen (1995) prefers the term 'irrealis', while Esser (1927, 1933) labels these 'future' pronouns. Because of the small amount of data, most of it elicited, I prefer to leave open the question of how these forms are best characterised. According to Vuorinen, a Padoe irrealis pronoun followed by the perfective marker *-to* indicates a present continuous aspect:

PAD *Lo'o-to mo-nahu inehu.*
 3SG-PERF UM/TRI-cook vegetable
 'She is cooking vegetables.' (Vuorinen 1995:103)

while an irrealis pronoun followed by *-po* is used for expressing a future conditional aspect:

PAD *Aku-po l[um]eko ai Tomata.*
 1PL-COND UM/go to Tomata
 '(Later) if I go to Tomata.' (Vuorinen 1995:104)

However, this understanding is not always reflected in Vuorinen's free translations, e.g. in this view it seems example (66) would be better translated 'Will he want to go to the garden with us?', conversely example (67) translated 'Are we receiving this?'

²⁰ According to Bybee, Perkins and Pagliuca (1994:251ff.), lexical verbs which can develop into future markers include those expressing desire (e.g. 'want'), obligation (e.g. 'must'), and movement (e.g. 'come', 'go'). The hypothesis which I deem most likely is that *lo* was earlier a deictic verb meaning 'be there (distant)', but the demonstration of this must be left for elsewhere; see especially Esser (1927:150ff.).

²¹ I am indebted to Erik Zobel for initially proposing a development along these lines. Irrealis pronouns are found in other Bungku-Tolaki languages besides Padoe, compare for example the following Mori Bawah constructions. Esser identified the (b) examples as 'future' forms (1927:98, 103):

This hypothesis receives further support in that a particle *lo* with future orientation is still to be found in both Mori Atas and the Watu dialect of Mori Bawah, where — as is shown in (68) through (70) — it appears to be conjugable like any ordinary intransitive verb (albeit having a semantically bleached meaning).

MRA

- (68) *Ka i-lo l[um]eko.*
 NEG 2PL-FUT UM/go
 'You two shall not go.' (Esser 1933:202)

- (69) *To-lo me-lere...*
 1PL.INC-FUT UM/INTR-dry.field
 'Whenever (supposing) we set out a dry field...' (Esser 1933:202)

MRB

- (70) *A-mo ro-lo men-tuwu.*
 NEG-PERF 3PL-FUT PL.SUB-live
 'They shall not live any longer.' (Watu dialect) (Esser 1933:202)

However, the situation has evolved in Padoe to the point where at present *lo* occurs only in third person forms — not first or second person — and even then only optionally.²² One can also say, for example:

PAD

- (71) *Iro l[um]eko ai Tomata.*
 3PL UM/go to Tomata
 'They will go to Tomata.' (Vuorinen 1995:103)

- | | | |
|-----|--|--|
| MRB | a. <i>Ku-lako.</i>
ISG-go
'I go.' | b. <i>Aku l[um]ako.</i>
ISG.FUT UM-go
'I will go.' |
| | a. <i>Ku-'ala-o.</i>
ISG-take-3SG
'I take it.' | b. <i>Aku um-ala-o.</i>
ISG.FUT UM-take-3SG
'I shall take it.' |

So far as is known to me, in Bungku-Tolaki so-called irrealis or future pronouns are restricted to a small set of geographically contiguous languages, namely Padoe, Mori Atas, Mori Bawah and Bungku. This limited distribution, as well as certain differences in the form of third person pronouns, suggest that the appearance of such pronouns is an areal feature unrelated to any period of common development.

Interestingly, parallel constructions are also to be found in Muna-Buton languages; compare for example the following data from Cia-Cia (van den Berg 1991c:317):

- | | | |
|-----|--|--|
| CIA | a. <i>O-'ala-'e.</i>
ISG.REAL-take-3SG
'I take it.' (realis) | b. <i>A-m-ala-'e.</i>
ISG.IRREAL-UM-take-3SG
'I take it.' (irrealis) |
|-----|--|--|

This similarity raises the possibility that irrealis pronouns are actually far older than assumed in this paper. Van den Berg, on the other hand, gives a different account of the development of the realis/irrealis distinction in Muna-Buton pronoun sets; in his view, for example, the Cia-Cia first person singular forms *o-* (realis) and *a-* (irrealis) both originate from a single form **a-*. For arguments in support of this analysis, see van den Berg (1991b:24-25). The whole issue requires further investigation.

²² The explanation given here also does not account for the third person singular future form *ono* — used in Padoe as an alternate to *lo'o* — but it likely has some connection to the form *ino* '3SG (future)' found in Mori Atas. The possible origins of *ino* have already been discussed at some length by Esser (1933:202).

Although the absence of *lo* in first and second person forms could be considered an argument against the proposed hypothesis, it should be remembered that these pronouns in and of themselves historically had no association with futurity. Therefore the development of *aku* '1SG' (< PMP **aku*), *kito* '1PL.INCL' (< PMP **kita*), *iro* '3PL' (< PMP **ida*), etc. into future markers must — in any account — have been via their first becoming well established in future contexts.

Vuorinen (1995:112) further mentions, without illustration, that the third person forms with *lo* also appear as optional variants in the paradigm of Set B2 pronouns. To the extent that this occurs, it appears to result from an analogical leveling whereby Sets B2 pronouns have come to more closely resemble Set B3 pronouns.²³

In Padoe, passive verbs which occur as the predicate of a main clause have come to be treated entirely as intransitives; they are agent-deleting, and the derived subject may be marked in the same ways as can the subject of an (ordinary) intransitive. This is illustrated in (72) through (74).

PAD

- (72) *H[in]enu-komiu-to kee ai sala?* (B2)
 PASS/hit-2PL-PERF INTERROG on road
 'Were you hit on the road?' (Vuorinen 1995:105)
- (73) *Ono t[in]anu owundu-olo.* (B3)
 3SG PASS/bury short-day
 'He will be buried in the afternoon.' (Vuorinen 1995:104)
- (74) *Inderio ai-h[in]enu?* (UNMARKED SUBJECT)
 where 2PL-PASS/hit
 'WHERE were you hit?' (Vuorinen 1995:110)

Although possessive pronouns are still used to mark the demoted agent, such marking is almost always restricted to passives used as nominalisations or in relative clauses; compare (75) through (77).

PAD

- (75) *io w[in]awo-mu*
 ART PASS/carry-2SG
 'your load (i.e. what is carried by you)' (Karhunen 1994:22)
- (76) *in-angga-no io sulia*
 PASS-work-3SG ART sulia
 'the sulia's work, what was done by the sulia' (Karhunen 1994:22)
- (77) *Henu in-angga-nggu ambo no-umari.*
 REL PASS-work-1SG not.yet 3SG-finish
 'That which I am doing is not yet finished.' (Karhunen 1994:39)

Vuorinen (1995:106) also gives a single example, repeated in (78), which supposedly shows the demoted agent marked on a passive verb functioning as verbal predicate.

²³ An analogical leveling in the reverse direction could also account for the absence of the formative *lo* in irrealis pronouns, but this is speculative.

PAD

- (78) *N[in]ahu-nggu inehu la'a.*
 PASS/cook-1SG vegetables that
 'Those vegetables were cooked by me.'

However, an interpretation of (78) as an equative or cleft ('My cooked thing is those vegetables, what was cooked by me are those vegetables') seems to me to be a more likely analysis.²⁴

4.3 Kulisusu

Kulisusu is a Bungku-Tolaki language spoken in the northeastern corner of Buton Island. Kulisusu has retained the PBT verbal morphology intact. The primary difference between Kulisusu and other Bungku-Tolaki languages lies in the use of pronoun sets: Kulisusu has simply lost the use of Set B pronouns to encode subjects (e.g. corresponding to the function of Padoe Set B2 pronouns), and thus emerges with nominative-accusative case marking. It is one of the few — perhaps the only — Bungku-Tolaki language to have done so. Data has been drawn from my own field notes. Kulisusu pronoun sets are as follows:

Table 8: Kulisusu pronoun sets

	Nominative	Possessive	Accusative	Independent
1SG	<i>ku-</i>	<i>-ngku</i>	<i>-'aku</i>	<i>ungkude</i>
2SG	<i>u-</i>	<i>-u</i>	<i>-ko</i>	<i>ingko'o</i>
3SG	<i>i-</i>	<i>-no</i>	<i>-ho, -'o, -o</i>	<i>inade</i>
1PL	<i>to-</i>	<i>-mai</i>	<i>-kai</i>	<i>ingkai</i>
2PL	<i>mi-</i>	<i>-miu</i>	<i>-komiu</i>	<i>ingkomiu</i>
3PL	<i>ndo-</i>	<i>-ndo</i>	<i>-'inda</i>	<i>indade</i>
1PL (COLL)	—	<i>-nto</i>	<i>-kita</i>	<i>ingkita</i>

Kulisusu is in the process of losing the first person plural inclusive/exclusive distinction, the forms *-nto*, *-kita* and *ingkita* now being employed only in a collective sense. The third singular accusative pronoun has three stem-conditioned allomorphs, the consonant of which (*h*, glottal or zero) reliably reflects an earlier stem final consonant (Mead 1998:76-79). Independent pronouns exhibit certain substitutions etymologically unrelated to the original PBT free pronouns.

Compare first the following clauses (all verbs are intransitive):

KUL

- (79) *La MisiMisikini i-lako-mo me-sango i raja.*
 La MisiMisikini 3SG-go-PERF UM/INTR-ask.permission at king
 'La Misi-Misikini went to ask permission of the king.'

²⁴ A passive which also had marking for the derived subject — if such were possible (e.g. **?no-n[in]ahu-nggu* 'it was cooked by me') — would more clearly be a main clause verb form.

- (80) *Be-ku-lako me-baho.*
 FUT-1SG-go UM/INTR-bathe
 'I want to go bathe.'
- (81) *Be-ku-pe-baho.*
 FUT-1SG-INTR-bathe
 'I want to bathe.'

As illustrated in (79) through (81), the features which characterise subject marking in Kulisusu main clauses in general are: (a) nominative pronouns are used to mark subjects,²⁵ which marking, as one will note, is incompatible with *um*; (b) a pronominal marker is present even when the subject is realised elsewhere in the clause, and (c) subject marking is restricted to the first verb of the clause. The common pattern in Kulisusu clauses is thus a conjugated verb (C-form) followed if appropriate by one or more M-forms.

This same pattern is found with transitive verbs, be the patient indefinite as in (82) and (83), or definite as in (84) and (85). In the latter case, an accusative pronoun — compare the third person singular suffix *-ho* in (84) and (85) — obligatorily follows the transitive verb in order to mark the patient thereof.

KUL

- (82) *Ana-mai, hiina i-lako mo-onto i e'e Molosu.*
 child-1PL NEG 3SG-go UM/TRI-see at water Molosu
 'Our son didn't go to look around at Lake Molosu.'
- (83) *Sabucuno ndo-po-onto-mo duka sa-mia cina...*
 thereupon 3PL-TRI-see-PERF also one-person female
 'Just then they also saw a woman...'
- (84) *Sadia ndo-lako um-onto-onto-ho.*
 always 3PL-go UM-REDP-see-3SG
 'They always went to watch him.'
- (85) *Sabucuno ndo-onto-ho-mo rapa-no Wangkinamboro...*
 thereupon 3PL-see-3SG-PERF head-3SG Wangkinamboro
 'Just then they saw Wangkinamboro's head...'

As may be seen in (86) and (87), stative verbs have an invariant form, whether or not a nominative pronoun precedes the stative stem.

KUL

- (86) *I-kowoo mo-bosi.*
 3SG-have.scent STAT-rotten
 'It smells rotten.'

²⁵ The only case where subjects are marked otherwise is in the construction *daa-ho* 'there is, there are' (which reflects the old use of Set B pronouns in their function of marking subjects), for example:

KUL *Daa-ho ika, daa-ho koila, daa-te duka garangga...*
 exist-3SG fish exist-3SG green.turtle exist-with also edible.seaweed
 'There are fish, there are green turtles, there is also edible seaweed...'

The phonologically more conservative Wolio form *daangia* 'there is, there are, it is the case that...' (Anceaux 1988:43) is cognate.

- (87) *I-mo-bosi.*
 3SG-STAT-rotten
 'It's rotten.'

Passive verbs in Kulisusu are marked by *-in-*; as predicates of main clauses they are agent-deleting, and used only when the agent is unknown or irrelevant. As illustrated in (88), in such cases the patient (derived subject) is marked as an ordinary subject, that is with a nominative pronoun.

KUL

- (88) *Mewangu mewangu, oloncuduo a'iso i-s[in]ambure.*
 morning morning yard that 3SG-PASS/sweep
 'Every morning that yard is swept.'

Only when a passive verb with *-in-* occurs as a participle — that is, when it serves as a noun as in (89) or noun modifier as in (90) — is expression of the demoted agent permitted, expressed by a possessive pronoun.

KUL

- (89) *Hina-mo i-da'a s[in]ikori-nto.*
 NEG-PERF 3SG-NEG.exist PASS/await-1PL.COLL
 'There's no longer anyone we are waiting for.' (lit: 'Our awaited one doesn't exist anymore.')
- (90) *ihi-no bawu in-ala-no itonia*
 flesh-3SG pig PASS-take-3SG just.then
 'the pig's flesh which he had taken earlier'

As further seen in (91) through (94), active and stative participles in relative clauses (either headed or headless) also now take the suffix *-no*, a pattern extended from the use of the third person singular possessive pronoun with passive participles.²⁶

KUL

- (91) *randaa-no raja mo-pii-no a'iso*
 daughter-3SG king STAT-sick-PART this
 'the king's daughter who was sick'
- (92) *mo-'ia-no ri'ai*
 UM/INTR-reside-PART here
 'the one(s) living here'

²⁶ In this use, however, *-no* is better regarded as a participle marker which has lost its connection with person and number. The same development of the third person singular pronoun is also found outside of Bungku-Tolaki, for example in both Muna and Wolio:

MUN *mie-hi niho r[um]ato-no ini*
 person-PL just UM/arrive-PART this
 'the people who had just arrived' (van den Berg 1989:232)

WOL *O ndoke-mo duka mo-hobuti-na bulu-na pani-na.*
 ART monkey-PERF also PART-pull.out-PART feather-3 wing-3
 'The monkey was also the one who pulled out the feathers of his wings.' (Anceaux 1988:56)

Given the close geographical proximity of Kulisusu to these other languages, this innovation has doubtless spread areally.

- (93) *ke-o unγκkude mom-potalo-no...*
 if-3SG 1SG UM/TRI-defeat-PART
 'If I win...' (lit: 'If it's me who defeats...')
- (94) *tama-no l[um]ingka-no itonia*
 father-3SG UM/set.off-PART just.then
 'her father who had recently gone off'

When a participle is transitive and has a definite patient, however, then patient marking supersedes marking with *-no*. Compare the intransitive participle above in (94) (from the underived stem *lingka* 'set off') with the transitive participle in (95) (from the transitive base *hohalu* 'seek, hunt'), which differ formally only in that the former is followed by the erstwhile possessive pronoun *-no*, but the latter by an accusative pronoun.

KUL

- (95) *io bawu-hako h[um]ohalu-'o a'iso*
 ART pig-COLL UM/seek-3SG that
 'that herd of pigs which was chasing him'

The verb morphology system of Kulisusu can thus be represented as shown in Chart 5. Accusative pronouns obligatorily follow transitive definite object forms, and mark the patient. Nominative pronouns mark agents and subjects, including the derived subject of passives. Note in Chart 5 the near identity of participle forms to the M-forms from which they historically derive.

	M-form	C-form	Participle
Stative	<i>mo-V</i>	NOM + <i>mo-V</i>	<i>mo-V-no</i>
Intransitive	<i>me-V</i> <i>mo-V</i> etc.	NOM + <i>pe-V</i> NOM + <i>po-V</i> etc.	<i>me-V-no</i> <i>mo-V-no</i> etc.
Transitive indefinite object	<i>moN-V</i>	NOM + <i>poN-V</i>	<i>moN-V-no</i>
Transitive definite object	<i>-um-V + ACC</i>	NOM + V + ACC	<i>-um-V + ACC</i>
Passive	<i>-in-V</i>	NOM + <i>-in-V</i>	<i>-in-V + POSS</i>

Chart 5: Kulisusu verbal inflection

5 Proto Kaili-Pamona and Proto Bungku-Tolaki compared

It should be apparent from the above discussion that the Proto Kaili-Pamona and Proto Bungku-Tolaki systems of verbal inflection are of very different natures. One major difference is that PBT maintained a living distinction between **maN-* and **⟨um⟩* on transitive stems, the former marking that the patient was indefinite, the latter that it was definite. This distinction had disappeared in PKP where only **maN-* (irrealis **naN-*) was retained as a marker of transitive (actor focus) verbs.²⁷

²⁷ I would submit, however, that the older indefinite object (antipassive) function of **maN-* was vestigially retained even in PKP in that the object of a **maN-* verb could not appear pronominally, just as is still the case for example in Da'a.

Were this the only difference between the two systems, it could be handled simply by revising our view of Proto Celebic verbal morphology by supplying the system with more ‘cells’, some of which (e.g. the cell for transitive **um*>) were then lost in Proto Kaili-Pamona. However, another difference concerns the place given to conjugated verbs in the respective protosystems. Conjugated verbs, or rather the Set A pronouns which characterise them, are not reconstructible for Proto Malayo-Polynesian, and therefore constitute one of the innovations by which these languages distinguish themselves therefrom. However in PKP conjugated verbs functioned as irrealis counterparts to realis goal focus verbs with **ni*-, in other words they occupied the functional slot formerly held by PMP **-en*. This is compatible with the view that Set A pronouns developed from fronted (possessive enclitic pronouns serving as) non-topic actor pronouns. This is illustrated in Chart 6, where a PMP box with a heavy line around it indicates a form which was continued in PKP, and a PKP box with a heavy line around it indicates an innovation.

PMP	Realised	Unrealised
	* <i>mina</i> -V	* <i>ma</i> -V
	* <i>minaR</i> -V	* <i>maR</i> -V
	* <i>minaN</i> -V	* <i>maN</i> -V
	* <i>⟨umin⟩</i> -V	* <i>⟨um⟩</i> -V
	* <i>⟨in⟩</i> -V	*V- <i>en</i>



PKP	Realis	Irrealis
Stative	* <i>na</i> -V	* <i>ma</i> -V
Intransitive	* <i>ne</i> -V * <i>no</i> -V etc.	* <i>me</i> -V * <i>mo</i> -V etc.
Transitive AF	* <i>naN</i> -V	* <i>maN</i> -V
Transitive GF	* <i>⟨in⟩</i> -V	SET A + *V

Chart 6: From PMP to PKP

In Proto Bungku-Tolaki, on the other hand, we find conjugated verbs serving as ‘counterparts’ to all kinds of verbs **except** *⟨in⟩* passives. This development is compatible with the view that conjugated verbs developed from gerunds (verbal nouns, atemporals) via fronting of an associated possessive enclitic pronoun, as illustrated in Chart 7.

PMP	Realised	Unrealised
	<i>*mina-V</i>	<i>*ma-V</i>
	<i>*minaR-V</i>	<i>*maR-V</i>
	<i>*minaN-V</i>	<i>*maN-V</i>
	<i>*⟨umin⟩-V</i>	<i>*⟨um⟩-V</i>
	<i>*in-V</i>	<i>*V-en</i>

⇒

PBT	M-form	C-form
Stative	<i>*mo-V</i>	Set A + <i>*mo-V</i>
Intransitive	<i>*me-V</i> <i>*mo-V</i> etc.	SET A + <i>*pe-V</i> SET A + <i>*po-V</i> etc.
Transitive indefinite object	<i>*moN-V</i>	SET A + <i>*poN-V</i>
Transitive definite object	<i>*⟨um⟩-V</i>	SET A + <i>*V</i>
Passive	<i>*in-V</i>	

Chart 7: From PMP to PBT

The difference between the Proto Kaili-Pamona and Proto Bungku-Tolaki systems is apparent for example in relative clauses. In Kaili-Pamona languages, conjugated verbs are used — indeed, in Uma and Pamona required — in order to relativise the patient of an ordinary transitive predicate, a context where conjugated verbs do **not** appear in Bungku-Tolaki languages.²⁸ For example, Pamona *tau nda-pepate* ‘a killed person, a person killed by them’ (Adriani 1931:455) can only be expressed in Kulisusu with an *-in-* passive, i.e. *mia p[in]epate-ndo*.

Furthermore, in Bungku-Tolaki languages where it is now possible for even *-in-* passives to be conjugated in main clauses (a pattern **not** found in KP languages), the pronoun marks not the agent of the action but rather the patient (derived subject). Compare example (96); see also examples (52), (74) and (88) above.

KUL

- (96) *Be-ndo-k[in]aa pokana-kana.*
 FUT-3PL-PASS/eat same
 ‘They will be eaten the same.’ (*eaten by them...)

²⁸ The rule that when the patient is relativised, an *-in-* form of the verb is required, is very nearly exceptionless in Bungku-Tolaki languages. Some ‘exceptions’ do exist, but almost always involving the applicative suffix *-ako*. For example in Mori Bawah it is standard for a conjugated form to be used instead of an *-in-* passive when a beneficiary is also cross-referenced on the verb of a relative clause:

MRB *kinaa anu i-binta-ako-ira*
 cooked.rice REL 3SG-leave-BEN-3PL
 ‘the cooked rice which she had left for them’ (Esser 1927:164)

For certain other exceptions in Mori Bawah, see Esser (1927:164).

This pattern is entirely consistent with the above hypothesis that actor (Set A) pronouns spread to passive forms from a source use of marking subjects and agents of non-passive verbs, but would be highly unexpected (unexplained) if their original use was to mark non-topic agents.

For these reasons, I cannot agree with van den Berg's proposal that a system akin to PKP could also have given rise to the Bungku-Tolaki and Muna-Buton languages. Although we do see the spread of conjugated verbs in Pamona, both Adriani (1931) and van den Berg himself (1996:103) considered this to be a recent development. Furthermore, in that Pamona uses the prefix *ka-* in the conjugated forms of stative verbs — see example (29) above — this innovation shows itself to be distinct from the process by which conjugated verbs developed in South-Eastern Celebic.

6 Conclusions

Because of the significant differences outlined above between Proto Kaili-Pamona and Proto Bungku-Tolaki, one hypothesis would be to assume that their common ancestor did not have conjugated verbs; rather conjugated verbs were innovated separately in Kaili-Pamona and South-Eastern Celebic. This would entail that this common ancestor was more Philippine-like than any of the present-day languages. Just such an hypothesis, for example, was proposed by the Dutch linguist Adriani, and later espoused by his compatriot Esser.

What was said by Dr Adriani in 1914 about the relationship of the Bungku and Mori languages to Bare'e [Pamona] (Adriani & Kruyt 1914:87, 90) has been corroborated by research conducted up to the present: however great the similarity between Bungku-Mori and Bare'e on the one hand their differences from Bare'e, and on the other their similarities with Loinan [Saluan] and its close relatives, are such that the language border between Bare'e and Mori must be taken to be very considerable. On this basis, among other things, Dr Adriani developed the following theory: "By its relationship with Bobongko (on the Togian Islands) and Gorontalo, the Loinan language points to a southward migration of the inhabitants of the north half of the northern peninsula, which divided near present-day Gorontalo: the Loinan branch proceeded via the Togian islands to the further shore east of Tanjung Api, spreading from there to the east (Balantak) and to the south (Bungku), while another branch proceeded to the west and then to the south, and then south of the equator a further branch separated back in an easterly direction. So one may consider that Bare'e, as the most eastern extension of this last-named side flow, came to a halt up against Loinan." (Adriani & Kruyt 1914:89) And one can add to this last statement: "...and further to the south, against Mori". The Mori and Toraja [Kaili-Pamona] language groups thus make up portions of two different language complexes, which have their common point of origin in the Philippine languages. (Esser 1927:7–8) (my translation)

Not all the details of this theory have stood the test of time. In particular we now know that the Gorontalo languages are part of an intrusive wave from the central Philippines which only recently reached the northern peninsula of Sulawesi (Blust 1991). It is also possible that the similarity of the Bungku-Tolaki languages with Saluan is more apparent (i.e. shared retentions) than actual.

It should further be noted that the present study has not **disproved** van den Berg's notion of a Proto Celebic macrogroup. This, of course, could only be accomplished by marshalling evidence — in the form of shared innovations — for some alternative subgrouping hypothesis. Rather, I have shown only that the evidence for a Proto Celebic macrogroup is

weaker than has heretofore been supposed. Although various sound and other shared changes have been proposed for establishing a close link between Kaili-Pamona and their neighbours to the southeast, the principle arguments have either been vitiated by the appearance of newer data, or have not stood up to rigorous investigation.

On the other hand, one possible link between the languages of central and southeastern Sulawesi — and which at the same time suggests that conjugated verbs were not separately innovated — concerns the form of the first person plural exclusive agent prefix. Although agent pronouns are in most cases identical to corresponding possessive pronouns, there is a discrepancy in the first person plural exclusive: in both KP and BT, the agent pronoun consistently shows up as *ki-*, regardless of the form of the corresponding possessive enclitic.

Table 9: Form of the first person plural (exclusive) prefix in various Kaili-Pamona and Bungku-Tolaki languages

	Agent	Possessive
Kulawi	<i>ki-</i>	<i>-kami</i>
Da'a	—	<i>-kami</i>
Uma	<i>ki-</i>	<i>-kai</i> (dialectally <i>-kami</i>)
Pamona	<i>ka-</i> (dialectally <i>ki-</i>)	<i>-mami</i>
Tolaki	<i>ki-</i>	<i>-mami</i>
Padoe	<i>ki-</i>	<i>-mami</i>
Mori Bawah	<i>ki-</i>	<i>-mami</i>
Bungku	<i>ki-</i>	<i>-mami</i>
Kulisusu	<i>ki-</i>	<i>-mai</i>
Moronene	<i>ko-</i>	<i>-mami</i>

If we assume that these language families separately innovated actor prefixes, why they should have reached for the same form **ki-* is then unexplained. On the other hand, assuming their common ancestor did have actor prefixes raises another set of other issues, including the following.

- (a) If having a reflex of the actor prefix **ki* '1PL.EXC' defines a macrogroup, what other languages would this encompass? A number of South Sulawesi languages, for example, also employ an agent pronoun *ki* covering the first person plural exclusive.
- (b) What position did conjugated verbs occupy in the verbal system of their common ancestor? Although Wolff's hypothesis that conjugated verbs developed in passive contexts dovetails with our understanding of Proto Kaili-Pamona verb forms, it would be a considerably further leap to derive the Proto Bungku-Tolaki system from it, i.e. by what diachronic pathway did conjugated verbs stop being passive forms?
- (c) What role did so-called verbal nouns (that is, *p*-forms of verbs, elsewhere variously termed nominalised verbs, verbal substantives or gerunds) play in the development and/or spread of agent pronouns? The role of gerunds in the spread of Set A pronouns to intransitive constructions has already been mentioned with respect to Pamona (§3.3). Furthermore, across Bungku-Tolaki languages, reflexes of Set A pronouns are everywhere incompatible with the occurrence of *<um>*. This also suggests, on the face of things, that a conjugated form such as Tolaki *ku-pongaa* 'I eat (something)' developed from an earlier verbal noun, e.g. **paN-kaen-ku* 'my eating'.

Because of the many different verb systems represented in Sulawesi, the languages of this area provide an ideal laboratory for anyone wishing to work on issues of focus and syntactic change in Austronesian. Although I raise certain unanswered questions, I hope the present study has at least presented a clearer picture of the Bungku-Tolaki languages than has previously been forthcoming, as well as brought into relief certain issues which in the past have perhaps not been fully appreciated. Given that we still know little about many of the languages of this area, exciting prospects await us as we follow the evidence where it leads.

References

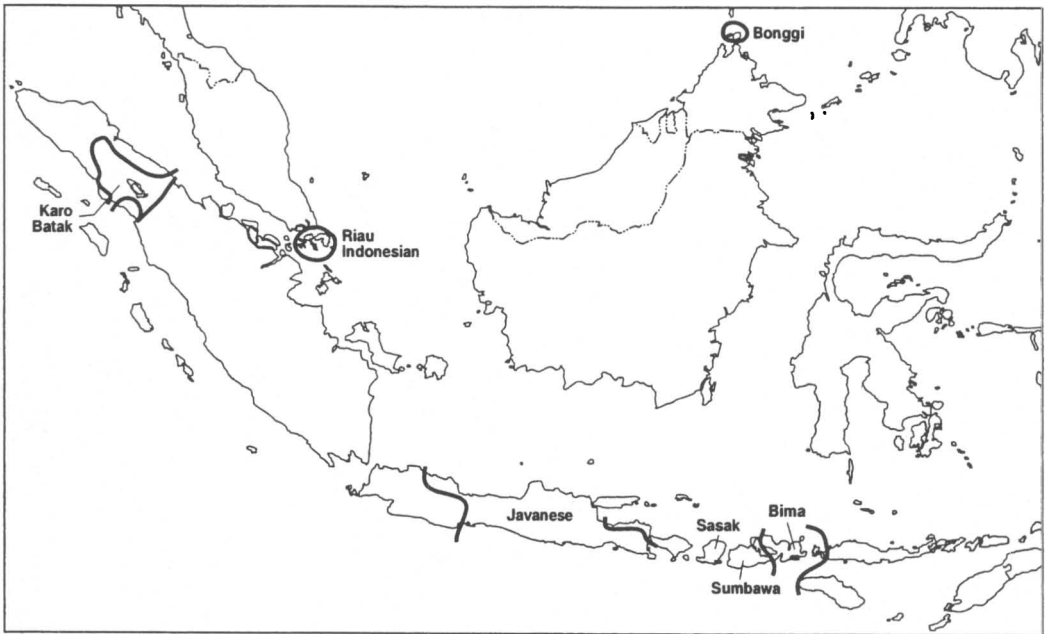
- Adriani, N., 1928, *Bare'e-Nederlandsch woordenboek met Nederlandsch-Bare'e register*. Leiden: E.J. Brill.
- 1931, *Spraakkunst der Bare'e-taal*. Bandoeng: Nix.
- Adriani, N. with Alb.C. Kruyt, 1914, *De Bare'e-sprekende Toradja's van Midden Celebes*, vol.3: *Taal- en letterkundige schets der Bare'e-taal en overzicht van het taalgebied Celebes-Zuid-Halmahera*. Batavia: Landsdrukkerij.
- Anceaux, J.C., 1988, *The Wolio language: outline of grammatical description and texts*, 2nd ed. *Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde*, 11. Dordrecht: Foris.
- Barr, Donald F., 1988a, Da'a verbal affixes and clitics. In Steinhauer, ed. 1988:11-49.
- 1988b, Focus and Mood in Da'a discourse. In Steinhauer, ed. 1988:77-129.
- Barr, Donald F. and Sharon G. Barr, 1988, Phonology of Da'a. In Steinhauer, ed. 1988: 131-151.
- Berg, René van den, 1989, A grammar of the Muna language. Dordrecht: Foris.
- 1991a, Muna dialects and Muncic languages: towards a reconstruction. In Ray Harlow, ed. *VICAL 2: Western Austronesian and contact languages: papers from the Fifth International Conference on Austronesian Linguistics*, 21-51. Auckland: Linguistic Society of New Zealand.
- 1991b, Muna historical phonology. In James N. Sneddon, ed. *Studies in Sulawesi linguistics*, part 2. *NUSA* 33:2-28. Jakarta: Universitas Katolik Indonesia Atma Jaya.
- 1991c, Preliminary notes on the Cia-Cia language (South Buton). In Harry A. Poeze and Pim Schoorl, eds *Excursies in Celebes: een bundel bijdragen bij het afscheid van J. Noorduyt als directeur-secretaris van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde. Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde*, 147:305-324. Leiden: KITLV Uitgeverij.
- 1996, The demise of focus and the spread of conjugated verbs in Sulawesi. In Steinhauer, ed. 1996:89-114.
- Blust, Robert, 1977, The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *Working Papers in Linguistics* 9/2:1-15. Honolulu: University of Hawaii.
- 1991, The Greater Central Philippines hypothesis. *Oceanic Linguistics* 30:73-129.
- Bybee, Joan, Revere Perkins and William Pagliuca, 1994, *The evolution of grammar: tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.

- Donohue, Mark, forthcoming, The pretenders to the Muna-Buton group. To appear in *Pacific Linguistics*.
- Esser, S.J., 1927, *Klank- en vormleer van het Morisch*, vol.1. *Verhandelingen van het Koninklijk Bataviaasch Genootschap van Kunsten en Wetenschappen*, 67/3. Leiden: Vros.
- 1933, *Klank- en vormleer van het Morisch*, vol.2. *Verhandelingen van het Koninklijk Bataviaasch Genootschap van Kunsten en Wetenschappen*, 67/4. Bandoeng: Nix.
- 1964, *De Uma-taal (West Midden-Celebes): spraakkunstige schets en teksten*. The Hague: Nijhoff.
- Himmelman, Nikolaus P., 1996, Person marking and grammatical relations in Sulawesi. In Steinhauer, ed. 1996:115-136.
- Hwang, Shin Ja, 1990, Foreground information in narrative. *Southwest Journal of Linguistics* 9.2:63-90.
- Karhunen, Marjo, 1994, The noun phrase in Padoe. In René van den Berg, ed. *Studies in Sulawesi linguistics*, part 3. *NUSA* 36:17-47.
- Martens, Michael P., 1988a, Focus and discourse in Uma. In Steinhauer, ed. 1988:239-261.
- 1988b, Focus or ergativity? Pronoun sets in Uma. In Steinhauer, ed. 1988:263-277.
- 1988c, Notes on Uma verbs. In Steinhauer, ed. 1988:167-237.
- 1989a, The Badaic languages of Central Sulawesi. In James N. Sneddon, ed. *Studies in Sulawesi linguistics*, part 1. *NUSA* 31:19-53.
- 1989b, Proto Kaili-Pamona: reconstruction of the protolanguage of a language subgroup in Sulawesi. Unpublished typescript.
- Mead, David E., 1996, The evidence for final consonants in Proto Bungku-Tolaki. *Oceanic Linguistics* 35:180-194.
- 1997, Reflexes of *aken across central and south-eastern Sulawesi. Paper presented at the Eight International Conference on Austronesian Linguistics, 28–30 December, Taipei, Taiwan.
- 1998, Proto Bungku-Tolaki: reconstruction of its phonology and aspects of its morphosyntax. PhD dissertation, Rice University, Houston, TX.
- 1999, Active, passive and antipassive in Bungku-Tolaki languages. In David Mead, ed. *Studies in Sulawesi linguistics*, part 5. *NUSA* 45:113-145.
- Muthalib, Abdul, Alimuddin D.P., Husen Chalik and Arsamid, 1985, *Kamus Tolaki-Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Muthalib, Abdul, Alimuddin D.P., J.F. Pattiasina, Haloma Balaka and Amir Kadir, 1985, *Kata tugas dalam Bahasa Tolaki*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Pawley, Andrew and Lawrence A. Reid, 1980, The evolution of transitive constructions in Austronesian. In Paz Buenaventura Naylor, ed. *Austronesian studies: papers from the Second Eastern Conference on Austronesian languages*, 103-130. Ann Arbor: University of Michigan.
- Ross, Malcolm D., 1995, Reconstructing Proto-Austronesian verbal morphology: evidence from Taiwan. In Paul Jen-kuei Li, Dah-an Ho, Ying-Kuei Huang and Cheng-hwa Tsang, eds *Austronesian studies relating to Taiwan*, 727-791. Taipei: Academia Sinica.

- Schachter, Paul and Fe T. Otanes, 1972, *Tagalog reference grammar*. Berkeley: University of California Press.
- Starosta, Stanley, Andrew K. Pawley and Lawrence A. Reid, 1982, The evolution of focus in Austronesian. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 145-170. Canberra: Pacific Linguistics.
- Steinhauer, Hein, ed., 1988, *Papers in western Austronesian linguistics*, No.4. Canberra: Pacific Linguistics.
- 1996, *Papers in Austronesian linguistics*, No.3. Canberra: Pacific Linguistics.
- Vuorinen, Paula, 1995, Person marking in Padoe. In René van den Berg, ed. *Studies in Sulawesi linguistics*, part 4. *NUSA* 37:97-121.s
- Wolff, John U., 1996, The development of the passive verb with pronominal prefix in Western Austronesian languages. In Bernd Nothofer, ed. *Reconstruction, classification, description: festschrift in honor of Isidore Dyen*, 15-40. Hamburg: Abera-Verlag.

— PART III —

*Languages of the rest of
Indonesia and Malaysia*



Map 3: Languages discussed in Part III

Voice and valency alternations in Karo Batak

CLODAGH NORWOOD

1 Introduction

Karo Batak, the second largest of five distinct Batak languages,¹ is a western Malayo-Polynesian (WMP) language spoken in an area of the north-western Sumatran highlands extending south-east from the city of Medan to the large volcanic Lake Toba. Partly due to the relatively inaccessible highland location, the language and cultural traditions of the Karo have been maintained to a remarkable extent. A large number of Karo people adhere to a Christian organisation which actively promotes the use of the Karo language, while skill in song and speech-making during traditional ceremonies is highly regarded even by younger people.

Of the twenty-four groups that the WMP languages are currently, though tentatively, divided into (Ross 1995), the Batak languages as a whole fall into the North-West Sumatra/Barrier Islands group that comprises Gayo, Batak, Nias, Mentawai and Enggano. Nothofer (1991) put forward the hypothesis that this group of languages may be more closely related to an older Paleo-Hesperonesian family that once covered the whole WMP area, than to the more centrally located Hesperonesian languages in this area. Superseded by the later Hesperonesian group, the geographically peripheral languages of the area, such as those of the North-West Sumatra/Barrier Islands group, and northern Sulawesi and southern Philippines languages, are purported to share certain features by virtue of their outlying location in the western Austronesian speaking area.

A detailed descriptive grammar of Karo Batak is available (Woollams 1996). The current paper is part of work in progress which aims to describe the forms and functions of two Karo verbal affixes in relation to the Philippines-type 'focus' system, and ultimately to place Karo Batak in a typological categorisation of WMP languages.

A semantically transitive verb in Karo can occur with either no prefix (indicated by \emptyset - in the gloss for expository purposes) as in (1), or with a homorganic nasal prefix, *N-*, as in

¹ The estimated number of speakers varies according to whether the relatively large numbers of speakers who live in other parts of Indonesia are included. The population of Taneh Karo itself is about 600,000.

example (2). In focus terminology the \emptyset -prefix would be an undergoer focus affix and the *N*- prefix an actor focus prefix, but as Karo does not exhibit a fully functioning focus system, these terms are not appropriate, whereas the less restrictive term *voice* is (see Fox & Hopper 1994). Other grammatical terms used in this paper are described in §3.2.²

(1) 352³

Ku-jumpai labe duana
I- \emptyset -meet first both
'I'll meet them both first.'

(2) 309

adi aku səkali n-jumpai ia
if I once *N*-meet him
'if I ever meet him'

This paper examines the distribution and function of the actor voice (*N*- verb form) constructions, and that of their undergoer arguments. It demonstrates that these constructions are not simply syntactically transitive alternatives to the much more common undergoer voice \emptyset - forms. Rather the distribution of the *N*- constructions is quite circumscribed. They are found most frequently (75% of all *N*- constructions) in subordinate clauses where the deleted pivot is the agent. Those found in main clauses fall into a number of discrete groups with clearly identifiable functions, most of which relate to detransitivisation. Other main clause functions, notably topicalisation of the agent, can be explained as relics of an earlier focus system.

This study is based on the examination of the text of a recorded conversation consisting of over 750 conversational turns.⁴ Transcribed and translated with assistance from a native speaker, it will henceforth be referred to as the text. Five hundred of these turns comprise spontaneous conversation between three participants about the raising of money through a proposed land sale and contain a number of semantically transitive verbs which are repeated throughout the conversation. Four other data sets were also checked in order to verify the conclusions drawn from analysis of the text, three written stories, and one short journalistic piece that included quoted conversation. Examples have been taken from the text wherever possible (text examples are preceded by a turn number).

2 Basic morphosyntax

2.1 Nominal marking

Table 1 shows the forms of the pronominal system.

² DAT dative; DEF definite; DEM demonstrative; EMPH emphatic marker; EXCL exclusive; EXIST existential; FAM familiar; INC inclusive; INTER interrogative; PERF perfective marker; PL plural; P7L particle; PURP purposive; REL relative marker; SG singular; TRS transitiviser.

³ Intonation breaks should be assumed at the beginning and end of all examples but are shown by double slashes (//) when they occur in the same line of an example.

⁴ I wish to thank Ramli Ginting and his family for assistance with data gathering, in particular his mother who has since died and to whom I here wish to record my respect. Personal names have been changed in the text. I also thank Barry Blake and Nikolaus Himmelmann for critical comments, and Andyda Meliala and Edimon Ginting. In particular I want to thank Fay Wouk for her generosity and conscientious editing throughout. Any errors in interpretation or analysis are entirely my own.

Table 1: Pronominal paradigms⁵

	Free pronouns (FF)	Actor affixes/clitics of \emptyset - forms	Genitive affixes/clitics
1SG	<i>aku</i>	<i>ku-</i>	<i>-(ng)ku</i>
2SG (formal)	<i>kam</i>	<i>-ndu</i>	<i>-ndu</i>
2SG (familiar)	<i>(əng)ko mu</i>	<i>-ko</i>	<i>-m</i>
3SG	<i>ia</i>	<i>-na</i>	<i>-na</i>
1PL.INCL	<i>kita</i>	<i>si-</i>	<i>-(n)ta</i>
1PL.EXCL	<i>kami</i>	<i>kami</i>	<i>kami</i>
2PL	<i>kena</i>	<i>kena</i>	<i>kena</i>
3PL	<i>kalak</i>	<i>-na/kalak</i>	<i>-na/kalak</i>

As Table 1 shows many of the A(ctor)⁶ suffixes of the \emptyset - forms are identical to the corresponding genitive forms. However first person singular and plural inclusive As of \emptyset - forms are bound prefixes, not genitive suffixes.⁷ Apart from the affixed As of the \emptyset - verbs, the remaining pronominal arguments of both verb forms (i.e. elsewhere) and the S(ingle) argument of intransitives are represented by independent pronouns, as shown in (3) for the P argument of a \emptyset - verb, and (2) above for both arguments of a *N*- verb.

(3) 635

Ku-idah ia rusur ku rumah ənda.

I- \emptyset -see her often at house this

'I see her often at this house'

Full NP arguments are not morphologically marked, but each voice form marks one argument by virtue of its constrained position vis-a-vis the verb. This marking follows the pronominal patterns in that the A of the \emptyset - verb forms must follow and be immediately adjacent to the verb, as must the P of the *N*- forms, except in those instances where the *N*-form is suffixed with *-sa* (§2.4 and §5.5).

Proper names in any function in Karo take the prepositional particle *si*, although it is mostly restricted to younger people's names as older people are addressed by relationship terms, which may include their clan name.

⁵ Third person pronouns, including possessives, are glossed in the examples according to the English meaning intended, not as 3SG or 3PL, and to be consistent, first and second person pronouns are also glossed in English.

⁶ The Dixon/Comrie terminology is used where A and P represent the arguments for Actor and Patient (or undergoer) respectively.

⁷ Himmelmann (1996:126) notes that in some South Sulawesi languages the occasional presence of a proclitic for first person in a paradigm that is otherwise comprised of enclitics is suggestive of incipient IRREALIS mode marking, while van den Berg (1996:91ff.) in the same volume takes the opposite view that the paradigm of proclitics represents reconstructable Goal Focus IRREALIS marking in the Proto Celebic focus system. The patterns in Karo and related languages need further investigation, the apparent lack of *i-/ni*-forms with first person *ku-/si-* being noteworthy, as are two prominent exceptions to first person proclitic forms. These are two commonly used lexical items which have a basically verbal function, but show nominal morphology by retaining the first person GEN suffixes (enclitics). They are *ate* 'wish' and *ni* 'say' (quotative). They will be glossed as nominals.

Dative marking with the preposition *man* (+*ba-* for pronominals) 'for/to', is used in ditransitive constructions. It is used to mark recipients, benefactives as in (6) below, and indirect objects of saying/telling etc., as well as the reflexive anaphor.

2.2 Verb morphology

In addition to the two voice prefixes under discussion, the other main verbal prefixes are *ər-* and *tər-*, *i-/ni-*, and a homorganic *N-* prefix which is used to derive most intransitive verbs. In this paper this intransitive *N-* prefix will be assumed to be merely homophonous with the *N-* voice prefix under discussion.

Intransitive verbs usually occur in one of three main lexically determined forms: they may be bare stem constructions, or marked with either the intransitive *N-* prefix as exemplified in (4) and (5), (which come from a written text), or with the *ər-* prefix. As noted above, the pronominal argument of intransitives is always an independent pronoun from Table 1. The preferred order is predicate initial, as in (4).

- (4) *E maka tedis mə si Naktaki, ng-andung ia, bərkət ia.*
 then stand *mə si Naktaki N-cry* he depart he
 'Then Naktaki stood up, crying, and departed.'

However, pragmatic and syntactic factors may cause inversion of the V(X)S pattern, as with *mulih* in (5).

- (5) *Ng-andung duana, sabab la i-sangka nande-na ia m-ulih.*
N-cry both because not i-expect mother-her she N-return
 'They both cried because her mother had not expected her to return.'

The very productive *ər-* prefix derives intransitive verbs from nominals and, with one exception, occurs with only one core argument, as in (6).

- (6) 38
Si ənggo ər-dahin kena
 REL PERF *ər-work* you
Sanggap kena ibas dahin kena.
 content you in work your
 'Those of you who are already working, be content in your work.'

The exceptional form with this prefix, which is possibly a relic from an erstwhile transitive paradigm, is derived from the verb *b(ah)an* 'do/make'. This form takes two arguments and frequently occurs with the *-sa* suffix as in (7). Further investigation is required to clarify what this might imply about the historical development of the *ər-* prefix.

- (7) 144
Uga ər-bahan-sa kita m-uat si məhuli?
 how *ər-make-sa* we.INCL *N-get* REL good
 'How do we make things good for ourselves?'

The *tər-* forms are basically passive verbs, as in (8). Accidental passives are included in this group.

- (8) 215
Pərən ku-idah kərina juma ah // lanai tər-dahi
 grass I- \emptyset -see all garden DEM never tər-tend
 'I noticed grass everywhere in that garden...not being tended.'

Second and third (but not first) person agents can be expressed with the *tər-* forms, second person using the same pronominal forms as \emptyset - verb forms, as in (9), and the third person using the enclitic *-sa*. However, two argument constructions which include the A argument are rare with *tər-*.

- (9) 141
aku lah gia tər-usih kena nak-ku
 I EMPH EMPH tər-follow you son-my
 'if you follow me (my advice), my son'

The intransitive nasal verbs (including a class that glosses as adjectives) and other prefixed forms such as the *ər-* and *tər-* forms, will not be considered further here. Where they occur in examples from henceforth their derivational prefixes are not included in interlinear glossing.

The morphosyntactic structure of the *i-* forms, illustrated in (10), is basically the same as that of the \emptyset - forms, with one difference. Either there is no *i-* prefix on first person forms, or first person *i-* forms have been conflated with the \emptyset - forms.

- (10) 209
O nande kai gundari i-suan-ndu i juma-ta Pindan ah?
 oh mother what now i-plant-you in garden-our name DEM
 'Oh Mum, what have you currently got planted in our Pindan garden?'

Speakers appear to regard *i-* as a freely variable allomorph of \emptyset .⁸ It is particularly prevalent in written genres, where \emptyset - forms are seldom used. But while there is a large degree of overlap, there is also evidence of a distinction between the two in the spoken varieties of the language.

Two criteria suggest that the *i-* forms, along with the now infrequently used *ni-* constructions, typically function as passives in spoken language. First, the *i-* forms have a low discourse frequency and second, the \emptyset - constructions occur with an A argument about 60% of the time, while the *i-* forms are much less frequently accompanied by an A argument. If there was once a modal distinction between the two forms, this no longer appears to be the case, although the prevalence of *i-* forms in written narratives is worth further investigation.

The *ni-* forms, which are said to be used only by older speakers, now occur rarely even in written genres; a comparison of passages of Neumann's translation of the Bible with the same passages from the modern version over fifty years later, for example, shows that where Neumann used *ni-* forms, the modern version uses *i-* forms or alternative phrasing.

There are also three valency increasing affixes: the causative prefix *pe-* and two applicative suffixes *-i*, and *-kən*. Verbs derived with *-kən* in particular have often become lexicalised, for example, *bəre* is often best glossed as 'allow', but *bərekən* means 'give'. One or two other verbal suffixes, including *-ən*, have limited productivity or very restricted distribution.

⁸ This observation is supported by Woollams (1996).

2.3 Emphatic particles

A possibly unusual feature in Karo is the presence of three frequent particles, *ngə*, *mə* and *pe*, which may be described as emphatic markers, although they are not optional in every context. Only one speaker uses them with very high frequency, mainly in contexts where their use is optional. Although they usually mark nominals, their specific distribution and function is the subject of ongoing research and they are presumed to be undergoing a functional change, moving perhaps from a vestigial form of case marking to discourse particles.⁹ In (11), from a written text, as in many other cases, *ngə* appears to be associated with the undergoer of the \emptyset -/i- forms, while *mə* appears to be associated with the agent of these forms and *pe* with the agent of the *N*- forms, but these are not absolute correlations as many counterexamples testify.

- (11) *tapi kərina ngə i-simbak-na*
 but all *ngə i-push.away-he*
 'but he pushed them all away'

2.4 The nasal verb suffix *-sa*

N- forms may take a suffix of the form *-sa* (or a phonologically determined allomorph *-ca*). Woollams (1996:115) claims this suffix is a third person object (undergoer) form for the *N*- forms. Certainly in some contexts it appears this way, as in (12), but in many other contexts the situation is not this simple.

- (12) *Maka ngadi mə ia ng-ajar-sa.*
 then stop *mə she N-learn-sa*
 'Then she stopped teaching (them).'

The suffix *-sa* cannot freely be used in all contexts with *N*- forms. As well as examples like (12), where it appears to substitute for the independent third person pronoun *ia*, it may also co-occur with this pronoun, as well as with first or second person undergoers in a few cases. This suffix is discussed in more detail in §5.5, where it will be argued that its function with the Karo *N*- forms is similar to that of the oblique case marker, *sa* in Tagalog which is used to mark specific, individuated undergoers of the actor focus forms.

3 The \emptyset - verb forms

3.1 Formal properties

The \emptyset - form clauses function as the basic active construction in Karo. They show an ergative morphological pattern whereby only the pronominal A(ctor) argument is a verbal affix. Table 2, which shows the forms of the affixes/clitics, is virtually a repeat of the second column of Table 1. In written forms of the language the prefix *i-* is usually used instead of \emptyset .

⁹ They are present in many of the examples where they are glossed with their own form to distinguish them from other particles.

Table 2: The \emptyset -/i- verb form affixes/clitics

	\emptyset -/i- form affixes/clitics = A
1SG	<i>ku-</i> (<i>i-</i> prefix not overt in first person form)
2SG	<i>-ndu/-ko</i>
3SG	<i>-na</i>
1PL.INCL	<i>si-</i> (<i>i-</i> prefix not overt in first person form)
1PL.EXCL	FF*
2PL	FF
3PL	<i>-na/FF</i>

* FF = free form pronouns as shown in Table 1.

Unlike the A argument of these verbs, their undergoer (P) argument is not a suffix or clitic. Pronominal Ps take the free forms shown in Table 1, while any lexical NPs are marked with *si* only if they are proper names. Examples (13) and (14) show pronominal A arguments in first and second person respectively, and example (15) a full NP A.

(13) 338

əna ku-jumpai si Neli.
let I- \emptyset -meet PN Neli
'Allow me to meet Neli.'

(14) 351

Jadi ate-ndu jumpai-ndu labe bibi nguda?
become wish-your \emptyset -meet-you first aunt young
'Do you want to meet younger auntie first?'

(15) 412

Uukur bapa-na ah motor ngə kap.
 \emptyset -buy father-his DEM car ngə EMPH
'His father bought a car (for him).'

3.2 Grammatical relations

If one assumes the existence in WMP languages of a coherent and underlying system of grammatical relations (but see Himmelmann 1996, especially p.121, who diverges from this point of view) the undergoer or patient argument (P) is the subject of the \emptyset - verb constructions.

Subject is defined here as the most privileged grammatical relation, sometimes called pivot (or trigger in Indonesian languages). It is not restricted to semantic agents, the A (or S) relations. Thus the P argument of the ergatively patterning \emptyset - forms, as well as the A argument of the accusatively patterning *N*- forms, and S, the single argument of intransitives, are all subjects. The other two arguments, the A of the \emptyset - forms, and the P of the *N*- forms will be called non-subjects. In one section (5.5), where an analysis of the suffix *-sa* is proposed, the non-subject P of the *N*- forms is referred to as a potential object.

These syntactic terms, subject and non-subject, are only used in discussion of syntactic functions. Otherwise as long as reference to the appropriate verb form is unambiguous, the semantically based terms actor and undergoer are used, often abbreviated as A and P respectively.

Returning now to the description of the \emptyset -forms, both the P and the A arguments of these constructions demonstrate a number of core properties, the P relation exhibiting the larger number of them. There is a strong tendency for this argument to be definite and specific but this is not an absolute constraint. In the following description of core properties some examples are elicited as no illustrative examples were available from the corpus.

3.2.1 Relativisation

The basic gap strategy used for relativisation in Karo is exemplified first of all with an intransitive S in (18). In (19) the relativised and gapped argument is a \emptyset -construction P. Square brackets, [], arbitrarily placed after the relativiser, represent the relativised argument.

- (18) 333
Lit danga ka bibi si [] nggəluh.
 EXIST still PTL auntie REL live
 'There is auntie who is still living.'

- (19) 773
Lit danga ngə sada kaset nari si [] ənggo ku-isi.
 EXIST still ngə one tape from REL PERF I- \emptyset -fill
 'There is another tape I have already filled.'

The A argument of the \emptyset -/i- constructions can also be relativised,¹⁰ as in the elicited example (20). With these arguments the strategy is one of pronoun retention. Woollams (1996:293) gives examples whereby possessor and dative arguments can also be relativised by this same strategy.

- (20) *Guru ng-ərgai murid si ingət-[na] pəlarəən.*
 teacher N-scold pupil REL forget-[she] lesson
 'The teacher was scolding the child who forgot her lessons.'

3.2.2 Quantifier float

A quantifier, such as *krina* 'all' in the following examples, will always be understood to belong with the P argument of a \emptyset -/i- construction. Thus in (21) whatever the arrangement of the constituents, *kərina* modifies the undergoer, not the actor argument *-na* of *ihamati* 'respect'.

¹⁰ Woollams (1996:293) claims this only occurs if there is not an equivalent *N*-verb to satisfy the requirements of an actor subject. However the example he gives uses a verb that does have an *N*-equivalent.

- (21) 773
jəlma kərina i-hamati-na (/ *i-hamati-na kərina jəlmal jəlma i-hamati-na kərina*)
 people all i-respect-he
 'He (the king) respected everyone.'
 * 'Everyone respected him.'
 (but for this gloss see example (34) below, the *N*- form equivalent)

3.2.3 Imperatives and reflexivisation

The \emptyset - construction is the basic form used for both imperatives and reflexivisation. Examples (22) (an elicited example) and (23) are reflexive constructions, in which the actor is always the antecedent of the reflexive anaphor. In (23) the anaphor is preceded by an optional dative pronoun. It should be noted, as Woollams (1996:205) has observed, that the reflexive anaphor agrees in person with its antecedent, unlike that of a *N*- form reflexive anaphor which has an invariant form (see §4.1).

- (22) *Ku-pəkəkək ba-ngku.*
 I- \emptyset -hit self-my
 'I hit myself.'
- (23) *Si-tandai-lah rusur man ba-nta.*
 we- \emptyset -know-EMPH often DAT self-our
 'We often indeed know ourselves.'

In imperatives the addressee, (the actor/agent), is freely omitted, as in example (24), unlike the situation in Tagalog, for example. The omission of the A argument in imperatives, and its control of agreement in reflexives, indicate its status as a core argument in these constructions.

- (24) 122
Pəsai ləbe dukut-na, agoi ləbe dukut-na.
 \emptyset -clean first grass-DEF // \emptyset -remove first grass-DEF
 'First clean up the grass, remove it...'

3.3 Ordering constraints

The basic predicate-initial order of Karo is reflected in the preferred **V A P** order of the \emptyset - constructions. Full nominal A arguments, like pronouns, must immediately follow the verb (with the exception of the first person prefixes). Most of the examples above represent the basic order, for example (15).

Unlike the cliticised A argument, the position of the subject, P, is extremely flexible. It frequently precedes the verb, as in (25). The use of the particle *pe* 'also' here is optional; its function is purely semantic. Nevertheless a particle, whether *pe* or another form of emphatic particle, marked an inverted subject in most of the examples I found of these constructions; that is, the pragmatic and semantic significance of the clause initial position often attracts extra marking. In (26), however, the pronominal undergoer preceding the verb is unmarked.

- (25) 189
Sen pe galang-galang b̄erek̄an-na man ba-ngku.
 money *pe* big-big Ø-give-he DAT self-my
 'He also gave a lot of money to me.'
- (26) *ənggo t̄alu kali kam ku-idah land̄ək.*
 PERF three times you I-Ø-see dance
 'I've seen you dance three times.'

3.4 Obligatoriness of arguments

Neither argument of the Ø- constructions, whether in main or subordinate clauses, is obligatory in spoken discourse, and once either A or P is an established entity it may be omitted in contexts where its reference will be understood. In (27b), which was uttered almost immediately after (27a), the undergoer subject has been omitted.

- (27)a. 635
Ku-idah ia rusur ku rumah ənda.
 I-Ø-see her often at house this
 'I see her often at this house.'
- b. 637
Ku-idah rusur // aku usur lewat arənda.
 I-Ø-see often // I often pass here
 'I often see [her], I often pass here.'

In (28) and (29) the non-subject (A) argument of the Ø- construction main clauses has been omitted. Example (29) could mean 'we'll sell the land', the most likely interpretation in the context, or it could be an imperative 'sell the land'.

- (28) 735
Tukur ndai b̄aras-na // tukur ndai ikan-na.
 Ø-buy PTL rice-their // Ø-buy PTL fish-their
 '(I) just bought (them) their rice and fish.'
- (29) 20
Emaka dayak̄ən juma e muat ənəm puluh juta.
 so Ø-sell land DEM N-get six ten million
 'So (I'll/we'll) sell the land to get sixty million.'

Apart from the zero anaphora of these examples, various particles may serve as anaphors.

Omission of arguments from subordinate clauses is discussed in §5.3.2.

3.5 Summary of Ø- verb forms

The Ø- forms, which have a morphologically ergative pattern of pronominal affixation, are the most frequent type in main clauses. Overall, the conclusion is that both A and P are core arguments of Ø- constructions, with P, the subject, ranking higher than A. The P of the

\emptyset - constructions (and the single argument of intransitives, S) can be relativised through a gap strategy, the primary strategy of relativisation in Karo. The non-subject A argument can also be relativised, if required, through a secondary strategy of pronominal retention in the relative clause. The A argument of the \emptyset - constructions also controls the reflexive anaphor, which agrees in person and number with it. In straightforward two-argument transitive clauses with a quantifier, the quantifier will be understood to modify the subject, the P argument, regardless of its position in the clause. Both A and P can be deleted, and frequently are in conversational discourse, and both can function as controllers or controllees in subordination with appropriate verbs. Finally, the \emptyset - constructions are the canonical form of imperatives.

4 The *N*- verb forms

4.1 Formal properties

The pronominal forms that the *N*- forms take are the free form (FF) pronouns seen in Table 1. Full NPs are unmarked (apart from the marking with *si* for proper names). P arguments, both pronouns and full NPs, usually follow the verb, although there are exceptions, including those constructions where the suffix *-sa* is present on the verb. An example of an *N*- construction with two pronominal arguments is (2) above, repeated here as (30). It is worth noting that, due to the distribution and syntactic conditions under which the *N*- form occurs, this example is one of the very few from the whole corpus that illustrate an *N*- form construction with two pronominal arguments. (31) shows both pronominal and full NP arguments and (32) full NP arguments only.

(30) 309

adi aku səkali n-jumpai ia
if I once *N*-meet him
'if I ever meet him.'

(31) 762

pərbahan aku muat uang sekolah
...because I *N*-get money school
'...because I have to earn money for school.'

(32) 446

səbab pərtambar-ku pe ənggo n-jutai ənda utang
because medicine-my *pe* PERF *N*-throw DEM money
'because my medicine has already wasted the money'

4.2 Grammatical relations

The grammatical subject of the *N*-form constructions is the A(ctor). It passes various tests for subjecthood as discussed below. On the other hand the undergoer, P, of these constructions passes none of the following tests for core status, except that under certain circumstances it may be omitted, and that it is able to control deletion in a lower clause, as shown in §5.3.1.

4.2.1 *Relativisation*

The actor argument of the *N*- verbs is readily relativised using the normal gap strategy.

- (33) 296
Lang aku la kah min aku si [] nukur e ya.
 not I not EMPH PTL I REL [] *N*-buy DEM EMPH
 'I am not the one who is buying it.'

But there is no strategy, including that of pronominal retention used for various other arguments, whereby the P of the *N*- form clauses can be relativised.

4.2.2 *Quantifier float*

With the use of an *N*- form, the actor, *jəlma* 'person', of (21) above, (*jəlma kərina i-hamati-na* ('he respected them all')), can be made the actor subject of the clause and can be quantified in this role. The quantifier *kərina* may appear in several positions but it always pertains to the subject, *jəlma* 'people'.

- (34) *Kərina jəlma ngə-hamati ia.* (OR: *jəlma ngəhamati ia kərina* OR:
 all people *N*-respect him *jəlma kərina ngəhamati ia*)
 'They all respected him.'

Note that **jəlma ngəhamati kərina ia* is not possible because P, *ia*, no longer directly follows the verb.

4.2.3 *Imperatives and reflexives*

Both imperatives and reflexives can be constructed from *N*- forms, but they are not canonical. The *N*- form in example (35), *nungkun*, is one of the very few that appear to occur in imperative mood and, like the more common \emptyset -form imperatives, can omit the actor/agent as in (35).

- (35) 353
Nungkun ləbe man pərmərintah ənda.
N-ask first DAT government official DEM
 'You'd better ask the government first.'

Reflexive anaphors of *N*- constructions have an invariant third person form showing no person agreement with their antecedents. Example (36) is from Woollams (1996:205); there were no examples in the corpus.

- (36) *Kam muji ba-na usur.*
 you *N*-praise self-3 always
 'You are always praising yourself.'

4.3 *Ordering constraints*

Examples (34) and (36) above illustrate the basic **A V P** word order of these constructions. The position of the actor subject is more flexible than that of the non-subject undergoer which in most instances directly follows the verb, unless the suffix *-sa* intervenes.

As example (37) shows, while the actor subject generally precedes the verb it need not come immediately before it. Here the optional particle *pe* follows the subject.

- (37) 275
Kalak e pe la nggit nukur bage saja-ng(ə).
 people DEM *pe* not *N*-want *N*-buy thus only-*ngə*
 'People may not want to buy (it), that's all.'

Generally, undergoer NPs do not precede the verb. However inversion has been observed, as in (38), where the undergoer is itself a complement clause. Here the complement clause, containing the verb *nukur* 'buy', precedes the main verb. In this instance the particle *pe* is not optional; it serves to disambiguate the main clause subject NP *kalak* 'people' from the unexpressed undergoer of *nukur*.

- (38) 276
Nukur pe kalak e la nggit.
N-buy PTL people DEM not *N*-want
 'Indeed others won't want to buy (it).'

Overall the order of the two alternative voice forms can be compared to those of Classical Malay (Cumming 1991:107 and 151ff.) and Toba Batak (Schachter 1984). They are:

undergoer voice \emptyset -/i- verb clauses: V(erb)~A [P]
 actor voice *N*- verb clauses: [A] V(erb)~P

where the tilde (~) indicates, for the argument that immediately follows it, a fixed and uninterruptible position with respect to the verb and, conversely, the square brackets ([]) indicate relative freedom of position for the argument they enclose. These two orders give an impression of a mirror image, typical of many WMP languages. This does not hold true for Karo however. In the case of the \emptyset - forms (undergoer voice), the pronominal A arguments are bound forms while both actor and undergoer pronominal forms for the *N*- forms are independent pronouns. Only when the suffix *-sa* is present on the *N*- forms, and no concurrent undergoer is present, is there a real mirror image. However as *-sa* is not an argument, as will be shown in §5, the impression of formal equivalence is false for Karo.

4.4 Obligatoriness of arguments

Although it is generally present in main clauses, an overt actor (the subject of the *N*-forms) is optional, as the following example shows.

- (39) 33
N-darami nakan i Jakarta ni-na.
N-seek rice in Jakarta word-his
 '(He is) looking for food in Jakarta he said.'

The undergoer may also be omitted. In many such cases the undergoer is not recoverable and the verb appears to be intransitive. But when the suffix *-sa* is present on the verb the undergoer is always recoverable, often from an immediately adjacent clause, sometimes from earlier in the discourse episode. For example, in one chain of utterances in the text with the verb *nukur* 'buy', the undergoer (the garden that was for sale, the main topic of discussion) was omitted in several clauses in succession, such as examples (37) and (38). Preceding this

series of utterances however was a solitary one *adi banta juma e, payo katamu ena?* ‘this is about us buying the garden, am I correct?’ which clearly established the identity of the undergoer. According to a native speaker these examples would be more grammatically correct if *-sa* had been suffixed to the verb. In Norwood (2000) I have demonstrated that undergoers of the *-sa* suffixed verbs, whether occurring in the same clause, or recoverable from a previous clause, are salient discourse entities, being both identifiable and specific.

Nevertheless, as indicated, a number of *N-* forms had no recoverable undergoer, nor was the addition of *-sa* judged possible in these circumstances. The syntactic status of these undergoers is discussed further in §5.5.

Omission of arguments in subordinate clauses is discussed in more detail in §5.3

4.5 Comparison of *N-* verb forms and \emptyset - verb forms

Table 3 sums up some of the main features of the two types of constructions described above. In addition to the features outlined in the table, various tests show that while the two arguments of the \emptyset - constructions participate in a number of syntactic processes, only the A argument of the *N-* forms satisfies various tests for core status. The non-subject A argument of the \emptyset - constructions may, for example, be relativised or be the omitted target of control, but the non-subject P argument of the *N-* forms cannot. The function of the *N-* form suffix *-sa*, which is of relevance to the status of the *N-* form undergoers, is discussed below in §5.5.

5 Functional analysis of nasal verb forms

The analysis was designed to establish the morphosyntactic status of the *N-* forms, their distribution and their function. It has been shown above that only one of their arguments, the agent subject, fulfils many of the criteria for core status. Compared to the \emptyset - forms these *N-* forms show a number of features of lowered transitivity.

Table 3: The morphosyntactic differences between the two voices
(based on the most common type of main and subordinate clauses for each voice)

	\emptyset - verb form		<i>N-</i> verb form	
	Main clause	Sub. clause	Main clause	Sub. clause
Unmarked order:	VAP	VAP	AVP	[] VP
Subjects:				
Semantic role:	P	P	A	A
Obligatory	no	no	(yes)	omitted
Pronominal form:	FF	FF	FF	–
Full NP:	unmarked and free position	unmarked	unmarked	–
Non-subjects:				
Semantic role:	A	A	P	P
Obligatory	no	no	?*	?*
Pronominal form:	affix	affix	FF enclitic	FF enclitic
Full NP:	verb enclitic	verb enclitic	variable	variable

* omitted under certain syntactic or discourse conditions – see §2.2.4 above.

5.1 Distribution of the *N-* prefix

The most striking feature of the *N-* forms is their distribution. Out of 109 examples of *N-* forms analysed from the text two stems occurred in this form only (that is, there were no \emptyset -equivalents). These were two complement taking predicates (CTPs) with 30 tokens between them. Out of the remaining 79 *N-* form clauses, less than 20 occurred in main clauses, or approximately one in four. There was a similar distribution for the written data set. This section examines the distribution of these verbs, the morphosyntactic features that are associated with them, and their functions in both main and subordinate clauses.

Table 4 shows the distribution of the main prefixes on semantically transitive verb stems from the text. Stems with more than five tokens, including those formed with valency increasing applicatives, are listed individually. There is a range of semantic types, including verbs of perception such as *idah* 'see', and of cognition such as *ətəh* 'know', and three common complement taking predicates (CTPs) (18, 21 and 22).

Two nominal forms noted in footnote 6, *ate* 'wish/like', *nin* 'word/say', and a third one *ukur* 'thought/think' were not included, both because of the nominal morphology of their first person forms and because they usually lack voice alternatives, although *N-* and *ər-* forms of *ate* and *ukur* are very occasionally found. If they had been listed with the \emptyset - forms, the figures would have been even more heavily weighted towards these constructions, especially given the high frequency of these three items.

The unequal distribution of the two voice prefixes clearly demonstrates that the *N-* forms do not function as basic transitive forms. This is particularly evident in looking at the first three stems, where the ratio of \emptyset - to *N-* forms is 100:3. Surprisingly, two verbs, *tukur* 'buy' and *buat* 'get', showed a greater frequency of *N-* forms. These are discussed below.

5.1.1 Semantic and syntactic parameters

A wide variety of factors were considered in interpreting the distribution given in Table 4. These included, among others, the specific parameters of transitivity from Hopper and Thompson (1980) shown to be relevant to other WMP languages, viz: (i) the number of participants, (ii) aspect, (iii) affirmation, (iv) mood and (v) individuation of O (P).

Only those parameters that proved significant are listed and discussed here.

1. Presence or absence of overt arguments in the clause.
2. The type and form of A and of P, whether a bound or FF clitic pronoun (as per Table 1), a demonstrative, or a definite, specific or indefinite NP
3. The presence in the containing clause of modals, tense/aspect adverbial particles, and complementisers and conjunctions.
4. Type of clause, whether (a) main clause — declarative, interrogative, negative or imperative — or (b) complement or other subordinate clause, including relative clauses. When a subordinate clause was recorded, the type and form of the main clause predicate was noted, for example desiderative, permissive, purposive, resultative or causative verbs, and whether the CTP was a verb, adjective or nominal.
5. The particular arguments involved in deletion, controller and deleted controllee.

Table 4: The distribution of prefixes on semantically transitive verbs in a conversational text

STEM	GLOSS	Ø-	N-	tər- /ər- /i-	TOTAL
1. <i>daya(kən)</i>	'sell'	34	1	0-2-4	41
2. <i>(ə)təh</i>	'know'	40	1	0-0-0	41 ¹¹
3. <i>idah</i>	'see'	23	1	1-0-0	25
4. <i>katakən</i>	'tell'	14	8	0-0-2	24
5. <i>tukur</i>	'buy'	6	13	0-0-1	20
6. <i>sungkun</i>	'ask'	15	3	0-0-0	18
7. <i>jumpai</i>	'meet'	15	4	0-0-0	19
8. <i>suan</i>	'plant'	8	8	0-0-1	17
9. <i>buat</i>	'get'	5	9	1-0-0	15
10. <i>təkən</i>	'sign'	7	7	0-0-0	14
11. <i>dungi</i>	'finish'	6	3	2-0-0	11
12. <i>darami</i>	'seek'	5	2	0-0-0	7
13. <i>pindo</i>	'ask'	5	2	0-0-0	7
14. <i>sampati</i>	'help'	4	1	1-0-1	7
15. <i>suruh</i>	'ask for'	6	0	0-0-0	6
Sub-totals:		193	63	5-2-9	272
16. <5 tokens per stem*		72	16	not counted	88
Ø- FORMS ONLY:					
17. <i>ban</i>	'do'	80		0-2-2	84
18. <i>akap</i>	'feel/think'	29		0-0-0	29
19. <i>bərekən</i>	'give'	14		0-0-0	14
20. <i>dat</i>	'get'	8		0-0-0	8
N- FORMS ONLY:					
21. <i>nggit</i>	'like'		20	0-0-0	20
22. <i>ngasup</i>	'be able'		10	0-0-0	10
TOTALS:		393	109	5-4-11	525

* These verbs were not counted with computer assistance and the actual numbers may vary slightly.

5.2 Argument structure – parameters 1–2

In Ø- verb clauses P, the subject of these constructions, was found to be overwhelmingly identifiable, with definite marking. P was most often either a pronoun or a NP marked with a possessive clitic or demonstrative. The identifiability or individuation of P accords with the findings of Hopper and Thompson regarding high discourse transitivity.

¹¹ Eleven further tokens of *ətəh* as well as numerous forms of an allomorph [*dah*] occurred in second person forms – all variations of a discourse marker comparable to English 'you know'.

On the other hand the *N*- forms had almost equal numbers of identifiable and unidentifiable undergoers. This can be explained partly in terms of undergoer incorporation, which is discussed in §5.4.3.

The \emptyset - verb form clauses were five times more likely to contain two overt arguments than the *N*-form constructions. Given that the majority of *N*- forms were in reduced complement clauses (where deletion of the A subject coreferential with a main clause argument is obligatory), these findings are not surprising. Typical examples of each type with the verb stem *dayaken* 'sell' are (40) which has a \emptyset - form and two overt arguments (in both of its clauses), and (41), with the *N*- form which has a deleted subject argument coreferential with a main clause argument.

(40) 5

Juma e adi anngo pagi dayakən-ndu,
garden DEM if PERF tomorrow \emptyset -sell-you

kuja ngə ate-ndu ban-ndu guna-na sen ənda?
where ngə wish-your \emptyset -make-you use-DEF money DEM

'If you sell the garden later on, what use will you make of the money?'

(41) 3

Mama // lit ate-ndu sura-sura-ndu
Uncle // exist wish-your wish-your

guna n-dayakən juma-ta ənda?
PURP *N*-sell land-our DEM

'Uncle, it is your intention to sell this piece of our land?'

Table 5 shows the results of parameters 1–2 above for a subset of the verbs listed in Table 4. The subset comprises only the first fifteen stems from Table 4. The results are divided to show the differences between clause types. Subordinate clauses themselves vary in type and deletion of subjects is not obligatory in all of them. Arguments that had zero expression are not included.

Table 5

	\emptyset - forms (No.=193)		<i>N</i> - forms (No.=63)	
	Main clause	Sub. clause	Main clause	Sub. clause
Overt A	106	29	9	9
Type of A:				
Identifiable NP	3	1	1	–
Non-identifiable NP	2	–	1	1
Pronominal	101	28	7	8
Overt P	93	24	6	28
Type of P:				
Identifiable NP	38	19	1	4
Non-identifiable NP	10	3	3	21
Pronominal	25	1	1	3
Clause	26	1	1	–
Enclitic <i>-sa</i>	–	–	14	11

5.3 Subordinate clauses

Analysis of parameters 3 and 4 showed that two in three *N*- form constructions in the text were subordinate clauses. A more casual analysis of semantically transitive *N*- forms in the written genres produced similar results. These findings tally with those of Woollams (1996:212), as do the relative proportions of *N*- versus \emptyset - forms in Table 2.

In a separate study, the *N*- form constructions were examined in the context of several inter-related syntactic and semantic parameters of subordination described by Lehmann (1988) and Hopper and Thompson (1980). These include the syntactic downgrading and desententialisation of the subordinate clause, as well as an increasing tendency to omit any explicit linking device as desententialisation increases. The majority of *N*- form clauses showed a high frequency of these features.

Some observations on the role of both the \emptyset - and the *N*- forms in subordination are recorded here, particularly in regard to control phenomena. Although relative clauses were discussed above, the descriptions of control and controllee relations applies equally to them as to other kinds of subordinate clause.

5.3.1 *N*- verb form control relations

When the main clause is an *N*-construction, the actor subject may control deletion in a subordinate clause. Due to the scarcity of *N*- form main clauses in general this is uncommon, with the exception of the two *N*- form CTPs (nos. 21 and 22 in Table 4), one of which, *nggit* 'want', is exemplified in (42), the other being *ngasup* 'be able'. Square brackets represent the deleted argument.

- (42) 295
Kalakai pe la ng-git [] nukur adi la bære Nini.
 people *pe* not *N*-want *N*-buy if not agree Nini
 'People won't want to buy if Nini doesn't agree.'

Control of deletion by the undergoer of an *N*- form main clause also appears to be possible. There were no instances of this across a wide range of data however, and (43) is an elicited example.

- (43) *Aku ng-idah ia sangana [] nuan-i galuh.*
 I *N*-see him while *N*-plant-TRS banana
 'I watched him while he planted bananas.'

With regard to the subordinate clause itself, when it is an *N*-form reduced complement clause, the actor subject must be deleted, as in the examples just shown and others scattered throughout the discussion. But in subordinate clauses which have a degree of independence, such as those introduced by a complementiser, as in (44), omission is not possible.

- (44) 309
Adi aku sækali n-jumpai ia.
 if I once *N*-meet him
 'If I ever run into him.'

Except under the particular discourse conditions described in §4.4, or when the suffix *-sa* is present on the verb, the undergoer of an *N*- clause is not deleted. In general therefore, it may

be concluded that only the actor subjects of the *N*- constructions may be the target of controlled omission in subordinate clauses.

5.3.2 \emptyset - verb form control relations

Most often in cases of subordination, especially those involving *N*-construction reduced complements and deleted actor pivots,¹² the main clause is a \emptyset - construction. Normally its actor controller is overtly expressed. In (45) this is *ku*-.

- (45) 252
Ku-dayakən nin-gku // sitik kəl kap e // [] muat dua juta.
 I- \emptyset -sell word-my // small very EMPH DEM // *N*-get two million
 'I'll sell I tell you, just a small piece, to get two million.'

But example (29), repeated here as (46), demonstrates that, even when absent, the \emptyset -construction actor can control deletion from the subordinate clause.

- (46) 20
Emaka dayakən juma e [] muat ənəm puluh juta.
 so \emptyset -sell land DEM *N*-get six ten million
 'So (we'll) sell the land to get sixty million.'

The \emptyset - construction subject, P, also controls deletion with a number of verbs, including verbs of speaking and cognition such as *suruh* 'tell/order' in (47) or *sangka* 'expect'.¹³

- (47) 577
Ku-suruh ma-nguda ləbe [] ngərana ras Joko.
 I- \emptyset -tell uncle-young first speak with Joko
 'I told younger uncle to speak with Joko first.'

When the subordinate clause itself is a \emptyset - construction, as in the second clause of (48), its P subject may be deleted. In this sentence *ia* 'he' is both intransitive subject of the main clause and deleted undergoer of the \emptyset - construction complement.

- (48) 485
Adi kari ia ng-git ndo ku-jumpai [] səkali nari ni-na.
 if later he *N*-want allow I-meet-TRS once again word-his
 'Later he would be willing if I met [him] another time, he said.'

Non-subject As in \emptyset - construction subordinate clauses are not normally deleted, but may be under some circumstances. Examples (49) (a) and (b) show a minimal pair using the same \emptyset - form CTP (*bere*) and the same embedded verb (*dayakən*) in the second clause. In the (a) example the A (*anak* 'child'/'he') is retained in the second clause; in the (b) example it has been deleted (as in fact has the understood undergoer in both utterances).

¹² Dixon (1994:17) points out that ergative languages always have a syntactic operation which feeds the need for an underlying A to function as a pivot.

¹³ Woollams (1996:313-314) provides an excellent and detailed discussion on the possibilities of a raising analysis for these verbs.

(49)a. 415

La pagi bære-na dayakən anak // pərutangkən.
 not tomorrow \emptyset -allow-he \emptyset -sell child // mortgage
 'If he doesn't agree to sell, mortgage (it).'

b. 320

Adi bage (ə)nda la bære-na dayakən [] // məsui kal ate-ku.
 if thus DEM not \emptyset -allow-he \emptyset -sell // sick very feeling-my
 'If he doesn't agree to sell, I'll feel quite ill.'

The most common CTP in the text is the nominal *ate* 'desire/wish'. It is generally, though not invariably, followed by a \emptyset -form complement. When this CTP is used, omission of core arguments from the complement clause with either \emptyset - or *N*-verb forms did not occur.

5.3.3 Summary

This brief discussion of the phenomena of control and deletion of arguments in complex constructions, while not complete, serves to illustrate the fact that either argument of both the *N*- and the \emptyset - constructions may be controller of deletion. However only the \emptyset -forms allow either of its arguments to be deleted controllees. That is, \emptyset -construction subjects (P) and non-subjects (A) may be deleted controllees, and, as well, *N*-construction subjects (A). But the undergoer of the *N*-construction may not be deleted as the controllee. Furthermore the form of the CTP is a significant factor in determining the following clause type and the possibility of deletion.

Table 6 sums up the possibilities for controller and controllee relations.

Table 6

	Controller	Controllee
\emptyset-verb forms		
Subject	yes	yes
Non-subject	yes	yes
Dative	yes	yes
<i>N</i>-verb forms		
Subject	yes	yes
Non-subject	yes	no *
Dative	?	?

* except when the suffix *-sa* is on the *N*-form in the subordinate clause.

5.4 The *N*-verb forms in main clauses

Out of 79 semantically transitive *N*-verb forms in the text (excluding numbers 21 and 22 in Table 4, i.e. the *N*-forms which were complement taking predicates) only twenty-two were in main clauses. This section describes the factors that condition the use of this relatively small number of *N*-form main clause constructions, i.e. irrealis mood (§5.4.1), topicalisation (§5.4.2), and undergoer incorporation (§5.4.3).

5.4.1 *Irrealis mood*

Anderson (1985:193-194) points out that, cross-linguistically, subordinate clauses often use the subjunctive form of the verb, this same 'subjunctive' form appearing in main clauses to mark functions such as irrealis mood. While the *N*- forms of Karo, so commonly found in subordinate constructions, are not the main form of marking for irrealis mood, in main clauses they are typically found in an irrealis context.

Realis and irrealis are defined, following Barr (cited in Himmelmann 1996:123)¹⁴ as:

REALIS: Past, completed action, a state or action already existing or occurring, a characteristic which is real, existing, fact, fully actualised.

IRREALIS: Non-past action, hypothetical, not yet realised action or state, a characteristic not yet real, not fully actualised. In this respect irrealis shares some features of subjunctive.

A feature of those *N*-forms found in the main clause subjunctive/irrealis context, as well as in subordinate clauses, was a lack of associated temporal marking. While tense/aspect adverbial particles were commonly found in \emptyset - form clauses, there were few in any of the *N*-clauses in the text. Examples (50) and (51) exemplify irrealis constructions.

(50) 535

Ei pensium bapa e nekolahken adek.

AFFIRM pension father DEM *N*-school-TR sister

'It's Dad's pension that is schooling (paying school fees for) sis.'

(51) 632

Aku n-dungi-sa ras ia.

I *N*-finish-*sa* with her

'I'll sort it out with her.' (*'I sorted it out with her.')

Although main clauses with *N*- forms may be used for irrealis mood it does not follow that irrealis is generally expressed by the use of the *N*- form. Most irrealis clauses are either future, marked by the future tense particle *kari*, or conditional, marked by *adi* 'if'. While either particle can occur with an *N*- form, they most frequently occur with \emptyset - verb forms. Additionally, as noted above, transitive imperatives, another class of irrealis, are always \emptyset -forms.

It was noted at the beginning of this section (5) that two stems showed disproportionate numbers of *N*- forms. These were the verbs *tukur* 'buy' and *buat* 'get'. That these two verbs appear most often in this form makes sense in the context, since the most uncertain elements of the whole endeavour are finding a buyer and getting the money. This does not totally explain why *dayaken* 'sell' has only one *N*- form, as the selling is also uncertain. But there is no uncertainty in the intention to sell and I presume that this is where the difference lies.

In a focus system, such as the alternative voice forms of Karo are based on, Hopper and Thompson's 1980 transitivity hypothesis predict that the *N*- forms would be used for imperfective and irrealis semantics, compatible with lowered transitivity. Except for the fact that the future marker is used with the \emptyset - forms, Karo conforms to their predictions regarding the association of irrealis marking and the use of the actor voice (*N*- forms).

¹⁴ Barr and Himmelman make the distinction in relation to Sulawesi languages.

5.4.2 Topicalisation

Two examples definitely did not fall into the category of irrealis mood. In (52) the speaker has just previously interrupted another speaker to correct a description of an event in which only the interrupting speaker had personally participated. With a particularly emphatic tone this speaker starts a long and uninterrupted passage, repeating the first person pronoun before proceeding.

- (52) 325
Aku // aku ras bapa-ndu mbarənda n-jumpai-sa ku pəngadilən ah.
 I // I and father-your past N-meet-sa at court DEM
 'It was ME, me and your late father who met (her) at the court.'

This is a topicalised construction; the emphasis on the first person pronoun subject is shown by its repetition after the intonation break. The use of the *N*- form *njumpai* allows the first person pronoun to be the subject of the sentence and to appear in sentence initial position.

Another example occurred further on in the same emphatic utterance:

- (53) 327
Kami n-jumpai-sa dua-dua ras katua mbarənda.
 we N-meet-sa together with older.sister late
 'We met together with (our) late sister.'

5.4.3 Undergoer incorporation

In the text a very high proportion of the *N*- forms were associated with just three stems. One of these, *nuan* (stem = *suan* 'plant'), had four tokens in one sentence with three further tokens in quick succession. Native speakers appear to regard these forms, consisting of *N*-form +indefinite P, as a form of nominalisation, indicated by the gloss in (54).

- (54) 105
Ku juma dahin-ku rusur nuan wartel //
 in garden job-my often N-plant carrots //
nuan kentang // nuan lacina əntah pe nuan markisah.
N-plant potatoes // N-plant chilli or pe N-plant passion-fruit
 'Frequently my job in the garden is carrot-planting, potato-planting, chilli-planting or passionfruit-planting.'

In describing these constructions Woollams (1996:179) points out that between the verb and its undergoer there can be no intervening material, not even a phonological pause; nor can the undergoer be modified in any way. This has the effect of creating an intransitive verb wherein the undergoer is incorporated into the verb. This is consistent with the proposal of Starosta, Pawley and Reid (1982) that the focus affixes were once nominalising affixes which became reanalysed as verbal focus affixes.

Table 7: Summarises the distribution of main clause functions of the *N*- forms

1. Irrealis mood	16
2. Topicalisation	2
3. Undergoer Incorporation	4 (plus 4 in subordinate clauses)
Total	22

5.5 The suffix *-sa*

As briefly discussed in §2.4 the *N-* forms may take a suffix of the form *-sa* (or *-ca*). Although this suffix cannot be used in all contexts with *N-* forms, in many instances where it is absent the judgement of a native speaker is that it could occur, and that its presence would make the construction grammatically more 'complete'. It has been claimed that this suffix is an allomorph of the independent form of the third person pronoun, *ia*, for the object¹⁵ or undergoer of the *N-* forms (e.g. Woollams 1996:115). Example (12) above, repeated here as (55) is an example where an analysis of third person undergoer is apparently unproblematic.

- (55) *Maka ngadi mə ia ng-ajar-sa.*
 then stop *mə* she *N-learn-sa*
 'Then she stopped teaching (?them).'

However, in addition to being found in constructions without an overt undergoer NP, such as (55), it may also co-occur with the usual free form third person pronominal undergoer *ia*, or with a simple third person NP in the same clause, as in example (56). In these cases it appears to cross-reference the undergoer NP. This creates problems for the third person analysis, but they are not serious ones. One could simply say that *-sa* is a third person object marker which can be used either to represent or to cross-reference the object.

- (56) 638
Ia kang ng-gorəng-sa kuning-ndu?
 she INTER *N-cook-sa* yellow-your
 'Did she cook up your yellow (medicine)?'

However, *-sa* also occurs with non-third person pronouns, as in (57), an example from Woollams (1996:219; his example 6.195). This presents much more serious problems for an analysis of *-sa* as a third person object marker. To retain this analysis, one would have to say that it was an object marker that was not restricted as to person, and also that it could be used either to represent or to cross-reference the object.

- (57) (Woollams 1996:219, e.g. 6.195)
Pitu wari nari rəh kami mərəng-sa kam kərina.
 seven day from come we *N-attack-sa* you all
 'In seven days we shall come and attack you.'

Woollams interprets these examples of *-sa* as a perfective marker, an analysis that is incompatible with his own (and my) analysis of the *N-* form constructions in main clauses as imperfective. Additional problems for a third person object analysis arise from the fact that whether the undergoer is overt or absent, replacing *-sa* with a free form pronoun causes the construction in question to undergo semantic and syntactic changes, as (58a and b) show.

- (58)a. 325
Dibatalah ng-idah-sa si bage e.
 God EMPH *N-see-sa* REL thus DEM
 'May God witness this, that he is thus (doing).'

¹⁵ The term object, a syntactic designation, is used with reference to the alternative theories being discussed. It is interchangeable with the semantic term undergoer.

- b. *Dibata lah ng-idah ia arbahan si bage.*
 God EMPH *N*-see him do REL thus
 'Indeed God is watching him do this.'

In (58a) *-sa* apparently refers to the event summed up in the following headless RC *si bage e*. If a free third person pronoun, *ia*, is used in this clause, as in (58b), this leads to changes in its syntactic structure and meaning such that the free form *ia* functions as an undergoer, followed by a complement clause; *ia* cannot cross-reference a following NP (such as the headless relative in (58a)).

Even if these syntactic differences in the use of the two forms are not taken into account, the presence of the suffix *-sa* with an undergoer which is either first or second person is sufficient evidence to refute the claim that *-sa* is an alternative form of third person undergoer, either as an argument in itself, or as a third person cross-reference form. The discourse properties of the undergoer that *-sa* represents or cross-references do, however, correlate with the presence or absence of this suffix. The undergoer of an *N*-verb suffixed with *-sa*, whether co-occurring in the same clause, or mentioned in a previous clause, is always an NP that is specific and identifiable; that is, individuated. It may be a personal pronoun, a lexical noun marked with a definite marker of some description, a proper name or a specific clausal undergoer. In every case it marks specific rather than general properties of the NP, in terms of the flow of information. It may not therefore be coincidental that the particular form of the suffix *-sa* is the same as the prepositional oblique case marker in certain Philippines languages, where its major function, apart from locative marking, is to mark specificity of the following NP (D. Zorc, pers. comm.). It seems plausible that it was once a prepositional marker of the undergoer NP that immediately follows the *N*-verb, and from this position was reanalysed as a verbal suffix. At the same time the overt mention of the NP itself immediately after the verb became redundant under certain conditions.

The suffix *-sa* also represents the third person pronominal agent of the stative passive *ter*-forms. While the contexts under which this occurs have not yet been investigated, it is possibly the case that these contexts are the same as those for its occurrence on the *N*-verbs, that is, it cross-references a specific and identifiable NP. This is a plausible explanation for the case of the passive agent in (59), who was identified by name in the previous clause.

- (59) 226
La ter-dahi-sa ban-na ka tomat-na je sitik.
 not *ter*-care for-*sa* \emptyset -do-he PTL tomato-his PTL little
 'It can't be looked after by him, he does the tomatoes.'

The suffix *-sa* also occurs as an object of certain prepositions. Woollams (1996:116) gives examples of this function with *kempak* 'to', and *ras* (comitative) 'with' as in (60) (his example 4.75).

- (60) *I-pindo-na galah banci ia iading ras-sa.*
i-ask-he PURP able he stay with-*sa*
 'He asked if he might be able to stay with him.'

In both sets of circumstances, either as passive agent (or cross-reference to passive agent), and as the object of prepositions, *-sa* marks an oblique role. This gives support for the analysis of *-sa* as an erstwhile oblique preposition which, having lost its following NP under certain circumstances, has been reanalysed as a verbal or prepositional suffix which itself represents the NP.

5.6 Summary of section 5

The *N*- verb constructions function primarily to allow argument sharing between the main and subordinate clause where the actor is pivot. They are also lower in transitivity than their \emptyset - verb counterparts. The syntactic status of their undergoers reflects this lowered transitivity; they are generally unaffected, unindividuated undergoers, and not tracked in the discourse. However, certain constraints can motivate the use of an *N*- construction when the undergoer is specific and identifiable (as indicated by formal marking patterns, such as definite marking, pronominalisation or zero anaphora, or a proper name). In those cases, the suffix *-sa* may function as a cross reference form on the *N*- verb, though its presence is not obligatory. This use of *-sa* points to it being related to the oblique marking preposition found in Tagalog and other Philippines languages.

N- constructions are uncommon in main clauses, constituting only twenty percent of all uses of *N*- verbs. When they occur, they either indicate irrealis mode or are used for undergoer incorporating constructions or for topicalisation of the A argument.

It may be argued that as a widespread feature of the Austronesian family, the syntactic pivot function of the *N*- constructions, along with the vestiges of the focus system, are inherited features. The focus system as described by Starosta, Pawley and Reid (1982) wherein the verbal focus affixes originate from earlier Proto Austronesian (PAN) nominalisers provides information about the semantic role of the focused NP, and its pragmatic prominence.

If pragmatic prominence of the NP was the major motivation for the use of the \emptyset - construction (undergoer voice or patient focus) in Karo, then there would be no need for explanations. However, in Karo, this is not the case. There are a number of \emptyset - construction undergoers that are not topical or referential, and not marked with identifying markers. It seems clear therefore that while the \emptyset - constructions are higher in transitivity by a number of the parameters of transitivity described by Hopper and Thompson (1980), the *N*- constructions are lower, by these same parameters. This suggests that discourse transitivity has been a major factor in conditioning the use of either construction type (see also Wouk 1984).

6 Conclusion

This paper has shown that the Karo *N*- verb constructions cannot merely be regarded as alternative syntactically transitive constructions, as in Cirebon Javanese, for example (Ewing 1999), nor does it maintain the focus system as found in Philippine languages such as Tagalog, although its functions are clearly derived from an older focus system. Rather, Karo *N*- forms have several divergent functions. This, I suggest, is to a large extent a reflection of historical change.

The most common use of the *N*- forms is to manipulate the argument structure of subordinated constructions so that A may function as a syntactic pivot. This function is reconstructable for PAN as described by Starosta, Pawley and Reid (1982). According to Starosta et al. the focus affixes originate from earlier PAN nominalisers, and the focus system served to provide information about the semantic role of the pragmatically prominent NP in a clause. And indeed the coreferential pivot NP of the *N*- constructions in subordinate clauses still has pragmatic prominence.

Two infrequent functions, topicalisation and object incorporation, are also reconstructable for PAN. Topicalisation is a main clause function of the *N*- forms, and one does not necessarily entail detransitivisation. The topicalised A is frequently associated with *pe*, one of three particular emphatic particles (*pe*, *ngə*, and *mə*), which may be vestiges of an older noun phrase marking system associated with the verbal focus system. The nominalising function of undergoer incorporation, which is found in either main or subordinate clauses, does involve detransitivisation. Both these functions can be seen to reflect pragmatic prominence, as topicalised As are highly pragmatically prominent, and As of incorporated undergoers are necessarily more prominent than their undergoers.

However, in Karo, as in many other Western Austronesian languages, the focus system (or the remnants of it) no longer serves chiefly to mark pragmatic prominence. Rather, it is a marker of discourse transitivity as described by Hopper and Thompson (1980), with *N*-forms associated with low discourse transitivity and \emptyset - forms associated with high discourse transitivity (see also Wouk 1984). *N*- forms were found, with respect to several semantic and syntactic parameters, to be low in transitivity. In main clauses they are a marker of irrealis mood, although periphrastic adverbs in \emptyset - constructions are more commonly used for this purpose. A further aspect of the lowered transitivity manifested by the *N*- constructions is seen in the non-individuation of the undergoers of an unaffixed *N*- verb. The affixation of the suffix *-sa* on the verb marks their undergoer as an individuated entity. This suffix, analysed here as an earlier prepositional case marker, appears to be similar to the obliquely marked arguments of certain Philippines languages such as Tagalog.

The findings of this paper, then, can be seen to provide support for the interesting and challenging hypothesis put forward by Nothofer (1991), described at the beginning of this paper, that Karo and its near relatives may be part part of an older Palaeo-Hesperonesion group, related to other outlying western Austronesian languages.

References

- Anderson, Stephen, 1985, Inflectional morphology. In T. Shopen, ed. *Language typology and syntactic description 3: Grammatical categories and the lexicon*, 150-201. Cambridge: Cambridge University Press.
- Berg, René van den, 1996, The demise of focus and the spread of conjugated verbs in Sulawesi. In H. Steinhauer, ed., 1996:89-114.
- Cumming, Susanna, 1991, *Functional change: the case of Malay constituent order*. Berlin: Walter de Gruyter.
- Dixon, R.M.W., 1994, *Ergativity*. Cambridge: Cambridge University Press
- Ewing, M., 1999, The clause in Cirebon Javanese conversation. PhD dissertation, University of California, Santa Barbara.
- Fox, B. and P.J. Hopper, eds, 1994, *Voice: form and function*. Amsterdam/Philadelphia: John Benjamins.
- Himmelman, N., 1996, Person marking and grammatical relations in Sulawesi. In H. Steinhauer, ed., 1996:115-136.
- Hopper, P. and S.Thompson, 1980, Transitivity in grammar and discourse. *Language* 56:251-99.
- Padia si Ndekah (si Nipegendek)*, The Old Testament (abbreviated version), 1953, tr. by J.H. Neumann. Djakarta: Lembaga Alkitab Indonesia.

- Padia si Badia ras Ende-Enden* (The Old and New Testaments and Hymns), 1977, Jakarta: Lembaga Alkitab Indonesia
- Lehmann, C., 1988, Towards a typology in clause linkage. In John Haiman and Sandra Thompson, eds *Clause combining in grammar and discourse*, 181-225. Amsterdam: John Benjamins.
- Norwood, C., 2000, The Kato Batak nasal verb suffix *-sa*: third person pronoun or transitiviser? *Melbourne Papers in Linguistics and Applied Linguistics*, Vol 0 Issue 1. 69-82.
- Nothofer, Bernd, 1991, The relationship between the languages of the Barrier Islands and the Sulawesi-Philippines languages. In Tom Dutton and Darrell Tryon, eds, 1992, *Language Contact and Change in the Austronesian world*, 389-409. Berlin: Mouton de Gruyter
- Ross, Malcolm, 1995, Some current issues in Austronesian linguistics. In D.T. Tryon, ed. *Comparative Austronesian dictionary* 1:45-120. Berlin: Mouton de Gruyter
- Schachter, Paul, 1984, Semantic-role-based syntax in Toba Batak. In Paul Schachter, ed. *Studies in the Structure of Toba Batak*, 122-149. UCLA Occasional Papers in Linguistics, 5.
- Starosta, Stanley, Andrew K. Pawley and Laurence Reid, 1982, The evolution of focus in Austronesian. In Amrin Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, 2:145-170. Canberra: Pacific Linguistics.
- Steinhauer, Hein, ed., 1996, *Papers in Austronesian Linguistics No.3*. Canberra: Pacific Linguistics.
- Woollams, G. 1996, *A grammar of Karo Batak*. Canberra: Pacific Linguistics.
- Wouk, F., 1984, Scalar Transitivity and Trigger choice in Toba Batak. In Paul Schachter, ed. *Studies in the Structure of Toba Batak*, 195-219. UCLA Occasional papers in Linguistics, 5.
- Yusmaniar, Lubis A. Hamid, S. Simorangkir and B. Surbakti, 1987, *Struktur Bahasa Karo*. Pusat Pembinaan dan Pengembangan Departemen Pendidikan dan Kebudayaan. Jakarta

Nominative and genitive case alternations in Bonggi

MICHAEL E. BOUTIN

1 Introduction¹

Bonggi is a western Malayo-Polynesian language spoken by approximately 1,400 people on Banggi and Balambangan islands in the Kudat District of Sabah, Malaysia. According to Blust (1998), the Sabahan languages, including Bonggi, form a primary branch of a North Borneo subgroup whose other primary branch is the North Sarawak languages. Bonggi has sentences like the following.

- (1) *Sia kiohol ulakng.*²
sia -in-kohol-∅ ulakng
3SG.NOM REAL-bite-ISA.UND snake
'He was bitten by a snake.'
- (2) *Mipa? nya kiohol?*
mipa? nya -in-kohol-∅
when 3SG.GEN REAL-bite-ISA.UND
'When was he bitten?'

The verb morphology in (1) and (2) is identical. In both examples, the verb morphology indexes the clause undergoer; however, in (1) the undergoer (*sia* '3SG.NOM') is in nominative case, whereas in (2) the undergoer (*nya* '3SG.GEN') is in genitive case. It is well known that actors in many western Malayo-Polynesian languages occur in genitive case

¹ I am grateful to Adrian Clynes, Paul Kroeger, David Moody, Bhuvana Narasimhan, Robert Van Valin, Jr. and Fay Wouk who provided valuable feedback on earlier versions of this paper. Responsibility for the current form and content, however, rests entirely with me.

² Abbreviations used: ACH achievement; ACT actor; ACY activity; ARG argument; DAT dative; DEF definite; EXC exclusive; GEN genitive; IRREAL irrealis; ISA induced state of affairs; LS logical structure; NOM nominative; NP noun phrase; NONACT nonactor; PL plural; PASS passive; PCS precore slot; PER perfect; PN personal noun; PSA privileged syntactic argument; REAL realis; RRG Role and Reference Grammar; SG singular; SR semantic representation; STAT stative and UND undergoer. The PSA in Bonggi occurs in the English free translation in **bold**. The symbol $_$ means 'linked.'

when they are not indexed by the verb; however, main clauses in which the argument indexed by the verb is an undergoer in genitive case are virtually unknown. This paper explains nominative and genitive case alternations in Bonggi.

The starting point for this paper is Silverstein's (1981) claim that case marking is a dependent variable which is contingent on the interaction of different independent variables. The following five semantic-pragmatic variables must be understood in order to account for case marking in Bonggi: (1) inherent lexical content of NPs (§3), (2) clause-level propositionality (§4), (3) clause linkage (§5), (4) tense (§6) and (5) the discourse pragmatic relations topic and focus (§7) (cf. Silverstein 1993:474).

The theoretical framework used to account for alternations between nominative and genitive case is Role and Reference Grammar (RRG).³ Section 8 summarises my explanation for these case alternations and reviews the implications of this paper for the study of western Malayo-Polynesian case systems.

2 Overview of voice in Bonggi

Voice is a mechanism that selects one nominal element in each clause for syntactic prominence. Because voice alternations consist of different ways of presenting a verb with its arguments, they are normally associated with verbs which have two or more arguments. In active voice, the actor is the syntactically prominent nominal; in passive voice, the undergoer is the syntactically prominent nominal. Verbs with one argument offer no alternative; thus, the syntactically prominent nominal is the single argument. However, single argument verbs in Bonggi do **not** all belong to the same verb class, as verb classes are defined semantically.⁴ Single argument verbs have a semantic valency of one and are labelled 'intransitive' voice in Table 1 following Nida (1949:168). Table 1 shows the underlying forms of indicative and imperative mood verbal affixes.

Induced states of affairs are complex in that one state of affairs brings about another. They are semantically transitive, having a valency of two. As seen in Table 1, five basic voice alternations can occur in induced states of affairs: active voice and four different passive voice forms, including a periphrastic passive formed with the auxiliary verb *anu*.⁵ All five voice alternations are morphologically marked, with the morphological marking occurring on the auxiliary verb in periphrastic passive constructions.

³ RRG theoretical notions are introduced in §4.1. For a brief overview of RRG see Kishimoto (1996:250-253) or Van Valin (1990:222-230, 1991:154-171, 1995). For a more extensive overview see Van Valin (1993).

⁴ Evidence for Bonggi verb classes can be found in Boutin (1994).

⁵ See Wolff (1996) for use of similar passive terminology for Philippine-type languages. See Boutin (forthcoming) for a description of the development of periphrastic passive constructions in Bonggi.

Table 1: Basic voice-related affixes in Bonggi

Verb class	Semantic valency	Voice	Syntactically prominent nominal	INDICATIVE		IMPERATIVE
				Irrrealis	Realis	
attributive state	1	intransitive	undergoer	m-		
achievement	1	intransitive	undergoer	mə-	in-	
accomplishment	1	intransitive	undergoer	-əm-	-in--əm-	
activity	1	intransitive	actor	-əm-	-in--əm-	(bare root)
induced state of affairs	2	active	actor	ng-	in-ng-	pəng-
	2	direct passive	undergoer	-ən	-in-	-a?
	2	periphrastic passive	undergoer	anu-ən	-in-anu	anu-a?
	2	local passive	marked undergoer	-an	-in- -an	-i
	2	instrumental passive	instrument (non-argument)	pəng-	-in-pəng-	
	1 or 2	periphrastic non-argument passive	non-argument	gien	gien	

The five voice alternations are exemplified below by various forms of the verb root *bagi* ‘to divide’: (3) illustrates active voice with the actor (*sia* ‘3SG.NOM’) being the syntactically prominent nominal; (4) illustrates a direct passive with the undergoer (*louk nyu* ‘your fish’) being the syntactically prominent nominal; (5) is a periphrastic passive with the undergoer (*louk nyu* ‘your fish’) being the syntactically prominent nominal; (6) is a local passive with the benefactive (*ou* ‘1SG.NOM’), which is a marked undergoer, being the syntactically prominent nominal. Finally, (7) is an instrumental passive with the instrument (*badi? ku* ‘my machete’) being the syntactically prominent nominal.⁶

(3) *Sia imagi louk nyu.*
sia in-ng-bagi louk nyu
 3SG.NOM REAL-ISA.ACT-divide fish 2PL.GEN
 ‘He divided your fish.’

(4) *Louk nyu biagi nya.*
louk nyu -in-bagi-0 nya
 fish 2PL.GEN REAL-divide-ISA.UND 3SG.GEN
 ‘Your fish was divided by him.’

⁶ Although instrumental passives occur, they are extremely rare in Bonggi.

- (5) *Louk nyu* ~ *inanu* *nya* *imagi*.
louk nyu *-in-anu-∅* *nya* *in-ng-bagi*
 fish 2PL.GEN REAL-PASS-ISA.UND 3SG.GEN REAL-ISA.ACT-divide
 'Your fish was divided by him.'
- (6) *Ou* *bigiadm* *nya* *louk nyu*.
ou *-in-bagi-an* *nya* *louk nyu*
 1SG.NOM REAL-divide-ISA.MARKED.UND 3SG.GEN fish 2PL.GEN
 'He divided your fish for me.'
- (7) *Badi?* *ku* *pimagi* *nya* *louk nyu*.
badi? *ku* *-in-pəng-bagi* *nya* *louk nyu*
 machete 1SG.GEN REAL-ISA.INSTRUMENT-divide 3SG.GEN fish 2PL.GEN
 'My machete is what he used to divide your fish.'

Three different forms of the passive auxiliary are used in periphrastic passive constructions to index undergoers: *nuan* (*anu* + *-ən*) occurs in irrealis modality as illustrated in (8); *inanu* (*-in-* + *anu*) occurs in realis modality as illustrated in (5); and *nua?* (*anu* + *-a?*) occurs in imperative mood as illustrated in (9). These three forms of the passive auxiliary are morphologically related, with the initial /a/ of the root *anu* being deleted when the root is suffixed (Boutin, forthcoming).⁷

- (8) *Louk nyu* *nuan* *nya* *magi*.
louk nyu *anu-ən* *nya* *ng-bagi*
 fish 2PL.GEN PASS-ISA.UND 3SG.GEN ISA.ACT-divide
 'Your fish will be divided by him.'
- (9) *Nua?* *na* *magi!*
anu-a? *na* *ng-bagi*
 PASS-ISA.UND.IMPERATIVE PER ISA.ACT-divide
 'Divide it up!'

Bonggi verbal affixes signal verb class and, for verbs with two or more arguments, the semantic role of the nominal indexed by the affix. For example, attributive states, achievements and accomplishments have a single argument which is an undergoer; however, each verb class is uniquely marked. Attributive states are marked by *m-*; achievements with irrealis modality are marked by *mə-*; and accomplishments with irrealis modality are marked by *-əm-*. If the primary function of these affixes was to signal the semantic role of the nominal, then we would expect all three classes to share the same affix. Similarly, the actor is the syntactically prominent nominal for both activities and active voice induced states of affairs; however, the two classes of verbs are uniquely marked. Activities are marked by *-əm-* whereas active voice induced states of affairs are marked by *ng-*. The prefix *ng-* has a dual function in that it signals both verb class (i.e. induced state of affairs) and the semantic role of the indexed nominal (i.e. actor). Similarly, *-ən* also has a dual function in that it signals both verb class (i.e. induced state of affairs) and the semantic role of the indexed nominal (i.e. undergoer).

⁷ The affixes *-ən* 'UND', *-in-* 'REAL' and *-a?* 'ISA.UND.IMPERATIVE' are the same affixes as are used with non-periphrastic passives when the pivot is an unmarked undergoer (cf. Table 1).

Bonggi, like many other western Malayo-Polynesian languages, not only allows semantic arguments of the predicate to be the syntactically prominent nominal, but it also allows non-arguments of the predicate to be the syntactically prominent nominal. For example, in (10) although the location is not an argument of the verb *milakng* 'to lie down', it is the syntactically prominent nominal. When the syntactically prominent nominal is not an argument of the verb, a periphrastic construction with *gien* 'place' normally occurs (cf. §5.2.1). The other option for making non-arguments the syntactically prominent nominal is an instrumental passive as in (7); however, instrumental passives rarely occur and are only used with instruments. As seen in the bottom of Table 1, periphrastic constructions with *gien* 'place' can occur in clauses with a semantic valency of either one or two.

- (10) *Kati? gien ku milakng.*
kati? gien ku -əm-ilakng
 here place 1SG.GEN ACY-lie.down
 'Here is where I lie down.'

3 Nominal marking

"Case is a system of marking dependent nouns for the type of relationship they bear to their heads" (Blake 1994:1;13). In Bonggi, the form of marking is contingent upon the type of nominal expression. Three types of nominal expressions are distinguished: common nouns (§3.1), personal nouns (§3.2) and personal pronouns (§3.3).⁸ Only personal pronouns and personal nouns receive overt case marking; personal pronouns are inflected, while personal nouns are preceded by one of two proclitics.

3.1 Common nouns

Bonggi, like many other western Malayo-Polynesian languages, makes a distinction between common nouns and personal nouns in terms of case marking. For example, in (11) the nominal expression *Umal* is case marked because it is a personal noun in nominative case; however, in (12) the nominal expression *daidn na* 'the trail' is not case marked because it is a common noun.⁹

- (11) *Si Umal miliug.*
si Umal m-liug
 PN.NOM Umal STAT-tall
 'Umal is tall.'

⁸ Some western Malayo-Polynesian languages include a fourth type of nominal expression, namely deictic pronouns. With respect to case marking, Bonggi deictic pronouns are treated like common nouns. For a discussion of case marking distinctions among deictic pronouns see Schachter and Otnes (1972:91-93) for Tagalog or McFarland (1974:148ff.) for Bikol.

⁹ In the majority of the languages of the Philippines, NPs including common nouns are case marked; however, core arguments are not case marked in the Sama Bajau languages (Pallesen 1985:97; Walton 1986:2).

- (12) *Ntimpad daidn na.*
m-timpad daidn na
 STAT-straight trail the
 'The trail is straight.'

Case is distinguished from case markers. Both the personal noun *Umal* in (11) and the common noun *daidn* in (12) receive nominative case. However, common nouns which occur in nominative case are not case marked (e.g. *daidn na* 'the trail' in (12)), whereas personal nouns which occur in nominative case are case marked (e.g. *si Umal* in (11)).

3.2 Personal nouns

Personal nouns include personal names (e.g. *Umal* in (11)), nicknames (e.g. *Lonti* 'hang down' in (13)), some kinship terms (e.g. *ama?* 'father' in (14)) and the indefinite substitute word *anu* when it means 'what's-his-name' as in (15).

- (13) *Si Lonti miliug.*
si Lonti m-liug
 PN.NOM Lonti STAT-tall
 'Lonti is tall.'
- (14) *Ntuhhal si ama?*
m-tuhhal si ama?
 STAT-thin PN.NOM father
 'Father is thin.'
- (15) *Si anu ntuhhal.*
si anu m-tuhhal
 PN.NOM what's-his-name STAT-thin
 'What's-his-name is thin.'

With the exception of vocatives, personal nouns are always preceded by a personal noun marker (cf. Schachter & Otnes 1972:95 for Tagalog). Personal nouns exhibit a two-way case distinction, with all non-nominative personal nouns being marked with a phonologically conditioned variant of *ny*. Unlike *si* 'PN.NOM' which has both a case marking function and a noun-class marking function, *ny* 'PN' only has a noun-class marking function in that it distinguishes personal nouns from common nouns. For example, in (16) the grammatical marker *ny* 'PN' occurs with both genitive case *ama?* 'father' and dative case *Umal*.

- (16) *Si Mual imori siidn ny ama? di ny Umal.*
si Mual -in-ng-bori siidn ny ama? di ny Umal
 PN.NOM Mual REAL-ISA.ACT-give money PN father to.DAT PN Umal
 'Mual gave father's money to Umal.'

When personal nouns are used as vocatives, usually only the last syllable of the noun occurs. For example, in (17) the personal name *Umal* occurs in the vocative form *Mal*, and in (18) the kinship term *ama?* occurs in the vocative form *ma?*. Vocatives are not dependents in constructions, but rather stand outside constructions. Since they do not mark the relation of dependent to head, they are not case marked (Blake 1994:9).

- (17) *Mal, kana? gulu!*
 Umal come.here first
 'Umal, come here first!'
- (18) *Ma?, kana? gulu!*
 Father come.here first
 'Father, come here first!'

The semantic basis for the distinction between personal nouns and common nouns is found in the animacy hierarchy (cf. Dixon 1994:85). Personal names and some kinship terms are higher on the hierarchy than common nouns which refer to people. In Bonggi, non-collateral consanguineal kinship terms in the first and second generation above the speaker and addressee (i.e. the speaker's and addressee's parents and grandparents) are treated as personal nouns. Furthermore, Bonggi and certain Bornean languages have what are known as 'death/mourning-names' which are used as terms of reference and terms of address for close relatives of the deceased.¹⁰ In terms of case marking, death/mourning-names are treated as personal names. This is illustrated in (19) by the presence of the nominative case marker *si* 'PN.NOM' before *obos* which is a death/mourning-name for the second oldest male child who has recently suffered the loss of a parent.

- (19) *Si obos ntuhul.*
si obos m-tuhul
 PN.NOM 2nd.oldest.male.orphan STAT-thin
 'That second oldest male who recently suffered the loss of a parent is thin.'

3.3 Personal pronouns

Like personal nouns, personal pronouns receive overt case marking. Unlike personal nouns, personal pronouns exhibit a three-way case marking distinction as shown in Table 2.

Table 2: Bonggi pronouns

	NOMINATIVE	GENITIVE ¹¹	ACCUSATIVE/DATIVE
1 singular	<i>ou</i>	<i>ku</i>	<i>diaadn</i>
1&2 singular	<i>kita</i>	<i>ta</i>	<i>dihita</i>
1 plural-inclusive	<i>kiti</i>	<i>ti</i>	<i>dihiti</i>
1 plural-exclusive	<i>ihi</i>	<i>mi</i>	<i>dihhi</i>
2 singular	<i>aha</i>	<i>nu</i>	<i>diha</i>
2 plural	<i>uhu</i>	<i>nyu</i>	<i>dihu</i>
3 singular	<i>sia</i>	<i>nya/na</i>	<i>nya</i>
3 plural	<i>siga lama</i>	<i>siga lama/nda</i>	<i>siga lama</i>

Nominative case is reserved for the syntactically prominent nominal in a clause. For example, in (20) *sia* '3SG.NOM' is inflected for nominative case because it is both a pronoun and the syntactically prominent nominal (cf. Table 2).

¹⁰ See Needham (1954a, 1954b) for a discussion of death/mourning-names in Borneo.

¹¹ Genitive case pronouns are enclitics which do not affect stress in the preceding word.

- (20) *Sia* *kindi* *bali nu.*
sia *ki-n-di* *bali nu*
 3SG.NOM GOAL-DIRECTIONAL-to house 2SG.GEN
 'She is going to your house.'

Genitive case takes its name from its function of encoding the possessor in NPs such as *bali nu* 'your (2SG.GEN) house' in (20). In such constructions, the genitive is an adnominal case which marks noun phrases as dependents of nouns. The most common non-adnominal use of genitive pronouns is to encode actors when they are not the syntactically prominent nominal (e.g. *nya* '3SG.GEN' in (4)). Other non-adnominal uses of genitive case (e.g. *nya* '3SG.GEN' in (2)) are discussed in §4.2, §5, §6 and §7.

The third class of pronouns is used for nonactors which are not the syntactically prominent nominal including: (a) undergoers which are not indexed by the verb, such as *diaadn* '1SG.NONACT' in (21); (b) non-macroroles such as *diaadn* '1SG.NONACT' in (22);¹² and topicalised pronouns such as *diaadn* '1SG.NONACT' in (23). The presence of *di* 'DAT' distinguishes dative (e.g. *diaadn* in (22)) from accusative (e.g. *diaadn* in (21)) pronouns.

- (21) *Sia* *mori* *diaadn* *siidn.*
sia *ng-bori* *diaadn* *siidn*
 3SG.NOM ISA.ACT-give 1SG.NONACT money
 'He gives me money.'
- (22) *Sia* *mori* *siidn di* *diaadn.*
sia *ng-bori* *siidn di* *diaadn*
 3SG.NOM ISA.ACT-give money to.DAT 1SG.NONACT
 'He gives money to me.'
- (23) *Diaadn,* *ndou* *melou.*
diaadn, *nd-ou* *m-lou*
 1SG.NONACT NEGATIVE-1SG.NOM STAT-embarrass
 'As for me, I am not embarrassed.'

3.4 Grammatical case marking versus adjunct marking

Argument case markers (including adpositions) are distinguished from adjunct markers (including adpositions). Adpositions which mark arguments of the verb have a case marking function, whereas adpositions which mark non-arguments (adjuncts) have an adverbial function.¹³ This is in accord with a basic distinction which is often made in the analysis of case systems; that is, the difference between grammatical case and adjunct marking.

For example, the preposition *di* 'to/at' can mark either locative arguments or locative adjuncts. Thus, *diaadn* '1SG.NONACT' in (22) is the locative-goal argument of the verb *mori* 'give', and the preposition *di* 'to.DAT' marks this argument as a syntactically oblique argument. On the other hand, prepositions which mark non-arguments or adjuncts have an adverbial function, e.g. *di* 'at' in (24). The locative phrase *di sungi na* 'at the river' in (24) is not an argument of the verb *lemongi* 'swim'; instead, it is a locative adjunct.

¹² Macroroles are the two primary arguments of a transitive predicate (cf. §4.1).

¹³ Arguments of the verb are defined by logical structures. See §4.1 for a discussion of logical structures.

- (24) *Sia lemongi di sungi na.*
sia -əm-longi di sungi na
 3SG.NOM ACY-swim at river the
 'He swims at the river.'

The following major points regarding case have been made in §3. A distinction was made between case and case markers. Although common nouns receive case, they are not case marked. Only personal nouns and personal pronouns receive overt case marking. Personal nouns are preceded by one of two proclitics: *si* 'PN.NOM' marks nominative case personal nouns, while *ny* marks all other personal nouns with the exception of vocatives. Personal pronouns are inflected for one of three different cases: nominative, genitive or accusative/dative. Prepositions which mark arguments of the verb have a case marking function, whereas prepositions which mark non-arguments have an adverbial function.

4 Clause-level propositionality

The clause is both the basic unit in syntax and the starting point for the analysis of case. Section 4.1 provides an overview of clause analysis in RRG, while §4.2 shows how Bonggi case marking reflects the relationships between predicates and their argument(s).

4.1 Predicate-argument relations in RRG

In RRG the semantic relationship between a predicate and its arguments is expressed by **Logical Structures (LSs)**. LSs provide a formal semantic representation for each verb and they consist of predicates, their arguments and a small set of operators (Van Valin 1990:223). Semantic representations in RRG are based on Dowty's (1979) theory of verbal semantics in which verbs are classified into four basic *Aktionsart* classes: states, achievements, accomplishments and activities.

Stative clauses are illustrated by attributive constructions in (11), (12), (13), (14), (15), (19) and (23). The LS for all attributive statives is shown in (25) where the first argument 'x' is the attributant and the second argument **pred'** is the attribute (Van Valin & LaPolla 1997:103). In Bonggi, attributive statives are prefixed with *m-*.¹⁴

- (25) LS for attributive statives: **be'** (x, [**pred'**])

Whereas the LS in (25) is the LS for all attributive statives, the semantic representations (SR) for (11) and (12) are shown in (26) and (27).

- (26) SR for (11): **be'** (Umal, [**tall**'])
 (27) SR for (12): **be'** (*daidn* 'trail', [**straight**'])

Basic *Aktionsart* classes depict spontaneous states of affairs; however, states of affairs can also be induced. Induced states of affairs are complex in that one state of affairs brings about another state of affairs. The LS for induced states of affairs is ϕ CAUSE ψ , where ϕ is a causal state of affairs which induces another state of affairs ψ . For example,

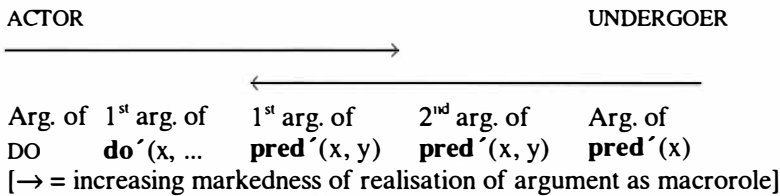
¹⁴ Regular morphophonemic alternations account for the variation in surface forms. Nasal assimilation makes the nasal *m-* homorganic with following nonsonorant consonants (e.g. (12), (14), (15) and (19)). Vowel epenthesis inserts vowels between prefixes and sonorant consonants (e.g. (11), (13) and (23)).

(16) is an induced accomplishment clause. "Accomplishments are coded by BECOME, which codes change over some temporal span, plus a state predicate" (Van Valin & LaPolla 1997:104). The LS for *bori* 'give' is shown in (28a), whereas the SR for (16) is shown in (28b). The ϕ portion of the LS in (28a) is an activity, while the ψ portion is an accomplishment. The second argument position in the ϕ portion of the SR in (28b) is \emptyset (i.e. not specified) since the causing activity is not specified (cf. Van Valin 1990:225).

- (28)a. LS for *bori* 'give': [do'(x, [predicate'(x)])] CAUSE [BECOME have'(y, z)]
- b. SR for (16): [do'(Mual, \emptyset)] CAUSE [BECOME have'(Umal, *siidn ny ama?* 'father's money')]

Actor and **undergoer** are the two primary arguments of a transitive predicate, either one of which may be the single argument of an intransitive verb (Van Valin 1993:43). 'Actor and undergoer are generalisations across classes of specific argument positions in logical structure' (Van Valin & LaPolla 1997:142). The relationship between macroroles and argument positions in LS is captured in the Actor-Undergoer Hierarchy in (29) (Van Valin & LaPolla 1997:146). This double hierarchy states that the argument position that is leftmost on the cline will be the actor and the argument position that is rightmost will be the undergoer. This is the unmarked situation; marked assignments to undergoer are possible.

(29) **Actor-Undergoer Hierarchy**



The number of macroroles a verb takes is either \emptyset , 1 or 2, and is largely predictable from the LS of the verb (Van Valin 1993:46-47). Default principles for macrorole assignment are shown in (30).

- (30) DEFAULT MACROROLE ASSIGNMENT PRINCIPLES:
 - a. Number: the number of macroroles a verb takes is less than or equal to the number of arguments in its LS.
 - 1. If a verb has two or more arguments in its LS, it will take two macroroles.
 - 2. If a verb has one argument in its LS, it will take one macrorole.
 - b. Nature: for verbs which take one macrorole,
 - 1. If the verb has an activity predicate in its LS, the macrorole is actor.
 - 2. If the verb has no activity predicate in its LS, the macrorole is undergoer.

In (25), because the second argument is a predicate, it cannot function as an argument. Thus, despite having two argument positions ('x' and 'y'), attributive statives have only one macrorole. This follows from the principle in (30a.2). The nature of the single macrorole is predictable from (30b); that is, the single macrorole in (11) is an undergoer since there is no activity predicate in its LS in (25).¹⁵

¹⁵ Activity predicates are predicates with **do'** in their LS.

According to (30a.1), the verb *bori* ‘give’ has two macroroles since its LS in (28a) has three arguments: ‘x’, ‘y’ and ‘z’. In (28a) **do**’ refers to a generalised unspecified activity predicate. **Do**’ has two argument positions. The first argument position in (28a) is occupied by ‘x’, the second by another LS, i.e. [**predicate**’ (x)]. The variable ‘x’ in (28a) refers to both the first argument of **do**’ and the only argument of **predicate**’. Because the same variable ‘x’ is used in both places, these arguments are coreferential. Coreferential arguments are counted as a single argument in LSs.

According to (29), ‘x’ in (28a) is linked to actor since ‘x’ is the first argument of **do**’ and ‘1st argument of **do**’ is leftmost on the cline in (29). Furthermore, according to (29), either ‘y’ or ‘z’ in (28a) can be an undergoer. In (16) ‘z’ (*siidn ny ama?* ‘father’s money’) is linked to undergoer. This is the unmarked choice for undergoer since ‘z’ is the second argument in the LS configuration **BECOME have**’ (y, z) and ‘2nd argument of **pred**’ (y, z)’ is rightmost on the cline in (29).

Macroroles provide the primary link between semantic representation and syntactic representation. The linking system works both from semantics to syntax and from syntax to semantics. This is indicated by the double-headed arrows in Figure 1 which links the syntactic representation for (16) with its semantic representation in (28b).

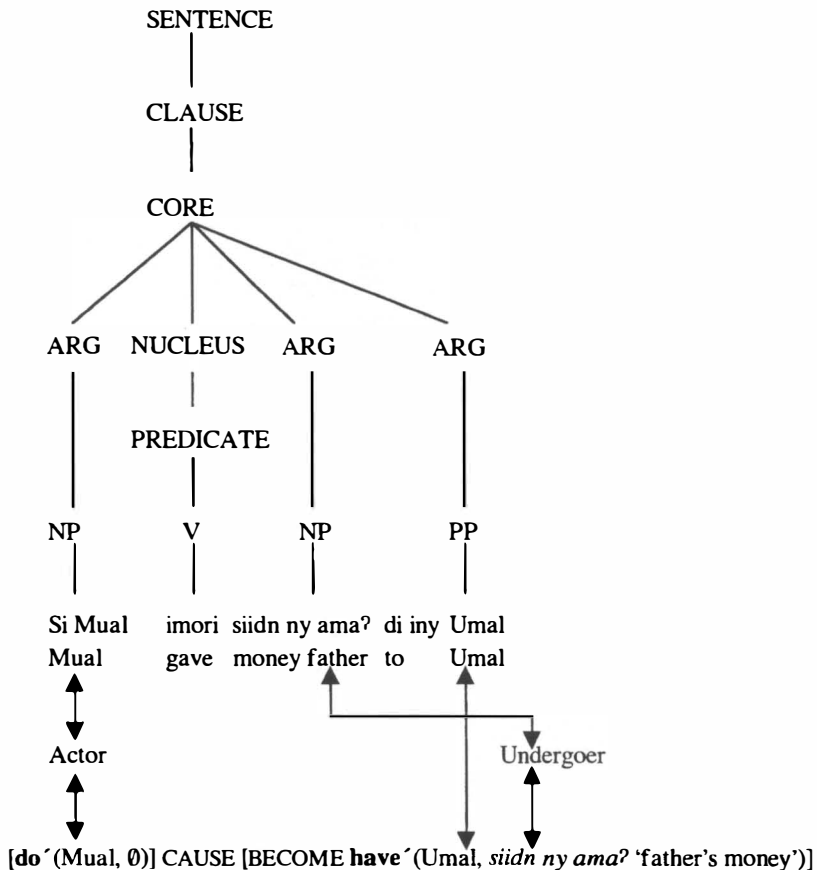


Figure 1: Linking syntax and semantics for (16)

Once arguments have been assigned to macroroles, actor and undergoer are assigned to specific morphosyntactic statuses (Van Valin 1993:76). The most important morphosyntactic status is the **privileged syntactic argument** (PSA) which includes both pivots and controllers. Pivots are construction-specific and are defined as a restricted neutralisation of semantic roles and pragmatic functions for syntactic purposes (Van Valin 1995:466). For example, there is a restricted neutralisation of semantic roles of the omitted argument in the dependent clauses in (31) and (32). The omitted argument in (31) is an actor, whereas the omitted argument in (32) is an undergoer. Since the actor is omitted in (31) and the undergoer is omitted in (32), the restriction cannot be stated in terms of semantic roles. The omitted NP in (31) and (32) is the pivot of the dependent clause. The omitted NP must be coreferential with the controller in the matrix clause; otherwise, the argument cannot be omitted (cf. (33)).

- (31) *Sia mingin kiliid diha.*
 3SG.NOM want see 2SG.NONACT
 'He wants to look at you.'
- (32) *Sia mingin midadn nu.*
 3SG.NOM want be.seen 2SG.GEN
 'He wants to be noticed by you.'
- (33) *Sia mingin diha kiliid nya.*
 3SG.NOM want 2SG.NONACT see 3SG.NONACT
 'He wants you to look at him.'

To summarise, the semantic relationship between predicates and their arguments is expressed by logical structures (LSs). An RRG analysis of clauses (e.g. (16)) includes a syntactic representation as in Figure 1, a semantic representation as in (28b), and a small set of principles for linking the two types of representation. These principles include the Actor-Undergoer Hierarchy in (29) which captures the relationship between macroroles and argument positions in LSs, and the default macrorole assignment principles in (30) which determine from LSs the number and nature of macroroles. Macroroles provide the primary link between LS and syntax. The most important syntactic status is the privileged syntactic argument (PSA). §4.2 discusses language specific principles for selecting a PSA and assigning case.

4.2 Bonggi case marking rules

Part of the process involved in assigning actor and undergoer to specific morphosyntactic statuses is case and preposition assignment. The case marking rules for Bonggi are given in (34) (cf. Van Valin 1991:171; Van Valin 1993:73; and Narasimhan 1995).

- (34) **Case marking rules for Bonggi**
- a. The PSA takes NOMINATIVE case.
 - b. Non-PSA actors take GENITIVE case.
 - c. Non-PSA undergoers take ACCUSATIVE case.
 - d. Non-macrorole arguments take DATIVE case as their default case.
 - e. Dependent clause PSAs take GENITIVE case.

Nominative case is defined in terms of the notion PSA. The case marking rules in (34) only apply to personal pronouns and personal nouns in Bonggi since common nouns are not case marked (cf. §3.1). The case marking rules in (34) are exemplified below.

- (35) *Sia imori siidn di diaadn.*
sia -in-ng-bori siidn di diaadn
 3SG.NOM REAL-ISA.ACT-give money to.DAT 1SG.NONACT
 'He gave money to me.'

- (36) *Sia imori diaadn siidn.*
sia -in-ng-bori diaadn siidn
 3SG.NOM REAL-ISA.ACT-give 1SG.NONACT money
 'He gave me money.'

Bonggi verbal affixes normally index one nominal per clause, and this nominal is usually the PSA. Examples (35) and (36) are active voice constructions in which the actor (*sia* '3SG.NOM') is indexed by the morphology on the verb (i.e. *ng-*). Furthermore, the actor is the PSA in both (35) and (36). As seen in (37) and (38), PSAs control coreferential deletion across clauses. In (37) the controller (*sia* '3SG.NOM') is an actor (cf. (36)), whereas in (38) the controller (*ou* '1SG.NOM') is a marked undergoer. According to (34a), the PSA is assigned nominative case. Thus, the main clause controller in (37) and (38) receives nominative case.

- (37) *Sia imori diaadn siidn, ma? minili?.*
sia -in-ng-bori diaadn siidn, ma? -in--əm-uli?
 3SG.NOM REAL-ISA.ACT-give 1SG.NONACT money and REAL-ACY-return.home
 'He gave me money and [he/*I] returned home.'

- (38) *Ou biniriadn nya siidn, ma? minili?.*
Ou -in-bori-an nya siidn, ma? -in--əm-uli?
 1SG.NOM REAL-give-ISA.MARKED.UND 3SG.GEN money and REAL-ACY-return
 'I was given money by him and [I/*he] returned home.'

Examples (39), (40), (41) and (42) are passive constructions in which the nominal indexed by the verb morphology is an undergoer. In Bonggi, two types of passive constructions are used to signal that the nominal indexed by the verb morphology is an undergoer: non-periphrastic passives (e.g. (39), (40) and (41)) and periphrastic passives (e.g. (42)). Periphrastic passives have a passive auxiliary such as *inanu* in (42), whereas non-periphrastic passives do not.

- (39) *Siidn biniri nya di diaadn.*
siidn -in-bori-∅ nya di diaadn
 money REAL-give-ISA.UND 3SG.GEN to.DAT 1SG.NONACT
 'Money was given to me by him.'

- (40) *Siidn biriidn nya di diaadn.*
siidn bori-ən nya di diaadn
 money give-ISA.UND 3SG.GEN to.DAT 1SG.NONACT
 'Money is given to me by him.'

- (41) *Ou biniriadn nya siidn.*
ou -in-bori-an nya siidn
 1SG.NOM REAL-give-ISA.MARKED.UND 3SG.GEN money
 'I was given money by him.'
- (42) *Siidn inanu nya mori di diaadn.*
siidn -in-anu-∅ nya ng-bori di diaadn
 money REAL-PASS-ISA.UND 3SG.GEN ISA.ACT-give to.DAT 1SG.NONACT
 'Money was given to me by him.'

As seen in (28a), the verb *bori* 'give' has two possible undergoers, either 'y' or 'z'. The predicate **have'** in (28a) corresponds to **pred'** in the LS configuration **pred'** (y, z); 'y' is the '1st argument of **pred'** (y, z)' and 'z' is the '2nd argument of **pred'** (y, z)'. According to (29), 'z' is the unmarked undergoer and 'y' is a marked choice for undergoer since the '2nd argument of **pred'** (y, z)' is further to the right on the cline than the '1st argument of **pred'** (y, z)'.

In realis modality, when an undergoer which is indexed by the verb is the 'z' argument (i.e. the unmarked choice for undergoer) as is *siidn* 'money' in (39), the verb is morphologically unmarked (∅). However, when an undergoer which is indexed by the verb is the 'y' argument (i.e. the marked choice for undergoer) as is *ou* '1SG.NOM' in (41), there is a corresponding morphological markedness in the verb morphology with the addition of the suffix *-an*.¹⁶ Thus, semantic markedness correlates with morphological markedness in realis modality.

In irrealis modality, the verb is morphologically marked even when the undergoer is the unmarked choice in terms of the hierarchy in (29). For example, in (40) the verb *biriidn* 'give-ISA.UND' is suffixed with *-an* indicating that the undergoer (i.e. *siidn* 'money') is an unmarked undergoer in terms of the hierarchy in (29).¹⁷ Stated in terms of Table 1, direct passives involve the unmarked choice for undergoer (i.e. the 2nd argument of **pred'** (y, z)), whereas local passives involve the marked choice for undergoer (i.e. the 1st argument of **pred'** (y, z)). As seen in Table 1, in irrealis modality both direct and local passives are morphologically marked; however, in realis modality only local passives are morphologically marked for undergoer.

As stated above, the nominal which is indexed by the verb morphology is normally the PSA. The verb in (39), (40) and (41) indexes the undergoer, which is the PSA in these clauses (cf. also (38)). According to (34a), the PSA is assigned nominative case. Thus, the undergoer in (39), (40) and (41) is assigned nominative case just as the actor is assigned nominative case in (35) and (36). However, because the PSA (*siidn* 'money') in (39) and (40) is a common noun, it is not case marked (§3.1). On the other hand, the PSA (*ou* '1SG.NOM') in (41) is inflected for case since it is a personal pronoun (§3.3). By (34b), the actor in (39), (40) and (41) is assigned genitive case since it is not the PSA.¹⁸

Example (36) is a dative-shift alternation in which the undergoer is the 'y' argument (i.e. *diaadn* '1SG.NONACT'), not the 'z' argument (i.e. *siidn* 'money'). According to (34c), the undergoer *diaadn* '1SG.NONACT' in (36) receives accusative case. In (35), (39) and (40) the 'z' argument (i.e. *siidn* 'money'), not the 'y' argument (i.e. '1SG'), is the undergoer.

¹⁶ *-an* is realised as *-adn* due to word-final nasals being preploded when preceded by non-nasal vowels.

¹⁷ *-an* is realised as *-idn* due to nasal prepllosion and vowel harmony.

¹⁸ Actors are considered core arguments and not oblique constituents in passive clauses (cf. Kroeger 1993:228-229).

According to (34d), the ‘y’ argument (i.e. ‘1SG’) receives dative case in (35), (39) and (40) since it is a non-macrorole argument. The preposition *di* ‘to.DAT’ marks the ‘y’ argument in the LS configuration ...BECOME **have**’ (y, z) when the ‘y’ argument is not the undergoer. Notice, however, that when the ‘z’ argument is not the undergoer as in (36), there is no overt dative case marker. Thus, (34d) applies only to the marked choice for undergoer in terms of the hierarchy in (29), never to the unmarked choice for undergoer. Or, stated in another way, the marked choice for undergoer receives dative case when it not the undergoer, whereas the unmarked choice for undergoer never receives dative case when it is not the undergoer.

In periphrastic passive constructions such as (42), two different arguments are indexed in the verb phrase. The passive auxiliary *inanu* in (42) indexes the undergoer (*siidn* ‘money’), whereas the main verb *mori* ‘ISA.ACT-give’ indexes the actor (*nya* ‘3SG.GEN’). The NP indexed by the passive auxiliary is the PSA in these constructions. This underscores an important fact about case marking in Bonggi. One cannot always determine which nominal is the PSA and consequently which nominal receives nominative case by simple reference to the morphology of the main verb.

One of the functions of passive constructions is to present non-actors as pragmatic pivots in order to maintain discourse topicality. **Pragmatic pivots** are syntactic pivots with pragmatic influence (cf. Van Valin 1993:65). In Bonggi, pragmatic pivots are determined by discourse topicality, as illustrated in (43). The speaker in (43) is a sultan who is angry at a group of people who keep bothering him. In (43) the sultan tells his guards what to do if they notice the people returning again. The discourse topic is the people who have angered the sultan. The NP which refers to these people is a pragmatic pivot in both clauses.

- (43)a. *Bakng midadn nyu pa malik si sida diti,*¹⁹
 if be.seen 2PL.GEN yet again PN.NOM plural here
- b. *nu-a?* *nyu ga mmati!*
anu-a? *nyu ga m-ng-pati*
 PASS-ISA.UND.IMPERATIVE 2PL.GEN EMPHATIC IRREALIS-ISA.ACT-die
 ‘If you notice **them** here again, kill **them**!’

According to (34a), the undergoer in periphrastic passive constructions receives nominative case since it is the PSA. In (44) the undergoer (*sia* ‘3SG.NOM’) is case marked since it is a personal pronoun; however, in (42) the undergoer (*siidn* ‘money’) is not case marked since it is a common noun. In (43b) the undergoer does not occur due to coreferential deletion. (44) shows that marked undergoers (i.e. the ‘y’ argument in the LS configuration **pred**’ (y, z)) can occur as the PSA in periphrastic passive constructions. By (34b), actors in periphrastic passive constructions receive genitive case (e.g. *nya* ‘3SG.GEN’ in (42), *nyu* ‘2PL.GEN’ in (43b) and *nu* ‘2SG.GEN’ in (44)).

- (44) *Gaabm pa sia nuan nu mori.*
gaabm pa sia anu-ən nu ng-bori
 better yet 3SG.NOM PASS-ISA.UND 2SG.GEN ISA.ACT-give
 ‘Moreover, you should give it to **him**.’

¹⁹ The plural marker *sida* is related to the third person plural Proto Austronesian pronoun **siDa*.

To summarise, Bonggi has pragmatic pivots which are determined by discourse topicality. In periphrastic passive constructions, the passive auxiliary indexes the PSA.

5 Clause linkage

Explanations for case marking are normally centered around the issues described in §4. The primary concern in §4 was to show how case is a reflection of the linking between semantic and syntactic representations. This section deals briefly with clause linkage.²⁰

Two basic aspects of clause linkage are the semantic relationship between the clauses being linked, and the syntactic relationship between them. The syntactic linkage may be ranked in terms of the strength of the syntactic bond between the units being linked. Similarly, the semantic linkage may be ranked in terms of the semantic relationship between the propositions being linked. In general, the closer the semantic relationship between two propositions, the stronger the syntactic bond (Van Valin 1993:111). Figure 2 provides an overview of the types of clause linkage and the degree of bonding presented in this paper (cf. Silverstein 1993:481; Van Valin 1993:112).

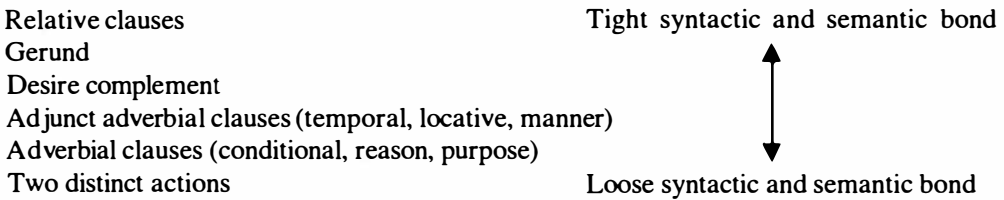


Figure 2: Types of clause linkage and degree of bonding

5.1 Loose linkage

Clauses which refer to two distinct actions are loosely linked, as in (45) where (45a) is linked to (45b) via the coordinate conjunction *ma?* ‘and’. In both (45a) and (45b) the PSA (*sia* ‘3SG.NOM’) is in nominative case. Two clauses joined by coordination function in the same way as equivalent simple clauses.

(45)a. *Inubu?* *sia* *ingengkabm* *karukng* *suga?* *kubal*
inubu? *sia* *i-ng-kengkabm* *karukng* *suga?* *kubal*
then 3SG.NOM REAL-ISA.ACT-grope gunnysack inserted skin

kerbou *na*
kerbou *na*
waterbuffalo the

‘Then he groped around for the gunnysack with the waterbuffalo skin

²⁰ My intent is not to present a detailed analysis of Bonggi clause linkage. Readers who are interested in a detailed RRG account of clause linkage are referred to Chapter 8 of Van Valin and LaPolla (1997).

- b. *maʔ sia miniliʔ kindi bali nya.*
maʔ sia -in-əm-uliʔ ki-n-di bali nya
 and 3SG.NOM REAL-ACY-return GOAL-DIRECTIONAL-to house 3SG.GEN
 and **he** returned to his house.'

Adverbial clauses divide into two types: clauses which are substitutable for by a single word, and those which are not (cf. Thompson & Longacre 1985:177ff.). Those which are substitutable for by a single word are referred to in Figure 2 as adjunct adverbial clauses which include time, location and manner clauses. The PSA in these clauses receives genitive case, as in (46), where the temporal adverbial clause in (46a) is linked to the main clause in (46b). The PSA (*gibatadn* 'dock') in the main clause (46b) receives nominative case according to (34a);²¹ however, the PSA in the temporal adverbial clause (46a) is in genitive case (i.e. *ku* '1SG.GEN') following (34e). The PSA always receives genitive case in adjunct adverbial clauses which function as temporal adjuncts (e.g. *ku* '1SG.GEN' in (46a)), locative adjuncts (e.g. \emptyset in (47b)),²² or manner adjuncts (e.g. *nyu* '2PL.GEN' in (48b)).

- (46)a. *Atakng ku mpanu,*
atakng ku -əm-panu
 while 1SG.GEN ACY-walk
 'While **I** was walking,
- b. *gibatadn irumbak na.*
gibatadn -in-rumbak na
 dock ACH.REAL-collapse PER
the dock collapsed.'
- (47)a. *Inubuʔ siga lama na igtimung*
inubuʔ siga lama na igtimung
 then some people DEF gather
 'Then **they** gathered together
- b. *nggien nual nya.*
nggien ng-sual nya
 place ISA.ACT-interrogate 3SG.NONACT
 where **they** could interrogate him.'
- (48)a. *Gaabm pa uhu mingisiadn lama leidn*
gaabm pa uhu m-ingisiadn lama leidn
 better yet 2PL.NOM STAT-pity people other
 'Moreover, **you** should pity other people
- b. *singgurua nyu mingisiadn deirdn nyu.*
singgurua nyu m-ingisiadn deirdn nyu
 like 2PL.GEN STAT-pity self 2PL.GEN
 like **you** pity yourselves.'

Adverbial clauses which are not substitutable for by a single word, including conditional, purpose, reason, concessive and substitutive clauses, are more loosely linked

²¹ Since *gibatadn* 'dock' is a common noun, it is not case marked (§2).

²² The actor in (47b) is deleted due to zero anaphora. Locative adverbial clauses have the shape of relative clauses (cf. Thompson & Longacre 1985:183).

to the main clause than temporal, locative and manner adverbial clauses. The PSA in these more loosely linked adverbial clauses receives nominative case, just like in loosely linked clauses found in coordinate constructions such as (45). For example, the PSA in the conditional clause in (49a) is in nominative case (i.e. *aha* '2SG.NOM') as is the PSA in the reason clause in (50b) (i.e. *ou* '1SG.NOM').

- (49)a. *Bakng ngua? aha kibori egas,*
bakng ng-kua? aha ki-bori egas
 if ACT-come 2SG.NOM ask-give rice
 'If **you** are coming to ask for rice,
- b. *ndaardn na egas mi.*
ndaardn na egas mi
 not.have PER rice 1PLE.EXC.GEN
 'our rice is gone.'
- (50)a. *Limidik ou*
-əm-lidik ou
 ACY-slash 1SG.NOM
- b. *pasal mingin ou nanam sikiou.*
pasal m-ingin ou ng-tanam sikiou
 because STAT-want 1SG.NOM ISA.ACT-plant cassava
 'I am slashing because I want to plant cassava.'

Desire complements were illustrated in (31), (32) and (33) of §4.1. The pivot of desire complements is also the undergoer of the matrix clause. If the pivot is not coreferential with the controller in the matrix clause, the pivot occurs in accusative case (e.g. *diha* '2SG.NONACT' in (33)) according to (34c). Thus, the pivot receives its case from its function as undergoer in the matrix clause, not from its function as PSA in the complement clause.

Gerunds are nominalised constructions which are often used as subordinate clauses with the meaning 'when/upon...' as in (51b) (cf. Shibatani 1988:99ff.). PSAs within gerunds always receive genitive case (e.g. *nya* '3SG.GEN' in (51b)).

- (51)a. *Inubu? sia minili?*
inubu? sia -in-əm-uli?
 then 3SG.NOM REAL-ACY-returned.home
 'Then **he** returned home.'
- b. *Pegdatakng nya di bali na, timeis na sia.*
pag-datakng nya di bali na -in-əm-teis na sia
 GERUND-coming 3SG.GEN to house the REAL-ACY-cry PER 3SG.NOM
 'Upon **his** coming to the house, **he** cried.'

5.2 Tight relative clause linkage

5.2.1 Relative clause formation strategies

Bonggi has two strategies for forming relative clauses. The most frequently occurring strategy involves deletion or gapping of the relativised nominal from the relative clause. For example, in (52) *siidn* 'money' is the head of the relative clause (relative clauses are in

curly brackets), but this argument is gapped in the relative clause itself. The verb *tiahu* 'stole' in the relative clause indexes the relativised nominal which is gapped. The NP which is gapped is the pivot of the relative clause. The gapped NP is a pragmatic pivot which must be coreferential with the head of the relative clause. Only pragmatic pivots can be relativised. The gapping strategy illustrated in (52) is the most common relativisation strategy in Philippine-type languages.

- (52) *Sia imori diaadn siidn*
sia -in-ng-bori diaadn siidn
 3SG.NOM REAL-ISA.ACT-give 2SG.NONACT money

{*tiahu nya.*}
-in-tahu-Ø nya
 REAL-steal-UND 3SG.GEN

'He gave me the money he stole.'

Gapped relative clauses can occur in active or passive voice. When the relative clause is in active voice as in (53), the gapped argument is an actor. When the relative clause is in direct passive voice as in (52), the gapped argument is an unmarked undergoer. When the relative clause is in local passive voice as in (54), the gapped argument is a marked undergoer.

- (53) *Sia nipu lama {moli gandubm.}*
sia ng-tipu lama ng-boli gandubm
 3SG.NOM ISA.ACT-cheat people ISA.ACT-buy corn
 'He cheats people who buy corn.'

- (54) *Sia nipu lama {biniriadn ku siidn.}*
sia ng-tipu lama -in-bori-an ku siidn
 3SG.NOM ISA.ACT-cheat people REAL-give-MARKED.UND 1SG.GEN money
 'He cheats people who have been given money by me.'

Gapping also occurs with periphrastic passives as seen in (55) where *siidn* 'money' is the head of the relative clause, and it is gapped in the relative clause. The passive auxiliary *inanu* indexes the gapped argument which is the pragmatic pivot. Gapped relative clauses with periphrastic passives show that the syntactic pivot (PSA) in these constructions is the argument indexed by the passive auxiliary (i.e. the undergoer), not the argument indexed by the main verb (i.e. the actor).

- (55) *Nubu? sia imori siidn {inanu*
nubu? sia -in-ng-bori siidn -in-anu-Ø
 then 3SG.NOM REAL-ISA.ACT-give money REAL-PASS-ISA.UND
inuga? di soig pahit na.}
-in-ng-suga? di soig pahit na
 REAL-ISA.ACT-insert to.DAT inside pocket 3SG.GEN
 'Then he gave the money which had been put inside his pocket.'

Example (56) illustrates that marked undergoers (i.e. the 'y' argument in the LS configuration **pred'** (y, z)) can occur as the PSA in gapped relative clauses with periphrastic passives (cf. (44)).

- (56) *Siga lama na ngedahap suhu lama {nuan nu*
siga lama na ng-dahap suhu lama anu-ən nu
 some people DEF ISA.ACT-arrest all people PASS-ISA.UND 2SG.GEN
mori.
ng-bori
 ACT-give
 'They arrest all the people who you give it to.'

The second strategy for forming relative clauses involves the use of the relative pronoun *nggien/gien* 'place'. In this strategy, the relative pronoun is the relativised nominal in the relative clause. For example, in (57) and (58) *nggien* 'place' is the relativised nominal in the relative clause. In (58) the relative clause is embedded in the main clause, whereas in (57) the relative clause is adjoined to the main clause.

- (57) *Inubu? sia ipanu ngirubm bunua {nggien nya limidik.}*
inubu? sia i-panu ng-irubm bunua nggien nya -əm-lidik
 then 3SG.NOM REAL-travel ACT-search area place 3SG.GEN ACY-slash
 'Then he traveled searching for an area **where** he could slash (for planting crops).'
- (58) *Inubu? sia ipanu ngirubm {nggien nya limidik.}*
inubu? sia i-panu ng-irubm nggien nya -əm-lidik
 then 3SG.NOM REAL-travel INS.ACT-search place 3SG.GEN ACY-slash
 'Then he traveled searching for a **place** he could slash (for planting crops).'

Nggien relative clauses can occur in active or passive voice. When the relative clause is in active voice as in (59), the argument indexed by the verb is an actor. When the relative clause is in direct passive voice as in (60), the argument indexed by the verb is an unmarked undergoer. When the relative clause is in local passive voice as in (61), the argument indexed by the verb is a marked undergoer. When the relative clause is in periphrastic passive voice as in (62), the argument indexed by the passive auxiliary is an unmarked undergoer.

- (59) *Sia mori siidn {nggien nanggukng diaadn.}*
sia ng-bori siidn nggien ng-tanggukng diaadn
 3SG.NOM ISA.ACT-give money which ISA.ACT-support 1SG.NONACT
 'He gives money by **which** to support me.'
- (60) *Sia iniit di bunua {nggien nya pineti.}*
sia in-iit-0 di bunua nggien nya -in-pati-0
 3SG.NOM REAL-bring-ISA.UND to area place 3SG.GEN REAL-kill-ISA.UND
 'He was brought to the area **where** he was killed.'
- (61) *Sia iniit di bali*
sia in-iit-0 di bali
 3SG.NOM REAL-bring-UND to house
{nggien nya biniriadn siidn.}
nggien nya -in-bori-a siidn
 place 3SG.GEN REAL-give-MARKED.UND money
 'He was brought to the house **where** he was given money.'

- (62) *Sia iniiit di bunua {nggien nya inanu
sia in-iiit-∅ di bunua nggien nya in-anu-∅
3SG.NOM REAL-bring-UND to area place 3SG.GEN REAL-PASS-UND
ngidipadn.}
ngi-dipadn
ACT-slave
'He was brought to the area **where** he was enslaved.'*

As stated above, only pragmatic pivots can be relativised in Bonggi. In both the gapping strategy (e.g. (52)–(56)) and the relative pronoun strategy (e.g. (57)–(62)) the pragmatic pivot of the relative clause must be coreferential with the head of the relative clause. The pragmatic pivot of the relative clause in (57)–(58) and (60)–(62) is the location. Bonggi, like other Philippine-type languages, allows non-macroroles to function as pragmatic pivots. However, unlike many of these languages, the locative suffix *-an* only occurs with core arguments (i.e. arguments represented in the LS of the verb) which are macroroles. When the pragmatic pivot is a non-macrorole, *nggien/gien* occurs. For example, (59) illustrates a relative clause in which the pragmatic pivot of the relative clause is an instrument.

When the pragmatic pivot is a non-macrorole in a monoclausal sentence, the clause takes the shape of a NP followed by a relative clause as in (63) and (64).²³ As seen in (65) *nggien* 'place' also functions as an interrogative pronoun. The LS for (65) is shown in (66) where the locative adjunct *nggien* 'where' takes the entire LS of the verb as one of its arguments.

- (63) *Nti {gien ku monsu?}.
nti gien ku ng-ponsu?
this which 1SG.GEN ACT-bathe
'This is what I bathe with.'*
- (64) *Sia {nggien ku mogot.}
sia nggien ku -am-ogot
3SG.NOM place 1SG.GEN ACY-hold
'It is where I am holding on.'*
- (65) *Nggien nu monsu? Di telaga.
nggien nu ng-ponsu? di telaga
where 2SG.GEN ACT-bathe at well.
'Where do you bathe?' 'At a well.'*
- (66) LS for (65): **where** '[do'(2SG, [bathe'(2SG))]

In summary, pragmatic pivots can be either macroroles or non-macroroles. Only pragmatic pivots can be relativised. Relative clauses are formed by either gapping the relativised nominal or using the relative pronoun *nggien* 'place'. The former strategy is used to relativise macroroles, whereas the latter strategy is used to relativise non-macroroles.

²³ Cf. also (10) in §2.

5.2.2 Case marking in relative clauses

Recall from §4.1 that syntactic pivots involve a restricted neutralisation of semantic roles for syntactic purposes. In gapped relative clauses there is a restricted neutralisation of semantic roles of the gapped argument. For example, the gapped argument in (53) is an actor, whereas the gapped argument in (52) is an undergoer. Thus, gapped NPs are syntactic pivots in relative clauses. Furthermore, gapped NPs must be pragmatic pivots; otherwise, they cannot be relativised.

In gapped relative clauses the gapped NP is both the syntactic pivot and the pragmatic pivot; however, in relative clauses formed using the relative pronoun *nggien* 'place', the syntactic pivot is the argument indexed by the verb morphology while the pragmatic pivot is the relative pronoun *nggien*. This fundamental difference influences case marking in the two types of relative clauses.

Relative clauses are tightly bound dependent clauses (cf. Figure 2) and, according to (34e), the PSA in dependent clauses receives genitive case. However, since the syntactic pivot is gapped in gapped relative clauses (e.g. (52)–(56)), the PSA is not available to receive genitive case. Case marking of the other nominals in gapped relative clauses is straightforward. For example, by (34b) the actor in the relative clause in (52) and (54) receives genitive case since it is not the PSA. Similarly, by (34c) the undergoer (*gandubm* 'corn') in the relative clause in (53) receives accusative case since it is not the PSA.²⁴

In *nggien* relative clauses, the syntactic pivot is distinct from the pragmatic pivot. The syntactic pivot is indexed by the verb morphology, while the pragmatic pivot is the relative pronoun. According to (34e), the syntactic pivot (i.e. the PSA) receives genitive case regardless of whether it is an actor (e.g. *nya* '3SG.GEN' in (57)) or an undergoer (*nya* '3SG.GEN' in (62)). Case marking of the other nominals also follows the rules in (34). When *nggien* relative clauses are in active voice, non-PSA undergoers occur in accusative case (e.g. *diaadn* '1SG.NONACT' in (59)) following (34c). The pragmatic pivot (*nggien*) is the non-macrorole NP which is relativised. Because relative pronouns are not arguments, they do not receive case nor can they be case marked.

Table 3: Correlation between voice and case in main and relative clauses

	PSA in main clause	PSA (pivot) in gapped relative clause	<i>Nggien</i> relative clause	
			Pragmatic pivot	Syntactic pivot
	nominative case	∅ (case marking)	∅ (case marking)	genitive case
Active voice	actor e.g. (35)	actor e.g. (53)	non-macrorole <i>nggien</i> e.g. (57)	actor e.g. (59)
Direct passive	undergoer e.g. (39)	undergoer e.g. (52)		undergoer e.g. (60)
Periphrastic passive	undergoer e.g. (42); marked undergoer e.g. (44)	undergoer e.g. (55); marked undergoer e.g. (56)		undergoer e.g. (62)
Local passive	marked undergoer e.g. (41)	marked undergoer e.g. (54)		marked undergoer e.g. (61)

Table 3 provides an overview of the correlation between voice and case in main clauses and both types of relative clauses.

²⁴ Recall from §3.1 that common nouns are not case marked.

In summary, case marking in Bonggi is sensitive to clause linkage. Loose linkage results in nominative case PSAs, whereas tight linkage results in either null or genitive case marking of PSAs within the dependent clause. As we move up the clause linkage hierarchy in Figure 2, the dependent clause becomes partially nominalised which results in null or genitive case marking in dependent clauses.

6 Tense

Bonggi has two tense auxiliaries: *bas* 'PAST' and *adak* 'almost'. This section briefly illustrates the relationship between tense and case marking.

Some languages have a split in their case system based on tense; that is, nominal case marking can be predicted from verbal tense marking. In Bonggi, however, one cannot always predict case from tense because past tense introduces an alternation between nominative and genitive case which is not available in nonpast tense. The PSA in nonpast tense, main declarative clauses is always in the nominative case; however, the PSA in past tense, main declarative clauses is sometimes in genitive case. This possibility only occurs when the PSA is a pronoun. When using past tense and a pronominal PSA, speakers have two choices: (1) a genitive case enclitic pronoun which follows the tense marker as in (67); or (2) a nominative case pronoun which precedes the tense marker as in (68).

With the exception of *sigalama* '3PL', the genitive case pronouns in Table 2 of §3.3 are enclitics. *Sigalama* is actually an NP 'some people', and not a true pronoun. Thus, it is not case marked; but instead, it is treated as a common noun. When genitive case enclitic pronouns follow a tense auxiliary (e.g. *nya* '3SG.GEN' in (67)), they contrast with non-clitic nominative case pronouns (e.g. *sia* '3SG.NOM' in (68)) and other nominals (e.g. *si Tagi* 'Tagi' in (69)) which precede tense auxiliaries. Genitive case enclitic pronouns which follow tense auxiliaries are **special clitics** (Anderson 1993:74) since other nominals cannot occur in this position as illustrated by (70).

- (67) *Bas nya nuud diaadn.*
bas nya ng-tuud diaadn
 PAST 3SG.GEN ISA.ACT-assist 1SG.NONACT
 'Already **he** assisted me.'
- (68) *Sia bas na nuud diaadn.*
sia bas na ng-tuud diaadn
 3SG.NOM PAST PER ISA.ACT-assist 1SG.NONACT
 '**He** has already assisted me.'
- (69) *Si Tagi bas na nuud diaadn.*
si Tagi bas na ng-tuud diaadn
 PN.NOM Tagi PAST PER ISA.ACT-assist 1SG.NONACT
 '**Tagi** has already assisted me.'
- (70) **Bas n Tagi nuud diaadn.*
 PAST PN Tagi assist 1SG.NONACT
 '**Tagi** already assisted me.'

Undergoer PSAs can also occur in genitive case following a tense auxiliary as in (71) where the undergoer (*ku* '1SG.GEN') is a genitive case enclitic pronoun. (71) contrasts with non-clitic constructions such as (72) in which the undergoer is in nominative case (cf. (68)).

- (71) *Bas ku kiohol nya.*
bas ku -in-kiohol-∅ nya
 PAST 1SG.GEN REAL-bite-ISA.UND 3SG.GEN
 'I have already been bitten by him.'
- (72) *Ou bas na kiohol nya.*
ou bas na -in-kiohol-∅ nya
 1SG.NOM PAST PER REAL-bite-ISA.UND 3SG.GEN
 'I have already been bitten by him.'

The effect of tense on case marking is different from that of clause linkage. On the one hand, different types of clause linkage evoke either nominative or genitive case marking of the PSA (§5). For example, coordinate clause linkage evokes nominative case marking (e.g. (45)). On the other hand, tense auxiliaries do not require a particular case marking. That is, Bonggi is not a language with a case marking split along the dimension of tense/aspect since both nominative and genitive case pronouns can occur in past tense (cf. Dixon 1994:97ff.). In the interaction of clause linkage and tense, past tense can override clause linkage in determining case marking. For example, in (73b) coordinate clause linkage evokes nominative case, but nominative case is overridden by the occurrence of past tense with a genitive case enclitic pronoun (i.e. *nya* '3SG.GEN').

- (73)a. *Onu bunua onu bunua biniaan na nya,*
onu bunua onu bunua -in-biaa?-an na nya
 what area what area REAL-follow-ISA.MARKED.UND PER 3SG.GEN
 'From place to place he had travelled,
- b. *tei? nda? bas nya iketomu.*
tei? nda? bas nya i-kə-tomu
 but not PAST 3SG.GEN REAL-NONCONTROL-meet
 but **he** did not find it.'

Example (74), like (73), involves coordinate clause linkage. However, the PSA in the linked clause in (73b) is in genitive case (i.e. *nya* '3SG.GEN'), while the PSA in the linked clause in (74b) is in nominative case (i.e. *sia* '3SG.NOM'). The reason for the use of genitive case in (73b) as opposed to nominative case in (74b) goes beyond clause linkage and tense; it has to do with discourse pragmatic reference which is the final variable that must be understood in order to account for case marking in Bonggi (cf. §7). Both the genitive case pronoun *nya* '3SG.GEN' in (73b) and the nominative case pronoun *sia* '3SG.NOM' in (74b) are topical (i.e. presupposed). According to Lambrecht (1994:119ff.) sentences can have more than one topic; thus, topics can be ranked in terms of degrees of topicality. Genitive case enclitic pronouns (e.g. *nya* '3SG.GEN' in (73b)) are more topical than nominative case non-clitic pronouns (*sia* '3SG.NOM' in (74b)).

- (74)a. *Bas ku mori nya*
bas ku ng-bori nya
 PAST 1SG.GEN ISA.ACT-give 3SG.NONACT
 'I already gave it to him'

- b. *ma?* *sia* *bas na mori* *saa na.*
ma? *sia* *bas na ng-bori* *saa na*
 and 3SG.NOM PAST PER ISA.ACT-give spouse 3SG.GEN
 and **he** has already given it to his spouse.'

The tense auxiliary *adak* 'almost' has both a temporal and a modal function. *Adak* 'almost' refers to a situation prior to the time of utterance that the speaker believes was possible, but that did not actually occur. In contrast, *bas* refers to past situations that really occurred. The implication associated with *adak* is that, had the situation happened, it would have had a negative consequence. *Adak* 'almost', like *bas*, can occur with genitive case enclitic pronouns as seen in (75).

- (75) *Adak ku kohoidn.*
adak ku kohol-an
 almost 1SG.GEN bite-ISA.UND
 'I was almost bitten.'

A detailed explanation for the occurrence of genitive case PSAs in past tense constructions is beyond the scope of this paper. Such an explanation would require a discussion of the scope of tense operators and a description of the relationship between tense and both time and temporal adverbial clauses. Furthermore, in order to account for differences in usage between clauses such as (71) and (72), reference must be made to discourse structure.

7 Discourse pragmatic relations topic and focus

This section is concerned with how the distribution of information in discourse affects case marking in Bonggi. Information status is described in terms of two pragmatic relations: topic and focus. **Topic** is what the proposition is about, whereas **focus** is the unpredictable or pragmatically non-recoverable element in an utterance (Lambrecht 1994:207).²⁵ The topical part of an utterance is presupposed; the focus is non-presupposed. In Bonggi the order topic-focus is the norm for statements, whereas in WH-questions the focus is in clause-initial position.

Lambrecht (1994:223ff.) makes a distinction between different types of focus structure. The fundamental contrast is between narrow and broad focus. In **narrow focus** the focus domain extends over a single constituent, while in **broad focus** it extends beyond a single constituent (Van Valin 1993:25). In **broad predicate focus**, which is the unmarked focus structure, the focus includes the predicate. Broad predicate focus is illustrated in (76b) which is a response to the question in (76a). The focus in (76b) is *kiohol ulakng* 'bitten by a snake', whereas *sia* '3SG.NOM' is the topic and the pragmatic pivot.

- (76)a. *Onu kusuat* *ny Abas?*
onu kə-suat *ny Abas*
 what NONCONTROL-incur PN Abas
 'What happened to Abas?'

²⁵ My use of the term **focus** follows its use in general linguistics and should not be confused with its use in Philippine linguistics.

- b. *Sia kiohol ulakng.*
sia -in-kohol-Ø ulakng
 3SG.NOM REAL-bite-UND snake
 'He was bitten by a snake.'

Narrow focus is illustrated in (77b) which is a request for further information prompted by the statement made in (77a). The focus in (77b) is the question word *onu* 'what', whereas the topic or presupposed information is *ngohol nya* 'bit him'.

- (77)a. *Si Abas kiohol.*
si Abas -in-kohol-Ø
 PN.NOM Abas REAL-bite-ISA.UND
 'Abas was bitten.'

- b. *Onu ngohol nya?*
onu ng-kohol nya
 what ISA.ACT-bite 3SG.NONACT
 'What bit him?'

Examples (76) and (77) illustrate that the argument indexed by the verb cannot be equated with pragmatic topic. In (76b) the indexed argument (*sia* '3SG.NOM') is the topic, but in (77b) the indexed argument (*onu* 'what') is the focus while *ngohol nya* 'bit him' is the topic.

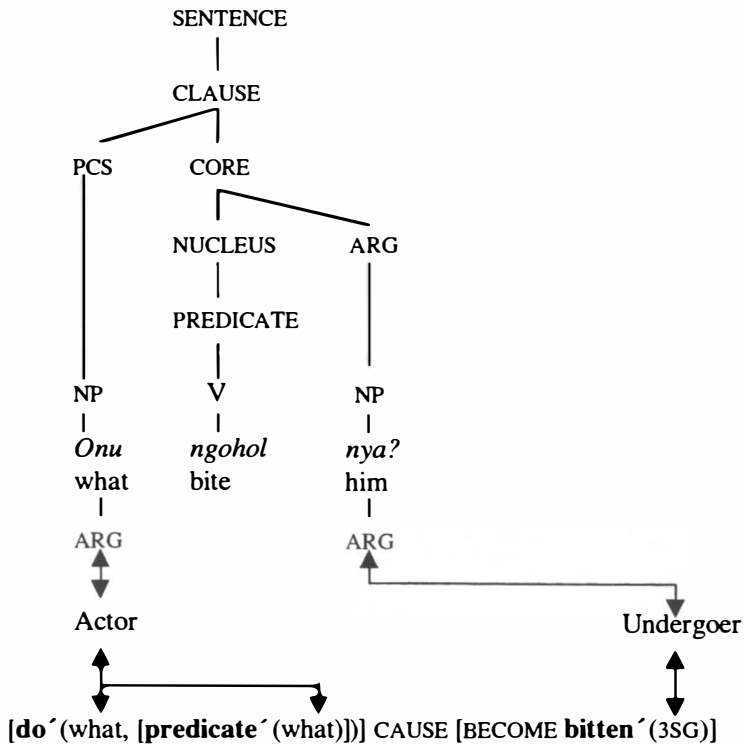


Figure 3: LS, constituent structure and pragmatic relations for (77b)

The relationship between the semantic structure (represented by LSs), constituent structure and pragmatic functions (topic and focus) is illustrated in Figure 3. Details for linking between LS and syntactic structure including the assignment of pragmatic functions are provided in Van Valin (1993). In narrow focus constructions (e.g. (77b)), the question word is in the precore slot (PCS). Figure 3 shows that the actor in (77b) is assigned the pragmatic function focus.

In (2) (repeated as (78)), the verb *kiohol* indexes the undergoer which is in genitive case. Both (77b) and (78) are narrow focus constructions with the focus being on the WH-word (*onu* 'what' in (77b) and *mipa?* 'when' in (78)). In (77b) the focused constituent is an argument of the verb (cf. Figure 3), whereas in (78) the focused constituent is an adjunct (cf. Figure 4).²⁶ Examples (77b) and (78) are equally good responses to (77a) and have the same topic; however, (77b) is an unmarked narrow focus construction, whereas (78) is a marked narrow focus construction.

- (78) *Mipa? nya kiohol?*
mipa? nya -in-kiohol-∅
 when 3SG.GEN REAL-bite-ISA.UND
 'When was he bitten?'

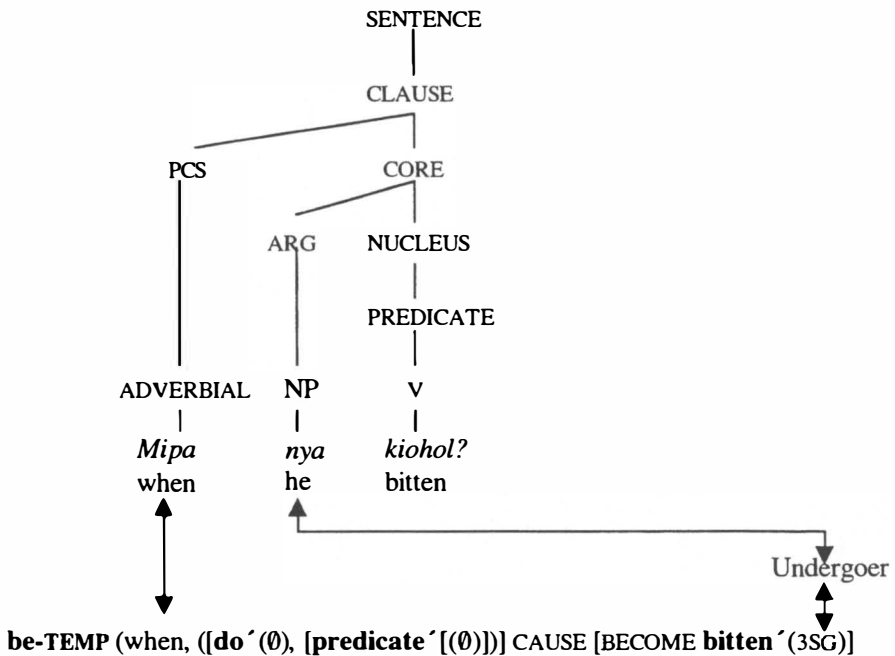


Figure 4: LS, constituent structure and pragmatic relations for (78)

²⁶ See Van Valin and LaPolla (1997:334-335) for discussion of the abstract temporal higher predicate *be-TEMP*':

Genitive case occurs in Bonggi WH-questions whenever the PCS is filled by an adverbial adjunct. Or stated another way, genitive case occurs in WH-questions when the focus is an adverbial adjunct. Thus, genitive case occurs in both (79a) and (79b) because the focus in (79a) is a locative adverbial adjunct and in (79b) a manner adverbial adjunct. Note that the use of genitive case marking here corresponds to that described in §5.1 for adjunct adverbial clauses.

- (79)a. *Nggién nya kiohol?*
nggién nya -in-kiohol-Ø
 where 3SG.GEN REAL-bite-UND
 'Where was **he** bitten?'
 b. *Pungga? buat nya kiohol?*
pungga? buat nya -in-kiohol-Ø
 how do 3SG.GEN REAL-bite-UND
 'How was **he** bitten?'

The analysis above accounts for most examples encountered. However, there are rare instances in which nominative case occurs in WH-questions when the focus is an adverbial adjunct. Compare the unmarked clause in (80a) in which the PSA (*nya* '3SG.GEN') is in genitive case with the highly marked clause in (80b) in which the PSA (*sia* '3SG.NOM') is in nominative case.

- (80)a. *Mipa? nya muli? ?*
mipa? nya -əm-uli?
 when 3SG.GEN ACY-return.home
 'When is **he** going home?'
 b. *Mipa? sia muli? ?*
mipa? sia -əm-uli?
 when 3SG.NOM ACY-return.home
 'When is **he** going home?'

The occurrence of nominative versus genitive case in (80) is pragmatically controlled. (80a) may be used whether or not the person being referred to is present. However, (80b) can only be used when the person being referred to is present and listening to the speaker and addressee. The use of nominative case in (80b) makes the referent prominent; thus, nominative case draws attention to the referent and implies something about the speaker's attitude toward the referent.²⁷

8 Conclusion

It is well known that actors in many western Malayo-Polynesian languages occur in genitive case when they are not indexed by the verb; however, main clauses in which the argument indexed by the verb is an undergoer in genitive case are virtually unknown. Thus, constructions such as those in (2), (71), (75), (78), (79a) and (79b) are particularly interesting. Genitive case marking in Bonggi is not lexically governed, idiosyncratic, or

²⁷ Paul Kroeger has suggested to me that the contrast in (80) may involve a deictic (80b) versus anaphoric (80a) use of the pronoun.

“quirky.”²⁸ Instead, the interaction of five independent variables accounts not only for the alternation between nominative and genitive case, but also for the case marking rules in (34). The inherent lexical content of NPs determines whether nominals are case marked or not (§3). Case marking then follows the language particular case marking rules in Bonggi (§4.2).

These rules may be modified by two different aspects of finiteness, tense (§6) and clause linkage (§5). Finiteness is here understood as a property of the clause rather than the verb; case marking is a nominal feature of finiteness, whereas tense is a verbal feature of finiteness (Givón 1990:853). In Bonggi, case marking is sensitive to clause linkage. The tighter the syntactic and semantic bond, the more likely the dependent clause becomes partially nominalised resulting in null or genitive case marking. This supports Givón’s claim that case marking of core arguments is most commonly modified toward the genitive (Givón 1990:498-499).²⁹ “The less finite a clause is, the more likely are its subject and object arguments to lose their normal case-marking, and to be coded instead by genitive morphology” (Givón 1990:503). Finally, the discourse pragmatic relations topic and focus interact in terms of case assignment (§7).

Table 4 provides a summary of case alternations in Bonggi. Adjuncts which are not pragmatic pivots are excluded from Table 4 because they are not case marked; instead, they are preceded by prepositions which have an adverbial function (cf. §3.4). First person singular pronouns are used to illustrate personal pronouns (cf. Table 2 in §3.3). The phrase ‘other nominals’ in Table 4 includes common nouns, relative pronouns and nominalisations. The dative case marker *di* ‘DAT’ which marks non-macrorole arguments is restricted to the ‘y’ argument in the LS configuration **pred’** (y, z) when the ‘y’ argument is not the undergoer. When the ‘z’ argument is not the undergoer, there is no overt dative case marker.

Table 4: Summary of case alternations in Bonggi

CASE	Privileged Syntactic Argument		Non-Privileged Syntactic Argument		
	main clause	dependent clause	Actor	Non-actor	
				Undergoer	Non-undergoer
	nominative	genitive	genitive	accusative	dative
personal pronouns ‘1SG’	<i>ou</i>	<i>ku</i>	<i>ku</i>	<i>diaadn</i>	<i>di diaadn</i>
personal nouns	<i>si</i>	<i>ny</i>	<i>ny</i>	<i>ny</i>	<i>di ny</i>
other nominals	∅	∅	∅	∅	<i>di</i>

Three apparent exceptions to Table 4 are: (1) WH-questions in which the pragmatic focus is an adverbial adjunct (cf. §7); (2) special clitics which follow tense auxiliaries (cf. §6); and (3) loosely linked adverbial clauses (cf. §5.1). The presence of genitive case in WH-questions whose focus is an adverbial adjunct is due to these being marked narrow focused constructions as opposed to unmarked narrow focused constructions in which the

²⁸ See Van Valin (1990), (1991) for Icelandic; Michaelis (1993) for Latin; and Narasimhan (1995) for Hindi.

²⁹ Originally claimed by Silverstein (1976).

focus of the WH-question is an argument. Similarly, because special clitics are marked constructions, the pivot is in genitive case. Finally, the presence of nominative case pivots in conditional, purpose and reason clauses simply underscores the incoherent nature of the traditional category of dependent clause (cf. Chafe 1988). That is, there is no reason to conclude that these clauses are more dependent than clauses linked by *ma?* 'and' (cf. Chafe 1988:20). Therefore, conditional, purpose and reason clauses are not exceptions to Table 4; instead, they are classified together with coordinate clauses in a single category of loosely linked clauses.

Many linguists view case as a mechanism for indicating grammatical relations such as subject, direct object and indirect object (e.g. Blake 1994:2; Spencer 1991:256). However, no reference has been made in this paper to any of these three relations. Even if we substitute PSA for subject, we are left without direct object and indirect object. Case marking in Bonggi, and by extension other Western Malayo-Polynesian languages, is not a direct indicator of grammatical relations. The case marking rules in (34) make reference to PSA, macroroles and core argument status, not grammatical relations.

References

- Anderson, Stephen R., 1993, Wackernagel's revenge: clitics, morphology, and the syntax of second position. *Language* 69:68-98.
- Blake, Barry J., 1994, *Case*. Cambridge: Cambridge University Press.
- Blust, Robert, 1998, The position of the languages of Sabah. In Ma. Lourdes S. Bautista, ed. *Paganánw: essays on language in honor of Teodoro A. Llamzon*, 29-52. Manila: The Linguistic Society of the Philippines.
- Boutin, Michael E., 1994, Aspect in Bonggi. PhD dissertation, University of Florida. forthcoming, A tale of two passives in Bonggi. Proceedings of the Fourth Biennial Conference of the Borneo Research Council, June 10-15, 1996, Universiti Brunei Darussalam.
- Chafe, Wallace, 1988, Linking intonation units in spoken English. In John Haiman and Sandra A. Thompson, eds *Clause combining in grammar and discourse*, 1-27. Amsterdam: John Benjamins.
- Dixon, R.M.W., 1994, *Ergativity*. Cambridge: Cambridge University Press.
- Dowty, David Roach, 1979, *Word meaning and Montague grammar*. Dordrecht: D. Reidel.
- Givón, T., 1990, *Syntax: a functional-typological introduction*, vol. 2. Amsterdam: John Benjamins.
- Kishimoto, Hideki, 1996, Split intransitivity in Japanese and the unaccusative hypothesis. *Language* 72:248-286.
- Kroeger, Paul R., 1993, *Phrase structure and grammatical relations in Tagalog*. Stanford: CSLI Publications.
- Lambrecht, Knud, 1994, *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- McFarland, Curtis D., 1974, The dialects of the Bikol area. PhD dissertation, Yale University.

- Michaelis, Laura A., 1993, On deviant case-marking in Latin. In Van Valin, ed. 1993:311-73.
- Narasimhan, Bhuvana, 1995, A lexical semantic explanation for 'quirky' case marking in Hindi. Paper presented at the Annual Meeting of the Linguistic Society of America.
- Needham, Rodney, 1954a, The system of teknonyms and death-names of the Penan. *Southwestern Journal of Anthropology* 11:416-431.
- 1954b, Batu Belah and Long Terawan: kinship terms and death-names. *Journal of Malayan Branch of Royal Asiatic Society* 27:215-217.
- Nida, Eugene A., 1949, *Morphology: the descriptive analysis of words*. Ann Arbor: The University of Michigan Press.
- Pallesen, A. Kemp, 1985, *Culture contact and language convergence*. Manila: Linguistic Society of the Philippines.
- Schachter, Paul and Fe T. Otones, 1972, *Tagalog reference grammar*. Berkeley: University of California Press.
- Shibatani, Masayoshi, 1988, Voice in Philippine languages. In Masayoshi Shibatani, ed. *Passive and voice*, 85-142. Amsterdam: John Benjamins.
- Silverstein, Michael, 1976, Hierarchy of features and ergativity. In R.M.W. Dixon, ed. *Grammatical categories in Australian languages*, 112-71. Canberra: Australian Institute of Aboriginal Studies.
- 1981, Case marking and the nature of language. *Australian Journal of Linguistics* 1:227-247.
- 1993, Of nominatives and datives: Universal grammar from the bottom up. In Van Valin, ed. 1993:464-498.
- Spencer, Andrew, 1991, *Morphological theory: an introduction to word structure in Generative Grammar*. Oxford: Basil Blackwell.
- Thompson, Sandra A. and Robert E. Longacre, 1985, Adverbial clauses. In Timothy Shopen, ed. *Language typology and syntactic description*, vol.2, *Complex constructions*, 171-234. Cambridge: Cambridge University Press.
- Van Valin, Robert D., Jr., 1990, Semantic parameters of split intransitivity. *Language* 66:221-260.
- 1991, Another look at Icelandic case marking and grammatical relations. *Natural Language & Linguistic Theory* 9:145-194.
- 1993, A synopsis of Role and Reference Grammar. In Van Valin, ed. 1993:1-164.
- 1995, Role and Reference Grammar. In Jef Verschueren, Jan-Ola Östman and Jan Blommaert, eds *Handbook of pragmatics manual*, 461-469. Amsterdam: John Benjamins.
- Van Valin, Robert D., Jr., ed., 1993, *Advances in role and reference grammar*. Amsterdam: John Benjamins.
- Van Valin, Robert D., Jr. and Randy J. LaPolla, 1997, *Syntax: structure, meaning and function*. Cambridge: Cambridge University Press.
- Walton, Charles, 1986, *Sama verbal semantics: classification, derivation and inflection*. Manila: Linguistic Society of the Philippines.
- Wolff, John U., 1996, The development of the passive verb with pronominal prefix in Western Austronesian languages. In Bernd Nothofer, ed. *Reconstruction, classification, description: festschrift in honor of Isidore Dyen*, 15-40. Hamburg: Abera Verlag.

The prefixes di- and N- in Malay/Indonesian dialects

DAVID GIL

1 The Malay/Indonesian dialectal landscape¹

Of the over one thousand Austronesian languages, Malay/Indonesian is the one with the greatest number of speakers, and by quite a large margin. Perhaps for this reason, it has a reputation of being well-studied, unexotic, maybe even a trifle uninteresting. However, this reputation reflects a fundamental ignorance with regard to some of the most elementary facts — sociolinguistic and also hardcore grammatical — about the Malay/Indonesian language, or languages.

When most people think of Malay/Indonesian, they generally have in mind one of the two ‘standard languages’ of Malaysia or of Indonesia — taught in elementary and high schools, and used in various formal or official situations, in politics, education, the media, and so forth. However, these standardised varieties are not the *real* Malay/Indonesian.

¹ Some of the field work described in this paper was supported in part by the National Science Foundation (grants SBR-9729519, INT-9423291 and SBR-9121167, with P. Cole, G. Hermon and C.-T.J. Huang as Principal Investigators). The ongoing field work in Indonesia is being supported by the Max Planck Institute for Evolutionary Anthropology in Leipzig and by the Universitas Katolik Indonesia Atma Jaya in Jakarta. Versions of this paper were presented at a guest lecture in the Faculty of Education, Language and Arts, Institut Keguruan dan Ilmu Pendidikan, Ujung Pandang, Indonesia, 10 July 1996, at the First Symposium on Malay/Indonesian Linguistics, Association for Linguistic Typology, Penang, Malaysia, 15 January 1997, and at the Panel on Focus in Western Austronesian Languages, Eighth International Conference on Austronesian Linguistics, Taipei, Republic of Taiwan, 30 December 1997: I am grateful to participants at all three events for helpful comments and suggestions. For helpful comments on a draft version of this paper I am indebted to Bernard Comrie, Uri Tadmor and Fay Wouk. In the course of my work in Malay/Indonesian I have benefited considerably from ongoing discussions with Peter Cole, Gabriella Hermon and Uri Tadmor, to whom my thanks are due. Finally, this paper would not have been possible without the unwitting assistance of numerous speakers of Malay/Indonesian dialects, who, in the course of their daily lives, contributed to my database of spontaneous speech specimens: for making it into the present paper, I am grateful to Ahmadsayuti, Aliawar, Amat, Anton, Anwar, Arifin, Arip, Badut, Dedi, Doni, Elly, Firdaus, Indra, Junjun, Junjun, Kairil, Kuduak, Nasar, Pai, Per, Riki, Riki, Rudi Candra, Supriyanto, Suriyadi, Tambuyang, Yanto, and others whose names I was not able to record.

Rather, they are artifacts of conscious, politically-motivated language engineering, rarified registers which few people speak 'properly' (whatever that means), and nobody acquires through the natural processes of first-language acquisition.

The *real* Malay/Indonesian is in fact not one language, or two, but rather a diverse set of dialects or language varieties, acquired natively, as a first or fluent early second language, by most of the inhabitants of the Malay/Indonesian archipelago, and spoken colloquially in most everyday contexts, at home, at work, in the marketplace, indeed almost everywhere. These myriad vernacular dialects or languages are of a low degree of mutual intelligibility; in their diversity they are perhaps comparable to the different varieties of Arabic, Romance or Slavonic. Sociolinguistically, and simplifying somewhat, one may roughly distinguish between two main types of Malay/Indonesian dialects: (a) 'classic' dialects, spoken by ethnically homogeneous speech communities, typically small localised populations of rural Malays; as opposed to (b) 'koiné' dialects, associated with ethnically heterogeneous speech communities, often consisting of large rural and/or urban populations which may be partly or entirely non-Malay.

This paper is concerned with some varieties of Malay/Indonesian belonging to the 'koiné' type. The main part of the paper deals with one such dialect, Riau Indonesian; but towards the end, four other dialects are examined, Jakarta Indonesian, Sulsel Indonesian, Irian Indonesian, and Kuala Lumpur Malay. These five varieties are quite different from each other: for the most part, speakers of one cannot understand speakers of any of the others without substantial experience. Nevertheless, all five dialects share certain structural features which may strike many readers familiar only with Standard Malay/Indonesian as novel, even surprising.

In general, the Austronesian family is renowned for the richness of its focus morphology, and although Malay/Indonesian is often considered to have undergone a reduction in this domain, it is still usually taken to have an active/passive distinction, expressed by the prefixes *meN-* and *di-* respectively. However, the results of this paper suggest a rather different state of affairs. Specifically, it is shown that in all of the colloquial Malay/Indonesian dialects under consideration, there is in fact no focus morphology at all, that is to say, no morphological active/passive distinction whatsoever.

But before examining the colloquial dialects in detail, let us first take a brief look at these two supposedly focus-marking prefixes in the standard language.

2 Standard Malay/Indonesian

The received view for Standard Malay/Indonesian is that there are two verbal focus markers: *meN-* for active voice, *di-* for its passive counterpart. But not everybody holds to that view. One sceptical voice is that of Benjamin (1993:361), who suggests that:

[...] such descriptions probably say more about the grammarians' theoretical preoccupations than about what motivates a Malay speaker to choose between one affix and another.

Indeed, even for the standard languages, a number of scholars have argued that the prefixes *meN-* and *di-* are, at best, atypical instances of voice morphology, and perhaps something else entirely.

One argument, put forward by Hopper (1979, 1983), alludes to the discourse function of the markers in question. In most languages active markers are associated with clauses that are

foregrounded and of high transitivity, while passive markers are associated with clauses that are backgrounded and of low transitivity (Hopper & Thompson 1980). Indeed, as has been noted by MacDonald (1976), Alsagoff (1992:12-13) and others, *meN-* tends to occur only on verbs which have an overt direct object. However, if a more general view of transitivity is taken, then a different, in fact diametrically opposite picture emerges. As shown by Hopper it is *meN-* which marks clauses that are backgrounded and of low transitivity, and *di-* which marks clauses that are foregrounded and of high transitivity. If anything, then, it is *di-* which exhibits the discourse function associated with active sentences, and *meN-* which displays the discourse function associated with passive sentences.

A second argument, pertains to a construction type which, for lack of a better term, might be dubbed 'funny control' — see Verhaar (1988:362-365), Abdul Chaer (1993:116-124), Rani (1996) and Tadmor (1996). The following examples, cited by Tadmor, are taken from Indonesian daily newspapers:²

- (1)a. Pejambret telah *berhasil* **ditangkap** oleh aparat kepolisian.
 AG-snatch PFCT MID-succeed **di**-capture by personnel ABSTR-police
 'The police succeeded in capturing the purse-snatcher.'
- b. Sebanyak 50 ekor gajah liar yang mengganggu ketenteraman
 one-many 50 CLF elephant wild REL me-N-disturb ABSTR-calm
 penduduk di Lampung *berhasil* **ditangkap**.
 AG-reside di Lampung MID-succeed **di**-capture
 'Fifty wild elephants who were disturbing the peace of the residents in
 Lampung were successfully captured.'

In both examples, an expression prefixed with *di-* is preceded by a so-called 'control verb' (italicised) — in fact, it is the same two forms in both sentences. However, if the constructions are translated into English, using a passive form of the verb, the results are obviously wrong: 'The purse-snatcher *succeeded in being captured* by the police', 'As many as fifty wild elephants who were disturbing the peace of the residents in Lampung *succeeded in being captured*'. The problem with the English translations is that they force the subject of the control verb *succeed* to be coreferential with the subject of the passive main verb *be captured*. However, as evidenced by the above examples, there is no similar constraint on coreferentiality in Standard Malay/Indonesian. Interestingly, though, Rani feels that there is something 'wrong' with these sentences, and argues that they should be prohibited in Standard Indonesian. But they occur, even in formal registers, and, as pointed out by Tadmor, they cast doubt on the characterisation of the prefix *di-* as a marker of passive voice.

A third argument, proposed by Saddy (1991, 1992) and Cole and Hermon (1998b), involves a particular kind of construction containing a sentence-initial NP which, to invoke the common 'movement' metaphor, appears to have been extracted and fronted from a

² In the examples presented in this paper, the prefixes *di-*, *N-* and *meN-* are indicated in boldface, when they are the focus of the discussion in the surrounding text. The interlinear translations make use of the following abbreviations: ABSTR abstract; AG agentive; APPL applicative; ASSOC associative; CLF classifier; CNJ conjunctive; COMP complementiser; DEM demonstrative; DIST distal; DISTR distributive; EXCL exclamation; FAM familiar; FUT future; IMP imperative; INTRG interrogative; MID middle (voice); NEG negation; OP operator; PERS personal; PROX proximate; PFCT perfect; REL relative; SG singular; TOP topic; 1,2,3 first, second, third person.

position deep within an embedded clause. Consider, first, the following simple-clause paradigm:³

- (2)a. Abang **menulis** surat ini.
 elder.brother **me-N-write** letter DEM:PROX
 'You wrote this letter.'
- b. *Surat ini abang **menulis**.
 letter DEM:PROX elder.brother **me-N-write**
 'This letter you wrote.'
- c. Surat ini abang tulis.
 letter DEM:PROX elder.brother write
 'This letter you wrote.'

Example (2a) illustrates a simple transitive clause containing the active prefix *meN-*. Example (2b) shows that in such a clause the direct-object NP *surat ini* 'this letter' cannot be fronted to sentence-initial position, as it can in English and many other languages. In order for fronting to occur, the active prefix *meN-* must be omitted, as in (2c). These and other facts have led some scholars to characterise the bare-stem construction in (2c) as instantiating a particular focus or voice, a kind of 'pseudo-passive', or 'second passive', contrasting with both the active, marked with *meN-*, and the passive, marked with *di-* — see, for example, Chung (1976), Abdul Hamid (1992), Alsagoff (1992), Guilfoyle et al. (1992), Nik et al. (1996:469-471), and Sneddon (1996).

However, as pointed out by Saddy and by Cole and Hermon, this analysis runs into difficulties when complex sentences are considered:

- (3)a. Dia **memberitahukan** saya bahwa abang **menulis**
 3SG **me-N-give-know-APPL** 1SG COMP elder.brother **me-N-write**
 surat ini.
 letter DEM:PROX
 'He informed me that you wrote this letter.'
- b. *Surat ini dia **memberitahukan** saya bahwa abang
 letter DEM:PROX 3SG **me-N-give-know-APPL** 1SG COMP elder.brother
menulis.
me-N-write
 'This letter he informed me that you wrote.'
- c. *Surat ini dia **memberitahukan** saya bahwa abang tulis.
 letter DEM:PROX 3SG **me-N-give-know-APPL** 1SG COMP elder.brother write
 'This letter he informed me that you wrote.'
- d. *Surat ini dia **beritahukan** saya bahwa abang
 letter DEM:PROX 3SG give-know-APPL 1SG COMP elder.brother

³ In (2), and in subsequent examples, the would-be prefix *meN-* is glossed as a concatenation of two independent prefixes, *me-* and *N-*. Some arguments in support of this analysis are presented in Benjamin (1993, 1997) and Stevens (1997); however, none of the claims made in this paper are dependent on such an analysis.

menulis.

me-N-write

'This letter he informed me that you wrote.'

- e. *Surat ini dia beritahukan saya bahwa abang tulis.*
 letter DEM:PROX 3SG give-know-APPL 1SG COMP elder.brother write
 'This letter he informed me that you wrote.'

Example (3a) is obtained from (2a) by embedding it in a matrix clause *Dia memberitahukan saya bahwa...* 'He told me that...'. And examples (3b)–(3e) show what happens when the embedded direct object *surat ini* 'this letter' is moved to the beginning of the sentence: the only grammatical construction is (3e), that in which both verbs, *beritahukan* and *tulis*, occur in base form, without the prefix *meN-*. The grammaticality of (3e) shows that the fronting of *surat ini* is not a clause-bound rule, such as the promotion of a direct object to subject position via passivisation, but rather an unbounded rule, more akin to topicalisation. Accordingly, it suggests that the simple bare-stem construction in (2c) might involve topicalisation, rather than a particular focus or voice — as indeed is assumed to be the case in traditional accounts of this construction, for example Payne (1970:88-90) and MacDonald (1976).

Now consider the contrast between (3e) and its ungrammatical variants (3b)–(3d), in which one or both of the verbs is prefixed with *meN-*. The ungrammaticality of (3b)–(3d) suggests that the prefix *meN-* cannot mark a verb over which an NP has moved. And as argued by Saddy and by Cole and Hermon, the sensitivity of *meN-*, in both matrix and embedded clauses, to whether an unbounded movement has taken place over it, suggests that whatever the function of this prefix, it is very different from that of a prototypical marker of actor focus or active voice.

Thus, the above three arguments suggest that all is not well with the characterisation of *meN-* and *di-* as active and passive prefixes even in Standard Malay/Indonesian. But this is hardly surprising, since, when we turn to consider the colloquial varieties of Malay/Indonesian, the portrayal of these prefixes as markers of focus or voice comes up against a much more substantial body of evidence.

3 Malay/Indonesian dialects

The prefix *di-* occurs, in the same form, in a wide variety of Malay/Indonesian dialects. However, the same is not true of *meN-*. Many dialects have no prefix in opposition to *di-*, and in most of the dialects that do, it is not *meN-* but rather the simpler *N-*. Accordingly — as suggested by the title of this paper — we shall be concerned mostly with a characterisation of the prefixes *di-* and *N-* in the some of the koiné-like dialects of Malay/Indonesian.

But first a point of methodology. When working with the colloquial varieties of Malay/Indonesian, it is often difficult or impossible to elicit reliable judgements from native speakers. What happens all too often is that the moment speakers realise that they are being questioned in a 'learned' context, they switch from whatever colloquial variety they had just been using into the standard language, or rather their sometimes imperfect variant thereof. And when speakers do provide judgements for ordinary or everyday language, they frequently make claims that are in gross conflict with their actual linguistic behaviour, for example by characterising as ungrammatical forms or constructions that they use all the time.

Faced with such obstacles, the study of Malay/Indonesian dialects reported on here makes use of an alternative method of data collection, based on the gathering of *spontaneous speech specimens*: actual utterances produced by native speakers in real live situations, written down on the spot and subsequently entered into a computerised database. All of the data presented below are of such a character. Because of the nature of such data, it is necessary, for each example, to include, in addition to the customary three lines (text, interlinear gloss, and free translation), an additional line describing the context in which the example was uttered, thereby justifying the translation that is provided, as opposed to any number of other translations potentially available for the same sentence had it been uttered in a different context. This additional line is enclosed in square brackets.⁴

3.1 Riau Indonesian

Riau Indonesian is the dialect of Malay/Indonesian spoken in informal situations by the inhabitants of Riau province in east central Sumatra. Contrary to what is suggested in various linguistic atlases, such as Moseley and Asher eds. (1994), the population of Riau is linguistically and ethnically heterogeneous. Although the indigenous population is Malay, a majority of the present day inhabitants are migrants from other provinces, speaking a variety of languages — mostly Minangkabau, but also Batak, Bugis, and others. Riau Indonesian is acquired as a native language by most or all children growing up in Riau province, whatever their ethnicity; and it is the language most commonly used as a *lingua franca* for inter-ethnic communication. In addition, like other colloquial varieties of Indonesian, it is gradually replacing other languages and dialects as a vehicle of intra-ethnic communication. It should be noted that Riau Indonesian is distinct from a set of dialects commonly subsumed under the term as Riau Malay, which are used, also in Riau province, by ethnic Malays, for intra-ethnic communication — see Kailani (1994), Kailani et al. (1983), Saidat et al. (1991) and Saidat et al. (1986). Riau Indonesian is also to be distinguished from another set of Malay dialects spoken by various indigenous or ‘para-Malay’ peoples in Riau province, known as Sakai, Akit, Orang Hutan or Orang Laut — see Kähler (1948, 1949, 1960). Finally, Riau Indonesian is also quite different from the variety of Malay/Indonesian used by the ethnic Chinese residents of Riau province when speaking to non-Chinese, and by non-Chinese when speaking to Chinese, sometimes referred to as ‘Bazaar Malay’. The Riau Indonesian data discussed in this paper are the product of several years of ongoing and continuous fieldwork in Riau province, beginning in 1992, and reported on in Gil (1994, 1999, 2000a,b, to appear a,c).

One of the most striking characteristics of Riau Indonesian is a widespread tolerance of underspecification of thematic roles. In general, in a basic three-word clause containing two participants associated with an activity, the thematic relationship between the participants and the activity is not specified; rather, it is left to be determined by the context in which the sentence is uttered. Following are some near minimal pairs illustrating the indeterminacy of

⁴ One disadvantage of this method of data collection is that it is not feasible to record the intonation contours of the spontaneous speech specimens. In work in progress, the possible effects of intonation are being examined in recordings of naturalistic texts in a variety of dialects of Malay/Indonesian. Preliminary results suggest that the structures and interpretations of sentences containing the prefixes *di-* and *N-* are not significantly affected by the choice of intonation contour, and therefore that the analysis proposed in this paper is not compromised by the absence of reference to intonation.

thematic roles. Each pair contains two sentences with the same activity (in boldface) in construction with two participants (in italics): in (4)–(7) the activity word occurs between the two participants, in (8) it precedes them both, and in (9) it follows them both. As evidenced by the respective contexts in which the sentences were uttered, the thematic roles associated with each member of the pair are diametrically opposed:

- (4)a. Kalau di Malaysia, *anak kecil-kecil* udah **bisa** *bahasa Inggris*.
 TOP di Malaysia child DISTR-little PFCT can language English
 [Small talk]
 ‘In Malaysia, even little children can speak English.’
- b. *Tiga belas* **bisa** *aku*.
 three over.ten can 1SG
 [Playing billiards on laptop computer]
 ‘I can hit the thirteen ball.’
- (5)a. *Saya* tak **tahu** *dia bilang apa*.
 1SG NEG know 3 say what
 [About person speaking in foreign language]
 ‘I don’t know what he’s saying.’
- b. *Siapa yang kena* tak **tahu** *aku*.
 PERS-what REL undergo NEG know 1SG
 [During horseplay in a dark room, speaker threw his sandal and hit somebody; later, recounting, he describes what happened]
 ‘I don’t know who got hit.’
- (6)a. *Aku* **pasang** *dua ribu*, *Rip*.
 1SG attach two thousand FAM-Arip
 [Playing cards and betting]
 ‘I’ll place two thousand, Arip.’
- b. *Bom* **pasang** *dia*.
 bomb attach 3
 [Watching a movie on TV]
 ‘They’re going to set off a bomb.’
- (7)a. *Dia* **kasi** *dua-dua*.
 3 give DISTR-two
 [In bar, somebody rings the bell for drinks all around]
 ‘They’re giving two each.’
- b. *Apa, ini-ini* **kasi** *dia*, *pantek?*
 what DISTR-DEM:PROX give 3 EXCL
 [Playing tetris on laptop computer, complaining that the wrong shape of object keeps on coming]
 ‘Damn it, why does it keep on giving me these?’
- (8)a. Nanti **tengok** *Arip poto* *anak gembel semua*.
 FUT look Arip photograph child beggar all
 [After I had taken some pictures of some beggar boys at a coffee shop, speaker is worried that a mutual friend will see them and be angry]
 ‘Later Arip will see all the pictures of the beggar boys.’

b. **Tengok** *tikus* *aku*.

look mouse 1SG

[Speaker learning to play a game of laptop computer billiards in which it is rather difficult to control the position of the simulated player with the trackpad, as a result of which the player often ends up under the table; the first time this happened, I jokingly asked him whether he was looking at the mice under the table; when this happened once again, speaker joked]

'I'm looking at the mice.'

(9)a. *Komputer* *David* **bawa**?

computer David take

[Getting ready to go out]

'Are you taking your computer with?'

b. *David* *komputer* **bawa**, Vid?

David computer take FAM-David

[Getting ready to go out]

'Are you taking your computer with you, David?'

In (4a) *bisa* 'can' is preceded by its experiencer *anak kecil-kecil* 'little children' and followed by its theme *bahasa Inggris* 'English', while in (4b) *bisa* is preceded by its theme *tiga belas* 'thirteen' and followed by its experiencer *aku* 'I'. In (5a) *tahu* 'know' is preceded by its experiencer *saya* 'I' and followed by its theme *dia bilang apa* 'what he's saying', whereas in (5b) *tahu* is preceded by its theme *siapa yang kena* 'who got hit' and followed by its experiencer *aku* 'I'. In (6a) *pasang* 'attach', 'place' or 'set off' is preceded by its actor *aku* 'I' and followed by its patient *dua ribu* 'two thousand', while in (6b) *pasang* is preceded by its patient *bom* 'a bomb' and followed by its actor *dia* 'he'. In (7a) *kasi* 'give' is preceded by its actor *dia* 'they' and followed by its patient *dua-dua* 'two each', whereas in (7b) *kasi* is preceded by its patient *ini-ini* 'these' and followed by its actor *dia* 'it'. In (8a) *tengok* 'look' is followed first by its actor *Arip* 'Arip' and then by its patient *poto anak gembel semua* 'all the pictures of the beggar boys', while in (8b) *tengok* is followed first by its patient *tikus* 'the mice' and then by its actor *aku* 'I'. And in (9a) *bawa* 'take' is preceded by its actor *David* 'you David' which in turn is preceded by its patient *komputer* 'computer', whereas in (9b) *bawa* is preceded by its patient *komputer* 'computer' which in turn is preceded by its actor *David* 'you David'. Thus, in a basic sentence containing an activity and two participants, the thematic roles of the participants are underspecified: context alone is what enables the hearer to assign particular thematic roles to participants in a particular utterance. In this respect, then, Riau Indonesian differs from the standard language, where, in such constructions, thematic roles are determined by word order.

Given such indeterminacy, one might expect Riau Indonesian to make use of various optional formal strategies to disambiguate thematic roles in those situations when there is a functional communicative need to 'zero in' on one possible interpretation and rule out other undesirable ones. However, this is not the case. Not only does Riau Indonesian *allow* such underspecification; in fact, it has at its disposal no grammatical devices, morphological or syntactic, which, when so desired, can be invoked to force a particular assignment of thematic roles and thereby eliminate the indeterminacy.

And what of the prefixes *di-* and *N-*? Can they not be used to associate certain thematic roles with certain participants? Indeed, these two prefixes occur quite frequently in Riau Indonesian. However, their function is quite different from that in Standard Malay/

Indonesian; in particular, they do not have the expected effect of assigning thematic roles and thereby eliminating the underspecification exemplified in (4)–(9) above. We shall now examine, in detail, the function of each of these two affixes in turn.

3.1.1 The prefix *di-* in Riau Indonesian

If the prefix *di-* were a marker of passive voice, one would expect forms marked with *di-* to be followed by actors, not by patients, and to be preceded by patients, not by actors. Sometimes, indeed, this is the case; but not always. Consider examples (10)–(12) below. (In each of the examples from here on, the participants which are of direct relevance to the subsequent discussion are italicised.)

- (10)a. Ndak bisa *dinaikkan itu*.
 NEG can *di*-ride-APPL DEM:DIST
 [At airport, man loading luggage onto conveyor belt encounters a damaged piece of luggage]
 ‘This can’t be loaded.’
- b. Sudah *diangkat barang* sama orang.
 PFCT *di*-lift thing accompany person
 [Landing at airport, arriving late at conveyor belt, passenger is worried]
 ‘The things may have already be taken by someone.’
- (11)a. *Aku digoreng*.
 1SG *di*-fry
 [Restaurant worker commenting to customer on the fried rice he had just served him]
 ‘I fried it.’
- b. *Aden disimer*.
 1SG *di*-polish
 [Shoeshine boy pointing to potential customer’s sandals, addressing other shoeshine boys, who are possible competitors]
 ‘I’m polishing them.’
- (12)a. Ah, *saya tak diganggu mister*.
 EXCL 1SG NEG *di*-disturb white.person
 [Playing Nintendo in turns; after I had played, speaker begins to play and I try to interfere, speaker observes that he didn’t disturb me when I was playing, implying that I shouldn’t disturb him now]
 ‘I didn’t disturb you.’
- b. I, *sakit engkau diituin aku*.
 EXCL hurt 2 *di*-DEM:DIST-APPL 1SG
 [During horseplay]
 ‘Eee, that hurts, you doing that to me.’

In (10), forms marked with *di-* are followed by a patient, rather than an actor (see also (13a) below). Such constructions are in fact attested also in the standard language. More surprising perhaps are the constructions in (11), in which forms marked with *di-* are preceded by an actor, rather than a patient. In Standard Malay/Indonesian, (11a) could only mean ‘I

was fried', and (11b) could only be understood as 'I was polished'. But such interpretations are quite obviously not intended here. Even more noteworthy are the constructions in (12), in which forms marked with *di-* are followed by a patient *and* preceded by an actor. Again, in Standard Malay/Indonesian, (12a) could only mean 'I wasn't disturbed by you' and (12b) could only be interpreted as 'Eee, that hurts, you having that done by me'. But although such readings are also available in Riau Indonesian, they are clearly not the ones that are intended in the actual contexts in question. (Additional constructions similar to these are given in (13b), (15a) and (15b) below.) Thus, examples (10)–(12) show clearly that the prefix *di-* does not function to discriminate actors from patients: neither does it mark a following participant as actor, nor does it mark a preceding participant as patient. In doing so, examples such as these underscore a fundamental difference between the prefix *di-* and prototypical markers of passive across languages.

In the absence of the usual grammatical functions associated with passives, it might nevertheless be expected that the prefix *di-* be associated with some of the discourse functions characteristic of passive constructions. In particular, one might anticipate that a form prefixed with *di-* marks its patient as being of high discourse prominence, for example the sentential topic, or otherwise definite. However, this expectation is also clearly belied by the facts:

- (13)a. *Jangan distop taxi, David.*
 NEG:IMP *di-*stop taxi David
 [Shopping, speaker isn't ready to go home yet]
 'Don't stop a taxi, David.'
- b. *Saya ditaruk enam.*
 1SG *di-*put six
 [Playing cards; interlocutor says to speaker that it's the speaker's turn; speaker says no it isn't]
 'I put down a six.'
- (14)a. *Bodoh, disimpan-simpan, tak mau dibagi-bagi.*
 stupid *di-*DISTR-put.away NEG want *di-*DISTR-give
 [At night market, complaining about friend who was pouring a drink, bit by bit, into a bottle, instead of giving it to the rest of his friends to share]
 'Stupid, *he's* putting it away, he won't share any of it.'
- b. *Buku pun dibawa.*
 book CNJ.OP *di-*carry
 [Speaker complaining that every time I go out I take lots of things with me]
 'You even take the notebook with you.'
- (15)a. *Saya dicari sepuluh lagi.*
 1SG *di-*seek ten CNJ.OP
 [Playing Mario, trying to get additional bonus points]
 'I'm trying to get ten more.'
- b. *Dia dikasi kad.*
 3 *di-*give card
 [At Kentucky Fried Chicken, in exchange for coupons]
 'They'll give us a card.'

In (13), the patient is indefinite, rather than definite. In (14), it is the actor, rather than the patient, that is the sentential topic (since it is not expressed overtly, its English gloss is italicised). And in (15), both properties obtain: the patient is indefinite *and* the actor is the sentential topic. From a discourse perspective, then, these are the last constructions that one would expect to find in passive form. Yet they contain the prefix *di-*. Thus, these examples show that the prefix *di-* does not possess the discourse properties characteristic of markers of the passive construction.⁵

So what, then, *is* the function of the prefix *di-*? As shown above, when attaching to a word, it does not pick out a patient associated with that word and assign it syntactic salience by requiring it to precede the host word, nor does it assign a patient discourse salience by marking it as the topic of the sentence. Nevertheless, the prefix *di-* is quite clearly a patient-oriented prefix. But its function is in fact much more straightforward. When attaching to a word, it marks that word, quite simply, as *having* a patient in its argument structure. In other words, it assigns a patient semantic or conceptual salience, by asserting its very existence.⁶

Constructions containing *di-* are of two quite distinct types, depending on the argument structure of the word to which *di-* attaches. Most commonly, the prefix *di-* applies to a word whose argument structure already contains a patient; in such cases, the effect of the prefix is to add emphasis to the patient. This first type of construction is illustrated by all of the examples in (10)–(15). In such constructions, the prefix *di-* is, narrowly speaking, redundant, since it does not change the interpretation of the construction *per se*, but only shades its perspective in a somewhat different way, drawing attention to the presence of the patient argument.

In this respect, the prefix *di-* is comparable to any number of other optional elements which may be added to a given construction. For example, a numeral may occur in direct construction with a noun, e.g. *tiga mangga* 'three mangoes'; however, an optional numeral classifier such as *biji* may be added to emphasise the unit of enumeration, e.g. *tiga biji mangga*. Similarly, like in other so-called 'pro-drop' languages, a single word, e.g. *lari* 'run',

⁵ Examples (10)–(15) above may also be examined in terms of Hopper and Thompson's (1980) notion of transitivity. Some are of low transitivity, e.g. (13a) with its irrealis mode and indefinite patient: such examples, although conflicting with Hopper's (1979, 1983) characterisation of Standard Malay/Indonesian *di-* as a marker of high transitivity, are nevertheless consistent with the cross-linguistic tendency for passive constructions to be of low transitivity. However, other examples are of high transitivity, e.g. (10b) with its perfective aspect and definite patient: such examples, while constant with Hopper's description of Standard Malay/Indonesian *di-*, are inconsistent with the cross-linguistic tendency for passive constructions to be of low transitivity.

⁶ To put it somewhat differently, the prefix *di-* may be thought of as some kind of an optional pronominal proclitic, referring to a participant bearing the thematic role of patient. In this context, it is worthy of note that the third person pronoun *dia* is sometimes suggested as the diachronic source for the prefix *di-* in Malay/Indonesian. Support for this claim is supposedly derived from the observation that in Standard Malay/Indonesian, the '*di-* passive' is grammatical only when the actor is third person: for first and second person actors the so-called 'pseudo-passive' or 'second passive' construction, as in (2c), must be used instead. Thus, the argument goes, the '*di-* passive' derives from a 'pseudo-passive' with third person pronoun *dia* – the form *dia* having been reduced to *di-*. Whatever the merits of this diachronic scenario for Malay/Indonesian in general, the above data show clearly that in Riau Indonesian, the third-person-actor constraint does not apply: forms prefixed with *di-* may have actors that are second person, as in (12b), or first person, as in (12a). Thus, if the prefix *di-* is to be thought of as some kind of a pronoun in Riau Indonesian, it is one that is marked for thematic role but not person.

may function as a complete sentence, e.g. 'He is running'; however, an optional pronoun *dia* may be added to emphasise the third person reference, e.g. *Dia lari*.

In certain cases, though, the presence of the prefix *di-*, albeit technically redundant, serves to enhance the discourse cohesion of the sentence. Consider the following example:

- (16) **Ditangkap dibotakkan disuruh** pulang kampung.
di-catch di-bald-APPL di-order go.home village
 [From narrative about a group of shoeshine boys, about what happens when they get caught by the police]
 'They're caught, they have their heads shaven, and they're told to go back home to their villages.'

In the above construction, a sequence of three activity words are prefixed with *di-*, the presence of the prefix suggesting that the three words share a common patient — in the case at hand, the sentential topic, the group of shoeshine boys.

Turning now to the second type of construction containing *di-*; here the argument structure of the host word does not contain a patient — generally, the word in question is one that refers to a thing or object rather than an activity. Although less frequently occurring than the first type, such constructions provide the most obvious evidence in support of the above-proposed analysis of the prefix *di-*: in this case, its function is to *introduce* a patient into the argument structure of its host word. Consider the following examples:

- (17)a. **Udah disambut.**
 PFCT **di-ant**
 [Interlocutor offers speaker drink; speaker complains that it's been lying around for a long time]
 'It's already got ants in it.'
- b. **Kentot dia dililin.**
 fuck 3 **di-candle**
 [Watching a movie in which woman in bathtub lowers a burning candle to her groin]
 'She's fucking with a candle.'

In (17a), the argument structure of the word *sambut* 'ant' does not contain a patient; the prefix *di-* thus introduces a patient, that is to say, something that is affected by ants, or 'anted' — in the case at hand, the drink. In (17b), the argument structure of the word *lilin* 'candle' does not contain a patient; once again, it is the prefix *di-* which adds a patient, somebody on the receiving end of the candle, or 'candled' — in this instance, the woman in the bathtub.⁷

The examples in (17), in which *di-* attaches to a word referring to a thing or object, call to mind another class of constructions:

⁷ The construction in (17b) is reminiscent of that in several dialects — e.g. Kelantan Malay (Abdul Hamid, 1994:202-205), Salako Dayak (Ina Anak Kalom and Hudson 1970:287-288), and Seraway Malay (Helfrich 1904:211) — in which, in passive sentences, *di-* may occur as a marker of the actor NP, or 'by-phrase'.

- (18)a. Saya enak **di** luar.
 1SG nice **di** outside
 [Swimming, speaker has had enough]
 'I want to go out.'
- b. Cari isteri **di** Pakning satu.
 seek wife **di** Pakning one
 [Conductor on bus telling unmarried passenger heading towards Sungai Pakning
 what he should do when he gets there]
 'Look for a wife in Pakning.'
- c. **Di** laut-laut, 'kan?
di DISTR-sea NEG
 [From a narrative about a sweet-water dugong which swam down to the sea,
 and, as a result of ingesting salt water, was caught by fishermen; speaker
 explaining where there is salt water]
 'In the sea, right?'
- d. **Di** bernang makan.
di swim eat
 [Asked whether he is hungry, speaker answers not, and explains, alluding to
 their earlier trip to the swimming pool]
 'I ate where we were swimming.'

Examples such as the above are generally considered to instantiate a different construction type, involving a 'different' grammatical marker, namely, a locative preposition *di*. This difference is reflected in the standardised orthography, used here, in which the locative *di* is written as a separate word.

However, the characterisation of the patient-oriented prefix *di-*, provided above, may be straightforwardly extended to account also for the locative preposition *di*, suggesting that the two may just be aspects of one and the same form. Specifically, whereas the prefix *di-* asserts that its host's argument structure contains a patient, the preposition *di* requires that its host's argument structure include an element of a somewhat different thematic role — that of a participant located in a place associated with the host expression. This thematic role may be referred to as locative theme.

In (18a) and (18b), *di* precedes words denoting locations, *luar* 'outside' and *Pakning*, the name of a small town. Such words already have a locative theme as part of their argument structure, namely whatever entities are located outside, or in Pakning. Accordingly, in examples such as these, the function of *di* is merely to emphasise an already existent argument — its effect is thus analogous to that of *di-* in (10)–(16). In contrast, in (18c), *di* precedes a word denoting a thing, while in (18d) it precedes a word denoting an activity. In both of these cases, the word in question does not have a locative theme as part of its argument structure; here, therefore, the function of *di* is to introduce a new locative theme into its host's argument structure — its effect is accordingly parallel to that of *di-* in (17).

Like any other theoretical constructs, thematic roles are not axiomatic, immutable objects, but rather entities that are posited in order to account for observed patterns of linguistic features. Recently, a number of scholars have suggested that thematic roles do not constitute an unstructured set of coordinate items, but instead may be hierarchically organised, with certain thematic roles grouping together to constitute superordinate roles, or hyperroles — see,

for example, Foley and Van Valin (1984), Kibrik (1997). Accordingly, throughout this paper, the thematic role of actor is understood as encompassing the two more specific roles of agent and experiencer. Thus, in the case at hand, rather than stipulating that *di-* marks the presence of either of two thematic roles, patient or locative theme, we may instead characterise the function of *di-* as asserting the existence of a single superordinate role, subsuming both patient and locative theme — a role which we may refer to as generalised patient.⁸

The semantic motivation of the thematic role of generalised patient is not too difficult to appreciate. Essentially, the generalised patient is that argument which is viewed as being in a *subordinate* relationship to the word marked with *di-*. In those cases where the generalised patient is an ordinary patient, subordination is realised more abstractly, in that the argument is viewed as undergoing an activity, or being affected in some way by the word prefixed with *di-*. And in those cases where the generalised patient is a locative theme, the subordinate relationship is manifest in concrete spatial terms, with the argument being physically enclosed or engulfed by the expression preceded by *di*. Note, however, that while subsuming the roles of patient and locative theme, the characterisation of generalised patient excludes ordinary themes, such as, for example, the single argument of property words such as *putih* 'white' and *besar* 'big'. And indeed, such words, when interpreted in the usual way as denoting properties, do not occur with the marker *di-*.

Further support for the unified analysis of *di-* is provided by examples such as the following:

- (19) Ini bisa **disinikan?** **dibawakan?**
 DEM:PROX can **di**-here-APPL **di**-down-APPL
 [Speaker learning how to use word processor, points to a lower position on
 the screen and asks if the cursor can be brought down there]
 'Can this be brought here? Brought down?'

If a distinction is to be maintained between a patient-oriented *di-* and a locative *di*, then constructions such as the above present an analytical puzzle: which of the two is it in (19)? Either analysis is plausible. If it is the patient-oriented prefix *di-* (as suggested, arbitrarily, by the orthography), then the appropriate structures are *di*-[*sinikan*] and *di*-[*bawakan*]: the prefix *di-* may then be viewed as asserting that *sinikan* 'bring here' and *bawakan* 'bring down' have a patient, which is the cursor, referred to by the demonstrative *ini* 'this'. If on the other hand it is the locative preposition *di*, then the resulting structures are [*di sini*]-*kan* and [*di bawa*]-*kan*: the preposition *di* may be taken to apply to the stems *sini* and *bawa*, asserting that both have a locative theme, which, once again, is the cursor, referred to by *ini* — the resulting expressions subsequently obtaining their causative meanings from the applicative enclitic *-kan*. Thus, both analyses end up with essentially the same meaning. While it is not impossible for a string of words to have two distinct structures which are then associated with the same meaning, such a state of affairs is not usual, and would not be posited without explicit motivation for the existence of two distinct structures. However, in the case at hand, there is no reason to posit such structural ambiguity: it is hard to imagine that the speaker actually had in mind one of the two structures to the exclusion of the other. Instead, the simplest course

⁸ In the preceding sentence, and in subsequent discussion, the single form in question is referred to as *di-* (with the hyphen), regardless of which of the two subtypes of generalised patient it refers to, and how it is accordingly written in the standard orthography.

isto simply discard the irrelevant distinction between a patient-oriented *di-* and a locative *di*. Once this is done, the puzzle disappears, and we are left, in (19), with a unitary underspecified *di-* marking the single role of generalised patient.

Yet additional evidence in favour of the unified analysis of *di-* is provided by typographical 'errors' committed by speakers of Riau Indonesian. As noted previously, in the standard orthography, the patient-oriented *di-* is written joined on to its host while the locative *di* is written as a separate word. Indonesian elementary school grammar textbooks often devote a section to explaining when to write *di-* or *di* — a fair indication, albeit not specific to the Riau dialect, that the distinction in question is unnatural, and lacking in psychological reality. And indeed, speakers, both uneducated and educated, often make 'mistakes', writing *di-* instead of *di*, or vice versa. Following is one example:

(20)

NOTICE

- *Dilarang Merokok ditempat tidur*
- *Please Refrain from smoking in bed*



THE MANAGEMENT
City Hotel



Dilarang merokok ditempat tidur.
di-forbid meN-smoke di-place sleep
 [Sign in room, City Hotel, Dumai]
 'Please refrain from smoking in bed.'

In (20) above, the first *di-*, in *Dilarang*, is the patient-oriented one, and is written correctly; however, the second *di-*, in *ditempat*, is the locative one, and should have been written as a separate word, rather than joined on, as it is above. 'Errors' such as these show that speakers of Riau Indonesian find it difficult to master the prescriptive distinction between two *di-*'s, thereby providing further support for the claim that, in their dialect at least, it is a spurious distinction.

3.1.2 The prefix *N-* in Riau Indonesian

Functionally, the prefix *N-* is a mirror-image of *di-*: while *di-* is patient-oriented, the prefix *N-* is oriented towards the thematic role of actor. Evidence for this is presented later in this section.

Before examining the relevant data, however, it is first necessary to take note of the fact that from a formal point of view, the two prefixes are of a rather different nature. Specifically, whereas *di-* displays regular agglutinative behaviour, *N-* is more closely bound

to its host, more irregular in its form and distribution, and would appear to reflect a greater degree of grammaticalisation. Some of the formal differences between *di-* and *N-* are summarised in Table 1, supported by the evidence in Tables 2 and 3 and example (21):

Table 1: *di-/N-* asymmetries

criteria	<i>di-</i>	<i>N-</i>	examples
productivity	productive	semi-productive	(22), (23)
form	immutable	phonologically conditioned	(23)
linearity	segmentable	mostly non-segmentable	(23)
reduplication	never undergoes reduplication	usually undergoes reduplication	(24)

The most important difference between *di-* and *N-* is with regard to their productivity: whereas *di-* can attach to any word whatsoever, *N-* can attach to some words but not others. The distribution of *N-* is partly arbitrary, and partly subject to phonological constraints. The arbitrary aspect of the distribution of *N-* may most appropriately be viewed as reflecting an ongoing cyclical diachronic process of lexicalisation, as represented in Table 2 below:

Table 2: Lexical distribution of *N-*: the cycle of lexicalisation

Stage	Stem	<i>N</i> -stem	Gloss
non-existent	(a) <i>pergi</i> <i>tidur</i> <i>susu</i>	* <i>mergi</i> * <i>nidur</i> * <i>nyusu</i>	'go' 'sleep' 'milk'
rare	(b) <i>taruk</i> <i>putih</i> <i>teh</i>	R <i>naruk</i> R <i>mutih</i> R <i>neh</i>	'put' 'white' 'tea'
common	(c) <i>tengok</i> <i>simer</i> <i>kopi</i>	<i>nengok</i> <i>nyimer</i> <i>ngopi</i>	'look' 'polish' 'coffee'
usual	(d) R <i>tampak</i> R <i>cilam</i> R <i>kentot</i>	<i>nampak</i> <i>nyilam</i> <i>ngentot</i>	'see' 'dive' 'fuck'
reanalysed as part of stem and therefore non-existent	(e) <i>nonton</i> <i>nangis</i> <i>ngantuk</i>	– (Standard Indonesian: <i>tonton</i>) – (Standard Indonesian: <i>tangis</i>) – (Standard Indonesian: <i>antuk</i>)	'watch' 'cry' 'sleepy'
*	(f) <i>makan</i> <i>minum</i> <i>masuk</i> (g) <i>maling</i> <i>nyamuk</i> <i>mata</i>	(Tagalog: <i>kain</i>) (Tagalog: <i>inum</i>) (Tagalog: <i>pasok</i>) (Proto Malayo-Polynesian: <i>maliN</i>) (Proto Malayo-Polynesian: <i>ñamuk</i>) (Proto Malayo-Polynesian: <i>mata</i>)	'eat' 'drink' 'enter' 'steal' 'mosquito' 'eye'

At stage (a) are words which never occur with the prefix *N-*. At stage (b) we find words which may take the prefix *N-*, but do so very infrequently, indicated by the symbol 'R(are)'. At stage (c) are words which commonly occur both in bare form and with the prefix *N-*. At

stage (d) we have words which usually take the prefix *N-*, and only rarely occur in bare form. And at stage (e) are words which, from a diachronic perspective, always occur with the prefix *N-*. However, from a synchronic point of view, these words can no longer be said to contain the prefix *N-*, since, in Riau Indonesian, the bare forms (*tonton*, *tangis*, *antuk*) do not occur. Rather, these words are now themselves in bare form, containing a stem which begins with a nasal consonant. From a Riau Indonesian perspective, they are thus no different from the words in (f), also in bare form, but which may be reconstructed as containing a reflex of a nasal prefix at a much earlier, pre-Malayic stage — as suggested by their Tagalog cognates, without the initial nasal. Nor for that matter are they any different from the words in (g), with a stem-initial nasal which is reconstructable all the way back to Proto Malayo-Polynesian, and hence without any connection whatsoever with the prefix *N-*, and therefore outside the cycle of lexicalisation.⁹

Thus, the prefix *N-* may be viewed as entering into usage, establishing a foothold, becoming more and more frequent, until finally it is used invariably, at which point it is reanalysed as part of the stem, and thus can be said to have disappeared. At any given point in space and time — such as, in the case at hand, present-day Riau Indonesian — different words reflect different stages in the ongoing cyclical process of lexicalisation.

As suggested by the data in Table 2 above, the distribution of *N-* is without obvious semantic motivation. For example, whereas *susu* 'milk' never takes *N-*, *teh* 'tea' does so but rarely, while *kopi* 'coffee' does so much more frequently. Similarly, whereas *tengok* 'look' often takes *N-*, *tampak* 'see' almost always does, and *tonton* 'watch' has already been reanalysed as *nonton*.

However, although partly arbitrary, the distribution of *N-* is also governed, in part, by systematic phonological factors. These, and also the variable phonological realisation of *N-* when it does occur, are spelt out in Table 3 below.

Table 3 specifies the place and manner of articulation of the initial consonant of the stem to which *N-* attaches, with filled cells representing actual phonemes, and empty cells corresponding to non-existent sounds. Within each filled cell, the first line provides an example of a stem taking the prefix *N-*, if such examples are attested; e.g. in the upper left-hand cell, *putar* takes *N-* to become *mutar*. The second line gives the gloss of the example, if present; e.g. for *putar*, 'revolve'. And the third line characterises the frequency with which *N-* occurs, with stems beginning with the consonant in question; e.g. for stems beginning with a *p-*, the prefix *N-* is common.

As evident in Table 3, and in particular the third lines of the various cells, the distribution of the prefix *N-* is phonologically conditioned by the first consonant of the stem. Specifically, if the first consonant is a voiced stop, an *h*, or a semi-vowel, *N-* is rare or unattested; otherwise, *N-* is common.¹⁰ (A special case is that of nasal consonants, for which the presence of *N-* is not detectable.) Together, then, the data in Tables 2 and 3 show that the

⁹ The Proto Malayo-Polynesian forms cited in (22) above are from the reconstruction by Dempwolff (1938), as cited by Adelaar (1992).

¹⁰ The difference between 'unattested' and 'rare' in Table 3 is probably not significant. Specifically, the absence of any examples of stems beginning with *d-* that take *N-* is presumably an accidental gap in my data, given that the corresponding examples with the other voiced stops are also very rare. Similarly, the absence any examples of stems beginning with semi-vowels that take *N-* most likely reflects the fact that such stems are, themselves, quite uncommon.

distribution of the prefix *N-* is semantically arbitrary, but partly governed by regular phonological conditioning.

Table 3 also underscores a second important difference between the prefixes *di-* and *N-*. In contrast to *di-*, which is formally immutable, the form of the prefix *N-* is phonologically conditioned, again by the first consonant of the stem. If the first consonant is an obstruent, it is replaced by the homorganic (or nearest) nasal, while if it is a liquid, the consonant remains, and the stem is prefixed with *me-*.¹¹

Table 3: Realisation of *N-*

	Labial	Dental	Palatal	Velar	Glottal
unvoiced stop	<i>putar</i> > <i>mutar</i> 'revolve' [common]	<i>tulis</i> > <i>nulis</i> 'write' [common]	–	<i>kopi</i> > <i>ngopi</i> 'coffee' [common]	<i>erti</i> > <i>ngerti</i> 'meaning' [common]
voiced stop	<i>baca</i> > <i>maca</i> 'read' [rare]	<i>d-</i> > [unattested]	–	<i>ganggu</i> > <i>ngganggu</i> 'disturb' [rare]	–
unvoiced affricate	–	–	<i>cuci</i> > <i>nyuci</i> 'clean' [common]	–	–
voiced affricate	–	–	<i>jual</i> > <i>nyual</i> 'sell' [common]	–	–
unvoiced continuant	–	–	<i>simer</i> > <i>nyimer</i> 'polish' [common]	–	<i>hisap</i> > <i>ngisap</i> 'suck' [rare]
sonorant	–	<i>rokok</i> > <i>merokok</i> 'cigarette' [common]	–	–	–
lateral sonorant	–	<i>ledak</i> > <i>me ledak</i> 'explode' [common]	–	–	–
nasal	<i>m-</i> > <i>m-</i> [untestable]	<i>n-</i> > <i>n-</i> [untestable]	<i>ny-</i> > <i>ny-</i> [untestable]	<i>ng-</i> > <i>ng-</i> [untestable]	–
semi-vowel	<i>w-</i> > [unattested]	–	<i>y-</i> > [unattested]	–	–

¹¹ The assimilation of *N-* in Riau Indonesian thus differs from that of *meN-* in Standard Malay/Indonesian with respect to the voiced stops and both affricates: whereas in the standard language, the homorganic nasal precedes the consonant in question (*memb-*, *mend-*, *mengg-*, *menc-*, *menj-*), in Riau Indonesian the homorganic nasal replaces the consonant in question — as it does for other obstruents. Work in progress suggests that the form of *N-* in Riau Indonesian is similar to, and hence is probably derived from, the form of the corresponding prefix in some of the Malay dialects spoken in Riau province. Of those Malay/Indonesian dialects that have been described in the linguistic literature, Riau Indonesian would appear to most closely resemble its counterpart in Iban (Asmah 1981), where the homorganic nasal also replaces all of the obstruent consonants. A further similarity between Riau Indonesian and Iban is that whereas initial obstruents are replaced, initial liquids trigger prefixation. However, whereas in Riau Indonesian the form of the prefix is *me-*, in Iban it is *nge-*.

Entailed by the above is a further difference between the two prefixes. Whereas *di-* is transparently segmentable, offering a clear example of agglutinative morphology, this is only true of *N-* in those cases where it is realised as *me-*. In most other instances, *N-* is non-segmentable, manifesting itself as a mutation of the initial consonant of the stem.

In turn, the segmentability of the two prefixes affects the way in which they interact with reduplication. As a rule, if the prefix is segmentable, it does not undergo reduplication, whereas if it is not segmentable, it reduplicates together with the stem. Accordingly, the prefix *di-* never undergoes reduplication, for example *dipinjam-pinjam*, not **dipinjam-dipinjam*, from *pinjam* 'borrow'. Similarly, the prefix *N-*, when realised as *me-*, does not undergo reduplication, as in (21a) below. However, when the prefix *N-* is manifest as a mutation of the initial consonant of the stem, it does undergo reduplication, as in (21b–e):

(21) *N-* and reduplication

a.	<i>melempar-lempar</i>	<i>*melempar-melempar</i>	'throw'
b.	<i>*minjam-pinjam</i>	<i>minjam-minjam</i>	'borrow'
c.	<i>*nembak-tembak</i>	<i>nembak-nembak</i>	'shoot'
d.	<i>*nyimer-simer</i>	<i>nyimer-nyimer</i>	'polish'
e.	<i>*ngopi-kopi</i>	<i>ngopi-ngopi</i>	'coffee'

Thus, as shown in Tables 2 and 3 and example (21), the prefix *N-* differs formally from *di-* in that it is less productive, more irregular in its behaviour, and more closely bound to its host word — in summary, more grammaticalised. Functionally, however, the prefix *N-* is a clear mirror image of *di-*, as we shall now see.

The traditional characterisation of the prefix *di-* as a marker of passive voice was disproved, in the preceding section, by the existence of constructions in which a form marked with *di-* is followed by a patient and/or preceded by an actor — see examples (10)–(12). Similarly, the characterisation of the prefix *N-* as a marker of active voice may be refuted by the existence of mirror-image constructions in which a form marked with *N-* is followed by an actor and/or preceded by a patient. Some examples of such constructions are given in (22)–(24) below:

(22)a. Dia tiap hari nyimer sama aku.
 3 every day N-polish accompany 1SG
 [Shoeshine boy pointing to regular customer]
 'I polish his shoes every day.'

b. Nampar sama komandan.
 N-slap accompany commander
 [About army life, and what happens to new recruits who are caught smoking]
 'They get slapped by the commander.'

(23)a. 'Mister' manggil.
 'mister' N-call
 [Shop attendants debating how to address me: 'bapak', 'abang', or...]
 'Call him "mister".'

b. Ada perempuan tadi, inrya nampak.
 exist woman PAST:PROX DEM:PROX-ASSOC N-see
 [Commenting on a woman with a low front to her blouse who had just passed by,
 speaker points to his own chest and says]
 'The woman before, her 'this' was showing.'

- c. *Saya lima puluh naruk.*
 1SG five ten N-put
 [Con men playing board game and trying to lure passersby to place bets;
 one of the con men calls out the bet that he has just placed]
 'I put down fifty.'
- d. *Simer aku tak boleh dia makai.*
 polish 1SG NEG can 3 N-use
 [Shoeshine boy gang chatter]
 'He's not allowed to use my shoeshine equipment.'
- e. *Ini siapa nulis?*
 DEM:PROX PERS-what N-write
 [Pointing to a receipt]
 'Who wrote this?'
- (24)a. *Putih 'kan nampak sama ikan.*
 white NEG N-see accompany fish
 [Watching TV program about fishing lures]
 'The white ones, the fish can see.'
- b. *Eddy Tansil tak bisa nangkap orang.*
 Eddy Tansil NEG can N-catch person
 [About an infamous criminal who escaped Indonesia to China]
 'Nobody can catch Eddy Tansil.'

In (22), the mirror image of (10), forms marked with *N-* are followed by an actor, rather than a patient. (See also (26a) below.) Whereas Standard Malay/Indonesian also permits forms marked with *meN-* to be followed by an actor, the constructions in (22) differ from their standard counterparts in the presence of *sama*, a multifunctional word, often used to mark an actor: in this usage it is reminiscent of the standard form *oleh* used to mark the actor of a form prefixed with *di-*. More divergent from the standard language are the constructions in (23), the mirror image of (11), in which forms marked with *N-* are preceded by a patient. In (23a) and (23b) the patient immediately precedes the form marked with *N-* and is its only overtly expressed argument. (Another similar example occurs in (27b) below.) In the analogous constructions with *meN-* in Standard Malay/Indonesian, the argument in question could only be understood as the actor: (23a) could only mean 'Mister calls (somebody)', and (23b) could only be interpreted as 'The woman before, her "this" was seeing (something)'. However, such interpretations are clearly not appropriate in the given context — indeed, the latter one is semantically incoherent. In (23c) the patient also immediately precedes the form marked with *N-*, but is itself preceded by the actor. In this case, the analogous construction with *meN-* in Standard Malay/Indonesian is simply ungrammatical. And in (23d) and (23e) the patient occurs at the beginning of the sentence, while the actor occurs immediately in front of the form marked with *N-*. Here, too, the analogous constructions with *meN-* in the standard language are ungrammatical, see example (2b) and subsequent discussion. Finally, in (24), the mirror image of (12), forms marked with *N-* are followed by an actor *and* preceded by a patient. Again, in Standard Malay/ Indonesian, the analogous constructions could only mean 'The white ones can see the fish' and 'Eddy Tansil can't catch anybody' — clearly not the intended interpretations here. Thus, examples (22)–(24) show that the prefix *N-* does not function to discriminate actors from patients: neither does it mark a following

participant as patient, nor does it mark a preceding participant as actor. In doing so, examples such as these highlight an essential difference between the prefix *N-* and prototypical markers of active clauses across the world's languages.

As with *di-* in the preceding section, it might be expected that, in the absence of the usual grammatical functions associated with active voice, the prefix *N-* might, nevertheless, be associated with some of the discourse functions characteristic of active constructions. In particular, one might predict that a form prefixed with *N-* marks its actor as being of high discourse prominence, for example the sentential topic, or otherwise definite. Once again, however, this expectation is also contradicted by the evidence:

- (25)a. Ada *orang* *nelepon* *saya*?
 exist person N-telephone 1SG
 [Telling what his boss would typically say when returning]
 'Were there any calls for me?'
 b. Ada *orang* *nyimer* *situ* *tak*?
 exist person N-polish DEM:DIST NEG
 [Shoeshine boy asking about Singapore]
 'Are there people shining shoes there?'
- (26)a. Kampung *saya* sering *nangkap* *ini* *saya*.
 village 1SG often N-catch DEM:PROX 1SG
 [Eating crabs, speaker points to crab and says]
 'In my village I would often catch ones like this.'
 b. Saya *makai*, 'kan, belum *rusak*, 'kan.
 1SG N-use NEG NEG:PFCT spoil NEG
 [Speaker boasting that he has used my laptop computer for a long time
 without anything going wrong]
 'When I've been using *it*, nothing has gone wrong yet, has it.'
- (27)a. *Siapa* *nyimer*?
 PERS-what N-polish
 [Shoeshine boy, pointing to my shoes]
 'Who polished *them*?'
 b. *Ini* *mutar* *ini*.
 DEM:PROX N-revolve DEM:PROX
 [Pointing to clock, suggesting that the time shown is wrong]
 'Somebody turned it around.'

In (25), the mirror image of (13), the actor is indefinite, rather than definite. In (26), the mirror image of (14), it is the patient, rather than the actor, that is the sentential topic. And in (27), the mirror image of (15), both properties obtain: the actor is indefinite *and* the patient is the sentential topic. From a discourse perspective, then, these are not the kinds of constructions that one would expect to find bearing a marker of active voice. Nevertheless, they contain the prefix *N-*. Thus, these examples show that the prefix *N-* does not possess the discourse properties characteristic of markers of the active construction.¹²

¹² As before, examples (22)–(27) may also be examined in terms of Hopper and Thompson's notion of transitivity. Some of the above examples are of high transitivity; e.g. (27b) with its perfective aspect

From the above data, it is clear that when the prefix *N-* attaches to a word, it does not assign syntactic salience to an actor associated with that word by requiring it to precede the host word, nor does it assign an actor discourse salience by marking it as the topic of the sentence. Nevertheless, *N-* is, quite obviously, an actor-oriented prefix. In fact, its function is completely parallel to that of the patient-oriented *di-*. Specifically, when attached to a word, it simply says that its host *has* an actor in its argument structure, thereby assigning the actor in question semantic or conceptual salience.

The distribution of the prefix *N-* points towards a characterisation of the thematic role of actor that is a mirror-image of that proposed in §3.1.1 above for generalised patient. Specifically, the actor may be taken to be that argument which is viewed as being in a *superordinate* relationship to the word marked with *N-*. In almost all cases, the argument in question is animate, and enjoys control over the activity in question, as agent or experiencer. Exceptions to this generalisation are few, and of the kind that actually prove the rule. For example, *meledak* ‘explode’ is one of those words that occurs very frequently with the prefix *N-*. Although its argument is usually inanimate, the presence of the prefix is explained by the highly active nature of the activity in question.

Like their opposite numbers with *di-*, constructions containing *N-* are of two different types, depending on the argument structure of the host word. Most commonly, the prefix *N-* attaches to a word whose argument structure already contains an actor — in which case the effect of the prefix is to add emphasis to the actor. This first type of construction is instantiated by all of the examples in (22)–(27). In such constructions, the effect of the prefix *N-* is generally quite subtle, merely contributing towards a more actor-oriented perspective. The following examples, also of the same type, illustrate the effect of the prefix *N-* more vividly:

- (28)a. Ah, masa melihat?
EXCL INTRG:EXCL N-look
[Playing cards, speaker accuses other player of trying to sneak a look at his hand]
‘Hey, why are *you* looking?’
- b. Aku nyimer ... simer sepatu dia.
1SG N-polish polish shoe 3
[Shoeshine boy beginning story about how he polished somebody’s shoes]
‘I polished...polished his shoes.’
- c. Mister, aku nyimer lagi ... simer, ayo.
white.person 1SG N-polish CNJ.OP polish EXCL
[At table with shoeshine boys; speaker takes leave from me...and then,
while walking off, calls out to other shoeshine boys ...]
(to me) ‘I’m going to shine shoes’ (to the other boys) ‘let’s shine shoes, come on.’

and definite patient: such examples, although conflicting with Hopper’s characterisation of Standard Malay/Indonesian *meN-* as a marker of low transitivity, are, nevertheless, consistent with the cross-linguistic tendency for active constructions to be of high transitivity. However, other examples are of low transitivity, e.g. (25b) with its yes/no-question mode, generic mood and covert patient: such examples, while consonant with Hopper’s characterisation of Standard Malay/Indonesian *meN-* as a marker of low transitivity, are, once again, inconsistent with the cross-linguistic tendency for active constructions to be of high transitivity.

In (28a), *lihat* is a verb that rarely takes the prefix *N-*; in the cycle of lexicalisation in Table 2, it would belong in Table 2(b). So the form *melihat* is exceptional, and demanding of an explanation. This is provided by the context, a card game in which the speaker accuses another player of sneaking a look at his hand. In this context, it is precisely the actorhood of the addressee which is at issue, and which triggers the use of the exclamatory interrogative particle *masa*: hence the prefix *N-*.

In (28b), *simer* is a verb that occurs equally freely with or without the prefix *N-*, as suggested, in Table 2 by its inclusion in Table 2(c). In fact, in this example, it occurs in both guises, first with *N-*, then without it. The example consists of a simple transitive sentence, in which the speaker pauses in mid-stream, and then, when resuming, repeats the word *simer*. In the first half of the utterance, the host word is adjacent to the actor *aku* and is accordingly marked with *N-*, while in the second half of the utterance, the same word is now further removed from the actor and closer to the patient, hence the prefix *N-* is absent. Moreover, in this particular example, the surface syntax closely reflects the semantics and the speaker's own perspective. Prototypically, transitive sentences describe activities that flow from actors to patients; when, as in (28b), the word order is *actor – activity – patient*, the linear order can be said to be iconic. Accordingly, in the first half of the utterance, the speaker's attention is closer to the actor, and *N-* is present, while in the second half, the speaker's interest has shifted away from the actor, and *N-* is absent.

In (28c), the same word, *simer*, once again occurs twice: first with the prefix *N-* and then without it. In this case, however, the conditioning factor is a change in the context itself. This example essentially consists of two utterances in close temporal succession. In the first, the speaker, who had been chatting with me at a coffee shop table, takes leave by telling me he's off to polish shoes: in this context, the actor is conceptually salient, as the speaker engaged in an act of leave-taking, and as the polisher of shoes about to resume plying his trade — hence the prefix *N-*. A moment later, however, the same speaker is calling out to his fellow shoeshine boys who had been hanging round the table: in this context, the actor, now including both speaker and addressees, is self-evident, and no longer important. Accordingly, it is not deemed worthy of explicit marking, and the prefix *N-* is absent.¹³

Moving on, now, to the second type of construction containing *N-*: here the argument structure of the host word does not contain an actor — in such instances, the word in question is usually one that refers to a thing or object rather than an activity. Although somewhat less common than the first type, such constructions provide the most straightforward evidence in support of the above-proposed function of the prefix *N-*, in that the function of the prefix *N-* is to introduce an actor into the argument structure of the word containing *N-*. The following examples are mirror-images of those in (17):

- (29)a. *Kita ngopi-ngopi aja.*
 1.2 DISTR-N-coffee NEG:CNJ.OP
 [Somebody suggests that the gang go and eat, speaker counters]
 'Let's just have coffee.'

¹³ An alternative account of the occurrence of *N-* on the first but not the second instance of *simer* in (29c) might make reference to the fact that the second instance is an imperative: in Standard Malay/Indonesian, imperatives typically don't take the prefix *meN-*. However, in Riau Indonesian, *N-* does occur freely in imperative constructions; therefore, its absence from the second instance of *simer* in (29c) must be due to other factors — specifically, those discussed above.

- b. Kenapa terus 'neh o beng'?
 why straight 'N-iced.tea'
 [Speaker and interlocutor have a conventionalised joke, in which speaker asks interlocutor why he is fat, to which interlocutor responds 'teh o beng', '(Because I drink lots of) iced tea'; after going through routine, speaker steps outside the routine and asks]
 'Why do *you* keep on saying "teh o beng"?'
- c. Ada tak untuk putih kulit di Malaysia?
 exist NEG for N-white skin di Malaysia
 [Speaker wants me to buy him some skin lotion on my next trip to Malaysia]
 'Is there anything for making one's skin white in Malaysia?'

In (29a), the argument structure of the word *kopi* 'coffee' does not contain an actor; the prefix *N-* thus introduces an actor, that is to say, somebody that acts in relationship to coffee, namely by drinking it. In fact, the prefixed form *ngopi* occurs frequently with the conventionalised meaning 'drink coffee'. Contrasting with (29a) in an interesting way is (29b) — an example of an idiosyncratic and creative use of grammar by a speaker who delights in various kinds of language play. Like *kopi*, *teh* 'tea' has no actor in its argument structure, and so, the prefix *N-* serves to introduce such an actor. Unlike *ngopi*, however, the prefixed form *neh* hardly ever occurs; recall the contrasting positions of *teh* and *kopi* in the cycle of lexicalisation in Table 2(b) and Table 2(c) above. Moreover, when the form *neh* does occur, it does not bear the corresponding conventionalised meaning 'drink tea'. Rather, in the case at hand, the actor associated with the tea is not the drinker, but rather the person who *says* 'tea': in a more logically felicitous (but awkward) orthography, the prefix *N-* would have had to have been written outside the quotation marks. Finally, in (29c), the argument structure of the word *putih* 'white' contains no actor; hence, here too, the prefix *N-* introduces an actor, which, in the case at hand, assumes the more specific role of cause or instrument. In general, words denoting colours or other properties rarely take the prefix *N-*; like *teh*, *putih* 'white' is also included in Table 2(b) above. Most often, causative forms of property words are marked with the applicative suffix *-kan*, for example *putihkan* 'whiten'. The rather unusual choice of the prefix *N-* in (29c) is a consequence of the particular context, in which the speaker explicitly asks about the existence and availability of the skin lotion, the actor marked with *N-*.¹⁴

3.1.3 Other constructions with *di-* and *N-* in Riau Indonesian

In Riau Indonesian, then, the prefixes *di-* and *N-* fully vindicate Benjamin's (1993:356-357) characterisation:

¹⁴ In conclusion to this section, it is necessary to acknowledge that Riau Indonesian does also possess a certain number of words bearing a prefix *meN-*, e.g. *mendarat* 'land' (for airplanes and so forth) from *darat* 'land', and *meninggal* 'die' (or 'leave the world') from *tinggal* 'leave'; however, these occurrences of the prefix *meN-* are clearly frozen, and the words containing them probably best considered to be loans from Standard Indonesian. The loan nature of these forms is supported also by the morphophonemic realisation of *meN-* before voiced obstruents, as in *mendarat* above: if the form were native to Riau Indonesian, the voiced obstruent would be replaced, resulting in **menarat*, in analogy to *maca* and *nganggu* in Table 3.

Malay-speakers, especially at the colloquial level [...] treat these affixes primarily as having meanings; any grammatical 'functions' they may have are secondary.

As shown above, the prefixes *di-* and *N-* do not pass the most basic test for passive and active markers: they do not even tell us which participant is the patient and which is the actor. Moreover, they do not assign definiteness or topicality to any of the clause's participants. Instead, their role is semantic, and quite straightforward, simply to assert the existence of a patient or an actor, and thereby mark it as conceptually salient in the given context.

The absence of any grammatical function associated with the prefixes *di-* and *N-* manifests itself in a number of additional constructions, which we shall now examine. To begin with, although constituting a clear-cut pair of semantically parallel items, the prefixes *di-* and *N-* do not form a paradigmatic set in the traditional sense. Less like voice markers, tense-and-aspect markers, and other such categories (which in a run-of-the-mill European language every verb is typically marked with once and exactly once), the prefixes *di-* and *N-* are more like adjectives, adverbs, or other such elements, whose presence is optional, and whose function, when present, is simply to add specificity of meaning. Thus, a word denoting an activity has the option of being marked with one of the two prefixes — as in (10)–(20) and (22)–(29), or not being marked, as in (4)–(9). Or, in fact, being marked by *both* prefixes simultaneously, as in the following examples:

- (30)a. *Dimotongnya!*
di-N-cut-ASSOC
 [Watching motorcycle rally on TV; two cyclists battling for lead; one finally overtakes the other on curve; speaker exclaims]
 'He overtook him!'
- b. *Bajunya diminjam.*
 garment-ASSOC **di-N-borrow**
 [Watching movie; scene in department store clothing section; man enters fitting room, woman enters after him, then woman emerges wearing man's clothes; speaker comments]
 'She borrowed his clothes.'

In the above constructions, the prefixes *di-* and *N-* occur on the same word in sequence. Such constructions are relatively infrequent, but no more infrequent than, say, stacked adjectives or stacked adverbs — to which they bear a principled resemblance. In each of the above examples, the double marking is triggered by a specific context, in which both the patient and the actor are conceptually salient and therefore worthy of overt morphological marking. In (30a), the dual prefixation underscores the closeness of the race and the total involvement of both participants, patient and actor, at the dramatic moment of overtaking. And in (30b), the presence of both prefixes befits a comical cross-dressing situation, in which the speaker is suddenly confronted with the unexpected sight of a man's clothes, marked with *di-*, on a woman, marked with *N-*.¹⁵

Further evidence for the lack of any grammatical function associated with the prefixes *di-* and *N-* comes from the interpretation of '*yang*-phrases', or phrases of the form *yang X*, where

¹⁵ Constructions containing both *di-* and *N-* are cited by Dunselman (1949:70) for Kendayan Dayak, and by Ina Anak Kalom and Hudson (1970:287-288) for Selako Dayak; however these constructions differ in that the two prefixes are separated by an additional expression, denoting the actor.

X is some arbitrary expression. In Standard Malay/Indonesian, a phrase of the form *yang X* is understood as referring to the subject of *X*.¹⁶ Accordingly, if *X* is headed by a word *Y* prefixed with *di-*, *yang X* refers to the patient of *Y*; for example, *yang dipukul Ali* 'REL **di**-hit Ali' means 'the one who was hit by Ali'. Similarly, if *X* is headed by a word *Y* prefixed with *meN-*, *yang X* refers to the actor of *Y*; for example, *yang memukul Ali* 'REL **me-N**-hit Ali' means 'the one who hit Ali'. However, since in Riau Indonesian, the prefixes *di-* and *N-* do not identify a particular thematic role as being the subject, it is not surprising to find that the interpretation of *yang*-phrases is not similarly constrained, as can be seen in the following examples:

- (31) Mamak kau *yang dicari* wang.
 mother 2 REL **di**-search money
 [Mother telling story about ungrateful son who wastes all their money, saying how she complained to him that she is the one who has to support the family on her own]
 'Your mother is the one who is looking for money.'
- (32) Ini *yang nembaknya*.
 DEM:PROX REL **N**-shoot-ASSOC
 [Kibbitzing laptop billiards, pointing to a specific ball]
 'This is the one to be shot.' OR: 'Shoot this one.'

In (31), the *yang*-phrase *yang dicari wang* refers to the actor of *cari* rather than the patient. (In Standard Malay/Indonesian, this sentence could only have the semantically bizarre interpretation 'Your mother is the one who is being looked for by the money'.) And in (32), the *yang*-phrase *yang nembaknya* refers to the patient of *tembak* rather than the actor. (In Standard Malay/Indonesian, a sentence like this could only mean 'This is the one to shoot it'.)

A final consequence of the non-grammatical nature of the prefixes *di-* and *N-* is the 'funny control' construction, discussed in §2 above, and illustrated in Standard Malay/Indonesian example (1). Similar examples are widespread also in Riau Indonesian:

- (33)a. Hidung aku *suka ditarik-tarik* sama abang aku.
 nose 1SG like **di**-DISTR-pull accompany elder.brother 1SG
 [About his elder brother]
 'My elder brother likes to pull my nose.'
- b. Doni sama Amat *mau ditumbuk* dia.
 Doni accompany Amat want **di**-hit 3
 [About a quarrel amongst the gang]
 'He wants to hit Doni and Amat.'
- (34) Ini *bisa dibuka*?
 DEM:PROX can **di**-open
 [Trying to open a locked door, having trouble with a bunch of keys]
 'Can this open it?'

¹⁶ It is this fact which led Keenan and Comrie (1977) to place Malay/Indonesian at the top of their accessibility hierarchy, characterising it as having a "subjects only" strategy of relativisation. However, even for the standard language, the facts are not that straightforward; for some alternative perspectives see Cole and Hermon (1998a).

The two examples in (33) correspond closely to those in (1). In (33a), the experiencer of *suka* 'like', namely *abang aku* 'my elder brother', is the actor, rather than the patient, of the word prefixed with *di-*, *tarik-tarik* 'pull'; clearly, the sentence does not mean 'My nose likes to be pulled by my elder brother'. Similarly, in (33b), the experiencer of *mau* 'want', namely *dia* 'he', is the actor, rather than the patient, of the form prefixed with *di-*, *tumbuk* 'hit'; in the given context, the sentence was not intended to mean 'Doni and Amat want to be hit by him'. Example (34) illustrates a similar if not identical kind of construction; here, the experiencer of *bisa* 'can', namely *ini* 'this', is the instrument, rather than the patient, of the expression prefixed with *di-*, *buka* 'open'; in the case at hand, the sentence is not understood to mean 'Can this be opened?'. A mirror-image funny-control construction occurs also with the prefix *N-*, as in the following example:

- (35) Orang tak *mau nyimer* lagi.
 person NEG want N-polish CNJ.OP
 [Gang of shoeshine boys sitting around, somebody suggests that speaker go
 and polish shoes, to which he responds]
 'People don't want to have their shoes polished any more.'

In (35), the experiencer of *mau* 'want', namely *orang* 'people', is the benefactive, rather than the actor, of the word prefixed with *N-*, *simer* 'polish'; in the context in question, the sentence was not intended to mean 'People don't want to polish shoes any more'. Thus, as shown by examples (33)–(35), the prefixes *di-* and *N-* do not necessarily establish a relationship of coreferentiality between the argument they pick out, that is to say the patient or actor of their host word, and the experiencer (or for that matter any other argument) of some other word in the clause — such as, in the above examples, *suka*, *mau* and *bisa*. Accordingly, the prevalence of the funny control construction may thus be viewed as yet another correlate of the absence of any grammatical functions associated with the prefixes *di-* and *N-* in Riau Indonesian.

In conclusion to the discussion of Riau Indonesian, it is worth considering briefly the interrelationship between the prefixes *di-* and *N-* and the inventory of syntactic and grammatical categories in Riau Indonesian. In the course of the preceding discussion, no mention was made of syntactic categories such as noun and verb, nor of grammatical categories such as subject and object (except of course in passing, to deny their relevance). Elsewhere (Gil 1994, 2000b) I have argued that such notions are unnecessary for an adequate description of Riau Indonesian. Indeed, the fact that it proved possible in the preceding pages to provide an adequate and comprehensive description of the prefixes *di-* and *N-* without recourse to notions such as these accordingly provides further support for the claim that Riau Indonesian does not distinguish between syntactic categories such as noun and verb, or between grammatical categories such as subject and object.

One specific corollary of the above is, however, of particular interest. Occasionally, it is suggested that in Standard Malay/Indonesian, one of the functions of these prefixes is to form verbs from nouns. Whatever the merits of such a claim may be for the standard language, in Riau Indonesian it is clearly false: the prefixes *di-* and *N-* are not associated with any verbalising functions whatsoever. Thus, forms prefixed with *di-* or *N-* enjoy the same syntactic distribution as their bare-stem counterparts, and, more generally, exhibit the same syntactic behaviour.

3.2 Other dialects

As shown in §3.1, the behaviour of the prefixes *di-* and *N-* in Riau Indonesian is quite different from that of their counterparts *di-* and *meN-* in Standard Malay/Indonesian. However, preliminary investigations into other dialects suggest that the Riau Indonesian pattern may in fact be the typical one, and that, if anything, it is the standard language that is the outlier with respect to the behaviour of these two affixes.

3.2.1 *The prefixes di- and N- in Jakarta Indonesian*

Jakarta Indonesian is the variety of Indonesian that is spoken in informal contexts in the capital city Jakarta. It is also the language of the Jakarta-based Indonesian entertainment industry, as a result of which it is gaining currency as a kind of colloquial lingua franca throughout the archipelago — where it is beginning to exert an influence on various regional varieties of Indonesian. In addition, an orthographically stylised version of Jakarta Indonesian can be found on the internet, for example in various chat groups used by Indonesian college students and other professionals in Indonesia and other countries. Some discussion of Jakarta Indonesian can be found in Wouk (1989, 1999). As a common language for interethnic communication, Jakarta Indonesian is distinct from the more extensively described Betawi Malay, the dialect of the indigenous Jakartan ethnic group known as Betawi, now constituting a minority within the metropolitan region — see Kähler (1966), Abdul Chaer (1976), Ikranagara (1980), Muhadjir (1981), Grijns (1991) and others. The Jakarta Indonesian data discussed in this paper are the product of ongoing field work since the beginning of 1999.

In Jakarta Indonesian, the prefixes *di-* and *N-* occur much more frequently than in Riau Indonesian — even though they are also quite common in Riau Indonesian. (Also, the morphophonemic realisation of the prefix *N-* in Jakarta Indonesian differs from that of its counterpart in Riau Indonesian, as in Table 3 above, but this is of no concern to us here.) Nevertheless, in spite of this readily observable difference in frequency, preliminary investigations suggest that the function of the prefixes *di-* and *N-* in Jakarta Indonesian is largely similar to that of these same prefixes in Riau Indonesian, as discussed at length in §3.1.

Examples (36) and (37) below show that in Jakarta Indonesian too, the prefixes *di-* and *N-* do not have the basic grammatical function of discriminating actors from patients:

- (36) Mister **diertiin**.
 white.person **di-**meaning-APPL
 [Speaker jokingly asking me to take him back to Europe to me, realises he doesn't speak English, but then says it doesn't matter]
 'You can translate for me.'
- (37) Ini **megang**.
 DEM:PROX **N-**hold
 [Speaker massaging me, extends his hand and asks me to hold it while he massages it with his other hand]
 'Hold this.'

In example (36), corresponding to Riau Indonesian (11), a word marked with *di-* is preceded by an actor rather than a patient. And in example (37), corresponding to Riau Indonesian (23a,b), a word marked with *N-* is preceded by a patient rather than an actor. Moreover, work in progress suggests that in Jakarta Indonesian too, the prefixes *di-* and *N-* do not identify any particular argument as having high discourse prominence, as sentential topic or simply as definite. Thus, although occurring with great frequency, the prefixes *di-* and *N-* would appear to be associated with neither the grammatical nor the discourse properties characteristic of passive and active constructions respectively. Rather, their primary function would appear, like in Riau Indonesian, to be semantic, namely, marking their host words as having a patient or actor in their argument structure.¹⁷

3.2.2 The prefix *di-* in Sulse Indonesian

Sulse Indonesian is the variety of Indonesian spoken throughout the province of Sulawesi Selatan ('South Sulawesi', commonly abbreviated to 'Sulse'), and, in particular, in its capital city Makassar, until recently known as Ujung Pandang. The primary function of Sulse Indonesian is as a lingua franca for communication between speakers of different ethnic groups, mostly Makassar and Bugis, though it is now increasingly being acquired as a native language by children growing up in the city of Makassar and its surroundings. Sulse Indonesian is also referred to occasionally as Makassar Malay. Some preliminary descriptions of Sulse Indonesian are provided in Aburaerah et al. (1983), Steinhauer (1988) and Tadmor (1998). My own data come from two field trips to the city of Makassar, lasting for about one month.

While all the regional varieties of Malay/Indonesian span a large range of registers, this is perhaps even more true of Sulse Indonesian. At its most basilectal, it appears to be a 'mixed language', with a largely Malay/Indonesian vocabulary alongside a mostly Makassar morphology and syntax, including features not usually associated with Malay/Indonesian such as case-marked pronouns, an elaborate set of pronominal affixes, and others. At this register, then, *di-* and *N-* are both absent, their place being taken by a variety of native Makassar morphemes. However, at a somewhat higher though still basilectal level, is a register with less Makassar and more Malay/Indonesian morphology. At this level, the prefix *N-* is still absent; however *di-* occurs frequently — impressionistically with greater frequency than in Riau Indonesian.¹⁸

¹⁷ Examples such as (36) and (37) are reportedly unattested in the corpus of Jakarta Indonesian texts collected by Wouk: this may be due to the fact that her texts are of a less basilectal register than the data discussed herein. Example (37) also conflicts with some of the generalisations regarding the semantics of the prefix *N-* made by Wallace (1976). Wallace argues that *N-* is associated with a general 'abstract-prospective' meaning; more specifically, he claims that in imperative constructions, the presence of *N-* marks a 'command to behave in a general manner', while its absence marks a 'command to perform a specific act'. But clearly, in (37), in spite of the presence of *N-*, the command is specific rather than general. However, although Wallace does not make it clear which dialect he is describing, it would appear to be the case that his conclusions pertain to Betawi Malay rather than Jakarta Indonesian.

¹⁸ Steinhauer's (1988) description, based on data in Aburaerah et al. (1983), makes reference also to a prefix *maN-/meN-/moN-*, clearly corresponding to Standard Malay/Indonesian *meN-*. However, Tadmor (1998) claims that this form does not exist in any real register of Sulse Indonesian, but is in

Preliminary evidence suggests that the function of *di-* in Sulse Indonesian is similar to that in Riau Indonesian. For example, (38) below bears a close resemblance, in form, meaning and context, to (11a) in Riau Indonesian:

- (38) *Kemarin bukan saya dibikin.*
 yesterday NEG 1SG **di-**prepare
 [Drink stall vendor makes customer avocado juice, and then asks which is better,
 this one or the one the same customer had yesterday; speaker diplomatically
 answers this one; vendor then boasts]
 'Yesterday it wasn't me who made it.'

Examples such as (38) above show that in Sulse Indonesian, as in Riau Indonesian and Jakarta Indonesian, a word marked with *di-* need not be preceded by the patient: it may, instead, be preceded by the actor. Thus, in Sulse Indonesian too, the prefix *di-* fails the most basic test of a passive marker, namely, discriminating between patients and actors.¹⁹

A larger corpus of data from the colloquial variety of Sulse Indonesian must be investigated before the discourse function of the prefix *di-* can be adequately characterised. However, available evidence suggests that the primary function of *di-* in Sulse Indonesian, as in Riau Indonesian, is semantic, namely, marking its host as having a patient in its argument structure.

3.2.3 *The prefix di- in Irian Indonesian*

Irian Indonesian is the variety of Indonesian spoken throughout the province of Irian Jaya, the western half of the island of New Guinea. Like the preceding varieties of Indonesian discussed above, the main function of Irian Indonesian is as a lingua franca for communication between speakers of different ethnic groups, of which there are hundreds. In addition, it is now increasingly being acquired as a native language by children growing up in Irian Jaya, and is thus poised to replace and thereby bring to extinction several hundred of the indigenous languages — this being the variety of Indonesian which is most responsible for Malay/Indonesian being ranked as one of the world's foremost 'killer languages'. To the best of my knowledge, there are no previous descriptions of Irian Indonesian in the linguistic literature. Impressionistically, Irian Indonesian bears a certain resemblance to the better-known Ambonese Malay, though there are also many differences between the two varieties. The data reported on in this paper come from a field trip to Irian Jaya lasting for approximately one month.

As in Sulse Indonesian, the prefix *N-* is absent in Irian Indonesian. However, unlike Sulse Indonesian, the prefix *di-* is also infrequent, at least in its guise as a patient-oriented

fact an unfortunate invention of Aburaerah et al., a hybrid of acrolectal Standard Malay/Indonesian morphology, with its prefix *meN-*, and basilectal Sulse Indonesian sound patterns, in which the standard schwa is realised, alternatively, as *a*, *e* or *o*. Tadmor's claim is thus consistent with my own observations to the effect that an actor-oriented prefix is absent from all but the most acrolectal registers of Sulse Indonesian.

¹⁹ The above observations differ from those of Steinhauer (1988), who, relying on Aburaerah et al.'s data, reports (p.147) that when a verb is prefixed with *di-*, the actor, if present, always follows the verb. Again, I can only suggest that Aburaerah et al.'s data represent, at best, a more acrolectal variety of Sulse Indonesian than is being described here.

prefix. In fact, preliminary investigation suggests that Irian Indonesian may be divided into two subdialects: dialect A, in which *di-* is completely absent, and dialect B, in which *di-* occurs, albeit infrequently. Impressionistically, dialect A may be associated with more basilectal registers and more remote inland regions, whereas dialect B is probably identified with less basilectal registers and more accessible coastal locations.

When present, the function of *di-* in Irian Indonesian would appear to be similar to that in Riau, Jakarta and Sulse Indonesian. For example, (39) below corresponds to (11) in Riau Indonesian, (36) in Jakarta Indonesian, and (38) in Sulse Indonesian:

- (39) *Kau dijaga.*
 2 *di-guard*
 [Two children shopping in market; older one gives basket to younger one and runs off, telling him]
 'You take care of it.'

Examples such as the above show that in Irian Indonesian too, a word marked with *di-* need not be preceded by the patient: it may, instead, be preceded by the actor. Thus, in Irian Indonesian as well, the prefix *di-* is obviously something other than a passive marker.

Again, a larger corpus of data is necessary before the discourse function of the prefix *di-* can be properly understood in Irian Indonesian. Once more, however, the facts at hand suggest that the main role of *di-* in Irian Indonesian is semantic, namely, marking its host as having a patient in its argument structure.

3.2.4 *The prefixes N- and men- in Kuala Lumpur Malay*

Kuala Lumpur Malay is the variety of Malay used by the ethnic Malay residents of the capital city of Malaysia, Kuala Lumpur, as a vehicle for colloquial intraethnic communication. Kuala Lumpur Malay is distinct from other colloquial varieties of Malay, also used in Kuala Lumpur, by Malays, Chinese and Indians, for interethnic communication. Since the Malay inhabitants of Kuala Lumpur are mostly relatively recent immigrants from a variety of other parts of Malaysia and Indonesia, Kuala Lumpur Malay has many of the characteristics of a koiné, in this respect resembling the varieties of Indonesian considered in the preceding sections. The existence of a local Kuala Lumpur dialect is not generally acknowledged in the linguistic literature; indeed, when I presented data from Kuala Lumpur Malay at linguistic conferences in Malaysia, some local linguists objected on the grounds that no such dialect exists. But it does, and it is almost completely mutually unintelligible with Standard Malay. (A large proportion of what makes Kuala Lumpur Malay different from Standard Malay is in the sound patterns, and is not reflected in the near-standard orthography adopted here, which, accordingly, downplays the divergent nature of the dialect.) The data discussed in this paper are the product of one and a half years of residence in a Malay neighborhood of Kuala Lumpur, corroborated by additional data elicited from my Malay students at the Universiti Kebangsaan Malaysia.

In its overall typological character, Kuala Lumpur Malay exhibits many of the properties of a mainland Southeast Asian language — see Gil (to appear b) for details. Among these properties is a largely 'isolating' nature, with very little morphology. In particular, *di-* is completely absent, both as patient-oriented prefix and as locative preposition. In contrast, the actor-oriented prefix is present, in two different guises: *N-*, as in the Indonesian koinés

considered above, and also *meN-*, as in Standard Malay/Indonesian.²⁰ However, the distribution of these two prefixes is very restricted, in terms of the number of lexical items to which they may attach, and the frequency with which they occur.

Due to the limited distribution of these two prefixes, and also to the subtlety of their functions, there are many basic analytical questions which I still have not been able to resolve. One is whether these two prefixes share all the same functions, or whether there are differences between them. Another is whether the prefix *meN-* may be analysed compositionally, as consisting of two prefixes, *me-* plus *N-* (as is implied, arbitrarily, by the interlinear gloss in (41) below). Answers to these questions must await further research. However, in the meantime, some preliminary observations regarding these prefixes can be made.

As in Riau Indonesian, here too the prefix *N-* does not mark a preceding argument as being an actor. This can be seen in the following example, closely paralleling Riau Indonesian (23a,b) and Jakarta Indonesian (37) above:

- (40) *Ni* **nengok.**
 DEM:PROX N-look
 [Playing laptop billiards; speaker kibbitzing, pointing to ball]
 'Look at this one.'

Thus, as suggested by the above example, the prefix *N-* is not associated with the grammatical function of a marker of active voice. Whether the same is true also of the prefix *meN-* must await further investigation.

Nevertheless, both prefixes would seem to be actor-oriented. The following example, with *meN-*, closely parallels (28b) in Riau Indonesian:

- (41) *Lama* *dia* **menangis** ... *nangis* *lama.*
 longtime 3 **me-N-cry** cry longtime
 [Playing game on laptop, in which man has to run course in limited amount
 of time; when he gets hit, he spends a few seconds stationary, crying; while
 man is crying, speaker says]
 'He's crying for a long time...he's crying for a long time.'

In (41) above, the word *nangis* 'cry' occurs twice, first with the prefix *meN-*, then in bare form. (As is the case in Riau Indonesian — see Table 2, part (e) above — Kuala Lumpur Malay does not have the Standard Malay form *tangis* 'cry'; hence, *nangis* may be considered to be the bare form.) The example consists of a simple intransitive sentence, repeated in a

²⁰ The claim that both prefixes, *N-* and *meN-*, occur in Kuala Lumpur Malay is complicated by the existence of an optional reduction rule, whose effect is to delete the first syllable of a word; in fact, this rule applies most frequently when the syllable in question begins with a labial consonant, e.g. [*pgi*] → [*gi*] for *pergi* 'go'; [*buket*] → [*ket*] for *bukit* 'hill'. This rule poses an analytical problem, in that application to a form prefixed with *meN-* results in a form seeming to bear the prefix *N-*, e.g. [*mñules*] → [*nules*] for *menulis* 'write'. Nevertheless, the claim that both prefixes, *N-* and *meN-*, occur in Kuala Lumpur Malay is supported by the following observation. Whereas speakers are typically unconscious of the reduction rule and generally deny producing the forms in question, they are usually quite willing to offer citation forms prefixed with either *N-* or *meN-* (when the forms in question exist).

somewhat different form. In the first part of the example, *nangis* is preceded by its actor, and is prefixed with *meN-*; while in the second part of the example, the actor is no longer overtly present, and *nangis* occurs in bare form. Example (41) suggests that the function of *meN-* in Kuala Lumpur Malay is similar to that of its counterpart *N-* in Riau Indonesian, namely, to assign the actor semantic or conceptual salience.

However, preliminary evidence suggests that the function of *meN-* in Kuala Lumpur Malay may be somewhat more elaborate than that of its Riau Indonesian counterpart. Specifically, whereas Riau Indonesian *N-* would appear to occur equally freely in clauses of high or low transitivity, the prefix *meN-* in Kuala Lumpur Malay tends to occur more readily in clauses of low transitivity. For example, in a number of informal questionnaires administered to native speakers of Kuala Lumpur Malay, subjects preferred the use of *meN-* in subordinate rather than main clauses; for ongoing as opposed to completed activities; for durative rather than punctual activities; and in the absence of an overtly expressed patient — all four of the above features being associated with backgrounded clauses of low transitivity. As a marker of low transitivity, the prefix *meN-* in Kuala Lumpur Malay thus resembles its counterpart *meN-* in Standard Malay/Indonesian, as pointed out in §2. However, in its preference for the absence of an overtly expressed patient, it would appear to differ from its Standard Malay/Indonesian counterpart.

The apparent function of *meN-* as a marker of low transitivity is closely related to its basic function as an actor-oriented marker. As noted previously, the prototypical two-participant activity may be viewed as moving from an actor to a patient. Whereas foregrounded clauses of high transitivity tend to focus on the latter part of this movement, namely the activity as affecting the patient, backgrounded clauses of low transitivity prefer to zero in on the initial part of this movement, that is to say the activity as originating with the actor. Accordingly, an actor-oriented marker is likely to assume various properties of a marker of backgrounded clauses of low transitivity — as is the case with *meN-* in Kuala Lumpur Malay.

3.3 Summary: *di-* and *N-* in Malay/Indonesian dialects

The characteristics of the prefixes *di-* and *N-* in the Malay/Indonesian dialects considered in the preceding pages may be summarised as in Diagram 1 below.

In Diagram 1, the vertical axis characterises the prefix *di-* and the horizontal axis the prefix *N-*. Each axis defines a five-point 'scale of strength' for the relevant prefix: *absent* < *infrequent* < *common and partially productive* < *common and fully productive* < *common and grammaticalised*. In conjunction, the two axes define a two-dimensional quadrant, within which the various Malay/Indonesian dialects may be situated.

Obviously, the five-point scale is a simplification of a much more complex picture. In particular, the scale conflates three logically independent criteria: productivity (how many forms the prefix may apply to), frequency (how often the prefix actually applies to the forms in question), and grammaticalisation (whether the prefix is associated with grammatical functions). A more accurate diagram would represent each of these criteria on an orthogonal axis — but of course would be impossible to construct on a single page. However, even the simplified diagram provided above is sufficient to reveal interesting patterns.

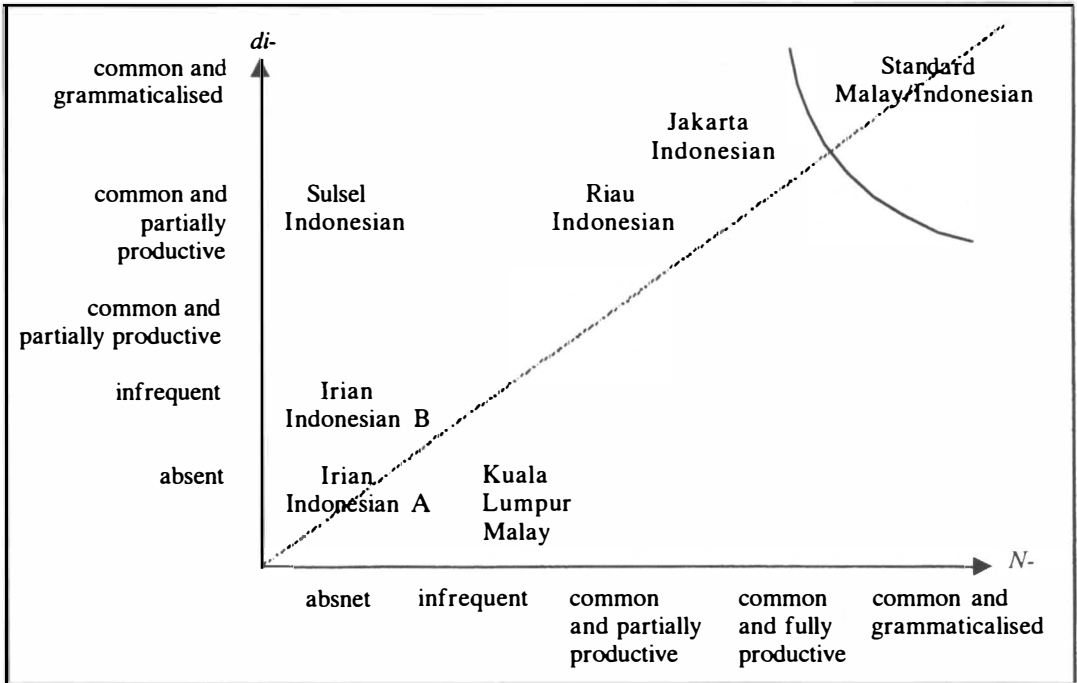


Diagram 1: The prefixes *di-* and *N-* in Malay/Indonesian dialects

The above diagram highlights the distinction between Standard Malay/Indonesian, off in the upper right corner, and all of the other dialects under consideration. Whereas in Standard Malay/Indonesian both of the prefixes are grammaticalised, in all of the other dialects neither of the two prefixes have any grammatical functions. From this perspective, then, it is Standard Malay/Indonesian that is the outlier — as represented by the arc in the diagram, which sets the standard language apart from all of the colloquial varieties.

The remaining dialects are spread out over the rest of the diagram, thereby reflecting the diversity of Malay/Indonesian dialects with respect to the two prefixes. Furthest away from Standard Malay/Indonesian is the A dialect of Irian Indonesian, in which both of the prefixes appear to be absent. Closest to Standard Malay/Indonesian, albeit still quite distinct, are Jakarta and Riau Indonesian, in which both of the prefixes occur with considerable frequency. (In Diagram 1, Jakarta Indonesian is positioned a little above and to the right of Riau Indonesian, reflecting the fact that although the two prefixes are classified similarly in both dialects, they occur more frequently in Jakarta Indonesian.)

Observe, now, the diagonal dotted line extending from the lower left to the upper right corner of the diagram. This dotted line passes through points in which both of the prefixes are of equal strength. As evident in the diagram, two dialects lie on the dotted line: the A dialect of Irian Indonesian, in which both prefixes are absent, and Standard Malay/Indonesian, in which both prefixes are common and grammaticalised. However, above and below the dotted line are dialects in which there is an asymmetry between the two prefixes, and one of them is stronger than the other. Above the dotted line are four dialects, Riau, Jakarta, Susel, and the B dialect of Irian Indonesian, in which the prefix *di-* is favoured, while below the dotted line lies a single dialect, Kuala Lumpur Malay, in which the prefix *N-* is dominant.

This patterning may be most insightfully understood in the context of 'patient prominence' — one of the most salient characteristic features of the Austronesian language family (Cena 1977; De Guzman 1976, 1992; Gil 1983, 1984). Whereas in most of the world's languages, actors are more important than patients in myriad ways, in Austronesian languages, a mirror-image state of affairs typically obtains, whereby patients generally take precedence over actors, in many respects. First, patients tend to be more referentially salient than actors, for example more likely to be definite, more likely to independently refer, more likely to have wide scope, and so on. Secondly, patients tend to be more likely than actors to occur in various specific constructions, for example as the 'topic', 'subject', or 'pivot' of the clause in 'Philippine-type' languages. Finally, patient-oriented focus markers tend to predominate over actor-oriented ones, by exhibiting the following properties: (a) earlier acquisition by children; (b) greater ease of processing; (c) greater frequency of occurrence; (d) greater productivity; and (e) greater formal simplicity.

Thus, the dialects above the dotted line, Riau, Jakarta, Sulsei, and the B dialect of Irian Indonesian, may be characterised as exhibiting the characteristic Austronesian pattern of patient prominence. Specifically, in these dialects, the patient-oriented focus marker *di-* occurs more frequently and is more productive than the actor-oriented *N-*. Indeed, in Sulsei and the B dialect of Irian Indonesian the patient-oriented prefix wins by default, since there is no actor-oriented prefix. In addition, in Riau and Jakarta Indonesian, patient prominence is also manifest in the greater formal simplicity of the prefix *di-* as opposed to *N-*, specifically in its immutable form and segmentable nature, see examples (19), (20) and Tables 1 and 2.²¹ In contrast, Kuala Lumpur Malay, below the dotted line, exhibits the crosslinguistically much more common pattern of actor prominence, with an actor-oriented prefix (in fact two) but no patient-oriented prefix.

The feature of patient prominence exhibits clear geographical patterning across the Malay/Indonesian varieties under consideration. The most highly patient prominent dialects are those in the core Austronesian areas of Sumatra (Riau Indonesian), Java (Jakarta Indonesian) and Sulawesi (Sulsei Indonesian). In the peripheries, patient prominence fades away and disappears, as the Malay/Indonesian dialects come into contact with non-Austronesian, non-patient-prominent languages. Thus, to the east, the Irian dialects of Indonesian take on some of the properties of the Papuan substratum languages. Similarly, in the northwest, Kuala Lumpur Malay, influenced by large numbers of migrants from the province of Kelantan bordering on Thailand, assumes many of the characteristics of mainland Southeast Asian languages such as Thai and Vietnamese. Finally, the weakening of patient prominence in Standard Malay/Indonesian may be argued to be a consequence of language contact of another kind, namely the effect of English, first as the language of the colonial power, and now even more cogently as the world language of commerce, technology and the media.

²¹ It should be noted that although Standard Malay/Indonesian is represented as straddling the dotted line, it actually shares these additional features with Riau, Jakarta, Sulsei and the B dialect of Irian Indonesian: in the standard language, too, *di-* is more productive than *meN-* and is of greater formal simplicity. Thus, in a more elaborate diagram, Standard Malay/Indonesian would also be characterised as somewhat patient prominent.

4 Conclusion

This paper has shown that in five colloquial dialects of Malay/Indonesian, there is no focus morphology, and no morphological marking of the active/passive distinction. Nevertheless, in the dialects under consideration, the prefixes *di-* and *N-* can indeed be characterised as patient-oriented and actor-oriented respectively. In other words, these prefixes, although not voice markers in the classical sense, still resemble voice markers to a certain degree.

These observations can be made more rigorous via the following definition of *generalised voice marker*:

- (42) A *generalised voice marker* is a marker M which, when applied to a form X, marks the argument of X bearing the thematic role T as having a set of properties P.

The force of the above definition can be appreciated through the consideration of a few examples, provided in Table (46) below:

Table 4: A typology of generalised voice markers

	Riau Indonesian Sulsel Indonesian Irian Indonesian	Standard Malay/Indonesian	Tagalog	English
Form:	<i>di-</i>	<i>di-</i>	<i>-in-</i>	<i>BE -en</i>
Thematic Role T:	patient	patient	patient	patient
Properties P:	existence	some subject properties	some subject properties	subject properties
	weak	←—————→		strong

In accordance with the definition in (42), a generalised voice marker may be oriented towards any thematic role T; for simplicity, Table (43) provides four examples of generalised voice markers oriented towards the thematic role of patient.

The difference between these four markers lies in the nature of the set P of properties which they associate with the patient: the size of this set provides for the characterisation of such markers in accordance with a scale of strength. At one extreme is the prefix *di-* in the three colloquial dialects of Indonesian considered in this paper, associating with the patient the minimal property of existence. At the other extreme is English *BE -en*, which associates with the patient a wide range of morphological, syntactic and semantic properties — in fact, those characteristic of English subject NPs. The *BE -en* complex in English is of course typical of passive markers across languages, which, in accordance with the above, may now be characterised as generalised voice markers which are patient oriented, and which assign to the patient those properties associated with subject NPs in the language in question.

In between these two extremes are generalised voice markers which assign to the patient a proper subset of the properties generally considered to be necessary in order for an expression to be characterised as a subject. Among such generalised voice markers are the prefix *di-* in Standard Malay/Indonesian, the infix *-in-* in Tagalog, and various other patient-oriented focus markers in Austronesian languages. Unlike *di-* in Riau, Jakarta, Sulsel and Irian Indonesian, such markers do have grammatical functions; however, unlike prototypical

passive markers, the set of properties that such markers associate with the patient falls short of those generally considered necessary for the patient to be considered to be a bona fide subject. Accordingly, such markers are frequently the focus of much debate in the linguistic literature, with some scholars arguing in support of their characterisation as passives, and others against — see, for example, Chung (1976) Hopper (1983) and Alsagoff (1992) for Malay/Indonesian; Schachter (1976, 1977), Payne (1982) and Kroeger (1998) for Tagalog; and Guilfoyle et al. (1992) for these and other Austronesian languages. However, the definition of a generalised voice marker proposed above provides for a resolution to such debates, by underscoring the obvious, namely that a typology of such markers must make reference not to a single categorial distinction but rather to a more elaborately structured continuum.

The concern of this paper up to this point has been exclusively synchronic; however, the definition of generalised voice marker sets the stage for some concluding remarks of a diachronic nature. The continuum of weak-to-strong generalised voice markers provides a parameter along which languages may change in the course of time. Let us now ask the obvious question: how did the affixes *di-* and *N-* in Malay/Indonesian get to be the way they are? Two plausible scenarios suggest themselves.

As is commonly known, a productive and well-entrenched system of morphological voice or focus is one of the hallmarks of Austronesian languages. Also, as is well reflected in other contributions to this volume, these voice systems become weakened and simplified as one moves out of Taiwan and the Philippines, south through Borneo and Sulawesi, and then south, west and east throughout the remainder of the Austronesian-speaking area. Accordingly, conventional wisdom has it that a productive system of morphological voice was characteristic of Proto Austronesian and Proto Malayo-Polynesian, but subsequently underwent decay and dissipation in various sub-branches of Malayo-Polynesian, including western Malayo-Polynesian languages and their well-known representative Malay/Indonesian.

In this context, the voice systems of the Malay/Indonesian dialects discussed in this paper may be construed as representing yet further stages in the ongoing historical dissolution of morphological voice in the Western Austronesian branch, culminating, at the end point of the process, with the A dialect of Irian Jaya, in which the old Austronesian morphological focus system is completely lost. In accordance with such a view, then, the colloquial dialects of Malay/Indonesian discussed in this paper constitute further simplifications of the system of morphological focus still present in Standard Malay/Indonesian.

This scenario is obviously consistent with the widespread sociolinguistic prejudices against the basilectal registers of Malay/Indonesian, reflected in various terms such as 'Low Malay', 'Bazaar Malay', and so forth — prejudices which view such registers as simplifications or corruptions of their acrolectal counterparts. Of course such prejudices are unfounded: witness Sulse Indonesian, in which, alongside a weaker version of the Standard Malay/Indonesian voicing system, there is another, parallel system of verbal morphology, largely borrowed from Makassarese, and of much greater complexity than anything the standard language has to offer.

Nevertheless, even if such prejudices are rejected, it is still not implausible to suggest that the voice systems of Riau, Jakarta, Sulse and Irian Indonesian as well as Kuala Lumpur Malay may have resulted from the simplification of an older system more closely akin to that of Standard Malay/Indonesian. Recall that all of these dialects are associated, to one degree or another, with the function of a lingua franca, or vehicle for interethnic communication.

Often, such language varieties undergo simplification, in order to fulfill their particular communicative function more efficiently. Even without attaching labels such as 'creole' or 'creoloid', it is still reasonable to suggest that the special function of these language varieties was what caused their systems of grammatical voice to be further weakened, thereby bringing to conclusion a process that had already been ongoing in western Malayo-Polynesian languages for thousands of years.

However, there is an opposite, equally plausible scenario accounting for the behaviour of *di-* and *N-* in Malay/Indonesian dialects. One proponent of this alternative is Benjamin (1993:363), who suggests that:

Their present-day status as syntactic-function markers is the result of their having been standardized into a single paradigmatic set, where previously they were independent elements used optionally for whatever nuance of meaning they could bring to an utterance.

According to this view, the situation in colloquial Malay/Indonesian dialects is the diachronically prior one, while the system of morphological voice marking characteristic of Standard Malay/Indonesian is the product of grammaticalisation of what was originally a purely semantically-based system.

Such a scenario posits a reversal in the unidirectional process of simplification of voice marking generally assumed to obtain within western Malayo-Polynesian languages. Specifically, whereas the semantically-based system of Proto Malayic would have resulted from an erosion of the more elaborate grammatically-based system characteristic of Proto Malayo-Polynesian, the contemporary system of Standard Malay/Indonesian would be the product of a subsequent regrammaticalisation of the purely semantic system associated with Proto Malayic. Such reversal cannot be ruled out on *a priori* grounds, as there is no reason to believe that linguistic change must maintain a constant direction over many thousands of years.

An arguably atypical feature of this scenario is that it characterises the acrolect as being innovative and the basilects as more historically conservative. However, such a state of affairs is by no means unknown. In fact, in the case at hand, there are plenty of possible social, cultural and historical reasons why the standard language should have gone in the direction that it has — factors amply dealt with in Benjamin's seminal article. (In fact, it may not even be the case that the behaviour of *di-* and *N-* described in this paper is restricted to basilectal varieties: constructions in which a word marked with *di-* is preceded by an actor rather than a patient are reported to occur also in the classic literary dialect of Kedah — Asmah 1995:54-56.)

An advantage of this scenario is that it is consistent with various proposals that have been made in the general linguistic literature to the effect that certain aspects of diachronic change are unidirectional, that is to say, can proceed in a particular direction but not in an opposite one. More specifically, it has been observed by many scholars (see, for example, Traugott 1990, to appear; Lehmann 1995; Haspelmath 1999) that change typically proceeds from concrete to abstract but not vice versa, and from semantic to grammatical but not vice versa — hence the numerous studies of grammaticalisation that can be found in the linguistic literature as opposed to the very few accepted examples of de-grammaticalisation. In the case at hand, a change from the semantically-based system of the colloquial dialects of Malay/Indonesian to the grammatically-based system of Standard Malay/Indonesian would be consistent with the proposed unidirectionality of diachronic change, providing a cross-linguistically typical instance of grammaticalisation.

However, at present, it is still difficult to make a clear decision in favour of one or another of the above two scenarios. Indeed, the reality is probably more complex than either scenario. In particular, there is no reason to believe that a diglossic situation, with semantically- and grammatically-based systems of voice marking coexisting alongside each other at different registers and interacting with each other in myriad ways, could not have been maintained since earliest times, when Malay first attained prominence as the language of a major civilisation. Ultimately, the choice between them will depend on the establishment of uncontested etymologies for the two prefixes in question, *di-* and *N-*, something which the field has still not been able to settle upon, especially with regard to the former — see De Casparis (1956:24), Teeuw (1959:141-144), and Adelaar (1992:161-163) for some conflicting proposals.

More generally, however, a further prerequisite to a satisfactory reconstruction of the history of focus and voice marking in Malay/Indonesian is an adequate synchronic description of the facts as they are in a range of contemporary Malay/Indonesian dialects. At present such descriptions are still sorely lacking. The present paper provides some first steps in this direction, but much more remains to be done.

References

- Abdul Chaer, 1976, *Kamus Dialek Jakarta*. Ende: Nusa Indah.
 1993, *Gramatika Bahasa Indonesia*. Jakarta: Rineka Cipta.
- Abdul Hamid Mahmood, 1992, *Ayat Pasif Bahasa Melayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
 1994, *Sintaksis Dialek Kelantan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Aburaerah Arief, Abdul Muthalib, Abdul Kadir Mulya and Zainuddin Hakim, 1983, *Kamus Bahasa Melayu Makassar – Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Adelaar, K. Alexander, 1992, *Proto-Malayic: the reconstruction of its phonology and parts of its lexicon and morphology*. Canberra: Pacific Linguistics.
- Alsagoff, Lubna, 1992, Topic in Malay: the other subject. PhD dissertation, Stanford University.
- Asmah Haji Omar, 1981, *The Iban language of Sarawak. A grammatical description*, Kuala Lumpur: Dewan Bahasa dan Pustaka.
 1995, *Bahasa Kesusasteraan Klasik Kedah*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Benjamin, Geoffrey, 1993, Grammar and polity: the cultural and political background to Standard Malay. In W.A. Foley, ed. *The role of theory in language description*, 341-92. Berlin: Mouton de Gruyter.
 1997, Affixes, Austronesian and iconicity in Malay. *Department of Sociology Working Papers* 132. National University of Singapore.
- Casparis, J.G. de, 1956, *Selected inscriptions from the 7th to the 9th Century A.D.* Prasasti Indonesia II. Bandung: Masa Baru.
- Cena, Resty M., 1977, Patient primacy in Tagalog. Paper presented at Linguistic Society of America Annual Meeting, Chicago, 28-30 December.
- Chung, Sandra, 1976, On the subject of two passives in Indonesian. In C.N. Li, ed. *Subject and topic*, 57-98. New York: Academic Press.

- Cole, Peter and Gabriella Hermon, 1998a, Subject and nonsubject relativization in Indonesian. Paper presented at the Second Symposium on Malay/Indonesian Linguistics, Association for Linguistic Typology, Ujung Pandang, Indonesia, 12 July.
- 1998b, The typology of WH movement: WH questions in Malay. *Syntax* 1:221-258.
- De Guzman, Videa P., 1976, Syntactic derivation of Tagalog verbs. PhD dissertation, University of Hawai'i, Honolulu.
- 1992, Morphological evidence for the primacy of patient as subject in Tagalog. In M. Ross, ed. *Papers in Austronesian Linguistics*, No.2, 87-96. Canberra: Pacific Linguistics.
- Dempwolff, Otto, 1938, *Vergleichende Lautlehre des Austronesischen Wortschatzes III, Austronesisches Wörterverzeichnis*. Zeitschrift für Eingeborenen-Sprachen, Beihefte 15, 17, 19. Berlin: Dietrich Reimer.
- Dunselman, P. Donatus, 1949, Bijdragen tot de kennis van de taal en adat der Kendajan-Dajaks van West-Borneo. *Bijdragen tot de Taal-, Land- en Volkenkunde*, 105:59-105.
- Foley, William A. and Robert D. Van Valin, 1984, *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Gil, David, 1983, Quantifier scope and patient prominence: some data from Batak, Buginese, and Tagalog. In S. Mordechay, ed. *UCLA Working Papers in Semantics*, 315-364. Los Angeles.
- 1984, On the notion of 'direct object' in patient prominent languages. In F. Plank ed. *Objects: towards a theory of grammatical relations*, 87-108. London: Academic Press.
- 1994, The structure of Riau Indonesian. *Nordic Journal of Linguistics*, 17:179-200.
- 1999, Riau Indonesian as a pivotless language. In E.V. Raxilina and Y.G. Testelec, eds *Tipologija i teorija jazyka, ot opisanija k objasnenuju, k 60-letiju Aleksandra Evgen'evicha Kibrika [Typology and linguistic theory, from description to explanation, for the 60th birthday of Aleksandr E. Kibrik]*, 187-211. Moscow: Jazyki Russkoj Kul'tury.
- 2000a, Riau Indonesian: a VO language with internally-headed relative clauses. *Snippets* 1.
- 2000b, Syntactic categories, cross-linguistic variation and universal grammar. In P.M. Vogel and B. Comrie, eds *Approaches to the Typology of Word Classes, Empirical Approaches to Language Typology*, 173-216. Berlin: Mouton.
- to appear a, Escaping eurocentrism: fieldwork as a process of unlearning. In P. Newman and M. Ratliff, eds *Linguistic Fieldwork*. Cambridge: Cambridge University Press.
- to appear b, Kuala Lumpur Malay as a mainland Southeast Asian language. In D. Gil, ed. *Papers from the Eighth Annual Meeting of the Southeast Asian Linguistics Society*. Tempe: Arizona State University.
- to appear c, Reflexive anaphor or conjunctive operator? Riau Indonesian *sendiri*. In P. Cole, G. Hermon and J. Huang, eds *Long Distance Reflexives, Syntax and Semantics*. New York: Academic Press.
- Grijns, C.D., 1991, *Jakarta Malay: a multidimensional approach to spatial variation*. Leiden: KITLV Press.
- Guilfoyle, Eithne, Henrietta Hung and Lisa Travis, 1992, Spec of IP and Spec of VP: two subjects in Austronesian languages. *Natural Language and Linguistic Theory* 10:375-414.
- Haspelmath, Martin, 1999, Why is grammaticalization irreversible? *Linguistics* 37:1043-1068.

- Helfrich, O.L., 1904, *Bijdragen tot de Kennis van het Midden Maleisch (Besemahsch en Serawajsch Dialect)*, Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen 53.
- Hopper, Paul J., 1979, Aspect and foregrounding in discourse. In T. Givón, ed. *Syntax and semantics*, vol.12: *Discourse and Syntax*, 213-241. New York: Academic Press.
- 1983, Ergative, passive, and active in Malay narrative discourse. In F. Klein-Andreu, ed. *Discourse Perspectives on Syntax*, 64-87. New York: Academic Press.
- Hopper, Paul J. and Sandra A. Thompson, 1980, Transitivity in grammar and discourse. *Language* 56:251-299.
- Ikranagara, Kay, 1980, *Melayu Betawi grammar*. NUSA 9.
- Ina Anak Kalom and Alfred B. Hudson, 1970, Selako traditional history: a story on the origins of Kampong Pueh. *Sarawak Museum Journal, Kuching* 18/36-37 (New Series): 281-300.
- Kähler, Hans, 1948, Ethnographische und linguistische Studien von den Orang laut auf der Insel Rangsang. *Anthropos*, 41-44:1-31.
- 1949, Ethnographische und linguistische Studien von den Orang laut auf der Insel Rangsang an der Ostküste von Sumatra (Schluß). *Anthropos* 41-44:757-785.
- 1960, *Ethnographische und linguistische Studien über die Orang darat, Orang akit, Orang laut und Orang utan im Riau-Archipel und auf den Inseln an der Ostküste von Sumatra*. Berlin: Dietrich Reimer.
- 1966, *Wörterverzeichnis des Omong Djakarta*. Berlin: Dietrich Reimer.
- Kailani Hasan, 1994, Morfonemik Bahasa Melayu Riau. In Zaharani Ahmad, ed. *Linguistik Melayu. Festschrift Tinta Kenangan Sempena Persaraan Azhar M. Simin dan Farid M. Onn*, 159-181. Bangi: Jabatan Linguistik, Fakulti Sains Kemasyarakatan dan Kemanusiaan, Universiti Kebangsaan Malaysia.
- Kailani Hasan, Mohd. Yunus R., Sugio H.M. and Nurbaiti, 1983, *Morfologi dan sintaksis Bahasa Melayu Riau*. Pusat Pembinaan dan Pengembangan Bahasa. Jakarta: Departemen Pendidikan dan Kebudayaan.
- Keenan, Edward L. and Bernard Comrie, 1977, Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8:63-99.
- Kibrik, Aleksandr E., 1997, Beyond subject and object: towards a comprehensive relational typology. *Linguistic Typology*, 1:279-346.
- Kroeger, Paul, 1993, *Phrase structure and grammatical relations in Tagalog*. Stanford: CSLI Publications.
- Lehmann, Christian, 1995, *Thoughts on grammaticalization*. Munich: Lincom Europa.
- MacDonald, R. Ross, 1976, *Indonesian reference grammar*. Washington DC: Georgetown University Press.
- Moseley, Christopher and R.E. Asher, eds, 1994, *Atlas of the world's languages*. London: Routledge.
- Muhadjir, 1981, *Morphology of Jakarta dialect, affixation and reduplication*. NUSA 11.
- Nik Safiah Karim, Farid M. Onn, Hashim Hj. Musa and Abdul Hamid Mahmood, 1996, *Tatabahasa Dewan, edisi baru*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Payne, E.M.F., 1970, *Basic syntactic structures in Standard Malay*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Payne, Thomas E., 1982, Role and reference related subject properties and ergativity in Yup'ik Eskimo and Tagalog. *Studies in Language*, 6:75-106.

- Rani, Abdul, 1996, Logika berbahasa Indonesia, Sebuah indikator penguasaan Bahasa Indonesia. Paper presented at Konferensi Internasional II Pengajaran Bahasa Indonesia Bagi Penutur Asing, Padang, Indonesia, 31 May 1996.
- Saddy, Douglas, 1991, Wh Scope mechanisms in Bahasa Indonesia. In L. Cheng and H. Demirdash, eds *More Papers on WH Movement*, 183-218. MIT Working Papers in Linguistics 15.
- 1992, A versus A-bar movement and Wh fronting in Bahasa Indonesia. MS. University of Queensland.
- Saidat Dahlan, Bustami Ramli, Abdul Razak, Asni Zainal and Syamsul Bahri, 1991, *Struktur Bahasa Melayu Riau*. Jakarta: Departemen Pendidikan dan Kebudayaan.
- Saidat Dahlan, Idrus Lubis, H. Said Mahmud Umar and Nursyam S., 1986, *Kedudukan dan fungsi Bahasa Melayu Riau*. Pusat Pembinaan dan Pengembangan Bahasa. Jakarta: Departemen Pendidikan dan Kebudayaan.
- Schachter, Paul, 1976, The subject in Philippine languages: topic, actor, actor-topic, or none of the above? In C.N. Li, ed. *Subject and Topic*, 491-518. New York: Academic Press.
- 1977, Reference-related and role-related properties of subjects. In P. Cole and J.M. Sadock, eds *Syntax and Semantics Vol.8: Grammatical Relations*, 279-306. New York: Academic Press.
- Sneddon, James Neil, 1996, *Indonesian, a comprehensive grammar*. London and New York: Routledge.
- Steinhauer, Hein, 1988, Malay in east Indonesia: the case of Macassarese Malay. In Mohd. Thani Ahmad and Zaini Mohamed Zain, eds *Rekonstruksi dan cabang-cabang Bahasa Melayu Induk*, 108-51. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Stevens, Alan M., 1997, Deverbal nominals in Indonesian. Paper presented at the Fourth Annual Meeting of the Austronesian Formal Linguistics Association, Los Angeles, 26 April 1997.
- Tadmor, Uri, 1996, Problems and strategies in teaching Indonesian passive constructions. Paper presented at Konferensi Internasional II Pengajaran Bahasa Indonesia Bagi Penutur Asing, Padang, Indonesia, 29 May 1996.
- 1998, Borrowed Makassarese function morphemes in Ujung Pandang Indonesian. Paper presented at the Second Symposium on Malay/Indonesian Linguistics, Association for Linguistic Typology, Ujung Pandang, Indonesia, 11 July 1998.
- Teeuw, A., 1959, The history of the Malay language. *Bijdragen tot de Taal-, Land- en Volkenkunde*, 115:138-56.
- Traugott, Elizabeth C., 1990, From less to more situated in language: the unidirectionality of semantic change. In S. Adamson, V. Law, N. Vincent and S. Wright, eds *Papers from the Fifth International Conference on English Historical Linguistics*, 497-517. Amsterdam: Benjamins.
- to appear, Constructions in grammaticalization. In R.D. Janda and B.D. Joseph, eds *Handbook of Historical Linguistics*. Oxford: Blackwell.
- Verhaar, John, 1988, Syntactic ergativity in contemporary Indonesian. In R. McGinn, ed. *Studies in Austronesian Linguistics*, 347-84. Athens: Ohio University Center for International Studies.
- Wallace, Stephen, 1976, Interplay of voice, aspect and mode: the semantics of verbal prenasalization in Malay. Paper presented at the Annual Meeting of the Linguistic Society of America, Philadelphia, 29 December 1976.

Wouk, Fay, 1989, The impact of discourse on grammar: verb morphology in Spoken Jakarta Indonesian. PhD dissertation, University of California, Los Angeles.

1999, Dialect contact and koineization in Jakarta, Indonesia. *Language Sciences*, 21: 61-86.

L

Voice in the languages of Nusa Tenggara Barat

FAY WOUK

1 Introduction

The dividing line between western Malayo-Polynesian and Central-Eastern Malayo-Polynesian languages runs down the middle of Sumbawa which, along with Lombok, comprises the Indonesian province of Nusa Tenggara Barat (NTB). Bima, a Central-Eastern language belonging to the Bima–Sumba subgroup, is spoken east of that line; Sumbawa west of it, and the latter’s closest relative, Sasak, throughout Lombok. These latter two form a branch of the putative Bali–Sasak–Sumbawa subgroup. This paper focuses primarily on Sasak, where the focus system is most strongly in evidence, but will also briefly explore the traces of it found in Sumbawa and Bima.

2 Sasak

There are approximately 1,657,000 speakers of Sasak distributed among a reported five major dialects (Teeuw 1958, Thoir et al 1986) although Herusantosa et al. (1987) identified only four, and none of those conclusions was based on a truly comprehensive dialect survey. A three-year project being carried out under the auspices of the Centre for the Study of Language and Culture (PPBK) at the University of Mataram, begun in 1998, should clarify the picture. The five commonly recognised dialects are named Ngeno-Ngene, Ngeto-Ngete, Meno-Mene, Meriak-Meriku and Ngene-Mene, after the words meaning “like that” and “like this”.

This study describes the Ngeno-Ngene dialect as spoken in the subdistrict of Selong, in the Lombok Timur Regency. Lombok, however, shows a tremendous amount of variation for such a small island with little in the way of natural barriers, and there are syntactic differences between the dialects of Sasak; thus some Selong constructions, for example, are not grammatical in Meno-Mene as spoken in Praya (Husni Muadz pers. comm.). It is not certain to what degree syntax differs at the subdialectal level, but a comparison of my data with data collected from a Ngeno-Ngene speaker from a different location (Austin 1996)

suggests that there are considerable differences.¹ Until the dialect survey of Lombok is completed, and a much more detailed study of the syntax of all dialects is made, it is impossible to say to what extent the following description is valid for speech from areas other than Selong.

My analysis is based on elicitation work with four native speakers from Selong, plus some consultation with a fifth, and a corpus of conversations by speakers from the subdistrict, collected during two visits to Lombok, December 1994 and November 1996–March 1997.² From this corpus, four conversations involving a total of fifteen different speakers from that subdistrict were chosen for close analysis for this paper.

2.1 Verb morphology

Sasak is in many ways a typical western Indonesian language. Thus, it has no nominal case marking. Neither aspect nor tense is marked on the verb, both being expressed by auxiliaries. The applicative suffixes *-in* (1) and *-ang* (2) are productive, and are found in both major transitive clause types (that is, those with or without a nasal prefix), not being restricted to a goal focus type.³

- (1) *ku=ny-(s)orak-in=e*
1SG=N-yell-APPL=3RD
'I yelled at him.'
- (2) *ku=peta-ang iye tangkaq*
1SG=look.for-APPL 3SG pot
'I'll look for a pot for it.'

Some intransitive verbs are marked with *be-* (3), cognate with Indonesian *ber-*. Often these verbs are stative, and instances of *be-* are so glossed in this paper. This simplification should not be taken to imply that they are all stative, as the actual semantics of *be-* are quite complex and beyond the scope of this paper.

- (3)a. *leq embe*
at where
'Where?'
- b. *siq taoq-taoq te=be-kedek ino*
RM place-place 1PL=ST-play that
'Where I usually play around.'

¹ For example, Austin describes a tense distinction indicated by word order variation, but I discovered no such phenomenon in my investigation.

² I am grateful to the staff of PPBK at the University of Mataram for their assistance during my time there, and to my three research assistants, Hadi, Ari and Ning, for their dedicated transcription and glossing efforts. Funding for the fieldwork was financed by the University of Auckland.

³ Abbreviations used in this paper include: 1PL first person plural, 1SG first person singular, 2SG second person singular, 3RD third person, 3SG third person singular, AGT agent, APPL applicative, ART article, CAUS causative, EMPH emphatic, EXIST existential, (F) feminine, FUT future, INTERROG interrogative, (M) masculine, N nasal prefix, NEG negative, NEG.IMP negative imperative, PASS passive, PAST (past), PERF perfective, PURPOSE purpose, RM relative marker, ST stative.

There are two types of passive found in Sasak, a non-volitional or stative passive, and an agentless passive. The non-volitional passives are marked by the prefix *ke-* (4).

- (4) *lawang ke-buke*
 door PASS-open
 'The door came open.'

Agentless passives take the prefix *te-* (5), probably related to Indonesian non-volitional *ter-*.

- (5) *bueq so iye te-endeng-endeng lonto, neng=ku*
 after.all EMPH 3RD PASS-ask-ask continuously word=1SG
 "After all, people keep asking for it all the time," I said.'

Most transitive verbs can occur either with or without a nasal prefix ((6) and (7)). This prefix appears as *nge-* before vowels and liquids. Elsewhere it is a single nasal phoneme, its place of articulation varying with the initial consonant of the root, while root-initial voiceless consonants are deleted. In examples, the deleted initial consonant is given in parentheses, thus *miaq*, from the root *piak* would be written *m-* (*p*)*iak*. Verb forms with the nasal prefix are referred to as nasal forms, while those without the nasal prefix are referred to as oral forms.

- (6) *ito aning saya⁴ n-jouq=e*
 there to 1SG N-take=3RD
 'That's where I took him.'
- (7) *wah-wah ku=jouq=e*
 already-already 1SG=take=3RD
 'I stopped taking him (along).'

2.2 Pronominal system

Sasak has a set of free-standing full pronouns which can refer to any semantic role, and a set of reduced pronominal forms that cliticise to both nouns and verbs, functioning as genitive markers on nouns, and with most verb types referring to actors. These forms are given in Table 1. The polite forms include borrowings from Balinese (*tiang, plinggih*), the use of first person plural *ite* for first person singular, and the use of third person (*s*)*ida/de* for second person. There are no separate plural forms for second and third person. Plurality is indicated, when desired, by adding the plural marker *pada* after any second or third person pronoun. When *pada* is not present, number is determined from context. There is also a verbal enclitic *-e* that refers only to third person patients (singular or plural). It can occur as an enclitic with both nasal and oral forms (see examples (6) and (7) above), though it is considerably more common with nasal ones (36 instances as compared to 13).

⁴ Note that the Indonesian pronoun *saya* is used in this example. Code-switching between Sasak and Indonesian, and nonce borrowing from Indonesian are common features of present-day Sasak.

Table 1: Sasak pronouns

	Full		Reduced	
	Basic	Polite	Basic	Polite
1SG	<i>oku</i>	<i>tiang, ite</i>	<i>ku</i>	<i>te</i>
1PL.INC	<i>ite</i>		<i>te</i>	
1PL.EXC	<i>kemi</i>			
2M	<i>epe, komu</i>	<i>(s)ide, plinggih</i>	<i>meq</i>	<i>de</i>
2F	<i>komu</i>	<i>(s)ide, plinggih</i>	<i>bi</i>	<i>de</i>
3	<i>iye</i>	<i>(s)ide, plinggih</i>	<i>ne</i>	<i>de</i>

The reduced pronouns may occur as proclitics with intransitive verbs, both unaffixed (8), and those with the stative prefix *be-* (9), and with both oral (10) and nasal transitive verbs (11).

- (8) *terus ku=taeq*
then 1SG=get.up
'Then I got up.'
- (9) *ite bilang te=be-deit, wajib ne=ngene ke rumah*
1PL each 1PL-ST-meet obligation 3SG=like.this to house
'Every time we met, she just had to say "come to my house".'
- (10) *epe meq=deit tono*
what 2SG(M)=find there
'What did you run into there?'
- (11) *ku=m-bait kepeng*
1SG=N-take money
'I took the money.'

If the verb is preceded by an auxiliary or a negative particle, or both, the reduced form appears as an enclitic on the first element of the verbal group instead of attaching to the verb ((12) and (13)).

- (12) *jemeq wah=ku lalo*
tomorrow PERF=1SG go
'I was going to go tomorrow.'
- (13) *ende=ku kewa n-jouq pepe-pepa*
NEG=1SG want N-take anything
'I don't want to take anything.'

With intransitive verbs and with oral and nasal transitives, the clitic represents the actor, as in (8)–(11) above, but when the verb takes the passive prefix *te-*, it represents the patient, and agrees with it ((14) and (15)).

- (14) *mule gen=ne te-kirim*
really FUT=3SG PASS-send
'Actually, they were going to send him.' (lit. he was going to be sent.)

- (15) *geng=ku te-kirim*
 FUT=1SG PASS-sent
 'I was going to be sent.'

The reduced pronouns may also occur as enclitics on verbs, representing the actor. Such enclitic actors are less frequent than proclitic ones, and are usually found on oral verbs (16). With nasal verbs enclitics are quite rare, being often, though not always, rejected in elicitation, and occurring only occasionally in texts (17).

- (16) *sakre telusur-in=meq doang keh*
 Sakre explore-APPL=2SG(M) just INTERROG
 'You just went exploring Sakre huh?'
- (17) *m-beng=ku ide se=tangkaq-se=tangkaq wah*
 N-give=1SG 2SG one=pot-one-pot already
 'I gave you a couple of pots already'

It should be noted that although proclitics can occur with all types of verbs, and enclitics with both nasal and oral transitives, they do not occur with equal frequency in all these environments, as is shown with respect to transitive verbs in Table 2. Not only are enclitics rare with nasal transitives, but proclitics are also less common with nasal forms than with oral forms. Overall, clitic actors are more likely to occur with oral forms. The reverse is true for pronouns; they are twice as likely to occur with nasal as with oral transitives.

Table 2: Actor clitic/pronoun distribution

	Total	Oral	Nasal
Proclitic	143	80	63
Enclitic	44	39	5
All clitics	187	118	68
Pronoun	52	17	35

While most clauses contain either a clitic or a free pronoun or NP, some contain both a clitic and a pronoun ((18) and (19)). However, this happens infrequently; there were only 9 instances in my database. This might be considered incipient verb agreement, but the rarity of the construction makes such an explanation dubious at best; instead it seems more logical to consider the free pronoun as simply an added emphatic element.

- (18) *oku ku=ng-gite loq Kirman*
 1SG 1SG=N-see AR(M) Kirman
 'I was looking at Kirman.'
- (19) *perhati-ang=ku laloq oku waktu ino loq Kirman*
 pay:attention:to-APPL=1SG very 1SG time that ART(M) Kirman
 'I was really paying attention to Kirman.'

2.2 Agent phrases

Sasak clauses may include an agent phrase preceded by the particle *siq*.⁵ These constructions are relatively rare, there being only 13 in my database. In 11 of them the verb is oral. When the agent phrase occurs with an oral verb, the verb also takes a third person clitic ((20) and (21)).

- (20) *loguq deit=ne oku siq to batur=ku*
 but meet=3SG 1SG AGT that friend=1SG
 'But that friend of mine ran into me.'

- (21) *meseq ndeq=meq giteq siq*
 impossible NEG=2SG(M) see RM
ne=awek-awek oku siq le Yeni ino
 3SG=pull-pull 1SG AGT ART(F) Yeni that
 'How could you not have see how that Yeni kept pulling at me?'

Note that the position of the clitic *ne* is independent of the position of the patient if overt; in both (20) and (21) the patient follows the verb, yet (20) has an enclitic and (21) a proclitic. However, if the patient is not overt, the clitic always seems to follow the verb (22).

- (22) *liwat aning jeding awek=ne siq le Lia*
 pass to toilet pull =3SG AGT ART(F) Lia
 'On the way to the bathroom, Lia pulls on them.' ('Them' here refers to the roots and leaves of a plant which is being discussed)

There is some question about the status of the clitic in these constructions. When the agent phrase occurs with an oral form, the third person clitic *ne* is always present, even though the patient may be first person. It thus does not appear that the clitic could represent the patient. In elicitation, examples with either a first or second person pronominal agent were judged grammatical, but the clitic on the verb always remained third person (23). Thus, it does not seem likely that the clitic represents the actor either.

- (23) *tiper ino popoq=ne siq oku*
 mat that wash=3SG AGT 1SG
 'I washed that mat.'

While agent phrases most often occur with oral verbs, speakers from Selong (but not from other areas, such as Praya [Husni Muadz pers. comm.]) also accept them with nasal verbs, in a construction that also includes both a pre-verbal patient and a patient clitic on the verb. There are two cases with nasal forms in my data base, one of which conforms to this construction (24), having both an initial pronoun which was interpreted by native speakers as referring to the patient, and a patient enclitic on the verb. The other does not (25).

- (24) *iye ne=nge-sir=e siq amaq Segep*
 3SG 3SG=N-desire=3RD AGT father Segep
 'Segep has the hots for her.'

⁵ *Siq* can also serve as a relative clause marker and or a personal article preceding human nouns, or introduce an instrument noun phrase.

- (25) *iyē ne=nge-sir komu siq loq Segep*
 3SG 3SG=N-desire 2SG AGT ART(M) Segep
 'Segep has the hot for you.'

In (25) *ne=ngesir komu* has the structure of a typical nasal clause with a clitic actor, that is, AVP. The agent phrase is clearly coreferential with the clitic *ne*, but the identity of the referent of the initial pronoun is not clear. This could be a standard nasal clause with two added agent expressions, one an emphatic pronoun in initial position and the other an emphatic agent phrase in final position. Or it could be some sort of blending of a standard nasal clause and an agent phrase construction, such as is found in (21). There, however, the free standing pronoun refers to the patient, while in (25) it cannot. Such blendings of two different clause patterns into a single unit, common in spoken language, can produce a construction which is formally ungrammatical (Ono & Thompson 1995).

It is possible that these agent phrase constructions are passives, marked by a combination of verbal morphology (*ne*) and agent demotion to an oblique phrase. However, this analysis raises a number of questions that cannot be answered based on the small number of examples available in this database. Why does the verbal affix change position from pre- to post-verbal for no obvious reason? Why is it possible with both oral and nasal verbs, and yet so much more frequent with oral ones? Can the instances with nasal verbs be considered passives? If this is a passive, and the agent is oblique and syntactically inaccessible, how does one explain the clause combining pattern found in (2), where the oblique agent appears to control deletion of the subject of the subordinate clause? This type of control pattern is the kind of evidence usually used in support of an ergative (active) analysis, not a passive one.

These problems suggest that a passive analysis is incorrect, and that it is better to consider the agent phrase independent of the voice system, a separate strategy for indicating something (as yet undetermined) about the status of the actor. Of course, this does not solve the problem of the presence of a clitic that agrees with neither agent nor patient. This is an area of Sasak grammar that needs considerably more study.

2.3 Syntactic constraints on verb type

There are a number of syntactic constructions where type of verb is constrained in Sasak. If, within relative clauses (26) and control structures (27), the absent, co-referential argument is an actor, a nasal verb must be used, while if it is a patient, an oral verb must be used.

- (26) *pire wah montor siq tabrakang=ne siq dengan siq*
 several PERF car RM crash=3SG AGT person RM
ny-(s)ewa iye wah
 N-rent 3SG PERF
 'Several cars were crashed by the person who had rented them.'
- (27) *ku=suruh inang=ku bareng amang=ku ng-(k)odu=e*
 1SG=order mother=1SG and father=1SG N-use=3RD
 'I told my parents to use it.'

When the patient is reflexive (28), or when the verb is used intransitively (29), a nasal verb must be used.

- (28) *iye m-(p)iak diri=ne*
 3SG N-make self=3SG
 'He made himself (do it).'
- (29) *lalo ng-endeng-ngendeng, kebiasaan*
 go N-ask-ask habit
 'Asking for things has become his habit.'

Imperatives employ oral verbs (30).

- (30) *ke, salin=e*
 go on change=3RD
 'Go on, change it (your shirt).'

All of these constraints are familiar ones from western Austronesian focus languages, first discussed in detail with respect to the question of subject properties and the identification of grammatical roles by Schachter in 1976.

2.4 Word order

Sasak allows considerable variation in major constituent word order, both in transitive and intransitive clauses. However, there are quite striking differences in occurrence among the different possibilities. It seems clear that nasal verbs have a basic word order of AVP, for both clauses with clitics and those with free-standing NPs, with a small amount of pragmatically conditioned variation. The distribution is given in Table 3. There were 162 nasal transitives in the data base; 12 occurred with no overt arguments, and thus present no evidence about word order. Of the remaining 150, only 14 had patients before the verb and/or actors after. Of course, when only one argument is present, one cannot speak definitively about the order of all major constituents, but since most of the clauses with only one argument present were either AV or VP, they provide no evidence of any significant tendency to employ other word orders.

Table 3: Word order in nasal clauses

2 arguments present	1 argument present	No argument present
A-V-P 81	A-V 21	V 12
V-A-P 3	V-P 35	
V-P-A 2	V-A 6	
	P-V 2	

It is, on the other hand, more difficult to see oral clauses as having a basic word order. When both arguments are present, four of the six possible orders are approximately equally frequent, while when only one argument is present, the result is consistent with more than one of the complete clause word orders.

Table 4: Word order in oral clauses

2 arguments present	1 argument present	No argument present
A-V-P 32	A-V 22	V 14
V-P-A 5	V-P 13	
V-A-P 30	V-A 15	
P-V-A 26	P-V 6	
P-A-V 20		

There are more 2-argument clauses with initial P than with either initial A or initial V, but I would hesitate to consider that even a hint of a basic order, as most initial P's seem to fall into certain very clear pragmatic classes which cross-linguistically are quite likely to be in clause initial position – question words, contrastive elements, and reintroductions of topical referents. Additionally, a comparison of VP and PV word orders for all clauses shows that VP is substantially more frequent (80 vs. 52). The data then suggest that clause final position is the basic one for a patient with an oral verb (just as with a nasal), but there is a fairly common pragmatically conditioned alternative. On the other hand, there is no tendency to order actors and verbs relative to one another, with VA and AV almost equally likely to occur (76 and 74 cases, respectively). I have not yet discovered any conditions which might determine verb initial as opposed to actor initial word order.

There proved to be no correlations between word order and the distinction between clitics and lexical NPs beyond the ones predicted by the rules of clitic placement; that is, if the argument is not in a position where a clitic could occur (patients not immediately after the verb, actors not adjacent to the verb), then of necessity that argument was lexical. Thus, all initial Ps were lexical. However, for word orders where cliticisation of either actor (AV, VAP, PVA, PAV, VA) or patient (VP, VPA) or both (AVP) were possible, there was no difference in the distribution of clitics and lexical NPs. Instead, both types of expressions were found with approximately the same relative frequencies in each of the frequent orders, although of course the absolute numbers for lexical NPs were much lower.

The distinction was relevant for clauses with agent phrases, which were predominantly VPA or VA. In these clauses all overt arguments were lexical. However, this relates not to word order, but to the nature of the construction itself.

2.5 Verb choice

2.5.1 Background

Given the presence of two verb forms, one obviously retaining the nasal morphology of actor focus, and the other with at least a strong association with the actor clitics found in patient focus in many Indonesian languages, it is natural to ask whether the Sasak forms have a function similar to actor focus and patient focus in other Austronesian languages, or whether they are being used in some other way. Initial examination suggested that the latter is the case. For one thing, nasal and oral constructions are almost equally frequent, both in general conversation and in narratives (Table 5), whereas in most languages where a conservative Austronesian focus system is still fully functional, patient focus is much more

frequent than actor focus in narratives (non-narrative data have not generally been investigated for those other languages.). For another, there is no patient focus affix, and the clitics occur freely with the nasal verbs, so that the alternation is not between a marked actor focus and a marked patient focus, but between a marked and an unmarked verb.

Table 5: Clause type frequency

	Narrative	Non-narrative
Nasal	38 (47%)	90 (46%)
Oral	43 (53%)	103 (54%)

In order to address the question of function, I have compared the discourse distribution of Sasak verb alternation to that seen in a number of well-studied Austronesian languages. Austronesian focus systems have been discussed within two discourse frameworks, that of discourse transitivity (Hopper & Thompson 1980), and that of topicality (first presented in Cooreman 1983, and then in more detail in Cooreman, Fox & Givón 1984). In both, certain discourse measures are tabulated, and conclusions are drawn from those tabulations. In this paper I am not arguing in favor of the conclusions they drew from the one or the other. I am simply applying their measures to Sasak, to see if the resulting patterns of frequencies are similar to those seen in the other languages.

Not all modern Austronesian languages with a functioning opposition between actor focus and patient focus display the same discourse profiles. For example, spoken Jakarta Indonesian does not show typical correlations between the focus forms and either topicality as defined by Givón and others, or transitivity as defined by Hopper and Thompson (Wouk, 1989). It could be argued that the unusual developmental history of Indonesian, both formal Standard Indonesian and the various colloquial varieties, has interfered with the natural development of the focus system, leading to an aberrant situation. However, the same is not true of Sasak, and the deployment of the focus system in Sasak shows substantial similarities to Indonesian in this regard, and differences from other, previously investigated western Austronesian languages.

I note that the discourse measures involved were all developed based on an analysis of narratives, and that their relevance to non-narrative data has not been demonstrated. However, I am not comfortable with an investigation that is restricted to narratives, for the simple reason that they are not in any way basic, or fundamental. They comprise one genre of natural speech, but only one, and not by any means the dominant one. For example, in this database of 345 clauses, only 105 (or less than a third) derive from narratives. The remainder exemplify all the other things people do with words. Any explanation or analysis of discourse function that is to have any real power must also take into account this other two-thirds of human interaction as well. Thus I have attempted, wherever possible, to apply these measures to all the speech in my database. I have also looked at those clauses which occurred only in narratives for purposes of comparison. In addition, previous work on voice has assumed the clause as the basic unit of speech; however, in spoken language the basic segment is the intonation unit, which may be a clause, and often is, but may also be less than or more than a clause. In measuring referential distance and topical persistence, I have used intonation units rather than clauses.

2.5.2 Transitivity

According to the transitivity hypothesis, voice forms correlate with the degree of discourse transitivity of the clause – high in ergative and active, low in passive and antipassive.

Measures of transitivity include, among others, mood, aspect, affectedness of O, individuation of O, and grounding within a narrative. Individuation of O is a complex measure, with many different dimensions involved, such as count/mass, concrete/abstract, animate/inanimate, and identifiable/non-identifiable. In the many Austronesian languages that retain much or all of the ancient focus system, actor focus correlates with low discourse transitivity, and patient focus with high. For most of these languages, individuation of O is the single most important variable in determining focus, and is itself determined by referential status. Actor focus co-occurs with unidentifiable patients, and patient focus with identifiable ones. This is true of such languages as Javanese, Tagalog, Toba Batak, (Wouk 1986a) and Early Modern Malay (Hopper 1979a, 1979b, 1982a, 1982b, 1988), and has been reconstructed for Proto Austronesian (Wouk 1986b). These languages are often called “discourse ergative” (Hopper 1979a, 1979b, 1982a, 1982b, 1988), since they clearly do not have the transitivity profile of an active/passive system.

If Sasak verbal alternation performed the same discourse function as focus systems in some of these other languages, we would expect the nasal clauses to be low in discourse transitivity, and the oral clauses high, and we would expect patient identifiability to be crucial in determining which verb form is used. This did not prove to be the case.

Table 6: Patient referential status

	Non-referential	Inter-rogative	First mention	Identifiable first mention	Later mention	Participant in conversation
All clauses						
Oral	17	14	23	15	99	34
Nasal	32	0	35	7	76	10
Narrative						
Oral	2	1	5	3	34	9
Nasal	9	0	10	2	25	5

As Table 6 shows, oral verbs are frequently used with non-referential and unidentifiable first mention patient referents, while nasal verbs are often used with identifiable patient referents, whether first mention, later mention or conversational participant. When only the narrative sections are considered, we still see some oral verbs with non-individuated patients, and nasal verbs with individuated patients remain quite frequent. Thus, referential status of the patients does not appear to play a strong role in choice of verb form, and in this regard Sasak does not then appear to be similar to languages such as Tagalog, Early Modern Malay, Javanese, or Toba Batak, where it does. Other measures of discourse transitivity investigated, such as aspect, mood, and grounding also showed little or no correlation with verb form.⁶

⁶ For a detailed discussion of transitivity in Sasak, see Wouk (1999).

2.5.3 Topicality

Cross-linguistically, different types of voice constructions have been shown to correlate with different degrees of relative topicality of actor and patient. In identifying voice constructions, clauses are not simply divided into two groups, actor focus and patient focus. Rather, criteria of verb morphology, nominal marking and word order are all employed to identify a full range of clause types with potentially different functions. The topicality patterns of a number of voice types are outlined in Givón (1994), and given in Table 7.

Table 7: Topicality

Active/ergative	AGT > PAT
Inverse	AGT < PAT
Passive	AGT << PAT
Anti-passive	AGT >> PAT

Topicality is indirectly assessed through two discourse measures, referential distance and topic persistence. Referential distance measures how far back the previous mention of a given referent is found. Early work counted up to twenty clauses back, assigning an arbitrary twenty to any referent not mentioned in that space. More recent work simply looks to see if the referent is found in the previous clause (highly topical) two or three clauses back (moderately topical), or more than three clauses back (low in topicality). Topic persistence looks to see how many times the referent is mentioned in the following ten clauses.

Other means have also been employed to determine the functions of different types of verbal clauses. One is the frequency of distribution of each construction, as a cross-linguistically stable pattern of such distribution has been found in narrative. Thus, it is reported that active/ergative typically accounts for 60–70% of all clauses, while passive and antipassive are typically under 10%, and inverse perhaps 20% (Givón 1994). Another is the frequency of non-anaphoric, or non-referential zero agents and patients. It has been found that non-referential zero agents are a characteristic of passives, while non-anaphoric zero patients are a characteristic of antipassives.

Based on these criteria, a number of Austronesian languages have been called discourse ergative, including Chamorro, Tagalog, and in fact all languages in which patient focus is the overwhelmingly most frequent form. In these languages, patient focus usually is identified with active/ergative and/or inverse functions, while actor focus is usually associated with antipassive.⁷

My investigation of topical persistence and referential distance in Sasak produced a very different picture.⁸ Since studies of some languages using the topicality approach have distinguished voices in part based on differences in word order, I calculated frequencies using only clauses with both arguments present. They were divided into four categories, based on a combination of word order and morphology: nasal clauses with AVP order, oral clauses with final P, oral clauses with initial P, and oral clauses with agent phrases. In Table

⁷ While the transitivity approach and the topicality approach both identify certain languages as discourse ergative, they do so on the basis of differing criteria (transitivity and topicality respectively). Thus, their definitions of discourse ergativity are not quite identical.

⁸ For a detailed discussion of topicality in Sasak, see Wouk (1999).

8 individual referential distance scores are combined to arrive at two groupings, high and medium (1, 2 or 3 clauses back) and low (more than three clauses back), and the relative topicality of actors and patients in each group of clauses is then compared on this basis. In Table 9 topical persistence data are presented in the same format. 0-2 is low topical persistence, 3-10 high. Table 10 includes the total number of each clause type, the relative topicality of agent and patient for each clause type with respect to referential distance (RD) and topical persistence (TP), and the conclusion that one would draw based on these results, using the topicality approach to voice. This is not meant to endorse this approach to determining voice, simply to allow comparison of Sasak and other languages that have been studied using this approach.

Table 8: Referential distance

	High	Low
P-final Oral		
Actor	33 (56%)	26 (44%)
Patient	45 (76%)	14 (24%)
P-initial Oral		
Actor	25 (57%)	19 (43%)
Patient	11 (25%)	33 (75%)
Agent Phrase Oral		
Actor	3 (27%)	8 (73%)
Patient	8 (73%)	3 (27%)
Nasal		
Actor	50 (62%)	31 (38%)
Patient	32 (40%)	49 (60%)

Table 9: Topical persistence

	High	Low
P-final Oral		
Actor	17 (29%)	42 (71%)
Patient	30 (51%)	29 (49%)
P-initial Oral		
Actor	19 (43%)	25 (57%)
Patient	9 (20%)	35 (80%)
Agent Phrase Oral		
Actor	6 (55%)	5 (45%)
Patient	9 (82%)	2 (18%)
Nasal		
Actor	30 (37%)	51 (63%)
Patient	25 (31%)	56 (69%)

Table 10: Overall topicality

	N	RD	TP	Conclusion
Nasal	81	a>p	a=p	active
Oral				
P-final	59			
AVP	31	p>a	p>a	inverse
VAP	28			
P-initial	44			
PVA	24	a>p	a>p	ergative (?)
PAV	20			
Agent phrase	11	p>>a*	p>a	passive

*If a larger sample bears out this preliminary indication.

The results make it clear that, according to the topicality criteria used, the nasal is not an antipassive, and is most likely an active. Oral clauses with final P seem to be inverse and those with an agent phrase passive. Oral clauses with initial P pattern like actives, which might suggest that they are 'ergative'.

This distribution makes Sasak appear rather different from other Austronesian languages that have been investigated with this methodology. Most striking, the nasal is not antipassive, and most of the oral clause types are associated with higher patient than actor topicality. Only when the patient is initial do we find higher actor topicality. This can be explained, however, without invoking ergativity, by the fact that the kinds of NPs that get fronted are characteristically discontinuous ones, which will show up as low in topicality when topicality is defined with reference to discourse continuity. However, the fact that they are fronted suggests some kind of importance for those discontinuous referents, a type of 'topicality' that cannot be captured by discourse measures such as RD and TP — the atypical position calls attention to them for a reason. It is thus logical that the verb form typically associated with greater patient topicality should be used with fronted patients. Calling these clauses ergative would mask this similarity.

2.6 Summary and historical speculation

To sum up the evidence from the discourse studies, the distribution of the two verb types in speech does not parallel that in other languages with a western Austronesian focus system, according to either of the analyses described. It is not surprising that both analyses should give anomalous results, since these two ways of looking at focus are necessarily complementary. If actor focus verbs are found in low transitivity clauses with non-identifiable patients, as in more typical western Austronesian languages, then actor focus clauses will, like antipassives, have patients of much lower topicality than their actors. When there is no strong correlation between verb type and patient referentiality, as in Sasak, the patterns of discourse transitivity and topicality that have been observed for other languages are not evident.

So what can we say about Sasak? Sasak seems to retain some aspects of the older Austronesian focus system, and gives evidence of having changed, at some point, to a western Indonesian type focus system, but is now in a stage somewhat intermediate between what is found in typical western Indonesian languages which retain a great deal of the focus system and the Oceanic languages which have lost all but traces in the form of frozen affixes.

Many of the characteristics of the PAN focus system are gone, such as marking of aspect and mood via verb morphology, and nominal case marking. Unlike Philippine type languages, but like western Indonesian ones, there are, first, only two major transitive 'voices', one of which is marked with a nasal prefix; and, second, two applicative suffixes which are found with both 'voices'. In addition, the old stative paradigm is largely gone, but like other western Indonesian languages, Sasak seems to have retained stative passive **ka-* as modern *ke-*. The presence of the prefixes *te-*, (probably cognate with Indonesian *ter-*) and *be-*, (probably cognate with Indonesian *ber-*)- are other similarities with western Indonesian languages.

The use of actor reduced forms shows similarities to and differences from both western Indonesian and PAN focus. As in PAN, the reduced forms can occur as enclitics on nouns, verbs, auxiliaries, and negatives, and as in western Indonesian languages they may occur as proclitics on verbs. That they appear as both proclitics and enclitics on verbs is surprising in a western Indonesian language, as generally, once the transition to verbal proclitic took place (for a given person), the enclitic was lost (Wolff 1997). It appears that, for some reason, Sasak, in retaining the enclitic, represents a phase which other western Indonesian languages must have passed through.

In addition, Sasak has undergone independent changes. There is no longer a clear association between clitics and 'patient focus', since the proclitics can occur with all verb types. The differential distribution of clitics and pronouns with the different verb types (clitics favor oral transitives, pronouns favor nasals and intransitives) is probably best explained by the fact that the proclitics were originally part of the focus system, and thus were restricted to oral (patient-focus) verbs, while other verb types occurred with pronouns. More recently, then, the clitics gained independent status and began to spread to other environments, as has the free pronoun to oral transitives. However, the recency of this spread leaves traces of the old system in the frequency differences.

The spread of enclitics to nasal verbs is a Sasak innovation, although one that hasn't yet gone very far. Finally, the identifiable patient constraint, which restricted actor focus verbs to clauses with unidentifiable patients, and patient focus verbs to clauses with identifiable ones, is gone. I cannot think of any logical connection between the spread of clitics to other verb forms and the loss of the identifiable patient constraint, beyond the fact that both represent a loosening of constraints on the use of verbal morphology. However, the combination of these two changes leaves Sasak very different from a typical western Austronesian or indeed western Indonesian language.

3 Sumbawa

Sumbawa, which has been the subject of a detailed dialect survey (Mahsun 1994), has approximately 300,000 speakers. There are four main dialects, Jereweh, Taliwang, Tongo and Sumbawa Besar, each with a number of sub-dialects. This discussion of two of those

dialects, Sumbawa Besar and Jereweh, is based on my work with a native speaker of the Sumbawa Besar dialect, a published description of that dialect by non-native speakers (Sumarsono et al. 1986), and a published description of the morphology of the Jereweh dialect by a native speaker of that dialect (Mahsun 1990). Sumarsono et al. has to be used with caution, as Mahsun (1990) points out that it contains certain errors in morpheme classification. There are grammatical differences among the dialects, some relating to the verbal system. Where relevant, these are described below.

In Sumbawa there is a productive alternation between a nasal-prefixed form and an oral unprefixing one, for many transitive verbs. However, the function of this alternation is quite reduced relative to what is found in a focus system. In Sumbawa, the nasal-prefixed form of the verb can only be used when the verb is formally intransitive, while the unprefixing form is used when the verb is used transitively. Because of the reduced distribution of nasal-prefixed verbs, in Sumbawa Besar texts (I have none from Jereweh), the nasal-prefixed verb occurs quite infrequently. In one personal narrative, there were two nasal-prefixed verbs as compared to 32 transitive verbs without the nasal prefix. The nasal shows a similar type of allomorphic alternation described for Sasak, but all initial stops are deleted when the nasal prefix is attached. Such verbs are represented in the same fashion as Sasak verbs with initial voiceless consonants.

The different functions of the two forms of the verb are illustrated in examples (31)–(34). In each set, the (a) example, which lacks a patient, contains a verb with a nasal prefix, while the (b) example, which has one, contains a verb without a prefix. Note that the (b) examples all display AVP word order, the basic order for transitive clauses in Sumbawa when both arguments are overt.

- (31) Sumbawa Besar dialect
- a. *Wayan de ya-m-(p)ukil*
Wayan RM FUT-N-hit
'It's Wayan who'll do the hitting.'
 - b. Sumbawa Besar dialect
Hamid ka-pukil Yono
Hamid PAST-hit Yono
'Hamid hit Yono.'
- (32) Sumbawa Besar dialect
- a. *Helmi ka-m-(b)ayar ko Wayan*
Helmi PAST-N-pay to Wayan
'Helmi paid Wayan.'
 - b. *Helmi ka-bayar mejang ta*
Helmi PAST-pay table that
'Helmi paid for that table.'
- (33) Sumbawa Besar dialect
- a. *Helmi ny-(s)apu pang kamar*
Helmi N-sweep in room
'Helmi is sweeping in the room.'

- b. *Helmi sapu kamar ta*
Helmi sweep room that
'Helmi is sweeping the room.'

(34) Jereweh dialect

- a. *na sama m-(p)lentong*
NEG.IMP 2PL N-throw
'Don't throw (things).' (Where the identity of what is not to be thrown is recoverable from context.)
- b. *na sama plentong plam so*
NEG.IMP 2PL throw mango that
'Don't throw that mango.'

Statives can be prefixed by *ba-* as in (35) (with allomorphs *bar-* before vowel initial roots, *bal-* before the single root *ajar*, and *ra-* before roots with initial labial consonants).

(35) Jereweh dialect

- nya so ba-kios ing karang*
3RD that ST-house in village
'He has a kiosk (small shop) in the village.'

Both dialects have an accidental passive marked by *ka-*, but in the Jereweh dialect it is not productive, being restricted to a very small set of roots (4 total). In SB dialect, on the other hand, its use is widespread (36).

(36) Sumbawa Besar dialect

- embang-embang ka-ku-teri ko aspal.*
suddenly PAST-1SG-fall to asphalt
- ka-belok nin-ku,*
PASS-scrape knee-1SG
'Suddenly I fell onto the road. My knee got scraped.'

Aspect is not a part of the Sumbawa verbal system, however, tense is marked by means of prefixes, *ka-* (alternating with *kam-* in SB dialect — see example (45)) marking the past, as in the first clause of (37), and *ya-* marking the future, as in (38). This seems to be an independent development in Sumbawa, and is not obligatory, being marked approximately half the time in the narrative I collected.

(37) Sumbawa Besar dialect

- ku-perasa ima-ku ya-sate terang*
1SG-feel hand-1SG FUT-want well
'I felt that my hand was going to recover.'

Sumbawa has both free standing pronouns and personal clitics (Table 11). The singular clitics *ku*, *mu* and *nya* can be encliticised to nouns as possessives. Singular first and second person *ku* and *mu*, and first person plural *tu*, but not third person singular *nya*, can be procliticised to transitive and intransitive verbs and to adjectives. Clitics may appear by themselves, as in (36), but they are often accompanied by a free standing pronoun, as in (38), suggesting that they might be considered a form of verb agreement, with optional pro-drop. In elicitation, examples with a pronoun but without the agreement clitic were judged

grammatical, but in a personal narrative my informant used a clitic for all verbs with first person subjects. (There were no clauses with second person subjects in the narrative.)

Table 11: Sumbawa pronouns

	Pronoun	Clitic
1SG	<i>aku</i>	<i>ku</i>
2SG	<i>kau</i>	<i>mu</i>
3	<i>nya</i>	<i>nya</i>
1PL.INC	<i>kita</i>	<i>tu</i>
1PLEXC	<i>kami</i>	<i>tu</i>

(38) Sumbawa Besar dialect

aku ka-ku-entek pang anar bungka
 1SG PAST-1SG-ascend in ladder behind
 'I climbed up the rear ladder.'

The three singular clitics can also appear on the ends of transitive verbs, representing the patient, as in (30), and (39) below.

(39) Sumbawa Besar dialect

untung ada tau ka-ingo-ku, teris angkat-ku, bawa aku
 luck exist person PAST-see-1SG then lift-1SG carry 1SG
ko sekolah
 to school
 'Luckily, there was someone who saw me, then lifted me up and carried me to the school.'

Sumbawa has a passive construction, in which the actor appears as an oblique NP, introduced by the prepositional phrase *ling* in Sumbawa Besar (40) or by *iN* in Jereweh (41). When the P is overt, the basic A-V-P word order may be altered to P-V-A, as in (40) and (41). Otherwise, the P comes immediately after the verb, with the agent phrase following, as in (44). In the Jereweh dialect the verb must take a prefix *i-*. Note that with a 3-argument verb either the patient (41), the recipient/beneficiary (42), or the location (43) can be fronted.

(40) Sumbawa Besar dialect

Andi pukul ling Iwan
 Andi hit by Iwan
 'Andi was hit by Iwan.'

(41) Jereweh dialect

baso uda ka-i-beli ing inang duman aku
 corn young PAST-PASS-buy by mother for 1SG
 'Mother bought me the young corn .' (lit. The young corn was bought for me by mother.)

(42) Jereweh dialect

Gait ka-i-beli lamung baru im bapak
 Gait PAST-PASS-buy shirt new by father
 'Father bought Gait a new shirt.' (lit. Gait was bought a new shirt for by father.)

(43) Jereweh dialect

bangkat kon ka-i-talat im bapak ke pade rau
 field that PAST-PASS-plant by father with rice field
 'Father planted that field with rice (of the type that doesn't need irrigation).'
 (lit. That field was planted with rice by father.)

In SB dialect the situation is a bit more complicated. In elicitation, clauses with an unmarked verb and an agent phrase were produced and glossed as passive, as in (40) above, but none occurred in my text. Instead, the text contained clauses with two different verb forms. Some were marked with *ka-*, as in (44). This is presumably an extension of the accidental passive mentioned above. Others took a prefix *ya-*, as in (45). Note that the *ya-* prefixed form can occur with neither an overt actor nor an overt patient, as in the first clause of (45). *Ka-* prefixed verbs, on the other hand, never appeared without a patient, and only appeared without an actor when there was no semantic agent; that is they were being used in the accidental passive meaning.⁹

(44) Sumbawa Besar dialect

ka-ajak-ku ling dengan-ku lalo ko Moyo
 PASS-invite-1SG by friend-1SG go to Moyo
 'My friend invited me to go to Moyo.'

(45) Sumbawa Besar dialect

teris ya-tampal mo ke air ade kam-ya-sedia
 then PASS-cover EMPH with bamboo RM PAST-PASS-prepare
ling sandro
 by traditional:healer
 'Then it was covered with bamboo which had been prepared by the
 traditional healer.'

According to Sumarsono et al. (1986), *ya-* also functions like the applicative suffixes found in many western Indonesian languages, allowing an indirect object to become a direct argument of the verb in active clauses, but my informant did not require any prefix in those cases, and always interpreted such clauses either as future or as passive. Thus, according to Sumarsono et al. (46a) means 'Mother salts the corn' while (46b) is ungrammatical, and (46c) means 'That man passes judgement'. For my informant, (46a) means 'Mother will salt the corn' while (46b) means 'Mother salts the corn', and (46c) means 'That man is judged'. This may be another case of misanalysis by Sumarsono et al., one that was not commented on by Mahsun (1990) since it relates to an affix not found in Jereweh dialect.

⁹ Sumarsono et al. (1986) also mention an archaic passive suffix *tu-*, homophonous with the first person plural clitic, used in proverbs and folk tales. My young informant consistently interpreted *tu-* as first person plural, and not knowing any appropriate Sumbawan proverbs, I was unable to check his interpretation of *tu-* in them. This suffix is not described as existing in Jereweh dialect.

- (46) Sumbawa Besar dialect
- a. *ina ya-sira jembrai*
mother ya-salt corn
 - b. *ina sira jembrai*
mother salt corn
 - c. *tau nən ya-ukum*
man that ya-judgement

To summarise, Sumbawa has diverged even further than Sasak from the typical western Indonesian language. It seems likely that Sumbawa first underwent the change from PAN type to western Indonesian type, and then lost focus. The evidence for this includes the change in basic word order to AVP, the presence of a passive prefix, and the fact that the first and second person clitics, although not restricted to transitive verbs, have become proclitics. The fact that third person did not become a proclitic suggests that the change may not have been complete, as is the case for some languages of Sulawesi (Wolff 1997). Sumbawa also utilises a nasal prefix like the actor focus affix of many western Indonesian languages, although not as part of a focus system. Like Sasak, Sumbawa has retained stative passive *ka-*, although in the Jereweh dialect its use has contracted almost to vanishing, and in the Sumbawa Besar dialect it seems to have expanded to an agented passive. Sumbawa also has a stative *ba-* prefix, something found only in some western Indonesian languages.

Sumbawa is unique in that the proclitics have spread to all verb types, and to adjectives, and seem to have developed into a form of verb agreement. Also, Sumbawa makes no use of the two applicative suffixes found in many western Indonesian languages, although reflexes of these suffixes are found in what are considered to be Sumbawa's two closest relatives, Balinese and Sasak. Finally, Sumbawa tense marking is a strictly local innovation.

4 Bima

The three NTB languages, even though they represent different subgroups of Austronesian, nonetheless still form a loss-of-focus continuum reflecting their geographic position, with Sasak closest to a traditional focus system, Sumbawa intermediate, and Bima furthest away.

There are approximately 400,000 speakers of Bima, and a preliminary study suggests that it has three major dialects: Bima, Wawo, and Kolo (Herusantosa et al. 1987); however, a careful dialect survey has not been done. This description of Bima is based on Jonker's (1896) grammar of Bima, data from two speakers (one from the village of Kore in the subdistrict of Sanggur in the Dompu Regency, the other from the Ra'ba regency) and a dictionary written by a speaker from the subdistrict of Wawo in the Bima Regency (Sahidu 1978).¹⁰ According to current best knowledge of Bima dialectology, Wawo and Sanggur speakers use the same dialect, while Ra'ba is in the Bima dialect area. My preliminary investigations suggest that while there may be some grammatical differences among the

¹⁰ I am grateful to Michael Dukes for sharing the Raba speaker data with me, and to Erik Van Rijn for assistance with translating portions of Jonker's grammar.

dialects and sub-dialects of this language, they would not appear to bear on our concerns here.

Bima, like all other Cenral-Eastern Malayo-Polynesian languages, has very little productive affixation. There is a causative prefix *ka-* (47), which Jonker states can also be used as a stative passive, although I have not run across this function in my data.

- (47) *sia ka-losa masampu 'di¹¹ uma*
 3SG CAUS-out trash from house
 'S/he removed the trash from the house'

There is a passive perfective prefix *ra-* (48), derived from the particle *ra'a* 'formerly'.

- (48) *ta'be dou ra-mbei buku 'ba la Ali*
 where person PERF-give book by ART Ali
 'Where is the person who was given a book by Ali?'

There is also an applicative suffix *-kai* (49), (homophonous with the preposition meaning 'with') and that's about it for morphology.

- (49) *depala cola-kai sakola landa es*
 but pay-APPL school sell ice
 'but I paid for school by selling ice'

Verbs and adjectives are frequently preceded by *ma* (50), which Jonker considers a prefix forming active participles. However, in current usage it is better considered a relative clause marker, because in addition to preceding verbs and adjectives, it also precedes prepositional phrases (51) and the existential particle (52).

- (50) *ti ba'de=ku cou ma weha*
 NEG know=1SG who RM take
 'I don't know who took it.'
- (51) *nahu ne'e doho 'di kadera ma 'di uma*
 1SG want sit in chair RM in house
 'I want to sit in the chair in the house.'
- (52) *nahu ne'e doho 'di kadera ma wara 'di uma*
 1SG want sit in chair RM EXIST in house
 'I want to sit in the chair which is in the house.'

In addition to free-standing pronouns, personal clitics are found, although only for singular referents. These are given in Table 12.

¹¹ In the Bima examples 'd' represents an alveolar implosive, and 'b' a labial one. Sahidu describes these sounds as a post-alveolar lateral and a bilabial non-plosive, while Jonker recognises only the alveolar one, which he considers to be similar to the Javanese retroflex alveolar stop. However, in my informant's speech they sounded like implosives. This may be dialectal variation.

Table 12: Bima pronouns

	Pronoun	Clitic
1SG	<i>nahu</i>	<i>ku</i>
2SG	<i>nggomi</i>	<i>mu, ta</i>
3SG	<i>sia</i>	<i>na</i>
1PL	<i>nami</i>	

The clitics occur as enclitics on nouns, marking possession, and as proclitics or enclitics on verbs, with enclitics occurring almost twice as frequently as proclitics. The clitics can co-occur with actor pronouns, as exemplified by (53) and (54). This usage makes them seem more like verb agreement markers than like anaphora. However, they are not obligatory, and (55) and (56) are equally grammatical. Additionally, in discourse, these clitics are more commonly used as possessives than as either anaphora or verbal agreement markers. Free pronouns and zero anaphora are considerably more common than clitics.

(53) *nahu lao-ku 'di amba*
 1SG go-1SG to market
 'I went to the market.'

(54) *nahu ngaha-ku oha*
 1SG eat-1SG rice
 'I eat rice.'

(55) *nahu lao 'di amba*
 1SG go to market
 'I went to the market.'

(56) *nahu ngaha oha*
 1SG eat rice
 'I eat rice.'

Thus, in Bima, as in other Central-Eastern Malayo-Polynesian languages, focus is gone, and little productive use of the old focus morphology remains. The modern causative prefix and relative clause marker may be derived from earlier PAn stative prefixes, but their functions have changed radically. The applicative suffix and passive prefix appear to be independent innovations, unrelated to developments in western Malayo-Polynesian languages.¹² Nominal case marking is gone. The clitics appear with both transitive and intransitive verbs.

Although the focus system has been lost, some of its functions are fulfilled by other strategies, either newly developed ones or retentions from the older system. Thus, for example, there is an agentless passive with the verb preceded by the morpheme *'di* (57), which also occurs as a preposition meaning in the same example, or as a purposive marker (58).

¹² It is possible that the applicative suffix was developed on the model of applicative suffixes in languages of Indonesia, but the morpheme does not appear to be cognate, and it is equally possible that no such influence was involved.

- (57) *ma loa 'di sakola 'di ngadu ese Mbojo*
 RM clever in school PASS send to Bima
 'whoever does well at school is sent to Bima'
- (58) *Ali kani balpoin 'di tunti-kai sura*
 Ali use pen PURPOSE write-APPL letter
 'Ali used a pen to write the letter.'

Word order variation is another such strategy. In situations of high patient topicality, basic A-V-P word order is replaced by P-V-A, with the actor expressed as a prepositional phrase, as in the following question, taken from an conversation about how to raise chickens for resale (59).

- (59) *janga re wa'a 'ba ita atau mai weha 'ba dou*
 chicken that carry by 1PL or come take by person
 'those chickens, do you get them yourself, or does someone bring them?'

The only clear trace of the focus system is found in the lexicon, where some transitive verbs have an initial homorganic nasal which is not part of their PAN root (*mbei* 'give', *ngaha* 'eat', *nangi* 'cry'). It would thus appear that in Bima the last remaining component of the focus system is the nasal prefix associated with actor focus.

5 Conclusion

The three languages examined show a gradual loss of focus, moving from west to east. In the westernmost language examined, Sasak, much of the focus morphology of the Indonesian type language remains, but it is no longer associated with focus. The suffixes, of course, lost their association with focus very early, in all western Indonesian languages, including Sasak, when the conjugated forms arose (Wolff 1997). In Sasak, those conjugated forms have in turn lost their association with focus as the clitics spread to other verb types, and pronouns spread to the oral verb. Additionally, the strong association between actor focus and patient non-individuation was lost, leading to a disruption of typical patterns of referent topicality and discourse transitivity. Thus, the nasal prefix remains, but no longer with the meaning it once had. Once this happens, the way is cleared for a gradual decrease in its use, presumably accompanied by a loss of the syntactic constraints that grew out of the focus system. This has not yet happened in Sasak, but may be the next stage.

In the middle language, Sumbawa, much of the focus morphology is lost, and what remains, the nasal prefix, plays a very minor role in the organisation of the language, having completely lost its association with focus. Presumably, after the focus system broke down, as it is doing in Sasak, the nasal prefix was used less and less frequently, since its functional load had been reduced. Eventually, it became restricted to the few uses we see today.

Finally, in the easternmost language, Bima, the focus morphology has, except for traces of the nasal prefix in the lexicon, completely disappeared.

Thus, the evidence from these three languages tells us a good deal about the nature and consequences of the loss of the focus system, and identifies the nasal prefix as its most robust component — retaining a role longest, and, when that has finally gone, leaving the longest-lasting trace.

Bibliography

- Austin, Peter, 1996, Ergativity, clitics and grammatical relations in Sasak. Paper presented at the First Conference on Lexical-Functional Grammar.
- Cooreman, Anne, 1983, Topic continuity and the voicing system of an ergative language: Chamorro. In Talmy Givón, ed. *Topic continuity in discourse*, 425-489. Amsterdam/Philadelphia: John Benjamins.
- Cooreman, Anne, Barbara Fox and Talmy Givón, 1984, The discourse definition of ergativity. *Studies in Language* 8:1-34.
- Givón, Talmy, 1994, The pragmatics of de-transitive voice: functional and typological aspects of inversion. In Talmy Givón, ed. *Voice and inversion*, 3-44. Amsterdam/Philadelphia: John Benjamins.
- Herusantosa, Suparman I made Gosong, I Wayan Bawa, Sumarsono, Razak Roediyanto and I Gusti Ptu Antara, 1987, *Pemetaan bahasa-bahasa di Nusa Tenggara Barat*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Hopper, Paul, 1979a, Some discourse sources of ergativity. *University of Hawaii Working Papers in Linguistics* 11/2:137-153.
- 1979b, Aspect and foregrounding in discourse. In Talmy Givón, ed. *Syntax and semantics 12: Discourse and syntax*, 213-224. New York: Academic Press.
- 1982a, Aspect between discourse and grammar. In Paul Hopper, ed. *Tense-aspect: between semantics and pragmatics*. Amsterdam: John Benjamins.
- 1988, How ergative is Malay? In R. McGinn, ed. *Studies in Austronesian linguistics*, 441-453. Athens, Ohio: Ohio University Center for International Studies – Center for Southeast Asia Studies.
- Hopper, Paul and Sandra Thompson, 1980, Transitivity in grammar and discourse. *Language* 56:251-299.
- Jonker, J.C.G., 1896, *Bimaneesche spraakkunst*. 'S Hage: M. Nijhoff.
- Mahsun, 1990, *Morfologi Bahasa Sumbawa Dialek Jereweh: Afiksasi dan reduplikasi*. Yogyakarta: Universitas Gadjah Mada
- 1994, *Penelitian dialek geographis Bahasa Sumbawa*. Yogyakarta: Universitas Gadjah Mada
- Ono, Tsuyoshi and Sandra Thompson, 1995, What can conversation tell us about syntax? In Philip Davis, ed. *Alternative linguistics : descriptive and theoretical modes*. Amsterdam Philadelphia: John Benjamins.
- Sahidu, Abdul Karim, 1978, *Kamus Bima Indonesia*. Mataram: UnRam
- Schachter, Paul, 1976, The subject in Philippine languages: topic, actor, actor-topic or none of the above. In Charles Li, ed. *Subject and topic*, 491-518. New York: Academic Press.
- Staff, Nell F., 1995, *Kamus Sasak Indonesia Inggris*. Mataram: Mataram University Press.
- Sumarsono, Ketut Seken, Nyoman Merdhena, Nengah Martha, 1985, *Kamus Sumbawa-Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Sumarsono, Made Nadera, Sunaryono Basuku, Nyoman Merdhana, Nengah Mertha, 1986, *Morfologi dan sintaksis Bahasa Sumbawa*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Teeuw, A., 1958, *Lombok, een dialect-geografische studie*. 's Gravenhage: Martinus Nijhoff.
- Thoir, Nazir, Ketut Reoni, I Ketut Karyawan, 1986, *Tata Bahasa Bahasa Sasak*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

- Wolff, John, 1999, The development of the passive verb with pronominal prefix in western Austronesian languages. In Bernd Nothofer, ed. *Reconstruction, classification, description: festschrift in honor of Isidore Dyen*, 15-40. Hamburg: Abera-Verlag.
- Wouk, Fay, 1986a, Transitivity in Batak and Tagalog. *Studies in Language*, 10:391-424.
- 1986b, Transitivity in PMP and PAN. In Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*, 133-158. Canberra: Pacific Linguistics.
- 1989, The impact of discourse on grammar: verb morphology in Spoken Jakarta Indonesian. PhD dissertation, University of California, Los Angeles.
- 1999, Sasak is different: a discourse perspective on voice. *Oceanic Linguistics*, 38:91-114.

Changes in word order and noun phrase marking from Old to modern Javanese: implications for understanding developments in western Austronesian 'focus' systems

GLORIA R. POEDJOSOEDARMO

1 Introduction

In modern Javanese SV(O), the dominant word order of English and many other European languages, is also dominant, though by no means exclusive. This causes Javanese (like modern Malay/Indonesian) to appear at first glance deceptively similar in structure to European languages. In Old Javanese, VS order was far more common than it now is and noun phrases were partially marked for 'case' or semantic role, making the language at that stage appear more like a 'focus' language, that is a language of the Philippine type.

This paper describes the basic features of the verb morphology, noun phrase marking and word order patterns of first modern Javanese and then Old Javanese. It then explores the nature of the changes which must have taken place to lead from the second to the first and looks tentatively at the degree to which it might be possible to reconstruct still earlier forms of the language and what implications might emerge from this endeavour for understanding the development (and loss) of 'focus' systems in western Austronesian languages.

Before proceeding, it should be noted that there are some problems with terminology, beginning with the word 'focus' itself, which is used in relation to Philippine languages to refer to a particular kind of grammatical system in which noun phrase initial particles indicate the semantic role of that phrase, a special particle indicates subject (or 'focussed on') status and the verb morphology indicates the semantic role of the subject. The term 'focus' is also used in linguistics to refer to an item in an utterance which is judged by the speaker to carry the newest and most salient information. This is indicated in English by making a word more prominent, that is by making its stressed syllable longer and louder and making it function as the pivot point for pitch movement. It is not entirely clear to me what the function

of the formal system of 'focus' in Philippine languages is but it almost surely is not to mark 'focus' in this second sense.

The use of the term 'focus' in the informational sense as well as the use of terms such as 'topic' or 'theme' may suggest a 'functional' as opposed to a 'formal' approach to linguistic analysis. Though it is not my purpose here to advocate a particular theory, it is my feeling that to ignore these functions, regardless of what they are called, can only obscure the reasons for changes which have taken place over the past two millennia in the morphology and syntax of the languages in the Malayo-Javanic subgroup of Austronesian. This is particularly true with regard to changes in word order which, as Cumming (1988) has pointed out in reference to Malay, came about due to gradual changes in functions of the orders, resulting in one which had previously been marked eventually becoming the most common.

In general in this article I use the term 'focus' in the Philippine sense and specify 'information focus' when the other meaning is intended.

2 Modern Javanese

2.1 Verb morphology

Though modern Javanese transitive verbs have clear active (i.e. agent/actor as subject) forms, indicated by the prefix *N-*, which are distinct from a set of passive (i.e. non-agent/actor as subject) forms, indicated by the prefixes shown in Table 1, to say that Javanese has 'voice', like European languages, rather than 'focus', like Philippine languages, is misleading.

Table 1: Passive prefixes in modern Javanese

	Prefix
First person agent	<i>dak-</i>
Second person agent	<i>kok-</i>
Third person agent	<i>di-</i>
(Unspecified agent)	<i>di-</i>

It is frequently pointed out, as Spitz has done in his contribution to this volume, that while European languages have a two-way active-passive voice distinction, Philippine languages have as many as four possible focuses for a given verb. In modern Javanese, though the European-language-like active-passive distinction exists, it is also possible to make multiple focus distinctions like those available in Philippine languages. Examples are given in sentences (1)–(4).

(1) **Actor focus** (i.e. actor = subject):

SUBJECT/ACTOR					VERB
<i>Mbok Marta</i>	<i>mau</i>	<i>esok</i>	<i>menyang pasar</i>	<i>tuku</i>	<i>beras.</i>
Mrs (name)	earlier.today	morning	go.to	market buy	rice.(uncooked)
'Mrs Marta this morning (went) to the market (to) buy rice.'					

(2) **Patient focus** (i.e. patient = subject):

SUBJECT/PATIENT		VERB	ACTOR
		(PASSIVE)	
<i>Beras sing neng pawon kae sing</i>		<i>dituku</i>	<i>mbok Marta</i>
rice which at/in kitchen that (is.that).which		be.bought	Mrs (name)
<i>mau esok.</i>			
earlier.today morning			

'The (uncooked) rice which is in the kitchen is the (rice) which was bought by Mrs Marta this morning.'

(3) **Benefactive focus** (i.e. benefactee = subject):

SUBJECT/ BENEFACTEE	VERB (PASSIVE- BENEFACTIVE)	PATIENT	ACTOR
<i>Ibu biasane ditukoke kain batik dening mbok Marta.</i>			
Mother usually be.bought.for cloth batik by Mrs (name)			

Literally: 'Mother is usually bought batik cloth by Mrs Marta,' (i.e. *ibu* 'mother' is the subject or 'focussed' noun phrase).
 Meaning: 'Mrs Marta usually buys batik cloth for mother.'

(4) **Locative focus** (i.e. location = subject):

SUBJECT/ LOCATION	VERB (PASSIVE- LOCATIVE)	ACTOR
<i>Mbok Marta kae sing biasane ditukoni ibu.</i>		
Mrs (name) that which usually be.bought.at mother		

'Mrs Marta is the one whom mother usually buys from,' OR:
 'Mrs Marta is the one whose shop mother usually buys at.'

In theory at least,¹ it is possible to make the focus choices in sentences (2)–(4) in either the active or passive voice. The examples given above are in the passive, that is the focussed item is the grammatical subject. It is possible to make the focussed item into the grammatical object, the grammatical subject being the actor/agent, as in sentences (5)–(7).

(5) **Actor as subject/Patient as object:**

OBJECT/PATIENT	VERB (ACTIVE)	SUBJECT/ACTOR
<i>Beras sing neng pawon kae, sing nuku Mbok Marta.</i>		
rice.(uncooked) which at/in kitchen that which buy Mrs (name)		

'The (uncooked) rice which was in the kitchen, the one who bought it is Mrs Marta.'

¹ In fact active voice is far less common than passive in Javanese and tends to occur only when the patient is indefinite and the continuing topic (that which has been and is being talked about) is the actor, or, as in the first example, where an actor is being introduced as a new topic and is thus the focus of information.

(6) **Actor as subject/Benefactee as object:**

SUBJECT/ACTOR VERB OBJECT/BENEFACTEE
 (ACTIVE-
 BENEFACTIVE)

Mbok Marta biasane nukokke ibu.

Mrs (name) usually buy.for mother

'Mrs Marta usually buys things for mother.'

(7) **Actor as subject/Location as object:**

SUBJECT/ACTOR VERB OBJECT/LOCATION
 (ACTIVE-
 LOCATIVE)

Ibu biasane nukoni mbok Marta.

Mother usually buy at Mrs (name)

'Mother usually buys (things) from Mrs Marta.'

One might point out that with English ditransitive verbs more than one object is also possible. However, the Javanese system is far more complex than the English one in that with English there is a maximum of two possible objects (usually called 'direct' and 'indirect') for any given verb whereas in Javanese there are frequently at least three choices. Furthermore, English ditransitive verbs are quite limited in number whereas the majority of Javanese transitive verbs can have more than one potential grammatical object. The Javanese system is additionally more complex than the English one in that, whether the verb is active or passive, the semantic role of the focussed entity is indicated by verbal suffixation or its absence. The verb morphology expressing this complex voice/focus system is shown in Table 2, where the generalised *di-* prefix is used to indicate passive. It should be remembered, however, that all of the options shown in Table 1 are available for all passive forms:

Table 2: The Javanese voice/focus system

	Patient-focus	Benefactive-focus	Locative-focus
Active voice	<i>N-</i>	<i>N- -ake</i>	<i>N- -i</i>
Passive voice	<i>di-</i>	<i>di- -ake</i>	<i>di- -i</i>

In fact, the system is more complex than that shown here in that forms in Table 2 only include those which might be labelled 'indicative mood'. There are partially comparable forms for the 'imperative/subjunctive' (Table 3) and the 'desiderative' (meaning something like 'I think I'll do X'—Table 4), though the active/passive distinction is not made with these forms. Actor focus in the imperative is used with intransitives or with potentially transitive verbs in contexts where the object is not relevant. The desiderative forms are, of course, all passive with first person agent.

Table 3: The Javanese imperative/subjunctive forms

Actor-focus	Patient-focus	Benefactive-focus	Locative-focus
<i>-a</i>	<i>-en</i>	<i>-(k)na</i>	<i>-ana</i>

Table 4: The Javanese desiderative forms

Patient-focus	Benefactive-focus	Locative-focus
<i>dak- -e</i>	<i>dak- -(k)ne</i>	<i>dak- -ane</i>

Example sentences for imperative/subjunctive are given in sentences (8)–(11) and for desiderative forms in sentences (12)–(14).

Imperatives

- (8) **Actor focus**
 (*Gaweane wis meh rampung.*) *Mangana dhisik.*
 work(definite) already almost finished eat first
 ‘The work is almost done. Eat first.’
- (9) **Patient focus**
 (*Peleme wis mateng.*) *Panganen.*
 mango(definite) already ripe eat
 ‘The mango(s) is(are) ripe. Eat it(them).’
- (10) **Benefactive focus**
 (*Kuwi lho. Korane wis teka.*) *Jupukna.*
 that (exclamatory newspaper.(definite) already arrive get
 particle)
 ‘There! The newspaper has come. Get it for me.’
- (11) **Locative focus**
 (*Adhikmu kuwi kudu ngerti.*) *Kandanana.*
 younger.sibling.your that must know tell
 ‘Your little brother has to know (i.e. understand). Tell (him).’

Desideratives

- (12) **Patient focus**
 (*Peleme wis mateng.*) *Dak-pangane.*
 Mango(definite) already ripe by.me-eat
 ‘The mango(s) is(are) ripe. (I think) I’ll eat (it/them).’
- (13) **Benefactive focus**
 (*Kuwi lho! Korane wis teka.*) *Dak-jupukne.*
 that (exclamatory newspaper already arrive by.me-get
 particle) (definite)
 ‘There! The newspaper has come. I’ll get (it for you).’
- (14) **Locative focus**
 (*Adhikmu kuwi kudu ngerti.*) *Dak-kandanane.*²
 younger.sibling.your that must know by.me-tell
 ‘Your little brother has to know (i.e. understand). I’ll tell (him).’

² The locative-desiderative form tends not to be used much by the present younger generation who substitute the ordinary indicative form (*dak-kandani*) though desiderative meaning is intended.

Though the terms 'patient-focus', 'benefactive-focus' and 'locative-focus' have been used above, it is true for Javanese as it is for Philippine languages that the actual semantic role of the focussed-on entity depends on the semantic structure of the verb. In particular, forms with *-ake* often have a 'conveyed-entity-focus'. The entity conveyed may or may not be an instrument but to interpret the form as 'instrumental' is misleading because only instruments which are conveyed away from the actor towards a goal can be focussed on with this suffix.

Though the *-i* suffix is usually locational or directional in some sense, the actual nature of the 'focussed-on' item is variable. In many cases, if there is a possible interpretation, more than one 'meaning' of either suffix can occur with a given verb root. Finally, even intransitive roots can be made transitive with the addition of one or both of these suffixes. When this happens, the semantic characteristics of the suffixes which have been outlined above are usually retained. Examples of some of these possibilities are given in sentences (15)–(21). (Some examples are active, some passive, as appropriate depending on the context.)

nulis – 'write'

(15) **Patient-focus:**

SUBJECT/ PATIENT	VERB (PASSIVE)	ACTOR	
<i>Karangan iki</i>	<i>ditulis</i>	<i>Ani</i> .	
composition this	be.written	(name)	
'This composition was written by Ani.'			

(16) **Benefactive-focus:**

SUBJECT/ BENEFACTEE	VERB (PASSIVE- BENEFACTIVE)	ACTOR	PATIENT	
<i>Adhike</i>	<i>ditulisake</i>	<i>Ani</i>	<i>layang dinggo</i>	<i>bapak</i> . ³
little.sibling.his/her	be.written.for	(name)	letter for	father
Literally: 'Her little brother (or sister) was written a letter for Father by Ani.'				
Meaning: 'Ani wrote a letter to Father for her little brother (or sister).'				

(17) **Conveyed-entity-focus:**

SUBJECT/ CONVEYED ENTITY	VERB (PASSIVE- CONVEYANCE)	ACTOR	GOAL
<i>Potlote</i>	<i>ditulisake</i>	<i>Ani</i>	<i>menyang kertas</i> .
pencil(definite)	be.applied.for.writing	(name)	to paper
'The pencil was (applied) to the paper by Ani (in order to) write.'			

³ The current younger generation especially, who use Javanese less and less for functions other than colloquial interpersonal social communication, tend not to construct sentences with more than two full noun phrases. They would thus, in order to convey the meaning of this sentence, collapse two of the noun phrases into a genitive construction: *Layange adhik sing dinggo bapak ditulis(a)ke Ani*, 'Little brother's letter for Father was written by Ani.' The example in the text of this paper (*Adhike ditulisake Ani layang dinggo Bapak*), which comes from data collected in the early 1970's from an informant who was at that time in his late 30's, is, according to a young informant currently in his 20's, interpretable; however, this young informant confessed that he would never say such a thing.

(18) **Locative-focus (human goal):**

SUBJECT/ HUMAN GOAL	VERB (PASSIVE-LOCATIVE)	OBJECT/ PATIENT	ACTOR	
<i>Tono</i>	<i>ditulisi</i>	<i>layang</i>	<i>dening</i>	<i>Ani.</i>
(name)	be.written.to	letter	by	(name)

Literally: 'Tono was written a letter by Ani.'

Meaning: 'Ani wrote a letter to Tono.'

(19) **Locative-focus (inanimate goal):**

SUBJECT/ INANIMATE GOAL	VERB (PASSIVE-LOCATIVE)		ACTOR	
<i>Kertas</i>	<i>kuwi</i>	<i>ditulisi</i>	<i>Ani.</i>	
paper	that	be.written.on	(name)	

'That paper was written on by Ani.'

туру – 'sleep'(20) **Conveyed-entity-focus:**

SUBJECT/ACTOR	VERB (ACTIVE- CONVEYANCE)	OBJECT/CONVEYED ENTITY	
---------------	------------------------------	---------------------------	--

Mbok Marta nurokake anake.

Mrs (name) put.to.sleep child.her

'Mrs Marta put her child to sleep.'

(21) **Locative-focus (inanimate goal):**

SUBJECT/LOCATION	VERB (PASSIVE- LOCATIVE)	ACTOR	
------------------	-----------------------------	-------	--

Kasur sing anyar kuwi dituroni Ani.

mattress which new that be.slept.on (name)

'The new mattress was slept on by Ani.'

2.2 Noun phrase marking

While Philippine languages indicate the focussed-on item with a special particle (*ang* in Tagalog), Javanese indicates focussed-on status of an item, whether that item is the grammatical subject or object, by lack of any prepositional marking. In the examples above the grammatical subject occupies initial position. The position of the subject is in fact variable, as will be explained in the following section. However, initial position for subject is the unmarked order in modern Javanese. The grammatical object, which is perhaps not a very appropriate term, as we will see shortly, usually follows the verb immediately and is also unmarked. In fact, the verb plus grammatical object normally form a unit, the elements of which cannot be moved in relation to each other. The only exceptions to this rule are found in sentences which have undergone 'double topicalisation', such as sentence (5). In this sentence the object is fronted to form a primary topic and the verb is nominalised to form a secondary topic, the subject of an equational sentence. The underlying subject becomes the predicate of the equational sentence.

Lack of marking on grammatical subject and object, with prepositions indicating the semantic role of other noun phrases, again, sounds deceptively like English and other

European languages. Javanese, however, differs from English and other European languages in that, whether a sentence is active or passive, it can have both a grammatical subject and a (not very appropriately named) grammatical object. That is, an entity which is unmarked prepositionally, usually follows the verb and, when in that position, together with the verb forms a unit. Since this latter entity can have almost any semantic role, perhaps a better term for it is 'verbal complement'. In passive sentences the verbal complement is frequently the agent but not necessarily. An example of a passive sentence cited above which contains such a verbal complement, which is not the agent, is sentence (18); where *Tono*, the goal of the action, is the grammatical subject and *layang* 'letter', the patient of the action, is the verbal complement.

The semantic role of any noun phrase other than the grammatical subject and verbal complement is marked by a preposition, except in the case of a patient or conveyed entity (which is not an instrument). This semantic role (patient or non-instrumental conveyed entity) is unmarked whether or not the phrase is focussed on as subject or verbal complement. The prepositions indicating the major non-focussed roles include the following:

<i>menyang</i>	– destination (inanimate)
<i>marang</i>	– destination (animate, usually human)
<i>neng</i>	– location
<i>dening</i>	– agent
<i>dinggo/kanggo</i>	– benefactee
<i>nganggo</i>	– instrument

2.3 Word Order

SV(O), as mentioned above, is the neutral or unmarked order in modern Javanese. The position of subject, however, as is true of any phrase having a non-focussed role, is variable. Most frequently, if the subject does not precede the verb, it occurs in clause-final position. As has been described elsewhere (Poedjosoedarmo 1977, 1986a, 1986b), the sequencing of phrases combines with placement of particular intonation contours to indicate the information status of each item in the sequence. Briefly, grammatical phrases in Javanese each constitute at least potentially distinct information units. Each information unit is marked by a particular intonation contour. There are three possible contours in modern Javanese: rising, falling (or falling-rising) and flat.⁴ If all types occur, they must be sequenced in this order. There may be more than one rising tone unit or none at all and more than one flat tone unit or none at all but every utterance must have one and only one falling tone unit.

As mentioned, these are information units. Each intonational contour signals a particular status of the phrase as an item of information. There are in fact four levels of importance of information. The newest and most important information is signalled by a falling tone. The second most important level is indicated by a rising tone. Relatively unimportant information

⁴ The actual contours are variable depending on many factors, such as whether the utterance is a statement or a question and other aspects of speaker intent or attitude. In Poedjosoedarmo (1977, 1986b), I use the terms *anticipatory* for the 'rising' tone unit, *focal* for the 'falling' (or falling-rising) tone unit, and *supplementary* for the 'flat' tone unit. The 'rising' one is nearly always actually rising and the 'flat' one is nearly always actually flat, though it can be at various relative pitches, depending largely on the actual contour of the focal unit.

is signalled by a flat tone. Completely recoverable information is normally indicated by ellipsis. Information structure and other elements of discourse are beyond the main topic of this article. I mention this analysis here because ranking of each phrase in terms of its importance as an information unit affects word order. It is also important because in modern colloquial Javanese (A)VS order (where A is an adverbial phrase) can have two quite different structures in terms of the information status of each phrase. Example (22) might occur in an orally told story.

- (22) *Dumadaan keprungu suarane macan.*
 ‘Suddenly (there) was heard the voice of a tiger.’

The phrase *suarane macan* ‘the voice of a tiger’, which is subject and occurs in final position, would be uttered with a falling tone, marking it as the most important bit of information in the utterance. *Dumadaan* ‘suddenly’ and *keprungu* ‘was heard’, an adverbial phrase and the verb, would each be uttered with a rising tone, marking them as important but not the focus of information.

Another example, sentence (23), with the constituent sequence (A)VS, might have a quite different structure in terms of the status of each element as an information unit:

- (23) *Banjur lunga wonge.*
 Literally: ‘Then left, the man.’
 Meaning: ‘Then, the man left.’

In this utterance, with the same sequence of constituent types as the first, the subject noun phrase *wonge* ‘the man’ is an established topic and relatively unimportant in terms of its information status. It would be uttered with a flat tone. The adverbial *Banjur* ‘Then’ would be uttered with a rising tone and, the most important information in the utterance, *lunga* ‘left’, would be uttered with a falling tone.

Except for the restrictions on the positioning of the ‘verbal complement’, the order of phrases other than the grammatical subject is also quite free. An example of a sentence in which the placement of every noun phrase is non-neutral or ‘marked’ occurs in sentence (24).

- (24) *Dening pak Kerta kuwi, nganggo watu gedhe, dibalang asune.*
 Literally: ‘By Mr Kerta, using a large stone, was hit/thrown at the dog.’
 Meaning: ‘The dog was hit with a large stone which Mr Kerta threw at it.’

3 Old Javanese

3.1 The language

The earliest inscription in the Javanese language is the Sukabumi Charter, which is dated 25 March 804 AD. There are earlier inscriptions which have been found on the island of Java but these are in Sanskrit. It is probable that Sanskrit was the principal language of literature on the island before the ninth century. A sixth century Chinese work, the Kao Seng Chuan, mentions a prince from Kashmir who came to Java and propagated Buddhist doctrine at the beginning of the fifth century (Zoetmulder 1974:6–11). This suggests that Sanskrit was the language of religion and literature on the island for nearly four centuries. It is thus not surprising that the Old Javanese language which has been preserved in both texts and inscriptions contains a large proportion of Sanskrit vocabulary. Juynboll’s *Woordenlijst*, cited

by Gonda (1952), lists 6790 Sanskrit words and 6925 indigenous lexical items occurring in Old Javanese. In other words, by this count, nearly half the vocabulary of Old Javanese was of Sanskrit origin. Gonda himself more conservatively estimates that in Old Javanese poetry in Indian metres (*kakawin*) 25–30% of the words are Sanskrit.

According to Zoetmulder, though the number of Sanskrit borrowings in Old Javanese was great, the words borrowed were mostly nouns and adjectives and they were borrowed, almost without exception, in their undeclined form. Zoetmulder (1974:11) feels that the grammar of Old Javanese was not affected by the massive importation of Sanskrit vocabulary. This is not to say that the Old Javanese language preserved in inscriptions and manuscripts is a close reflection of the spoken language in Java in the ninth century. Sanskrit, the only Indian language to have influenced Javanese, was spoken colloquially nowhere in India during the first millennium AD. It was, however, the language of science, literature and religion in most of India at that time. Similarly, classical Old Javanese, called Kawi, became the language of science, literature and religion in Java and the language of the Sukabumi Charter remained little changed as a literary language throughout a period which extends from the ninth to the fifteenth century (Zoetmulder 1974:7). It is the features of this literary language which are described here. The description is based on Zoetmulder and Poedjawijatna's description in *Bahasa Parwa* (1954, reprinted 1993).

3.2 Verb morphology

Though the Javanese language in its colloquial form has surely changed over the past millennium and though we have evidence that many features of the literary language have changed, it is quite amazing that the basic semantic categories represented by the verb morphology appear to have remained quite stable. Old Javanese, like modern Javanese, had both a 2-way voice distinction between active and passive and a three-way focus distinction cutting across it, for which the basic semantic categories which could be focussed on were (in addition to actor in all the active forms) patient, conveyed object and location or goal. With the addition of a prefix, to be described below, the form indicating conveyed object could also take on benefactive meaning. The forms are not exactly identical to those of modern Javanese. In some cases suppletion appears to have occurred. In others, one of two competing forms has been lost or fossilised or a form has undergone phonological change. The forms of Old Javanese are shown in Table 5.

Table 5: The Old Javanese voice/focus system

	Patient-focus	Conveyed-object-focus	Locative-focus
Active voice	<i>-um-</i> , (<i>m</i>) <i>aN-</i>	<i>-um-</i> <i>-aken</i> , (<i>m</i>) <i>aN-</i> <i>-aken</i>	<i>-um-</i> <i>-i</i> , (<i>m</i>) <i>aN-</i> <i>-i</i>
Passive voice	<i>-in-</i>	<i>-in-</i> <i>-aken</i>	<i>-in-</i> <i>-an</i>

The infixes *-um-* and *-in-* only occur in fossilised forms in modern Javanese. (*m*)*aN-* has been reduced to prenasalisation. *-aken* still occurs in the Krama or polite speech level form but in Ngoko, the unmarked level, it has been reduced to *-ake*. The differentiation in form between the locative suffix for active (*-i*) and passive (*-an*) has been neutralised, *-i* now being used for both.

Examples of some Old Javanese sentences illustrating the various verb forms (taken from Zoetmulder & Poedjawijatna 1954) are given in (25)–(30).

(25) Actor as subject/patient as object

	SUBJECT/ ACTOR	VERB ACTIVE/ PATIENT-FOCUS	OBJECT/ PATIENT	
	<i>Tan dadi ri sīṣya</i>	<i>mangan</i>	<i>drawya ning guru.</i>	
	not fitting to student	eat	that.owned by teacher	
	'It is not fitting for a student to eat that which belongs to a/his teacher.'			

(26) Patient as subject

	SUBJECT/PATIENT	VERB PASSIVE/ PATIENT-FOCUS	
	<i>Ikang nāga</i>	<i>pinangan ing apuy...</i>	
	that dragon	eaten in fire	
	'The/(that) dragon (which was) eaten by fire . . .'		

(27) Actor as subject/conveyed-entity as object

	VERB ACTIVE/ CONVEYED-ENTITY-FOCUS		SUBJECT/ACTOR		OBJECT/ CONVEYED- ENTITY
	<i>Umarpanakĕn</i>	<i>ta</i>	<i>mahārāja</i>	<i>Janamejaya</i>	<i>lĕmbu</i>
	give	post-verbal particle	king	(name)	cow
		LOCATIVE/ HUMAN GOAL			
	<i>satus</i>	<i>ri sang</i>	<i>Brahmana.</i>		
	one.hundred	o particle.of.respect	Brahman(s)		
	'King Janamejaya gave one hundred cows to the Brahman(s).'				

(28) Conveyed-entity as subject

	VERB PASSIVE/ CONVEYED ENTITY-FOCUS		SUBJECT/ CONVEYED ENTITY		LOCATIVE/HUMAN GOAL
	<i>Inarpanakĕn</i>	<i>ta</i>	<i>lĕmbu satus</i>	<i>ri sang</i>	<i>Brahmana.</i>
	be.given	post-verbal particle	cow one.hundred	to particle.of.respect	Brahmans
	'One hundred cows were given to the Brahman(s).'				

(29) Actor as subject/locative (human goal) as object

		SUBJECT/ACTOR		VERB ACTIVE/ LOCATIVE FOCUS	OBJECT/ LOCATIVE (HUMAN GOAL)
	<i>Mahyun</i>	<i>ta</i>	<i>mahārāja</i>	<i>Janamejaya</i>	<i>umarpane</i>
	want	post- verbal particle	king	(name)	give
					<i>sang</i>
					particle
					Brahman(s)
					of
					respect

CONVEYED ENTITY

lěmbu satus.

cow one.hundred

'King Janamejaya wants to give the Brahman(s) one hundred cows.'

(30) Locative (human goal) as subject

VERB	SUBJECT/LOCATIVE	CONVEYED ENTITY
PASSIVE/LOCATIVE FOCUS	(HUMAN GOAL)	

<i>Inarpanan ta</i>	<i>sang</i>	<i>Brahmana</i>	<i>lěmbu satus</i>
given	post-verbal	particle.of.	Brahman(s) cow one.hundred
	particle	respect	

ACTOR

de mahārāja Janamejaya.

by king (name)

'The Brahman(s) was/were given one hundred cows by King Janamejaya.'

In addition to differences in form evident from Table 5, the formation of the passive in Old Javanese also differed from modern Javanese in that the proclitic pronoun forms indicating person of the agent shown in Table 1 did not exist in the *Parwa* literature which is the basis of Zoetmulder and Poedjawijatna's description. While modern Javanese has three sets of pronouns — independent forms, the proclitic forms used to indicate person of the agent of passive verbs shown in Table 1, and enclitic forms used in genitive constructions and following prepositions, Old Javanese had only two sets. Like Philippine languages, enclitic forms were used both in genitive constructions and to indicate person of the agent of passive verbs. These Old Javanese enclitic pronouns are shown in Table 6. Note that the two rows of third person pronouns, though cognate with forms indicating singular and plural respectively in other Austronesian languages, did not indicate number differences in Old Javanese. Similarly, the first person *-mami* has the independent form *kami*, which is first person plural exclusive in other Austronesian languages. Both first and second person have the forms *-ta* and *-nta*, which have the independent form *kita*. This a first person plural inclusive pronoun in other Austronesian languages. However, in Old Javanese, none of these forms had explicitly plural meaning. The forms *-mami*, *-ta*, *-nta*, *-ira*, *-nira* were used to refer to persons of status while *-ku*, *-ngku*, *-mu*, *-nyu*, *-ya*, *-nya* were unmarked for status. Unlike in modern Javanese, the third person enclitic forms were only used when no nominal reference to the agent occurred.

Table 6: Old Javanese post-cliticised pronouns indicating person of agent of passive verb

	Enclitic pronouns
First person	<i>-ku, -ngku;</i> <i>-mami; -ta, -nta</i>
Second person	<i>-mu, -nyu;</i> <i>-ta, -nta</i>
Third person	<i>-ya, -nya;</i> <i>-ira, -nira</i>

There are no examples from the data I analysed of the enclitics occurring immediately following a passive verb but *-nira* occurs as enclitic to a preposition in sentence (31).

- (31) *Salikur kweh ning ratu pĕjah de nira.*
 twenty-one quantity of king(s) die by him
 'Twenty-one kings died by his (hand).'

As mentioned above, the suffix *-aken* alone did not have benefactive meaning in Old Javanese but in combination with a prefix *pa-* it did have this meaning. The active form of the prefix *pa-* was prenasalised, producing *ma-*; the passive form contained the infix *-in-*, producing a prefix *pina-*. Examples of these forms are given in sentences (32)–(33).

(32) **Actor as subject/benefactee as object**

SUBJECT/ACTOR	VERB
	ACTIVE/BENEFACTIVE FOCUS
<i>Mangkana ling bhagawān Waiśampāyana,</i>	<i>macaritākĕn</i>
thus said (title) (name)	tell a story
OBJECT/BENEFACTEE	
<i>mahārāja Janamejaya.</i>	
king (name)	
'Thus said Bhagawan Waisampayana, (who then) told a story to/for King Janamejaya.'	

(33) **Benefactee as subject**

VERB	SUBJECT
PASSIVE/ BENEFACTIVE	BENEFACTEE
<i>Mamalaku pinājarakĕn i sang Kuntī sira.</i>	
be.asked.to.do.sth be.spoken.for to title.of. name he	
(on.one's.behalf) respect	
'He asked that he be spoken for concerning that matter to Kuntī.'	

Also, in addition to the passive forms with *-in-*, Old Javanese had a second set of passive forms with the prefix *ka-*. Where locational meaning was involved, a suffix *-an* also often occurred. These forms often had an accidental connotation, or described the result of an event without reference to the agent, or described ability to do something. Examples are *katon* 'able to be seen'; *katĕkan* 'be hit by something, have something befall one'. A probably historically related form *maka-* also described ability. This latter form also had a passive counterpart *pinaka-*. Examples of these affixes occur in *makawāhana* 'have as one's vehicle' and *pinakaśiṣya* 'happen to be the student of (someone)'.¹

The prefix *ka-/ke-*, with or without an accompanying suffix *-an*, still occurs in Modern Javanese but is probably not as productive as it was in Old Javanese. The prefix *maka-* occurs only in fossilised forms and *pinaka-* no longer occurs outside of preserved literature.

3.3 Noun phrase marking

Though Old Javanese did not have obligatory noun phrase marking as Philippine languages do, there was a much more highly developed system of marking noun phrase functions than what remains in modern Javanese. A particle occurring variously as *ng*, *ang*, *ing* often marked a subject, as in examples (34) and (35).

- (41) *Ikang wwang yan manapak sabhântara ndātan hana*
 that man if attend.meeting although not exist
umaritrāna ya, kewala tĕka ri kĀwakanya, . . .
 escort him but come by=at himself
yeku tan wruh ring lokasthiti.
 that not know to custom/manners
 'If a man attends a meeting and no one has brought him there but he comes
 of his own accord, that man does not know good manners.'

In addition to noun phrase marking particles which indicated the role of the following phrase, Old Javanese had a rather large number of particles which marked references to humans, regardless of the grammatical role of the phrase. These particles indicated something about social rank. Ordered here from lowest to highest, they were *si*, *pun*, *sang*, *sang hyang* and *ḍang hyang*.

3.4 Word order

Word order in Old Javanese in independent clauses was almost without exception VS(O). Of the examples given above having SV order, the verb phrase following the noun constitutes a dependent clause in all cases. A main verb, that is a verb in an independent clause, was usually followed by the particle *ta*. Some examples given above containing dependent clauses with a subject preceding the verb are found in sentences (25), (26), (29) and (32). A possible exception occurs in sentence (31). However, the actual meaning of this is 'Twenty-one is the number of kings who died by his hand'. The verb which follows its subject is thus also in a dependent clause.

Though already mentioned, it should probably be stressed that pronominal agents of passives were enclitics in the *parwa* literature rather than proclitics as in modern Javanese. Also, though not many verbal particles appear in the data presented here, adverbial particles such as *ta* also followed the verb in Old Javanese rather than preceding it as many adverbial elements do in modern Javanese.

4 Discussion, hypotheses and implications

As seen from the preceding presentation, the Javanese language has over the past millennium maintained almost unchanged the basic characteristics of its verb morphology: that is, the form of any transitive verb simultaneously indicates a two-way voice distinction between active and passive and a (maximally) three-way focus distinction between (neutral) patient, conveyed entity/benefactive, and locative. What has changed over this long period of time is (1) noun phrase marking and (2) word order. These have changed from a system which very much resembled Philippine languages to the modern Javanese one which, at least superficially, is reminiscent of European languages.

If we assume, as most linguists working in this field who have speculated on the topic do (e.g. Wolff 1973, 1980, 1996), that the Philippine languages are the most conservative western Austronesian languages and most closely reflect the morphological and syntactic systems of Proto Malayo-Polynesian, then what changes led from this Philippine-like system

to the Old Javanese one? And an even more frustrating question since, having at our disposal nearly 1500 years worth of records, we feel that we should know the answer to it: how did the Old Javanese system evolve into the modern Javanese one? As explained in the introduction to the description of Old Javanese here, since the literary language remained essentially unchanged for nearly a millennium, what we have in the records are samples of the beginning and end of a path with little evidence of what happened to the language in between.

In the remainder of this paper I will speculate and hypothesise — for we can do nothing more — concerning possible sequences of changes and causes of those changes which might have led (1) from a Philippine-type system to that of Old Javanese and (2) from Old Javanese to modern Javanese.

4.1 A possible path from a Philippine-type system to Old Javanese

The major difference between a Philippine-type system and that of Old Javanese is the presence in the latter of an active vs. passive voice distinction in addition to the focus system. Another perhaps less significant difference is the non-ubiquitousness of noun phrase marking in Old Javanese, including the fact that human referents are marked with particles indicating social status rather than grammatical or semantic role.

In attempting to discover a possible cause for the changes which took place, we must not underestimate the importance of the fact that Sanskrit was the language of literature, science and religion in Java for probably at least four hundred years before Old Javanese began to be used for these purposes. Zoetmulder marvelled that, despite the enormous influx of Sanskrit vocabulary, Old Javanese remained essentially 'Indonesian' (i.e. Austronesian) in character. This is true, but we know from studies of contemporary and better-documented historical contact situations that transfer tends to occur more frequently on the discourse level than on the clause or sentence level (Odlin 1989; Gass & Selinker 1983). We know also that contact between colloquial languages can differ in its effects from the influence of one literary language on another evolving one. In colloquial contact situations from which pidgins and creoles tend to evolve, users of the emerging contact language extract vocabulary from the foreign source but give it structure and meaning inherent in their own first language. Where a foreign literary language has been used for a long time and a local language then begins to be used for literary purposes, almost the opposite can happen: that is, speakers of the local language redefine forms in their own language to express concepts inherent in the foreign language which has become familiar to them for literary purposes (Thomason & Kaufman 1988). This is particularly likely to happen if massive translation occurs. Baker (1992) refers to forms of a language which evolved due to massive translation as 'translatese'. That the *parwa* literature, if not translated in its entirety, at least followed closely the Sanskrit originals, is, according to Zoetmulder (1974:68), a matter about which there can be no doubt. To quote him, "The *parwas* are adaptations in prose of parts of the Sanskrit epics and show their immediate dependence by Sanskrit quotations throughout the text."

When Javanese began to be used for the purpose of writing literature which had previously existed only in Sanskrit, it is quite possible that an attempt was made to express grammatical concepts felt to be important in the Indian language, including the distinction between active and passive and the concept of definiteness. In the existing Austronesian focus system, the form for agent focus differed from the forms for the other focuses in having a prefix ending

in a nasal or the infix *-um-* instead of an infix *-in-*. Sanskrit scholars, whether or not they were first language Javanese speakers, might have interpreted this chance formal distinction as a means of conveying the active vs. passive concept and applied the newly identified active marker to all verb bases, including ones containing suffixation to indicate non-patient focus.

An attempt to express definiteness vs. indefiniteness might similarly have resulted in reinterpreting the role-marking particles as markers of definiteness, resulting in their deletion when indefiniteness was intended.⁵ Finally, personal noun phrase markers which had formerly carried role information, such as *si*, were reinterpreted in the construction of a system for indicating social rank, an important concept in Indian culture.

All of this is, of course, pure speculation. However, some such sequence of innovations on the part of Javanese scholars attempting to use their native language to convey the content of Sanskrit literature could have occurred.

4.2 From Old Javanese to Modern Javanese: what lies in between?

Reconstructing the changes which led from Old Javanese to modern Javanese is, in a way, more difficult than guessing at the pre-history of the language, and the difficulties inherent in the task are far more frustrating since we appear to have a continuous record of written evidence. However, since Old Javanese as a literary language, like the Sanskrit language that its literary works must originally have been translated from, was preserved in its original literary form nearly unchanged for many centuries, we have no record of the language that was actually spoken. Our written records therefore jump from this form which is more than a millennium old through a few phases with minimal innovations to a modern literary language which, in grammar, differs little from the contemporary spoken variety.

Becker (1979) has hypothesised that word order changes in the Malayo-Javanic group of Austronesian were due to influence from European languages, beginning with Sanskrit and other Indian languages spoken colloquially in the archipelago during the first millenium AD and later including Portuguese, Dutch and English. However, another explanation for the word order changes seems to me more plausible. Lehmann (1973) and others have pointed out that one syntactic change in a language often triggers another and that certain constellations of patterns tend to cooccur. Simplifying and generalising the essence of these claims, it appears that there may be a tendency in language for elements which have strong syntactic links to the verb to be located next to the verb. The focus systems of Philippine languages and, presumably of pre-Javanese, allowed for only one noun phrase with strong syntactic links to the verb: the focussed element. However, with the rise of literary Old Javanese and the concomitant superimposition of a voice system on top of the native focus system, there arose the possibility of having two noun phrases (which we are calling for lack of better terms 'subject' and 'object') with close syntactic links to the verb. As long as these were marked with identifying particles there was no problem in interpretation but as the use

⁵ As mentioned above, *ng*, *ang*, *ing* often marked the grammatical subject, but even at this stage of Javanese language history, the exact role-related meaning of many of the particles was beginning to deteriorate. This tendency to collapse meanings of particles has continued into the present. In modern colloquial Javanese, the particle *karo* can have at least the following meanings:

'with'	<i>Aku neng pasar karo ibu.</i>	'I (went) to the market with mother.'
'to'	<i>Aku kanda karo bapak.</i>	'I told (it) to father.'
'by'	<i>Kuwi digarap karo adhikmu.</i>	'That was done by your little brother (or sister).'

of the role marking particles declined, problems in interpretation may have arisen. The solution was to use a syntactic pattern found in dependent clauses, SVO order. This additionally by chance made sense because the affixation indicating the underlying role of the subject (actor or other) was located at or near the front of the verb while the morphology indicating the underlying role of the object of active verbs came at the end of the verb.

This, again, is pure speculation but evidence that the hypothesis might be correct comes from Classical Malay. Malay is, of course, a different language from Javanese but a closely related one and one which, during the past two millennia, has gone through a series of syntactic changes similar, though not identical, to those affecting Javanese. Based on our knowledge of the history of the archipelago and of the many lexical borrowings which occurred between Javanese and Malay over the centuries, we can postulate that there had always been a sizable number of bilinguals in these two languages and that the two languages continuously influenced each other, both in their spoken and literary forms.

Classical Malay dates from a period beginning just before the last of our Old Javanese manuscripts and continuing for several centuries. Like Old Javanese, the literary form of the language changed little during these years: the language of *hikayat* composed in the nineteenth century shows little difference from ones written in the fifteenth century. We might thus regard Classical Malay as a language containing grammatical features which might also have been part of an intermediate stage of spoken Javanese. In Classical Malay, though word order was variable, the most common word order for intransitive and passive sentences was VS. For active transitive sentences, however, the most common pattern was SVO. This agrees with the hypothesis presented above about the syntactic changes which took place in Javanese. It was the necessity of showing a close link between the verb and two noun phrases associated with it which initially prompted the word order change.

In Javanese, the adoption of the new order for passive sentences was probably related to loss of obligatory marking of the agent and associated with the option of having a non-agent (such as patient of a passive benefactive verb) follow the verb. It was also probably in association with these developments that the modern set of proclitic agent markers on the passive verb developed. The end result was an order which was the mirror image of Old Javanese and Philippine languages: from Passive-Verb + Enclitic-Agent + Subject-of-Passive, the eventual pattern to emerge was Subject-of-Passive + Proclitic-Agent + Passive-Verb.

The dominance of these gradually evolving patterns in Malay, as Cumming (1988) has suggested, resulted from a gradual shift in function of the possible word orders and, in the colloquial language, the development of a complex interplay between intonation and order to mark these functions. A parallel development almost certainly occurred in Javanese.

One final note to the hypothesis has to do with the proclitic agent pronouns in Javanese.⁶ The second person form *kok-* is the most transparent, probably deriving from the first syllable of *kowe* which is probably cognate with Malay *kau*. The epenthesised final glottal stop in the proclitic is a frequent sporadic innovation in Javanese phonology. The third person proclitic agent pronoun *di-* is more problematic. It has been suggested in the case of Malay that the *di-*

⁶ As noted, there are no proclitic pronouns in the *parwa* literature (tenth century). I am grateful to Erik Zobel for pointing out to me that proclitic pronouns did occur in the Ramayana *kakawin*. The word order pattern was thus introduced fairly early (though not before the development of voice as distinct from focus. However, the forms of the proclitic pronouns found in the Ramayana are not those of modern Javanese. They are: *k-*, *m-*, *n-*, *t-*, *kam-*, *r-*. It is also not clear if the verbs with which these forms occurred were passive. Kern (1898) refers to them as *de kortste vorm* or 'the short form' of the pronouns.

prefix came from *dia*, but in Javanese by the Old Javanese period, before proclitic agents had evolved, sequences of two vowels had merged into single intermediate vowels. Thus /i/ + /a/ ⇒ /e/ and *dia*, if it had occurred, would have become **de*. If the verbal prefix *di-* cannot be derived from a Javanese pronoun, another possible source is borrowing from Malay. As mentioned earlier, the two languages were in more or less constant contact and there was a great deal of translation and borrowing between them during all literary periods. It should be noted that the *kidungs*, literature of the late Majapahit period, contain the passive prefix *depun*, composed of *de* (genitive marker) + *pun* (honorific), both being elements found in the *parwa* language. The modern Javanese Krama of *di-* is *dipun-*. A suggested explanation might be that *di-* as a passive marker was a borrowing from Malay and that *depun*, 'contaminated' by *di-* became *dipun-*.

Finally, the most perplexing form of all the proclitic agent forms is the first person *dak-*. The following hypothesis is pure conjecture but, to my knowledge, no more plausible one has been proposed. The hypothesis is that *dak-* may be cognate with Malay *hendak* 'wish, will' and that it came to be used as a proclitic form first in the desiderative. Though its original meaning was not pronominal, as the agent of the desiderative is always first person, it took on first person meaning and was later generalised and used with the indicative passive forms.

5 Final words

Wolff has suggested a rather different sequence of changes from those proposed here as the ones leading to proclitic agent of the passive in other western Austronesian languages. Specifically, he suggests that in some of these languages the fronting of the cliticised agent pronoun occurred first before the development of the dual voice/focus system in the verb morphology (my terminology). This clearly was not the case in Javanese but I do not mean to imply that changes couldn't have happened in that order in other languages. In language change as in other processes, there are often multiple paths to a final destination. What is more important than the order of the changes is perhaps that if one of the changes discussed here occurred, the others were very likely to follow. Either a Philippine-like system or a Malayo-Javanic-like system appears to be relatively stable; intermediate systems may be more likely to change. If this is so, it would explain the fact that similar sets of changes appear to have taken place independently in different locations, not necessarily for the same reasons or in the same order.

A sequence of changes similar to those outlined in this paper may have led from a Philippine language-like focus system to the voice/focus system of Old Javanese, and a sequence of changes leading from Old to modern Javanese may have been motivated by factors similar to those suggested here. However, if we are really to understand each of the systems we are describing, whether they form historical chains as in the case of Old and modern Javanese, or contemporary variants (such as the many existing varieties of modern Malay/Indonesian), it is not enough to simply describe the forms. We must understand what it means for a given noun phrase to be focussed on or, in a voice/focus system, to be selected as subject or object. Under what circumstances is each form likely to occur? In addition, in all of these languages, though a single word order pattern usually dominates, there always appear to be alternative possibilities. What does it mean to select one pattern rather than another? Under what circumstances is each pattern most likely to occur?

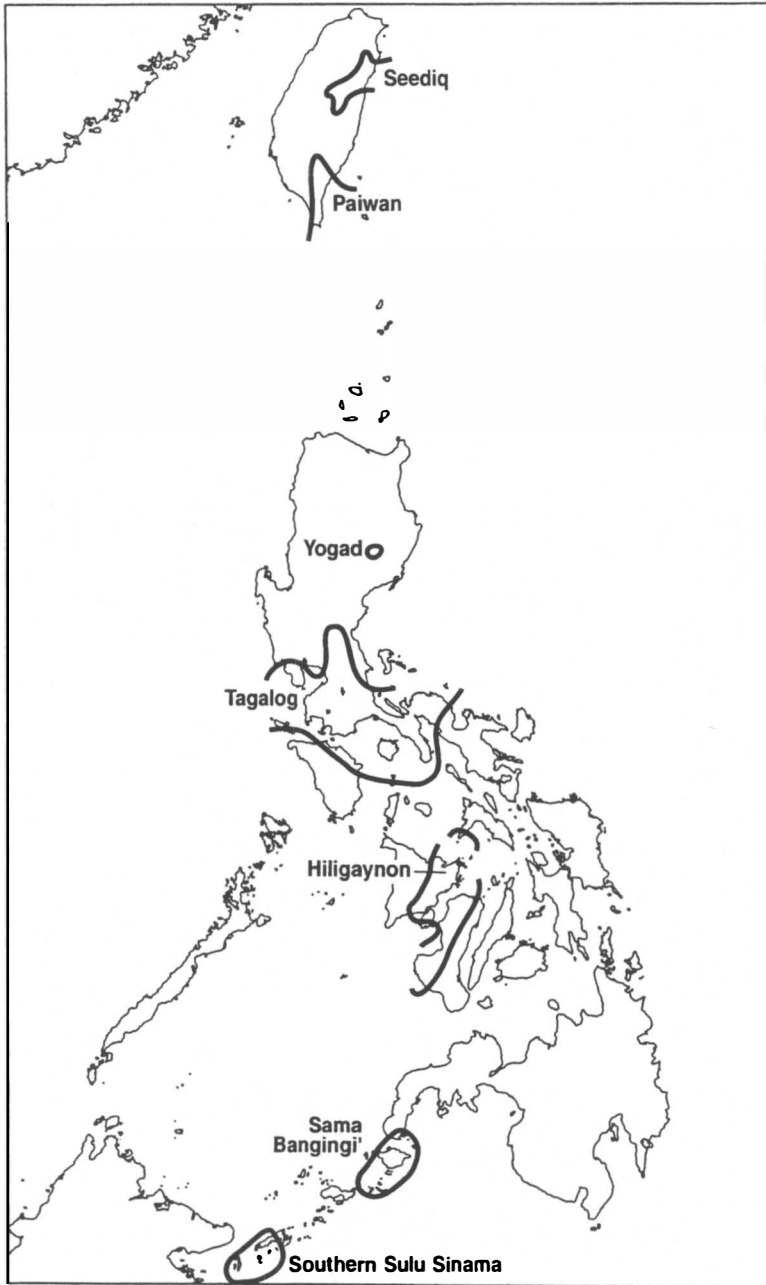
It is only when these questions have been answered that we will truly understand how the languages we are describing actually work and why they change when they do.

References

- Baker, M., 1992, *In other words: a coursebook on translation*. London: Routledge.
- Becker, A.L., 1979, The figure a sentence makes: an interpretation of a classical Malay sentence. In T. Givón, ed. *Syntax and semantics*, vol.12: *Discourse and syntax*, 243-259. New York: Academic Press.
- Cumming, Susanna A., 1988, Syntactic function and constituent order change in Malay. PhD dissertation, University of California at Los Angeles.
- Gass, S. and L. Selinker, 1983, Language transfer: a conference report. In F. Eppert, ed. *Transfer and translation in language learning and teaching*. Singapore: Singapore University Press for SEAMEO Regional Language Centre.
- Gonda, Jan, 1952, *Sanskrit in Indonesia*. Nagpur.
- Kern, Hendrik, 1898, Bijdragen tot de Spraakkunst van het Oudjavaansch. *Bijdragen tot de Taal-, land- en Volkenkunde*, 49:635-654.
- Lehmann, W.P., 1973, A structural principle of language and its implications. *Language*, 49:47-66.
- Odlin, T., 1989, *Language transfer: cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.
- Poedjosoedarmo, Gloria, 1977, Thematization and information structure in Javanese. *NUSA* 3:34-43.
- 1986a, *Role structure in Javanese*. *NUSA* 24.
- 1986b, Subject selection and subject shifting in Indonesian. *NUSA* 26:1-17.
- Thomason, Sarah Grey and Terrence S. Kaufman, 1988, *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Wolff, John U., 1973, Verbal inflection in Proto-Austronesian. In A. Gonzalez, ed. *Parangal kay Cecilio Lopez*, 71-91. Quezon City: Linguistic Society of the Philippines.
- 1980, Verbal morphology and verbal sentences in Proto-Austronesian. In P.B. Naylor, ed. *Austronesian studies: papers from the Second Eastern Conference on Austronesian Languages*, 153-167. Ann Arbor: University of Michigan, Center for South and Southeast Asian Studies.
- 1996, The development of the passive verb with pronominal prefix in western Austronesian languages. In Bernd Nothofer, ed. *Reconstruction, classification, description: festschrift in honor of Isidore Dyen*, 15-40. Hamurg: Abera.
- Zoetmulder, P.J., 1974, *Kalangwan: a survey of Old Javanese literature*. The Hague: Marinus Nijhoff.
- Zoetmulder, P.J. and I.R. Poedjawijatna, 1954, *Bahasa Parwa*. Djakarta: Obor.

— PART IV —

*Languages outside Indonesia
and Malaysia*



Map 4: Languages discussed in Part IV

The morphology and syntax of Seediq focus

ARTHUR HOLMER

1 Introduction

The Seediq language of Taiwan is spoken in the mountainous areas of Nantou County and in coastal areas of Hualien County, stretching from central Taiwan to the Pacific coast.¹ Seediq is an Atayalic language, and the Seediq tribe is commonly referred to as a subgroup of the Atayal people. Consequently, figures as to the number of speakers vary considerably, ranging from 5,000 (Chen 1992) to around 20,000 (Ferrell 1969). The dialect described in this paper is the Paran dialect, also referred to as the *Tgdaya* (literally 'uphill') dialect by speakers of other dialects. It is spoken in the Seediq-speaking areas of Nantou County, its original centre being the village of Paran (Wushe in Chinese).

2 The morphology of focus

Seediq has basically a four-focus system. The foci are the following: AF (Actor Focus, indicating that an actor is clause subject), PF (Patient Focus, indicating that a patient is clause subject), LF (Locative Focus, indicating that a location or a partially affected patient is clause subject) and IF (Instrument Focus, indicating that an instrument, a beneficiary, an object given or the patient of a causative is clause subject). These forms can in turn appear in various moods and tenses, as illustrated in Table 1.² The abbreviations in the left column refer to the following moods and tenses:³

¹ This paper is the result of fieldwork conducted in Taiwan in 1993, 1995 and 1998. I gratefully acknowledge the financial assistance rendered by the Swedish Council for Research in the Humanities and Social Sciences, the Lundberg Ido Foundation, and the Swedish Tercentenary Foundation, as well as practical assistance from the David C. Lam Institute for East-West Studies, Hong Kong Baptist University, and the Institute of History and Philology, Academia Sinica, Taipei. Naturally, I am most deeply indebted to my Seediq consultants, especially Ms Temi Nawi, of Puli, Taiwan. Any mistakes are mine and mine alone.

² Certain forms are rather uncommon and possibly no longer productive. I have, however, included all forms attested in my corpus. As concerns the forms given under IF, I am grateful to Lin Hsui-Hsu for

PRES stands for Present Tense, the unmarked form of the verb, also the one used if preceded by an auxiliary bearing temporal information. The reference of a PRES verb is not always present tense, however. With PF it has future reference in isolation, and with other non-AF foci it is tense-neutral in meaning, i.e. it can have present, past or future interpretation, depending on context.⁴

IMP stands for Imperative, the form used in commands, and also the form used immediately following the sentential negator *ini*.

PRET stands for Preterite Tense. This is the form of the verb used to refer to a punctual action at a time earlier than a certain reference point — note that it does not necessarily refer to a time earlier than the time of utterance, so it is an instance of relative tense rather than absolute tense. This category can only be used to refer to a completed event in the (relative) past, and never to indicate a progressive action in the past, nor a perfective view of a (relative) future event. Thus, this tense/aspect category necessarily combines the meaning of relative past tense with an aspectual meaning of perfectivity. Neither of these elements can be ignored.⁵

IMMED FUT and DIST FUT stand for Immediate Future and Distant Future respectively. These are only distinguished in AF, and only for some verbs — thus, verbs which form AF PRES in *m-* have no Immediate Future form. Immediate Future often has connotations of an action being scheduled and planned in advance, whereas the Distant Future has implications of an action taking place at some indefinite time in the future. The Seediq future appears to be an absolute rather than a relative tense. At any rate, I have not succeeded in eliciting examples where a future form is used in an absolute past context.

Note also that the forms listed as FUT for LF and PF are very specialised in meaning (see §2.2 and §2.3) and are certainly not prototypical future forms. They are listed as FUT simply to allow maximum symmetry in the table.

IRR stands for Irrealis, the verb form used for warnings or suggestions. Its meaning varies from something like 'Let us...' to 'Careful! X might happen!', depending on context. The most commonly seen form is PF IRR. AF IRR is very uncommon and is perhaps not productive. IRR is the category referred to in Holmer (1996) as Subjunctive.

The table above illustrates the various forms of the Seediq verb. The contemporary use of these focus forms is, however, not necessarily related simply to focus, but also has some connotations of aspect. Moreover, there are some other points which deserve comments. The details of usage of each focus will be presented in the following sections.

having brought certain facts to my attention which have led to a more complete picture. Moreover, while the exact status of some of the forms given in this table is not certain, the table includes major revisions of inaccuracies and mistakes present in Holmer (1996).

- 3 Another abbreviation given in the table is RED-, which stands for reduplication of the initial consonant of the root.
- 4 Such examples will be translated in this paper as English present tense so as not to have to choose a marked tense category for the translation.
- 5 Clearly, then, neither 'tense' nor 'aspect' is a particularly suitable term to describe this category. I choose the term 'tense' because there is a usage of focus itself which behaves more prototypically like aspect (see §2.2 and §2.3).

Table 1: The Seediq focus affixes

	AF	PF	LF	IF
PRES	-m-/m-	-un	-an	s-
IMP	∅	-i	-i	-ani
PRET	mn-/mn-	-n-	-n-an	sn- (?)
IMMEDFUT	m-			s-
		RED-un	RED-an	
DIST FUT	mp-			∅/p- (?)
IRR	m- -a	-o	-e	-ane/-ano

2.1 Actor focus

The distinctive AF morpheme is an *m*, appearing either as an infix (1a) or as a prefix (1b). It is not entirely predictable which verbs have an infix *-m-* [um] and which have a prefix *m-* [mu],⁶ although verbs with an active meaning tend to have the infix form and stative verbs tend to have the prefix form. There is even one transitive/unaccusative pair which differs only in this respect (1c,d). However, such examples are very rare (I have found no such other minimal pair), so it appears that the active/stative distinction is no longer productive in Seediq, although it still to a certain extent determines whether *-m-* is realised as a prefix or an infix with a given verb. Some verbs, such as *beebu* 'to beat', do not have an *-m-* in the AF form. Adjectives, which behave syntactically as verbs to a certain extent, do not have any *-m-* form either.

- (1)a. *q-m-alang* [qumálaN] (< *qalang*)
'to fence in'
- b. *m-sepi* [musépi] (< *sepi*)
'to dream'
- c. *t-m-utuy* [tumútuy] (< *tutuy*)
'to wake someone'
- d. *m-tutuy* [mutútuy] (< *tutuy*)
'to awaken'

The AF IMP form is given in Table 1 as a \emptyset -affix, i.e. as being identical to the verb root. This holds for both action verbs and process verbs (2a). However, adjectives have an imperative or negatable form beginning in *k-/q-* (2b,c).⁷

- (2)a. *Ini imah sino Lubi.*
NEG (AF.IMP)-drink wine Lubi
'Lubi doesn't drink wine.'

⁶ This paper follows the orthographic principles outlined in Li (1992). Pre-stress vowels are predictable and therefore not written. The orthographic pre-stress consonant clusters are resyllabified with an epenthetic vowel, which is realised as [u] unless separated from the stressed vowel by [ʔ] or [h], in which case it copies the stressed vowel. For more details see Yang (1976); Holmer (1996).

⁷ The choice of *k-* or *q-* depends on whether the root contains the phoneme *q*. If it does, the prefix is *q-*. If not, the prefix is *k-*.

- b. *Ini k-paru sapah -mu.*
 NEG IMP-big house 1SG.GEN.
 'My house is not big.'
- c. *K-paru hari p-n-atis -su!*
 IMP-big a.bit -PF.PRET-write⁸ 2SG
 'Write a bit bigger!' (lit. 'Let that which you write be a bit bigger!')

2.2 Patient focus

In isolation, PF PRES is interpreted as having a future meaning (3a). However, it may be combined with the past tense auxiliary *wada*, giving a past tense interpretation (3b). When PF is used, it always has perfective connotations, with the action being viewed as a whole rather than as an ongoing process.

- (3)a. *Mah-un -mu sino kiya.*
 drink-PF 1SG.GEN wine that
 'I'll drink up that wine.'
- b. *Wada -mu mah-un sino kiya.*
 PRET 1SG.GEN drink-PF wine that
 'I drank up that wine.'

PF FUT is given in Table 1 as reduplication of the initial consonant combined with the *-un* suffix. While this treatment allows a maximum symmetry within the table, it should be noted that this form emphasises the truth value of the proposition (4).

- (4)a. *Biq-un -su -mu.*
 give-PF 2SG.NOM 1SG.GEN.
 'I'll give it to you.'
- b. *B-biq-un -su -mu!*
 -FUT-give-PF 2SG.NOM 1SG.GEN
 'I will give it to you!'

PF PRET (formed by the *-n-* infix) normally refers to the product of an action (5a,b), and in many cases is primarily nominal in nature. Moreover, not all verbs have a PF PRET form (5c). Thus, the use of PF PRET is not directly parallel to the use of PF PRES with a past tense auxiliary (5d).⁹

- (5)a. *hlama t-n-ekan*
 rice.cake -PF.PRET-pound
 'pounded rice-cake'
- b. *t-n-abus*
 -PF.PRET-winnow
 'chaff'

⁸ -X- in a gloss (such as -PF.PRET-) indicates that the morpheme concerned is an infix.

⁹ Thus, the meaning of PF.PRET is in this case less obviously temporal than aspectual. However, given that *-n-* and *wada* cannot cooccur, there is motivation to treat them as instances of the same category.

- c. **N-imah -mu sino nii.*
 PF.PRET-drink 1SG.GEN wine this
 (intended reading: 'I have drunk up this wine.')
- d. *Wada -mu mah-un sino nii.*
 PRET 1SG.GEN drink-PF wine this
 'I drank up this wine.'

2.3 Locative focus

LF PRES is normally interpreted as a locative focus when used in isolation (6a). However, LF PRET does not necessarily have a locative connotations, although it may (6b). LF PRET is often used as a straight passive (6c), usually with a partitive or imperfective meaning, in contrast to the completive or perfective interpretation of PF with a preterite auxiliary (6d).

- (6)a. *Tkan-an -mu beras duhung nii.*
 pound-LF 1SG.GEN rice mortar this
 'I pound rice in this mortar.'
- b. *P-n-uq-an -mu damac pngerax kiya.*
 -PRET-eat-LF 1SG.GEN food plate that
 'I ate food from that bowl.'
- c. *P-n-uq-an -mu damac nii.*
 -PRET-eat-LF 1SG.GEN food this
 'I ate this food (there may be some left).'
- d. *Wada -mu puq-un damac nii.*
 PRET 1SG.GEN eat-PF food this
 'I ate up this food (there's none left).'

LF FUT is entirely nominal in meaning, and refers to the possibility of an action occurring. It is usually used together with the verbs *niqan* 'there is' and *uka* 'there is not'.

- (7) *Uka m-mah-an sino kiya.*
 not.have -FUT-drink-LF wine that
 'That wine is undrinkable.'

LF IRR is listed as *-e* while PF IRR is listed as *-o*. The reason for this is that *-e* sometimes has partitive/imperfective connotations as opposed to completive/perfective connotations for *-o* (see 8a,b).¹⁰ However, *-e* and *-o* are, according to my consultants, often freely interchangeable (8c).

- (8)a. *Mah-e -ta sino...*
 drink-LF.IRR 1PL.INC.GEN wine
 'Let's drink some wine...' (line of a song)

¹⁰ This classification also tallies well with the cognate forms *-ay* and *-aw* in Atayal (Egerod 1965; Huang 1995). Note, however, that *-ay* in the Mayrinax dialect of Atayal may also be used in AF (Huang 1995:79-81), whereas this does not occur in Seediq.

- b. *Mah-o -mu sino nii.*
 drink-PF.IRR 1SG.GEN wine this
 'I might drink up this wine (so if you want any, drink it now!).'
- c. *Qta-i hari! Mah-e / mah-o -daha sino kiya!*
 see-PF.IMP a bit drink-LF.IRR drink-PF.IRR 3PL.GEN wine that
 'Careful! They might drink that wine!'

LF IMP is formed by suffixation of *-i*. Note that this form is identical to the form for PF IMP. Thus, the PF/LF distinction is not realised in the imperative or negatable mood, as illustrated by examples (9a–d).

- (9)a. *Ini -mu tkan-i beras nii.*
 NEG 1SG.GEN pound-PF.IMP rice this
 'I haven't pounded this rice.'
- b. *Ini -mu tkan-i beras duhung nii.*
 NEG 1SG.GEN pound-LF.IMP rice mortar this
 'I do not pound rice in this mortar.'
- c. *Ini -mu puq-i damac nii.*
 NEG 1SG.GEN eat-PF.IMP food this
 'I haven't eaten this food.'
- d. *Ini -mu puq-i damac pngerax kiya.*
 NEG 1SG.GEN eat-LF.IMP food plate that
 'I haven't eaten food from that plate.'

This lack of distinction between LF and PF IMP is shared with the closely related language Mayrinax Atayal (Huang 1995).

2.4 Instrument focus

The tense/mood neutral form of IF involves the prefixation of an *s*-morpheme. This morpheme does not recur in any of the other forms of IF, with the possible exception of IF PRET, if the form *sn-* given in Table 1 is to be considered the regular IF PRET affix. While IF prototypically indicates that the clause subject is an instrument or beneficiary, a couple of verbs, notably *s-qada* 'IF-discard' (cf. *q-m-ada* '-AF-discard') and *sa-apa* 'IF-carry' (cf. *m-apa* 'AF-carry') use IF forms to create a straight passive, parallel in function to a PF form for other verbs (10).¹¹

- (10)a. *S-qada -na ka qyqeya.*
 IF-throw.away 3SG.GEN NOM thing
 'He throws the thing away.'
- b. *Sa-apa -mu laqi kiya.*
 IF-carry 1SG.GEN child that
 'I carry that child.'

The IF IMP suffix is *-ani* ((11a); compare with affirmative (11b)). However, it appears that some verbs use the LF/PF IMP form *-i* in alternation with or instead of *-ani* ((11c);

¹¹ I thank Naomi Tsukida for having brought this fact to my attention.

compare with affirmative (11d)) when the meaning is clearly instrumental. Interestingly enough, if *-ani* is used with such verbs, the implication is often benefactive rather than instrumental (see (11e)).

- (11)a. *Ini -daha dmt-ani ido damac kiya.*
 NEG 3PL.GEN eat.with-IF.IMP rice food that
 'They don't eat that food to accompany rice.'
- b. *S-damac -daha ido damac kiya.*
 IF-eat.with 3PL.GEN rice food that
 'They eat that food to accompany rice.'
- c. *Ini -mu lbu-i klabuy nii.*
 NEG 1SG.GEN wrap-PF/LF.IMP paper this
 'I don't wrap (things) with this paper.'
- d. *S-labu -mu klabuy gaga.*
 IF-wrap 1SG.GEN paper that
 'I wrap (things) with that paper.'
- e. *Ini -mu lby-ani.*
 NEG 1SG.GEN wrap-IF.IMP
 'I haven't wrapped it up for him/her.'

This seems to indicate that the situation is more complex than it appears at first sight. If this contrast can be shown to be systematic, it is possible that the imperative/negatable mood displays a distinction between a clearly instrumental focus and a clearly benefactive focus, a distinction which I have not found with other mood/tense forms (it does not hold for all verbs, as can be seen from (11a)).

IF PRET is given as *sn-* (12a,b). This form is actually quite rare in normal usage, and it often alternates with the PRET forms of other (non-actor/passive) foci, such as PF PRET (12c) or LF PRET (12d). This does not, however, change the basic configuration of the clause: (12d) is still an IF clause in that the instrument is subject.

- (12)a. *S-n-damac -daha ido ciga qcurux nii.*
 IF-PRET-with 3PL.GEN rice yesterday fish this
 'They ate this fish with rice yesterday.'
- b. *S-n-qada -na Pawan qyqeya nii.*
 IF-PRET-throw.away 3SG.GEN Pawan thing this
 'Pawan threw away this thing.'
- c. *Q-n-ada -na Pawan qyqeya nii.*
 -PF.PRET-throw.away 3SG.GEN Pawan thing this
 'Pawan threw away this thing.'
- d. *S-n-bet-an -mu ricah btakan nii.¹²*
 -PRET-beat-LF 1SG.GEN plum bamboo this
 'I used this bamboo pole to knock down plums (from the tree).'

¹² This example and its paraphrases ((13a,b), (16c), (36a,b)) are inspired by a paraphrase in Chang 1997b. However, I have adapted the word order in a couple of the examples to what my principal consultant feels is the least marked order.

IF FUT is given as either *s-*, \emptyset or *p-*. The simplest option is to use *s-* as a tense-neutral form (13a), or, as with PRET, to borrow a form from another focus, such as PF PRES, which has a future interpretation ((13b), see also (15c)). Again, this does not affect the configuration of the clause. The \emptyset form seems to be some kind of nominalisation rather than a regular IF FUT, although it is used in contexts where an IF form, particularly FUT, would be expected (13c). The status of the *p-* prefix (not to be confused with a causative *p-*)¹³ is not clear either (13d), although its use seems to be similar.

- (13)a. *S-sebuc -mu ricah btakan nii.*
 IF-beat 1SG.GEN plum bamboo this
 'I am knocking/will knock the plums out of the tree with this bamboo pole.'
- b. *Sbet-un -mu ricah btakan nii.*
 beat-PF 1SG.GEN plum bamboo this
 'I will knock the plums out of the tree with this bamboo.'
- c. *Ngal-un -daha sudu tabu -daha dapa.*
 take-PF 3PL.GEN grass fodder/feed.IF(?) 3PL.GEN cow
 'They take hay to feed cows.'/'They take hay as their cow-fodder.'
- d. *P-tabu -mu dapa sudu `nii.*
 IF.FUT(?)=feed 1SG.GEN cow grass this
 'I shall feed cows with this grass.'

IF IRR is given as *-anel-ano* (14). The former (*-ane*) is the regular reflex of forms appearing in other Atayalic languages, such as *-anay* in Mayrinax Atayal (Huang 1995). The status of *-ano* in Seediq is more unusual. According to my consultants, it can be used optionally in place of *-ane*.

- (14)a. *Qta-i hari! Sbt-ane/sbt-ano -daha laqi qhuni kiya!*
 look-PF.IMP a.bit beat-IF.IRR 3PL.GEN child wood that
 'Careful! They might beat a child with that piece of wood!'
- b. *Qta-i hari! Qda-ane/qda-ano -daha lukus -su!*
 look-PF.IMP a.bit discard-IF.IRR 3PL.GEN clothes 2SG.GEN
 'Careful! They might throw away your clothes!'

The alternation of *-ane* and *-ano* is reminiscent of the alternation between LF IRR *-e* and PF IRR *-o*, and it is quite likely that *-ano* is formed by analogy to the latter (taking the *-an* suffixed form to be a derived stem to which either *-e* or *-o* can be affixed). This is probably strengthened by the fact that IF IMP is *-ani*, which superficially also resembles the PF *-i* morpheme suffixed to a stem ending in *-an*. In view of this, we would expect the *-an* morpheme to occur in other types of formation as well. This is in fact the case. Thus, a parallel PF form for *q-m-ada* 'to throw away' is *qda-an-un*, comprising both the *-an* suffix and the *-un* suffix (15a,b). This form has an unambiguous future interpretation, typical of PF PRES, which is normally lacking in the simple IF form (see (15c)). In fact, even a 'double' LF form *q-n-da-an-an* is attested, where LF *-an* is affixed to a form which already contains the *-an* suffix (15d).

¹³ The unmarked causative in Seediq is AF in function, and could thus not cooccur with the GEN clitic *-mu*. Non-actor-focussed forms of the causative combine the *p-* with the focus morphology concerned (e.g. *p-un* 'CAUS--PF'; *p-an* 'CAUS--LF'; *s-p-* 'IF-CAUS.')

- (15)a. *Qda-an-un -na kusun ka qyqeya kiya.*
 discard-AN-PF 3SG.GEN tomorrow NOM thing that
 'He will throw away that thing tomorrow.'
- b. *Pa-an-un -mu laqi kiya.*
 carry-AN-PF 1SG.GEN child that
 'I'll carry that child.'
- c. *Sa-apa -mu laqi kiya.*
 IF-carry 1SG.GEN child that
 'I carry that child.'
- d. *Q-n-da-an-an -daha huling kdere kiya.*
 -PRET-discard-AN-LF 3PL.GEN dog cliff that
 'They threw a dog over that cliff.'

Thus, the affixes in use with IF cannot be subsumed into a regular pattern, rather, they appear to be the result of two different systems operating simultaneously; one following regular Atayalic reflexes, and the other overgeneralising parts of these by treating the *-an*-section in IF forms as an intermediate stem-forming suffix to which further suffixation is possible. It is possible that this is the result of some kind of change which the verb system in Seediq is undergoing.

3. The syntax of Seediq

3.1 Word order

Seediq is basically a VOS language, although a certain amount of word order variation does occur. A full NP subject typically appears clause-finally (16a), optionally followed by a time adverb (but note that such an adverb may also occur immediately after the verb). In a passive (i.e. non-AF) clause, the Agent usually precedes the subject (16b). However, it is also rather common for the Agent to be postposed after the subject, particularly if there are more than two arguments in the clause (16c).

- (16)a. *M-n-imah sino Pawan.*
 -AF-PRET-drink wine Pawan
 'Pawan drank wine.'
- b. *Wada puq-un qolic ka damac -su.*
 PRET eat-PF rat NOM food 2SG.GEN
 'Your food was eaten up by rats.'
- c. *S-sebuc -na ricah ka btakan Pawan.*
 IF-strike 3SG.GEN plum NOM bamboo Pawan
 'Pawan strikes plums (down from the tree) with a bamboo.'

Another possible word order is SVO, which is derived by topicalisation of the subject to pre-verbal position. Such a topic may be, but need not be, followed by a topic marker *ge*.

- (17) *Boyak ge m-ntena babuy (heya).*
 boar TOP AF-resemble pig 3SG.NOM.LONG
 'A wild boar is similar to a pig.'/'A wild boar, it is similar to a pig.'

Auxiliaries, negators and subordinators precede the verb, in the following order: subordinator, tense-marker, negator, modal or Aktionsart-auxiliary and main verb.

3.2 Arguments

Indefinite/non-individuated nouns are not case-marked in Paran Seediq.¹⁴ However, definite agents in passive clauses (including proper nouns) are obligatorily preceded by the agent marker¹⁵ *na* (which can conveniently be described as a genitive/ergative¹⁶ determiner). Additionally, subjects of both active and passive clauses are optionally preceded by the subject marker¹⁷ *ka* (which can, for its part, be described as a nominative determiner).

- (18)a. *Wada puq-un *(na) Pawan (ka) bunga -su.*
 PRET eat-PF GEN Pawan NOM sweet.potato 2SG.GEN
 'Pawan ate up your sweet potato.'
- b. *M-n-ekan bunga (ka) Pawan.*
 AF-PRET-eat sweet.potato NOM Pawan
 'Pawan ate (some) sweet potato.'

Pronoun morphology, on the other hand, is richer. Seediq has two sets of pronouns, so-called long pronouns, which occur in the same position in the clause as full NP's, and clitic pronouns, which appear cliticised to the first verbal element in the clause (be it a subordinator, an auxiliary, a negator or a main verb). In modern Paran Seediq, long pronouns can distinguish two cases:¹⁸ Nominative and Genitive. Genitive long pronouns are primarily used as possessive predicates (19a). Nominative long pronouns can be used either as subjects (19b) or objects (19c), for all focus types.

- (19)a. *Naku ka sapah nii.*
 1SG.GEN.LONG NOM house this
 'This house is mine.'

14 In the Taroko dialect spoken in Hualien County, however, nouns referring to humans do have an object case suffix: *-an*.

15 This marker is homophonous with the third person singular genitive clitic *na*. Often it is impossible to determine which is which in a given context. In doubtful cases I have glossed *na* as the clitic. However, the clitic always directly follows an auxiliary verb if there is one, so in 18a we are clearly dealing with the agent marker.

16 I adhere to traditional terminology in referring to this case as 'genitive' (GEN), although I am well aware that its function together with a verb is undeniably that of ergative case. Whether or not this implies that Seediq is an ergative language (as suggested for other focus languages — including the very closely related language Atayal — in an increasing number of works, see Huang 1994; Starosta 1986; Starosta 1997) is a matter of debate, and largely depends on how we choose to define terms such as 'ergativity' and 'transitivity'. Certainly, Seediq does have clear ergative characteristics, although other features seem to place it almost equally clearly in the accusative camp. I shall not address this question here.

17 It has a wider use, in fact, since it can be used to indicate high individuation of other NP's as well, although its most common function is as a subject marker.

18 Specific object/oblique pronouns are not used today in the Paran dialect (as opposed to the situation in the Taroko dialect), except in deliberately archaic speech.

- b. *Q-m-n-ita -ku Pawan (yaku).*
 -AF-PRET-see 1SG.NOM Pawan 1SG.NOM.LONG
 'I saw Pawan.'
- c. *Q-m-n-ita -ku heya.*
 -AF-PRET-see 1SG.NOM 3SG.NOM.LONG
 'I saw her/him.'

Clitic pronouns also distinguish two cases: Nominative and Genitive, although the distinction has fused for most persons. There is also a small set of portmanteau clitics which combine one agent with one patient. If a desired configuration does not exist as a portmanteau form, two pronouns are chosen from the table. GEN clitics are used for agents of passives (non-AF), and NOM clitics are used for subjects, irrespective of the focus of the verb. The entire set of pronouns is illustrated in Table 2.

Recent research (Chang 1997a) has suggested that the clitic pronouns should be viewed as agreement morphemes instead. There are certain advantages to this suggestion, as it explains the co-occurrence of clitic pronouns with coreferent clausemate long pronouns or NP's, as well as other problematic facts. At the same time, this view is not entirely unproblematic, since other types of suffixation in Seediq regularly lead to a stress shift connected with other morphophonemic changes (for details see Yang 1976 or Holmer 1996). This type of stress shift is hardly ever triggered by the presence of a clitic pronoun.¹⁹ In this paper I have chosen not to address this question, but simply to adhere to traditional terminology and refer to the relevant category as clitic pronouns, while noting that they in certain ways behave like agreement morphemes, perhaps being an intermediate stage in the development of agreement.

Table 2: Pronouns in Seediq

	Long pronouns			Clitic pronouns	
	NOM	GEN	(OBJ)	NOM	GEN
1SG	<i>yaku</i>	<i>naku</i>	<i>(kenan)</i>	<i>ku</i>	<i>mu</i>
2SG	<i>isu</i>	<i>nisu</i>	<i>(sunan)</i>	<i>su</i>	
3SG	<i>heya</i>	<i>nheya</i>	–	–	<i>ra</i>
1PL.INC	<i>ita</i>	<i>nita</i>	–	<i>ta</i>	
1PLEXC	<i>yami</i>	<i>namami</i>	–	<i>nami-mian</i>	
2PL	<i>yamu</i>	<i>namamu</i>	<i>(munan)</i>	<i>namu</i>	
3PL	<i>dheya</i>	<i>ndheya</i>	–	–	<i>daha</i>
Portmanteau clitic:					
	1SG.GEN, 2SG.NOM		2SG.GEN, 1SG.NOM		1SG.GEN, 2PL.NOM
	<i>misu</i>		<i>saku</i>		<i>maku</i>

¹⁹ It can occur optionally with the verb *máha* 'to go, get going', as in [maháku] 'I'm going to...' or [mahásu] 'You're going to...' and in certain set phrases, such as *haani ta* 'go-IF.IMP 1PL.INC' > [níta] 'let's'. In all of these examples the stress shift is optional. With other pronouns and most other verbs, the stress shift is impossible.

If two clitics cooccur with a verb (this is only possible with non-AF verbs²⁰) the order is NOM-GEN (20). However, the portmanteau forms, in so far as they are morphologically transparent, contain atomic elements in the opposite order (compare for instance the relative positions of the *m-* and *-s-* in *misu* '1SG.GEN, 2SG.NOM' and the relative positions of *s-* and *-k-* in *saku* '2SG.GEN, 1SG.NOM').²¹

- (20)a. *Q-n-ta-an -ku -namu.*
 -PRET-see-LF 1SG.NOM 2PL.GEN
 'You (PL) saw me.'
- b. *Q-n-ta-an -namu -daha.*
 -PRET-see-LF 2PL.NOM 3PL.GEN
 'They saw you (PL).'
- c. *Qta-un -su -mu.*²²
 see-PF 2SG.NOM 1SG.GEN
 'I will see you.'

Enclitic pronouns are attached to the first verbal element in the clause. This can be a subordinator (21a), a tense-marker (21b), a negator (21c) or a main verb (21d). However, they may not cliticise to topics and other preverbal elements (21e). A clitic pronoun may also cliticise to a noun, if it is used predicatively (21f) or if the clitic is the possessor of the noun (21g).

- (21)a. *Netun -ku -na wada ini qta-i*²³...
 if 1SG.NOM 3SG.GEN PRET NEG see-PF.IMP
 'If he didn't see me...'
- b. *Wada -na puq-un ka bunga.*
 PRET 3SG.GEN eat-PF NOM sweet.potato
 'He ate up the sweet potato.'
- c. *Ini -ku kela m-bahang kari mukan.*
 NEG 1SG.NOM (AF.IMP)-know AF-listen language Taiwanese
 'I can't understand Taiwanese.'
- d. *M-n-ekan -ku bunga ciga.*
 AF-PRET-eat 1SG.NOM sweet.potato yesterday
 'I ate sweet potatoes yesterday.'

²⁰ Note also that clitics can only co-occur if at least one of them is unambiguous as to case (see §4.3).

²¹ Chang (1997a) suggests instead that bound pronoun ordering can be generalised as first/second person – third person (i.e. third person follows first and second person – the generalisation says nothing about the order of first and second person pronouns). This generalisation, while technically correct, and while it also covers (or is not contradicted by) the morpheme ordering within the portmanteau pronouns, is actually not particularly illuminating, since it does not address or account for the relative ordering of bound pronouns which are not third person (NOM-GEN for regular clusters, and GEN-NOM for atomic elements within portmanteaux).

²² *-su-mu* '2.SG.NOM-1.SG.GEN' is in free variation with *misu* '2.SG.NOM,1.SG.GEN', which apparently has a rather archaic flavour. The other portmanteau clitics, however, do not have non-portmanteau alternants.

²³ Recall that the negator *ini* is obligatorily followed by a verb in IMP mood.

- e. **Ciga -ku m-n-eyah hini.*
 yesterday 1SG.NOM AF-PRET-come here
 (Intended reading: 'I came here yesterday.')
- f. *Seediq -ku yaku.*
 Seediq 1SG.NOM 1SG.NOM.LONG
 'I am a Seediq.'
- g. *tama -mu*
 father 1SG.GEN
 'my father'

4. Syntactic properties of Seediq focus

4.1 Auxiliaries and focus

In a Seediq clause, various types of auxiliary may receive focus affixation. Some of these auxiliaries correspond in meaning to adverbs of manner (22a–d), others are directional in meaning (22e–h). In Holmer (1996), such auxiliaries are termed 'focus auxiliaries' by virtue of their ability to carry a focus distinction, as opposed to 'tense auxiliaries' which can only carry a tense/aspect distinction. I shall adhere to this practice in this paper.

- (22)a. *Hde-un m-ekan ngiyo ka qolic.*
 finish-PF AF-eat cat NOM rat
 'The rats will be finished off by cats.'
- b. *Ma h-m-n-edu m-ekan damac laqi nii!*
 and -AF-PRET-finish AF-eat food child this
 'But this child finished off all the food!'
- c. *Bleq-un -daha m-ekuy ka dapa.*
 properly-PF 3PL.GEN AF-tie NOM cow
 'They tie the cow securely.'
- d. *Nme-un -daha t-m-uting ka qmegi.*
 powder-PF 3SG.GEN -AF-beat NOM soapwort
 'They beat the soapwort to a powder.'
- e. *Ha-un -mu m-angan qedin -mu.*
 go-PF 1SG.GEN AF-take wife 1SG.GEN
 'I'll go catch my wife.'
- f. *Yah-o m-ekan qolic ka bunga.*
 come-PF.IRR AF-eat rat NOM sweet.potato
 '(Careful!) Rats might come and eat the sweet potatoes.'
- g. *M-n-eyah m-ekan bunga qolic kiya.*
 AF-PRET-come AF-eat sweet.potato rat that
 'That rat came to eat sweet potatoes.'
- h. *Sa-i pusa uyung bebe ka parih nii!*
 go-PF.IMP put back eaves NOM hoe this
 'Go and put this hoe under the eaves behind the house!'

An interesting point to note is that if the auxiliary is in a non-AF focus (22a,c–f, h), the main verb is obligatorily in AF.²⁴ This is despite the fact that it is the main verb which is primarily affected by the diathetic change. Thus, in (22e), the clause subject is clearly the patient of the verb *mekan* ‘to eat’ rather than of *yah-o* ‘come-PF.IRR’, although the PF morphology is realised on the latter. In this sense, AF is a default focus form which in itself does not make a clause necessarily AF. Rather, a clause is AF if and only if the first verb capable of focus-marking (be it a focus auxiliary or a main verb) is AF. It is always the focus of this first verb which determines the focus interpretation of the entire following clause.

A small set of constructions is exempt from this principle. Thus, passive complements of certain control verbs appear in a non-AF focus (23a,b). Likewise, so-called ‘tough constructions’ (such as ‘good to eat’, ‘nice to hear’) also have a non-AF second verb (23c).

- (23)a. *Ani ima m-qaras kux-un seedaq.*
 even who AF-happy like-PF person
 ‘Everyone wants to be liked.’ (lit: ‘Everyone is happy to be liked.’)
- b. *M-qaras Pawan qta-an seedaq.*
 AF-happy Pawan see-LF person
 ‘Pawan likes to be seen.’
- c. *Malu puq-un damac nii.*
 good eat-PF food this
 ‘This food is good to eat.’

What these constructions all have in common is that the patient of the embedded verb is simultaneously the subject of the matrix verb or adjective. Thus we can generalise that focus is realised on the auxiliary rather than on the main verb, as long as the underlying actor of both verbs is the same. If the underlying actor is different, each verb must take care of its own focus interpretation.

The interaction of focus forms with auxiliaries has another interesting implication. In most cases, the form of the auxiliary is identical to the form which the main verb would have if the auxiliary were absent. Thus, if the auxiliaries were absent in (22), the main verbs would have the following forms: (22a) *puq-un* ‘eat-PF’; (22c) *bkey-un* ‘tie-PF’; (22e) *ngal-un* ‘take-PF’ etc.

However, the situation is different for IF. While we would expect an instrument/oblique subject to be cross-referenced by an auxiliary in IF (in analogy to the PF/LF auxiliaries in the examples under (22)), such an IF auxiliary is ungrammatical (24a). If the meaning of the auxiliary is required (for instance if we wish to express the idea of ‘going to do something’), an auxiliary in PF/LF may be used instead (24b). Note that this does not change the configuration of the clause, namely that the oblique is still clause subject. This use of a PF/LF auxiliary in what is structurally equivalent to an IF clause is rather marked, however. While it can be elicited, I have yet to find it in spontaneous examples. Instead, the least marked construction would simply involve not using any auxiliary at all (24c).

²⁴ *Pusa* in (22h) is not an exception. It is a causative form and as such never overtly marked with the AF *-m-* morpheme (for more details see Holmer 1996:48ff.). However, the syntactic function of a bare causative is always AF.

However, with the two verbs *s-qada* 'IF-throw.away' and *sa-apa* 'IF-carry', which, as we recall from §2.4, are IF in form although arguably PF in function, the opposite holds. These two verbs are commonly accompanied by PF/LF auxiliaries: examples like (24d) and (24e) are relatively unmarked, and are often spontaneously produced by Seediq speakers.

- (24)a. **Sa-aha -daha q-m-alang lmiq̄u ka dapa.*
 IF-go 3PL.GEN -AF-fence.in forest NOM cow
 (Intended reading = 24b)
- b. *Ha-un -daha q-m-alang lmiq̄u ka dapa.*
 go-PF 3PL.GEN -AF-fence.in forest NOM cow
 'They'll go and fence is some forest for the cow.'
- c. *S-qalang -daha lmiq̄u ka dapa.*
 IF-fence.in 3PL.GEN forest NOM cow
 'They fence in a section of forest for the cow.'
- d. *Asi -daha sa-i q-m-ada huling d-m-edux.*
 just 3PL.GEN go-PF.IMP -AF-discard dog -AF-howl
 'They just get rid of dogs that howl.'
- e. *S-damux -daha sapah*
 IF-roof 3PL.GEN house

ka btunux n-sa-an -daha m-apa.
 NOM stone PRET-go-LF 3PL.GEN AF-carry
 'They make the roof of the house with the slate they have carried.'

Thus, while the distinction between AF and non-AF seems to be a property of the clause as a whole (in that it can be realised on an auxiliary rather than on the verb itself), and while the further distinction between LF and PF is often related to the aspectual interpretation of the action rather than to the focus of the verb, IF seems to be a category which is only connected with the verb to which it belongs. It cannot be realised on an auxiliary,²⁵ and its function is usually expressed without resorting to an auxiliary at all (although such examples are possible, see (24b)).

4.2 Focus with ditransitives and causatives

Another important aspect of the focus system is its use with verbs which have more than two arguments. Such verbs are ditransitives and causatives.²⁶ A ditransitive verb in AF indicates that the clause subject is the agent, i.e. the giver (25a), and an AF causative²⁷ indicates that the subject is the causer (25b).

- (25)a. *M-bege -ku sapah nü Pawan.*
 AF.FUT.IMMED-give 1SG.NOM house this Pawan
 'I'll give Pawan this house.'

²⁵ The IF IMP affix *-ani* can, however, be used with auxiliaries. Baudhin (undated, approx. late 1960's, early 1970's) quotes many examples with *ha-ani* 'go-IF IMP', meaning 'let's'.

²⁶ Crucially, this refers to causatives of transitives rather than causatives of intransitives.

²⁷ Note that AF causatives do not carry the *m*-affix.

- b. *Pi-imah -ku sino seediq kiya.*
 (AF),CAUS-drink 1SG.NOM wine person that
 'I invite that person to drink wine.'

A ditransitive verb in IF, on the other hand, indicates that the subject is the object given (26a), and a causative IF verb indicates that the subject is patient of the action (26b,c).

- (26)a. *S-bege -mu Awi lukus -mu.*
 IF-give 1SG.GEN Awi clothes 1SG.GEN
 'I give my clothes to Awi.'
- b. *S-pi-imah -mu seedaq sino nii.*
 IF-CAUS-drink 1SG.GEN person wine this
 'I invite someone to drink this wine.'
- c. *S-p-qita -na Awi ka patis Pawan.*
 IF-CAUS-see 3SG.GEN Awi NOM book Pawan
 'Pawan shows Awi the book.'

Both AF and IF are unambiguous in this respect. AF can only refer to the causer/giver, and IF can only refer to the patient/object given. The use of LF and PF, on the other hand, is often ambiguous. While the use of either LF or PF is obligatory if the subject is the causee of a causative verb (27a) or the recipient of a ditransitive verb (27b), these foci can also be optionally used for objects given (27c) or patients of causatives (27d), preferably (but not exclusively) if the recipient is not overtly expressed.

- (27)a. *P-n-mah-an -mu sino seedaq kiya.*
 CAUS-PRET-drink-LF 1SG.GEN wine person that
 'I invited that person to drink wine.'
- b. *B-n-iq-an -mu lukus -mu heya.*
 -PRET-give-LF 1SG.GEN clothes 1SG.GEN 3SG.NOM.LONG
 'I gave him/her my clothes.'
- c. *Ani lukus ini -daha ngal-i b-n-ege -daha.*
 even clothes NEG 3PL.GEN take-PF.IMP -PF.PRET-give 3PL.GEN
 'They do not accept even clothes if they are given by them.'
- d. *P-n-mah-an -mu sino nii.*
 CAUS-PRET-drink-LF 1SG.GEN wine this
 'I invited someone to drink this wine.'

The generalisation seems to be that LF/PF can be used whenever the subject is not the agent (i.e. the causer or giver) and that IF is optionally used to clarify that the subject is a patient or object given. This type of clarification is particularly important if both the recipient/causee and the patient are overtly expressed.

4.3 Factors underlying focus choice

We have hitherto looked at the distribution of various foci with various configurations of arguments. We have not, however, addressed the question of how a given focus is chosen, or rather, why a given argument is chosen as subject. Which of these choices is more basic is a matter of debate (for varying points of view see Chang 1997b and Holmer 1996). However,

from a descriptive point of view (given particularly the ambiguity of focus choice illustrated in §4.2) it is simpler to assume that focus choice reflects subject choice rather than vice versa.

If only one argument is present in the argument structure (note that it need not be overtly expressed), it serves as clause subject, and the focus used is AF. AF is also used with zero-place verbs such as weather verbs (28). Furthermore, AF is used as a default form for any verb following an auxiliary which is capable of carrying focus (so-called ‘focus auxiliaries’).

- (28) *Gaga q-m-uyux.*
 PROG -AF-rain
 ‘It is raining.’

If a one-place predicate is used with an extra NP, such as an oblique of some kind, it is usual for the oblique to be realised as the subject, and then the focus of the verb is changed accordingly. Thus, if the oblique is the reason for an action, the verb appears in PF (29a, quoted from Chang 1997b:71). If the oblique is a location, the verb appears in LF (29b, quoted from Chang 1997b:71).

- (29)a. *Lngis-un na Pawan ka Temi.*
 cry-PF GEN Pawan NOM Temi
 ‘Pawan will cry for Temi.’
- b. *Lngis-an na Pawan ka pray-an.*
 cry-LF GEN Pawan NOM cook-LF
 ‘Pawan cried in the kitchen.’

Another possibility is that the agent is chosen as subject and the oblique can still be expressed. This is grammatical in some cases (30a) but ungrammatical in others (30b, quoted from Chang 1997b:71). It is difficult to assess the reason for this difference, but presumably it has to do with the possibility of interpreting the oblique.

- (30)a. *Gaga l-m-ingis yqeya q Pawan.*
 PROG -AF-cry wet.field Pawan
 ‘Pawan is crying in the field.’
- b. **L-m-ingis Temi ka Pawan.*
 -AF-cry Temi NOM Pawan
 (Intended reading: ‘Pawan is crying for Temi.’)

If a two-place predicate is used, either the agent or the patient may be clause subject. If one of the arguments is definite, it is likely to appear as clause subject (31a,b). If both are definite, the patient is more likely to be realised as subject, particularly if it is also totally affected (31c). This is presumably because a definite agent can be marked with a GEN determiner *na*, whereas there is no corresponding object determiner. With personal names, on the other hand, there seems to be no obvious preference either way (31d,e), probably because both are inherently definite. The focus of the verb varies accordingly, AF with agent subjects and LF/PF with patient subjects (recall that LF/PF is regularly used as a past tense straight passive with no locative connotations).

- (31)a. *Wada puqun qolic ka bunga.*
 PRET eat-PF rat NOM sweet.potato
 ‘The sweet potatoes were eaten up by rats.’

- b. *M-n-ekan bunga ka qolic.*
 -AF-PRET-eat sweet.potato NOM rat
 'The rat ate sweet potatoes.'
- c. *P-n-uq-an na qolic ka bunga.*
 -PRET-eat-LF GEN rat NOM sweet.potato
 'The rat ate the sweet potatoes.'
- d. *Q-m-n-ita Awi ka Pawan.*
 -AF-PRET-see Awi NOM Pawan
 'Pawan saw Awi.'
- e. *Q-n-ta-an na Pawan ka Awi.*
 -PRET-see-LF GEN Pawan NOM Awi
 'Pawan saw Awi.'

If the agent (but not the patient) is pronominal, it generally only appears as subject if the patient is either indefinite or partially affected (32a). A definite or totally affected patient is regularly realised as a subject (32b). Conversely, a pronoun patient is almost always realised as clause subject (32c). A long pronoun serving as an object with a full NP subject is questionably grammatical (32d).

- (32)a. *M-n-ekan -ku ido.*
 -AF-PRET-eat 1SG.NOM rice
 'I ate rice.'
- b. *P-n-uq-an -mu damac -su.*
 -PRET-eat-LF 1SG.GEN food 2SG.GEN
 'I ate your food.'
- c. *Q-n-yut-an -ku -na quyu.*
 -PRET-bite-LF 1SG.NOM 3SG.GEN snake
 'I was bitten by a snake.'
- d. *?Q-m-n-iyuc yaku ka quyu.*
 -AF-PRET-bite 1SG.NOM.LONG NOM snake
 'The snake bit me.'

Finally, if both arguments are pronominal, the favoured configuration is that the patient is subject, with the agent realised as a genitive clitic. This is, however, only possible if at least one of the pronouns is unambiguously either NOM or GEN (33a,b), i.e. if one of the pronouns is *ku* '1SG.NOM', *mu* '1SG.GEN', *na* '3.SG.GEN' or *daha* '3PL.GEN'. Co-occurrence of two morphologically ambiguous clitics is ungrammatical (33c). If this configuration is required, the patient is realised as an object, with the agent as subject of an AF construction (33d). In other words, the NOM-GEN ordering evident in clitic clustering is not sufficient to determine the interpretation. Instead, a morphological distinction is also required.²⁸

²⁸ Assuming a person-based ordering (such as Chang 1997a does, with 1/2-3) would imply that position would not be expected to indicate grammatical function, thus explaining the facts in (33c-e). If this is the case, however, then the fact that regular clitic clusters of first and second persons obligatorily are in NOM-GEN order, whereas the reverse order only occurs with portmanteau clitics, which have a clearly idiosyncratic form (featuring vowel changes etc.), seems to be a quite remarkable coincidence.

- (33)a. *Q-n-ta-an -ku -namu.*
 -PRET-see-LF 1SG.NOM 2PL.(GEN)
 'You (PL) saw me.'
- b. *Q-n-ta-an -namu -daha.*
 -PRET-see-LF 2PL.(NOM) 3PL.GEN
 'They saw you (PL).'
- c. **Q-n-ta-an -namu -nami.*
 -PRET-see-LF 2PL.NOM/GEN 1PL.EXC.GEN/NOM
 (Intended reading (?): 'You saw us.'/'We saw you.')
- d. *Q-m-n-ita -namu yami.*
 -AF-PRET-see 2PL.NOM 1PL.EXC.NOM.LONG
 'You saw us.'
- e. *Q-m-n-ita -nami yamu.*
 -AF-PRET-see 1PL.EXC.NOM 2PL.NOM.LONG
 'We saw you.'

If the verb is a two-place predicate, and there is an oblique NP present, the subject may be either the oblique NP or one of the other arguments. If the oblique is the subject, the focus of the verb is LF if the oblique is a location (34a), and IF if the oblique bears any other role (34b). The oblique is generally chosen as subject either because it must (34b,c),²⁹ or because it is the most salient or definite argument present (34d).

- (34)a. *Tkan-an -mu beras duhung nii.*
 pound-LF 1SG.GEN rice mortar this
 'I pound rice in this mortar.'
- b. *S-qalang -daha lmiqu ka dapa.*
 IF-fence.in 3PL.GEN forest NOM cow
 'They fence some forest for the cow.'
- c. *?Q-m-alang lmiqu dapa dheya.*
 -AF-fence.in forest cow 3PL.NOM.LONG
 (Intended reading: 'They fence in some forest for a cow.')
- d. *Sa-angan -daha qcurux atak nii.*
 IF-take 3PL.GEN fish chopsticks this
 'They pick up fish with these chopsticks.'

If the oblique is not clause subject, the focus is either AF or PF/LF, depending on whether the agent or the patient of the two-place verb is subject. The criteria determining this are probably the same as for two-place predicates in general, although in practice it appears that PF/LF is greatly preferred.³⁰

²⁹ In other words, because it cannot be interpreted unless in subject position cross-referenced by the focus of the verb.

³⁰ I have yet to find spontaneous AF examples corresponding to (35) in the Paran dialect, although they can be elicited.

- (35) ...*sruk-an -daha puniq, snaw-an -daha qsiya...*
 singe-LF 3PLGEN fire wash-LF 3PLGEN water
 '...they singe it with fire and wash it with water...'

Finally, with three-place predicates such as causatives or ditransitives, it is generally definiteness or discourse saliency which determines which argument is realised as subject. It seems to be impossible to combine an oblique with a three-place predicate, presumably because of parsing difficulties. In fact, even examples of three-place predicates with all three arguments overtly realised as full NP's (rather than clitic pronouns) are hard to come by, especially in non-AF. Moreover, such non-AF examples are usually agent-final rather than subject-final (36a,b), as opposed to the strict subject-final order for two-place predicates (36c,d).

- (36)a. *S-sebuc -na ricah ka btakan Pawan.*
 IF-strike 3SG.GEN plum NOM bamboo Pawan
 'Pawan strikes plums (down from a tree) with the bamboo.'
- b. *?S-sebuc -na ricah na Pawan ka btakan.*
 IF-strike 3SG.GEN plum GEN Pawan NOM bamboo
 (Intended reading = 36a)
- c. *Puq-un qolic ka bunga.*
 eat-PF rat NOM sweet.potato
 'The sweet potatoes will be eaten by rats.'
- d. *#Puq-un bunga ka qolic.*
 eat-PF sweet.potato NOM rat
 (Intended reading = 36c)
 (Only possible reading: 'The rats will be eaten by sweet potatoes.')

Subject choice, and consequently focus choice, is thus determined by a rather complex set of factors. An important factor determining subject choice is definiteness: a definite NP has priority to be subject above an indefinite NP. However, this is tempered by other factors, such as the combinatory possibilities of clitic pronouns. Some types of NP (such as certain obliques) can only appear in subject position, thereby directly forcing focus choice. Another type of NP which can only appear in subject position is a relativised noun: thus, regardless of whether a given NP is agent (37a) or patient (37b), it must be the subject of the relative clause to which it belongs, with the appropriate focus expressed on the verb (see (37c), which is ungrammatical with the intended reading), unless it is represented by a resumptive pronoun within the relative clause itself (37d).

- (37)a. *seedaq m-n-apa btunux*
 person AF-PRET-carry stone
 'a person who carried a stone'
- b. *btunux n-apa na seedaq*
 stone PF.PRET-carry GEN person
 'a stone carried by the person'
- c. *#btunux m-n-apa seedaq*
 stone AF-PRET-carry person
 (Intended reading: 'a stone which a person carried')
 (Only possible reading: 'a stone which carried a person')

- d. *kari ini dehuk dungus na*
 word NEG reach meaning 3.SG.GEN
 'a mistake' (lit. 'a word, its meaning doesn't reach')

Likewise, in a wh-question with a clause-initial wh-word (as opposed to so-called 'in-situ' wh-questions), the focus of the verb must correspond to the wh-word being clause subject (38).

- (38)a. *Ima q-n-ta-an -su?*
 who? -PRET-see-LF 2SG.GEN
 'Whom did you see?/'(I know you saw someone. Who was it?)
- b. *Ima q-m-n-ita isu?*
 who? -AF-PRET-see 2SG.NOM.LONG
 'Who saw you?'
- c. **Ima q-m-n-ita -su?*
 who? -AF-PRET-see 2SG.NOM
 (Intended reading: 'Whom did you see?')
- d. *Q-m-n-ita -su ima?*
 -AF-PRET-see 2SG.NOM who?
 'Whom did you see?/'(Did you see anyone? In that case, whom?)

In cases like the above, the definiteness of the argument and the non-favouring of pronouns as direct objects ((38b), see (32d)) are of little impact, as these factors are overridden by the requirement that the extracted element be subject of the clause from which the extraction has taken place.

5 Summary

In this paper I have addressed some questions concerning the use of focus in Seediq: how the various foci are formed, which affixes are used, how focus and aspect interact, and which criteria underlie the choice of focus for a given context in Seediq. However, I have deliberately avoided analysing the facts presented here within the framework of any particular theory – the purpose of this paper has been to present an as exact as possible descriptive account of the Seediq focus system.

I have attempted to deal exclusively with questions which are relevant for the behaviour of focus in Seediq. However, given that the focus system is intertwined with most other aspects of the grammar, this has necessarily involved describing rather large adjacent areas of Seediq syntax in general. It should be noted that these descriptions can not be, and are not intended to be, exhaustive.

References

- Baudhin, R., undated, Glossary of Kari Sejik, ms. Wushe, Taiwan.
 Chang Yung-li, 1997a, Seediq bound pronouns: pronominal clitics or agreement affixes?
 Paper presented at the Eighth International Conference on Austronesian Linguistics,
 Taipei.

- 1997b, Voice, case and agreement in Seediq and Kavalan. PhD dissertation, National Tsinghua University, Hsinchu, Taiwan.
- Chen Kang, 1992, *Taiwan gaoshanzu yuyan* [Languages of the aboriginal peoples of Taiwan]. Beijing: Central Minorities Institute.
- Egerod, Søren, 1965, Verb inflection in Atayal. *Lingua* 15:251–82.
- Ferrell, Raleigh, 1969, *Taiwan aboriginal groups: problems in cultural and linguistic classification*. Institute of Ethnology Monograph No. 17. Taipei: Academia Sinica.
- Holmer, Arthur, 1996, *A parametric grammar of Seediq*. Travaux de l'institut de linguistique de Lund 30. Lund: Lund University Press.
- Huang, Lillian, 1994, Ergativity in Atayal. *Oceanic Linguistics* 33:129–43.
- 1995, *A study of Mayrinax syntax*. Taipei: Crane Publishing Company.
- Li Jen-kuei, 1992, *Taiwan nandao yuyan de yuyin fuhao xitong* [Orthographic systems for Formosan languages]. Taipei: Ministry of Education.
- Starosta, Stanley, 1986, Focus as recentralisation. In Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: papers from the Fourth International Conference on Austronesian Linguistics*, 73–95. Canberra: Pacific Linguistics.
- 1997, Transitivity, ergativity and the best analysis of Atayal case marking. Paper presented at the Eighth International Conference on Austronesian Linguistics, Taipei.
- Yang Hsiu-fang, 1976, Saideyu Wushe fangyan de yinyun jiegou [The phonological structure of the Paran dialect of Seediq]. *Bulletin of the Institute of History and Philology* 47:611–706.

The Sinama derived transitive construction

JUN AKAMINE

1 Introduction¹

This paper describes two types of transitive sentences in Manuk Mangkaw Sinama (henceforth MNK).² MNK is spoken in Manuk Mangkaw Island, Tawi-Tawi Province, in the south-west of the Philippines and is a dialect of Southern Sulu Sinama. Discussion of the typology of Sinama vis-à-vis other Philippine languages is beyond the scope of the paper. However, it will be shown below that Sinama is a morphologically ergative language.

In MNK, there are two kinds of construction that appear to be semantically passive. One, marked by the verbal infix <in>, seems to be similar to what Filipinists call the *goal-topic* construction. The second construction contains a prefix-like form *leq* attached to the verb.³ To my knowledge, this construction has not yet been the subject of any discussion, except in Akamine (1996). In that work I analysed the *leq* construction within the framework of ‘traditional grammar’ in Philippine linguistics, employing the traditional concepts of grammatical subject, active voice and passive voice. In the present analysis, I will re-analyse the construction from the viewpoint of discourse transitivity and show the semantic differences between the two constructions.

¹ The data presented here are results of several periods of fieldwork in Manuk Mangkaw Island. I am grateful to Hadji Amilhamja S. Juaini who accommodated me. This research was supported in part by (a) the Asian Studies Scholarship Program from the Ministry of Education, Science and Culture, Japan and (b) a separate grant from the same ministry for the special project ‘Culturo-Ecological Structure of Network Society in Wallacea’ (#07041057), headed by Tanaka Koji of Kyoto University. I want to express my special thanks to Lawrence Reid for his many comments on an earlier version of the paper. I also would like to thank participants in the workshop on ‘Transitivity and Ergativity/Accusativity in Syntactic Typology’ at ILCAA, Tokyo University for Foreign Studies, who gave me deeper insight into the topic of this paper.

² Abbreviations used: AF actor focus; D dual person; DET determiner; FUT future; GEN genitive; GF goal focus; IMP imperative; IMPF imperfective; MN minimal; NML non-minimal; NOM nominative; PERF perfective; PRD predicative.

³ The form is homophonous with the preposition which marks genitive NPs.

In §2, I will illustrate the Sinama case marking system for prepositions and pronouns. Lexical NPs have no distinctive morphological case marking — case identification depends upon word order. Verbal affixes are described in the first part of §3. The main part of the third section is devoted to presenting the so-called {*leqN-*} construction. In the fourth section, I will describe the semantic differences between *in*-type sentences and {*leqN-*} type sentences.

2 Sinama case marking system

For the sake of brevity, I will use the traditional terms of Philippine linguistics in the present analysis. By the term *subject*, I mean ‘grammatical subject’ — that is, the equivalent of the *ang* phrase in Tagalog. However, I stay with the term *focus* as used by Schachter and Otnes (1972:69) to refer to the feature of a verbal predicate that determines the semantic relationship between that verb and its grammatical subject.

2.1 Prepositions

MNK has five case-marking prepositions, namely *leq*, *ma*, *ni*, *min*, and *maka*. Some linguists label such forms as *case markers* because they indicate the semantic relationship between the verb and its complements. Such an interpretation, however, ignores an important aspect of the language. These forms often appear as the predicates of non-verbal sentences (see example (1); in these positions, they are not functioning as case markers. To ignore this fact leads to a misunderstanding of their function.

- (1) *Ma aku duyan.*
 at I.NOM durian
 ‘The durian is mine.’

Even when such forms function to mark case, the question of determining their lexical category remains. There are two possible categories for those forms: prepositions or determiners. They can be distinguished syntactically — prepositions typically precede any type of NP, including pronouns, as in (2), while determiners typically do not precede pronouns.

- (2)a. *leq ku* ‘by me’
 b. *leq si Abdul* ‘by Abdul’
 c. *leq anak-anak* ‘by the child’

There is only one determiner in Sinama, *si*, which occurs before every proper noun regardless of its case. It is not a nominative case marker as in Tagalog.

Though I have rejected the term case marker above, I will reserve the term *case* for the semantic relationship between the verb and its complements. Each preposition in MNK carries case-meaning: *leq* ‘agent’, *ma* ‘location’, *ni* ‘goal’, *min* ‘source’, *maka* ‘instrument’. Their English translations are ‘by’, ‘at/in’, ‘to’, ‘from’, and ‘with’ respectively.

Table 1: Prepositions in MNK

	<i>leq</i>	<i>ma</i>	<i>ni</i>	<i>min</i>	<i>maka</i>
case relations	agent	location	goal	source	instrument

The preposition *leq* marks an NP as being the agent of a transitive construction.

- (3) *K-in-akan leq tabi duyan nu.*
 GF-eat by we.GEN durian your
 'We ate your durian.'

The preposition *ma* marks an NP that indicates location.

- (4) *Bey aku ngiskul ma UP.*
 PERF I.NOM AF-study at UP
 'I studied at the University of the Philippines.'

The preposition *ni* marks an NP that indicates the end point of the event.

- (5) *Bey aku piqiq ni Sabah.*
 PERF I.NOM go.there to Sabah
 'I have been to Sabah.'

The preposition *min* marks an NP that indicates the starting point or source of the event.

- (6) *Bey b-in-illi leq na tinapay min danakan ku.*
 PERF GF-buy by she.GEN bread from sister I.GEN
 'She bought the bread from my sister.'

The preposition *maka* marks an NP that indicates instrument.

- (7) *Si Abdul bey mappot kayu maka bariq.*
 DET Abdul PERF AF-cut wood with bush.knife
 'Abdul cut wood with a bush knife.'

2.2 Personal pronouns

MNK Sinama has three basic sets of pronouns as shown in Table 2: nominative, predicative and genitive.⁴

Nominative pronouns occur as the grammatical subject of a sentence. Their function is almost the same as that of the *ang* form pronouns in Tagalog. Predicative pronouns occur as the predicate of equational or identificational sentences. They also follow all prepositions

⁴ Like many other Philippine languages, Sinama pronouns have *dual* person forms that include the one spoken to. The dual person is an independent personal category that always contains more than one person. This is why I have avoided the binary opposition of *singular* and *plural* and have used the term *minimal* and *non-minimal* instead. The minimal set refers to the minimal number of members in the set, while non-minimal refers to anything above what is required of the minimal set. For instance, the minimal set for the dual person is the speaker and the hearer. Anyone added to this minimal set is called the non-minimal set. In traditional Philippine linguistics, a pronoun in the first person non-minimal set is called the *first person plural exclusive* and the dual person non-minimal form *first person plural inclusive*. The same dichotomy as suggested here was probably first coined by Harold Conklin (Lawrence Reid, pers. comm. 1998).

other than *leq*. All of the predicative pronouns except the second person are the same as those in the nominative set. The genitive pronouns occur as the possessor in a possessive construction. In addition, the genitive pronoun may appear as the agent of a transitive construction preceded by the preposition *leq*.

Table 2: MNK personal pronouns

	Person	Nominative	Predicative	Genitive
Minimal	1	<i>aku</i>	<i>aku</i>	<i>ku</i>
	2	<i>kaw</i>	<i>kaqaw</i>	<i>nu</i>
	D	<i>kita</i>	<i>kita</i>	<i>ta</i>
	3	<i>iya</i>	<i>iya</i>	<i>na</i>
Non-minimal	1	<i>kami</i>	<i>kami</i>	<i>kami</i>
	2	<i>kam</i>	<i>kaqam</i>	<i>bi</i>
	D	<i>kitabi</i>	<i>kitabi</i>	<i>tabi</i>
	3	<i>sigala</i>	<i>sigala</i>	<i>sigala</i>

The second person predicative pronouns are exemplified below.

- (8)a. *Kaqaw iya mayad.*
2.MN.PRD the AF-pay
- b. **Kaw iya mayad.*
2.MN.NOM the AF-pay
'You are the one who is going to pay.'
- (9)a. *Ili ma kaqaw, sikeyya itu.*
that at 2.MN.PRD not this
- b. **Ili ma kaw, sikeyya itu.*
that at 2.MN.NOM not this
'That (one) is yours, not this (one).'

2.3 Word order in MNK

In Sinama, basic clauses are generally predicate-initial. There seems to be no fixed order of argument occurring after the verb as shown in (10a) and (10b).

- (10)a. *Bey niqadjal manuk leq si Abdul.*
PERF *in*cook chicken by DET Abdul
- b. *Bey niqadjal leq si Aliq manuk.*
PERF *in*cook by DET Aliq chicken
'Ali cooked the chicken.'

The subject of the transitive sentence can be topicalised (11). The subject of an intransitive can also be topicalised (12a) or remain untopicalised (12b).

- (11) *Manuk bey niadjal leq si Abdul.*
chicken PERF *in*cook by DET Abdul
'Abdul cooked the chicken.'

- (12)a. *Anak-anak bey nengge.*
 child PERF *N*-stand.up
- b. *Bey nengge anak-anak.*
 PERF *N*-stand.up child
 'The child stood up.'

A complete discussion of word order would require text analysis, as subjects are often omitted in texts. However, such a study has not yet been carried out.

2.4 MNK as an ergative language

It is not my purpose to discuss whether Sinama is syntactically ergative or accusative. In this section, however, I will discuss morphological ergativity in Sinama.⁵ Let us consider the actancy system.

- (13) *Bey iya paragan diqilaw.*
 PERF he.NOM *pa*-run yesterday
 'He ran yesterday.'
- (14) *Bey iya b-in-onoq leq si Markos.*
 PERF he.NOM *ni*kill by DET Markos
 'Markos killed him.'

In sentence (14), the patient *iya* 'he' in the transitive sentence is nominative, just as the subject in the intransitive sentence in (13). On the other hand, the agent of the transitive sentence is marked by the preposition *leq*. These data indicate that Sinama has a morphologically ergative case-marking pattern. In the following section, I will refer to an agent complement, genitive NP marked by *leq* as an 'ergative complement'.⁶

3 The {*leqN*-} construction

3.1 Sinama verbal affixes

To begin with, I will summarise the verbal affixes in Sinama that I have encountered so far.

⁵ There has been a long history of discussion about whether Philippine languages are accusative, ergative, or neither of the two, as argued by Shibatani (1988). A fuller study of the issue lies outside the scope of the paper but a few words should be said. De Guzman (1988) supports an ergative analysis (EA) of Philippine languages. Within EA, the so-called actor focus is considered intransitive or antipassive. The main reason is that both intransitive and antipassive verb forms appear to be morphologically unified in contrast to transitive ones; i.e. *mag*- and *-um*- occur with the intransitive and antipassive group and *-in*-, *-an*-, and *i*- with the transitive in Tagalog (De Guzman 1988:340-341). The EA would provide reason to think that the Sinama *leqN*- prefix derives a transitive from an intransitive *N*- verb.

⁶ The term 'nominative' is traditionally used only in the analysis of accusative languages. I will, however, employ this term, rather than 'absolute', in the present analysis.

Table 3: MNK verbal affixes⁷

	Indicative	Aptative	Imperative
Actor	<i>pa-</i> , <i>N-</i>	<i>maka-</i>	<i>N-</i>
Goal	<i>-in-</i>	<i>ta-</i>	<i>-un</i>
Locative	<i>paN-</i> <i>-an</i> , <i>-in-</i> <i>-an</i>	<i>kapaN-</i>	* <i>paN-</i> <i>-in</i>
Beneficiary	<i>-in-</i> <i>-an</i>	<i>ka-</i>	<i>-in</i>
Instrument	<i>paN-</i>	<i>tapaN-</i>	<i>paN-</i> <i>-un</i>

The symbol *N-* represents a prefixed nasal that assimilates in various ways to the initial phoneme of the stem.⁸ It simply indicates that the grammatical subject of the sentence is an actor NP (15). Some motion verbs such as *palaksu* 'to jump', *palabey* 'to pass by', *palege* 'to lie down' are marked by the prefix *pa-* as seen in the sentence (13) above. An infix *in* marks goal-focus (16).⁹

- (15) *Bey ngadjal ingkalla.*
 PERF *N-*cook bachelor
 'The bachelor cooked.'
- (16) *Bey ni-qadjal deing kurapug leq ingkalla.*
 PERF *in*cook fish grouper by bachelor
 'The bachelor cooked the grouper fish.'

The infix *in* never indicates perfective aspect, as it does in most Philippine languages, but it simply indicates goal-focus. As seen below, the *in* construction can carry a future sense.

- (17) *P-in-abillihan leq ku halong pahal.*
*in*sell by I.GEN charcoal tomorrow
 'I will sell charcoal tomorrow.'

Instead, Sinama employs auxiliary verbs to show aspect: i.e. *bey* perfective (18), *lay* imperfective (19) and *song* contemplated (20).

- (18) *Bey ni-liqis leq na pinggiq kayu.*
 PERF *in*grate by he.GEN tuber wood
 'He grated the cassava.'
- (19) *Lay na matey.*
 IMPF already *N-*die
 '(He) died already.' ('He is already in the state of being dead.')

⁷ The notation (*) indicates that this form is quoted from Pallesen (1985:99) because I have no direct data at present.

⁸ Preceding /p, b, t, s, k, (g)/, *N-* assimilates to the place of articulation of the stem initial consonant and that consonant is deleted. In other environments it has the following realisations: *ng-* occurs preceding /q/, but /q/ drops, *nga-* occurs preceding /h, l/, *ngan-* occurs preceding /d, j/ and, *ngang-* occurs preceding /g/. Some examples are as follows: *qaq > ngaq* 'to get/take (AF)'; *hinang > ngahinang* 'to work/make (AF)'; *linggiq > ngalinggiq* 'to throw a fishing net (AF)'; *doleq > ngandoleq* 'to anger (AF)'; *janjiq > nganjanjiq* 'to promise (AF)'; *gamot > nganggamot* 'to grow (AF).

⁹ It has an allomorph *ni-* preceding *l* and *q*.

- (20) *Song aku tau nganad Sinama inut-inut.*
 FUT I.NOM know.how.to *N*-learn Sinama slowly
 'I will learn to speak Sama language slowly.'

The difference between the perfective and imperfective senses is illustrated in (21).

- (21) *Lay aku lango bey nginum bir.*
 IMPF I.NOM drunk PERF *N*-drink beer
 'I am (still) drunk (because I have) drunken beer.'

3.2 {*leqN*-} construction

In MNK, there are cases where a form *leq* occurs with an actor-focus verb, which is marked by *N*-. This form is preposed to the verb, appearing similar to a prefix, and gives the sentence a perfective sense.

- (22) *Leq ngadjal leq ku manuk itu.*
leq N-cook by I.GEN chicken this
 'I have cooked the chicken.'

Compare this with (23), which is an actor-focus sentence whose prefix is symbolised by *N*-.

- (23) *Ngadjal aku manuk.*
N-cook I.NOM chicken
 'I cook chicken.'

Although the verb *ngadjal* (< *N-qadjal*) in (22) appears to be actor focus, marked by *N*-, the grammatical subject in the sentence is not the actor. The agent is marked by the preposition *leq* while the unmarked grammatical subject is *manuk* 'the chicken'. It should be noted that the notional object in this sentence is definite whereas the notional object of an actor-focus *N*-verb is indefinite as in (23).

The most appropriate interpretation of (22) in English would be 'I have cooked the chicken', but not 'I cooked the chicken', 'I cook the chicken', or 'I will cook the chicken'. The perfective interpretation is apparently triggered by the prefix-like form *leq* preceding the *N*-marked verb. This suggests that the verbal affix is made up of *leqN*-.

This hypothesis seems to be supported by examples (24) and (25), showing a modification relation between a noun and a verb. Thus, when a verb modifies a noun, *leqN*- functions like a past participle in English.

- (24) *deing leqmila*
 fish *leq-N*-split
 'split fish'
- (25) *Bilahi aku isi sapi leqngalunok.*
 like I.NOM flesh cow *leq-N*-soften
 'I like tenderised beef.'

If *leqN*- functions as a real prefix, it should not occur separated from the verbal stem. Consider the following examples. The perfective aspect marking auxiliary verb *bey* can occur with the *leqN*- construction as in (26a). Example (26b) indicates that *leqN*- is a combined

verbal form. Thus, the form *leq* is to be interpreted as a prefix, which attaches to the *N-* stem or *leqN-*.

- (26)a. *Bey leqmong siya leq si Abdul.*
 PERF *leq-N-break* chair by DET Abdul
 'Abdul has broken the chair.'
- b. **Leq bey mongsiya leq si Abdul.*

There are sentences, however, which appear to contradict the foregoing observation. In sentence (27), *leq* seems to be separated from the verb by clitic pronoun *ku*.

- (27) *Leq ku ngadjal manuk.*
leq I.NOM *N-cook* chicken
 'I have cooked the chicken.'

One might assume that *leq* in (27) is a kind of auxiliary verb, since a clitic pronoun follows right after it.¹⁰ However, there are data that indicate the *leq* in (27) is neither a verbal prefix nor an auxiliary verb, but simply a preposition. In sentence (28), a proper noun phrase occurs between *leq* and the verb. An auxiliary verb allows only pronouns to be cliticised to it and not a proper noun, nor a common noun.

- (28) *Leq si Abdul mong siya.*
 by DET Abdul *N-break* chair
 'Abdul has broken the chair.'

Sentences like (27) and (28) thus show the same case marking pattern as (22), an unmarked patient and an ergative complement with an actor-focus verb. What is missing is the *leq-* prefix. I argue that (27) and (28) are the same construction as (22), and that the missing prefix can be explained as follows. Sentence (27) may be derived from sentence (29a) by preposing the ergative complement *leq ku* 'by me' before the verb. In such cases, the verbal prefix *leq-* is obligatorily dropped to yield sentence (27) as indicated in (29b) and (29c).

- (29)a. *Leqngadjal leq ku manuk.*
 b. *Leq ku ngadjal manuk.* (= 27)
 c. **Leq ku leqngadjal manuk.*
 'I have cooked the chicken.'

The grammatical subject, *manuk* 'chicken', can be topicalised as shown in (30a) without affecting the *leq-* prefix. When an ergative complement precedes the verb, the prefix *leq-* must be dropped as shown in (30b) and (30c).

¹⁰ This phenomenon is widely observed in most of the Philippine languages. In basic sentences, an enclitic pronoun normally follows the first word of the predicate clause. The following pairs of examples from Tagalog quoted from Schachter and Otnes (1972:183), indicate that the third person singular genitive pronoun *niya* obligatorily occurs right after the verb.

- a *Nakita niya si Ben.* 'He saw Ben.'
 a' **Nakitasi Ben niya.* 'He saw Ben.'
 b *Hindi niya nakita si Ben.* 'He didn't see Ben.'
 b' **Hindinakita niya si Ben.* 'He didn't see Ben.'

- (30)a. *Manuk leqngadjal leq ku.*
 b. *Manuk leq ku ngadjal.*
 c. **Manuk leq ku leqngadjal.*

Dropping of the verbal prefix happens whenever an ergative complement is preposed to the verb, as in (31a), where the ergative complement is a proper noun.

- (31)a. *Leq si Jam ngadjal manuk.*
 by DET Jam *N*-cook chicken
 b. **Leq si Jam leqngadjal manuk.*
 'Jam has cooked the chicken.'

Note that in the usual goal-focus sentence, an ergative complement cannot be topicalised as shown in (32a). It must occur in post-verbal position (32b). This is one of the characteristics that makes the *leqN*- construction unique.

- (32)a. **Leq si Abdul bey sinipaq kambing.*
 by DET Abdul PERF *in*kick goat
 b. *Bey s-in-ipaq leq si Abdul kambing.*
 PERF *in*kick by DET Abdul goat
 'Abdul kicked the goat.'

The synchronic general rule goes something like this.¹¹ An ergative complement, once preposed to the verb, will trigger the deletion of the prefix *leq*-. It is evident, therefore, that *leq*- can only be a verbal prefix attached to the *N*- stem and this is the reason I call the construction a 'derived' transitive sentence.

In both *leqN*- and *in* constructions, the agent can be omitted.¹² But, as observed above, topicalisation of the ergative complement in the *leqN*- construction is unique in that it requires that the prefix *leq*- be omitted. Thus, the *leqN*- construction is syntactically different from the goal-focus sentences.

4 Semantic functions

In this section I investigate some of the features of *leqN*- constructions from the standpoint of scalar transitivity proposed by Hopper and Thompson (1980). Of the ten criteria for transitivity which they proposed,¹³ I will discuss four.

¹¹ It is not my purpose to speculate how the construction is diachronically derived. The *leq*- prefix and the *leq* preposition are possibly both derived from an earlier verb via different grammaticalisation paths.

¹² Comrie (1988:18) states that "in the prototypical cases, the agent is not omissible in the ergative construction, but is omissible and is in fact normally omitted from the passive construction." Shibatani (1988:91-94) also states that in the passive sentences, the agent is normally dropped but in the Philippine type goal-focus sentences, the agent is less likely to be omitted than is observed in the passive constructions of other languages. He claims that this is one of the important differences between Philippine goal-focus constructions and passive constructions. I have only limited information from text analysis but my impression is that both *in* type construction and *leqN*- type construction normally require the ergative complement to appear.

¹³ These criteria are (a) participants, (b) kinesis, (c) aspect, (d) punctuality, (e) volitionality, (f) affirmation, (g) mode, (h) agency, (i) affectedness of O(object), and (j) individuation of O.

4.1 Aspect

As previously noted, the construction is perfective.

- (33) *Kakan-un intollo ku. Leqmalla na.*
 eat-IMP egg I.GEN *leq-N-cook* already
 'Eat some of my eggs. (They are) already boiled.'

4.2 Mode

The *leqN-* construction seems to be restricted to realis mode. Thus, only adverbs with a past sense can occur in *leqN-* construction (34) and (35).

- (34) *Leq ku milli iti ma Sambuwangan diqilaw.*
 by I.GEN *N-buy* this at Zamboanga yesterday
 'I bought this one at Zamboanga yesterday.'
- (35) **Ayyan leq nu milli pahalū?*
 what by you.GEN *N-buy* tomorrow
 'What are you going to buy tomorrow?'

4.3 Affirmation

The construction seems not to occur in negative sentences. This may bear some relation to mode as discussed above. This would explain why *leqN-* construction cannot be negated by the auxiliary verb *maha* (36). In the case of negatives, the *<in>* type goal-focus construction will be employed (37).

- (36) **Maha leqmong siya leq si Abdul.*
 not *leq-N-break* chair by DET Abdul
 'Abdul did not break the chair.'
- (37) *Maha bey p-in-ong siya leq si Abdul.*
 not PERF *<in>break* chair by DET Abdul
 'Abdul did not break the chair.'

4.4 Affectedness of patient

The grammatical subjects of *leqN-* constructions are interpreted as being totally affected while the grammatical subjects of the 'goal-focus' sentences are only partially affected.

- (38) *Leqmangan leq kuting.*
leq-N-eat by cat
 'The cat has eaten (it) up.'
- (39) *K-in-akan leq kuting.*
<in>eat by cat
 'The cat ate (some of it).'

Table 4 compares the two constructions with respect to semantic transitivity.

Table 4: Semantic transitivity

	Aspect	Mode	Affirmation	Affectedness
<i>leqN-</i>	+telic	+realis	+affirmative	total
<i>in</i>	-telic	-realis	-affirmative	partial

From this table, it can be seen that these constructions are both syntactically transitive but they differ in degrees of semantic transitivity. The *leqN-* construction is higher in transitivity than the *in* infix transitive one.

5 Concluding remarks

I have described two kinds of transitive constructions in Sinama. Of these two, the *leqN-* type construction is higher in semantic transitivity than the *in* infix type construction. The present descriptions of these constructions are entirely synchronic. I will comment here on three problems for future study:

- (a) text analysis is needed,
- (b) the dialectal distribution of the construction is not well known, and
- (c) a diachronic explanation of the evolution of the constructions is necessary.

Quantitative text analysis is necessary to clarify the situations in which the *leqN-* construction appears instead of the *in* construction. Whether the preferred position of the ergative phrase is preverbal or postverbal also requires text analysis.¹⁴

Cross-dialectal distribution of the construction needs investigation because not all Sinama dialects have this construction. For example, Sitangkay Sinama does not have the *leqN-* construction. It employs only the infix *in* for all transitive expressions. Both Sikubung Sinama and Sapa-Sapa Sinama have the prefix *leqN-* together with the infix *in*. Texts from the Sibutu Sinama indicate that it also has the *leqN-* construction as well as *in* infix type transitive sentences (Allison 1977). Though Pallesen (1985) gave no clear statement on differences between Central Sulu Sinama and Southern Sulu Sinama, it seems to me that they form a 'dialect chain'. The *leqN-* construction, however, will be one of the criteria which distinguishes the two.

Surprisingly, the *leqN-* construction is also found in Indonesian Sinama around Dondo Bay in the Buol-Tolitoliarea (40) and (41).¹⁵

¹⁴ As for productivity, it appears that the *leqN-* construction is a productive sentence pattern in MNK. It also occurs in causative sentences.

Leq ku maragan ma anak-anak.
leq I.GEN *N-pa*-run at child
 'I have made the child run.'

¹⁵ I conducted research on Indonesian Sinama in twenty-one speech communities. Only two of them have the *leqN-* prefix. All the others have the prefix *di-* that appears to function like the *in* infix in the Philippine-type Sinama.

- (40) *Dayah ore lemarikkit le anaq ku.* (Santigi)
 fish that *le-N-cook* by child I.GEN
 'My child has cooked that fish.'
- (41) *Munihi ore lengadakaq leq polisi.* (Labuan Lobo)
 person that *le-N-catch* by policeman
 'Policeman has caught that man.'

I am not yet sure whether there is complementary distribution between the *leqN-* construction and the goal-focus construction in Indonesian Sinama (the Indonesian Sinama *di-* prefix being equivalent to the *<in>* infix in Sinama) as observed in MNK. Note that MNK and Indonesian Sinama are different languages — speakers from each speech community cannot understand speakers from the other. However, the presence of the *leqN-* construction in two speech communities in Northern Sulawesi suggests a period of close contact between the speakers of Southern Sulu Sinama and some speakers of Indonesian Sinama after the separation of the two languages. Further study of Indonesian Sinama would be useful, for socio-linguistic and descriptive-linguistic purposes.

References

- Akamine, Jun, 1996, A grammatical analysis of Manuk Mangkaw Sinama. PhD dissertation, University of the Philippines, Diliman.
- Allison, E. Joe, 1977, Discourse analysis of a southern Samal text: Abunawas and the beautiful widow. *Studies in Philippine linguistics*, 1:143-69.
- Comrie, Bernard, 1988, Passive and voice. In Shibatani, ed. 1988:9-23.
- De Guzman, Videa P., 1988, Ergative analysis for Philippine languages: an analysis. In Richard McGinn, ed. *Studies in Austronesian linguistics*, 323-45. Athens: Ohio University.
- Hopper, Paul J. and Sandra A. Thompson, 1980, Transitivity in grammar and discourse. *Language* 56:251-299.
- Pallesen, A. Kemp, 1985, *Culture contact and language convergence*. Manila: Linguistic Society of the Philippines.
- Schachter, Paul and Fe Otones, 1972, *A Tagalog reference grammar*. Los Angeles: University of California Press.
- Shibatani, Masayoshi, 1988, Voice in Philippine languages. In Shibatani, ed. 1988:85-142.
- Shibatani, Masayoshi, ed., 1988, *Passive and voice*. Typological studies in language 16. Amsterdam: John Benjamins Publishing Co.

Some aspects of 'focus' in Sama Bangingi'

JOANN GAULT

1 Introduction¹

Sama Bangingi' (also called Balangingi') is one of several Sama languages which are spoken throughout the Sulu Archipelago in the south-western Philippines. The Bangingi' language is spoken from the Samales Island Group south of Basilan Island north to the coastal areas and islands of Basilan and the southern Zamboanga Peninsula including the coastal areas of Zamboanga del Sur. It is a member of the Sama-Bajaw subgroup of Malayo-Polynesian, which includes also the Sama-Bajaw languages of Sabah and Sulawesi (Pallesen 1985:2-3).

Bangingi' is an ergative language which shares many of the features of 'focus' that are characteristic of Philippine languages, but which also evidences some features of grammar and discourse similar to Bahasa Malay, especially the language as it was spoken in the last century as analysed by Hopper (1983). Historically, the Bangingi' have had extensive contact with speakers of Bahasa Malay, being traders and pirates until this century, ranging in their exploits from Sulu to Singapore and east to the Moluccas (Pallesen 1985:9).

2 Case marking

2.1 Pronouns

The Sama Bangingi' nominal system distinguishes three cases: absolutive, ergative and oblique.² These cases are most easily seen in the three pronoun sets given in Table 1 below.

¹ This understanding of focus (as yet incomplete) is a result of contact with and study of the language which began in 1974. Many Bangingi' have had input into that study. The examples in this paper have been given or checked by Mr Abdulmashir (Bong) B. Kasim of Taluksangay, and Mrs Jana Amping of Campo Islam, both in Zamboanga City.

² Abbreviations used: ABS absolutive case; ERG ergative case; GEN genitive case; OBL oblique case; AF agent focus; PF patient focus; RF recipient focus; IF instrument focus; LF location focus; SG singular; DU dual (i.e. 1+2SG person); PL plural; ABL abilitative mood; CP completive particle; DET determiner;

Table 1: Pronoun sets

Person	Singular			Plural		
	ABS	ERG	OBL	ABS	ERG	OBL
1	<i>aku</i>	<i>-ku</i>	<i>ma-aku</i>	<i>kami</i>	<i>-kami</i>	<i>ma-kami</i>
1+2	<i>kita</i>	<i>-ta</i>	<i>ma-kita</i>	<i>kitabí</i>	<i>-tabí</i>	<i>ma-kitabí</i>
2	<i>kaa</i>	<i>-nu</i>	<i>ma-kaa</i>	<i>kaam</i>	<i>-bi</i>	<i>ma-kaam</i>
3	<i>iya</i>	<i>-na</i>	<i>ma-íya</i>	<i>sigaaam</i>	<i>-sigaaam</i>	<i>ma-sigaaam</i>

The absolutive pronouns are free standing pronouns and function as the ‘focussed’ noun phrase, which has been variously called in Philippine linguistics the ‘(syntactic) topic’, ‘syntactic pivot’, or ‘subject’.

The ergative pronouns are phonologically bound to the preceding morpheme. They function both as agents of transitive verbs (i.e. non-agent focus) and as possessive pronouns. As agents they are cliticised either to the verb root or to the ergative particle *e’ni*, which occurs with certain inflected verbs. As possessive pronouns they are bound to the noun which they modify. In the examples below the possessive pronouns will be labelled as genitive.

The oblique pronouns consist of the general purpose oblique marker *ma* and the free standing pronouns. The marker plus pronoun is generally pronounced as one phonological word. Free standing pronouns may also be oblique when they occur in a prepositional phrase, that is, the general purpose *ma* may be replaced by a more specific preposition such as *ni* ‘to, towards’, or *min* ‘from’.

2.2 Nominal case markers

An absolutive full noun phrase is more often than not unmarked morphologically; however, there is an optional preposed particle *in* which appears to have been borrowed from Tausug, the prestigious trade language of the Sulu Archipelago. This particle tends to occur in more formal speech, especially to introduce a topicalised or fronted noun phrase. It also occurs frequently on a fronted noun phrase in an embedded clause, thereby setting that clause off from the main clause.

Ergative full noun phrases are always marked by the ergative particles *e’ni*. Oblique full noun phrases may be marked with the all purpose *ma* or with prepositions *ni* ‘to’, *min* ‘from’, *maka* ‘with’ or *sampay* ‘including’. Non-focussed patients are generally less specific than a focussed patient, and may be marked as oblique by the morphology.

3 Transitive constructions

3.1 Focus affixation

There are five focus or voice orientations in Sama Bangingi’ transitive clauses: agent, which is the semi-transitive antipassive, and four true transitive voices: patient, recipient,

EX existential; IMP imperative mood; IND indicative mood (basic); IND-P indicative with *ni/-in-* affix; INT intransitive; P passive or passive-like; PM person marker; PST past (i.e. previous to the time of the utterance); RD reduplication; ST stative.

instrument and location. The affixes employed by these transitive voices, along with the intransitive affixes, are listed below in Table 2. For the purposes of this paper, only the indicative forms will be referred to, except for a brief discussion in §3.2.

Table 2: Focus affixes

	Intransitive	Antipassive		Transitive		
		Agent	Patient	Recipient	Instrument	Location
IND	<i>aN</i> ³ <i>pa-</i> <i>-um-</i>	<i>aN-</i>	\emptyset	<i>-an</i>	<i>paN-/</i> <i>pag-</i>	<i>paN--an/</i> <i>pag--an</i>
IND-P			<i>-in-</i>	<i>-in--an</i>		
ABL	<i>maka-</i>	<i>maka-</i>	<i>ta-</i>	<i>ta--an/</i> <i>ka--an</i>	<i>tapaN-/</i> <i>tapag-</i>	<i>tapaN--an/</i> <i>kapag--an</i> ⁴
IMP	<i>aN-</i> <i>pa-</i>	<i>aN-</i>	<i>-un</i>	<i>-in-un</i>		

The mapping of semantic role onto grammatical focus is generalised as follows: the agent/experiencer NP is the 'focussed NP' of agent focus (AF); the patient-like/undergoer NP is the 'focussed NP' of patient focus (PF); the recipient/beneficiary/goal (in Fillmore's sense as the end point of the action) is the 'focussed NP' of recipient focus (RF); the instrument/means is the 'focussed NP' of instrument focus (IF); the spatial and logical location is the 'focussed NP' of location focus (LF). Verbal roots affixed for location focus and, to a lesser extent, for instrument focus often function as nominal arguments.

Note that the location focus affix appears to be a combination of instrument and recipient affixes. There is often a semantic correlation to this, as the location of an action is often just as much a means to the accomplishment of the action as is an instrument. The choice between *paN-* and *pag-* in instrument and location is frequently that between a single action and habitual, durative or reciprocal action. According to Pallesen (1985:117), **paN-* was Proto Sama-Bajaw, with the *paN-/pag-* distinction coming as a result of convergence between the Sama languages and Tausug. Indeed, the instrumental and locative use of the affix *pag-* seems to be subsidiary to its durative/reciprocal function. As a marker of durative/reciprocal function, it frequently occurs as an aspectual marker along with agent, patient and recipient focus affixations. Following are examples of the basic indicative focus affixes with the root *kehet* 'to cut/slice'. The absolutive or 'focussed' NP is in bold type. The morphophonemics of N will be specified for this set of examples only.

3 The phonological derivations of N are: $aN \rightarrow$

<i>aNi-</i> / -- <i>C_i</i>	<i>C_i = p, b, t, k</i>
<i>angaN_j-</i> / -- <i>C_j</i>	<i>C_j = d, g</i>
<i>anga-</i> / -- <i>C_k</i>	<i>C_k = l, h</i>
<i>ang-</i> / --V	

$$C_i \rightarrow \emptyset / aNi-$$

aN+allang → *angallang*;

aN+bowa → *amowa*; *aN+sulat* → *anular*; *aN+kehet* → *angehet*;

aN+dakdak → *angandakdak*; *aN+geret* → *angangeret*;

aN+lingan → *angalingan*; *aN+hinang* → *angahinang*

4 *Ka--an* seems to be replacing *ta--an* in RF and LF forms, whether or not the LF is *paN-* or *pag-*.

- (1) *Kehet-ku mampallam itu maka laring-laring si Kakah.*
 PF.cut-1SG.ERG mango this with small.knife PM Kakah
 'I'll cut **this mango** with Kakah's small knife.'
- (2) *Kehet-an-ta kaa mampallam.*
 cut-RF-1DU.ERG 2SG.ABS mango
 'I'll cut **you** some mango.'
- (3) *Laring-laring si Kakah ya pangehet-ku mampallam itu.*
paN-kehet-ku
 small.knife PM Kakah DET IF-cut-1SG.ERG mango this
 '**Kakah's small knife** is what I'll cut this mango with.'
- (4) *Laring-laring itu pag-kehet-ku mampallam sakahaba' aku*
 small.knife this IF-cut-1SG.ERG mangoes whenever 1SG.ABS
abaya' amangan.
 desire eat
 '**This small knife** is what I cut mangoes with whenever I want to eat (them).'
- (5) *Sainggahan pangehet-ku bang insa'-niya' laring-ku?*
paN-kehet-ku
 how IF-cut-1SG.ERG when not-any knife-1SG.GEN
 'How can I cut (it) when I have no knife?'
- (6) *Papan-papan itu ya pag-kehet-an-ku sinayul.*
 small.plank this DET LF-cut-LF-1SG.ERG vegetables
 '**This piece of wood** is where I'll cut up vegetables.'
- (7) *Ya pag'-isun-an-sigaam bang sainggahan pangollo'-sigaam*
paN-kollo'-sigaam
 DET LF-discuss-LF-3PL.ERG if how IF-fetch-3PL.ERG
ma aa pananambal
paN-taN-tambal.
 OBL person IF-RD-medicine
 'What they discussed was how they could fetch a healer.'
- (8) *Kamaya' kaa angehet mampallam pasal laring-u ato:m.*
aN-kehet
 be.careful 2SG.ERG AF-cut mango because knife-that sharp
 '**You** be careful cutting mangoes because that knife is sharp.'

Example (1) is patient focus; example (2) is recipient focus. Note that the absolutive NP, *mampallam itu* 'this mango' in (1) has a specific referent, while *mampallam* in (2) does not. In general the 'focussed NP', called hereafter the 'subject', must be specific and referential, that is, information which is recoverable by the hearer from the linguistic or situational context. When the recipient/beneficiary is the subject, the patient may or may not be specific.

The recipient focus *-an* may have as its 'subject' either an inanimate goal of the action or an animate beneficiary. In example (16) below, the goal of the action is the receptacle 'jar'. In (17) the goal is an animate recipient/beneficiary. Notice that when an animate recipient is the subject of a RF verb, the inanimate receptacle is expressed as an oblique NP. Note also that the word order changes.

Examples (3) and (4) above are instrument focus. The distinction between *paN-* as expressing a single action in (3), and *pag-* as expressing habitual action in (4) is clear. As in these examples, instrument focus constructions are frequently cleft-like, since one of the main functions of IF is to specify the instrument used in the action. Thus instrument focus frequently contains an element of contrast, either implied as in (3) or explicit as in example (19) below. *Pangehet* in (5) and *pangollo'* in (7) are IF which express means rather than an implement. *Pananambal* in (7) is IF affixation plus reduplication of the initial CV of the root. It indicates a person whose occupation is expressed in the root.

Example (6) is location focus. Note that the habitual *pag-* is used rather than *paN-*. *Pag'isunan* in (7) is an example of LF affixation used as a logical location, in this case the topic of discussion. Example (8) is agent focus. Here it is the action itself that is important (and its effect on the agent), rather than its effect on the patient. Thus agent focus has been called semi-transitive, because although a patient (whether specific or not) may be present in the sentence, the degree to which it is affected by the action of the verb is often unimportant.

3.2 Focus and ergativity

Sama Bangingi' is an ergative language with the agent focus functioning as an antipassive. The agent focus is often used to introduce participants or props into a discourse, that is, when the patient is first mentioned. It is also used when the patient is non-specific or non-referential, or deleted entirely. Where the patient is specific, the agent focus is often used to express incomplete or partitive action. As illustrated in the preceding example, it is also used when the speaker is talking about the action itself rather than the effect of the action on the patient. The agent focus construction is the only focus which significantly reduces the topicality of the patient. The patient is often expressed as an oblique NP in agent focus, always so when it is a pronoun.

The other focus constructions are transitive. The high degree of transitivity is clear in patient and recipient focus. These are the most frequent constructions and are employed when the patient and/or recipient are specific and given information. The degree of transitivity is admittedly not as clear in cases of instrument and location focus. But all four transitive focuses/foci pattern alike in several ways. They all employ the ergative agent, and they all share the same set of aspect and mood affixations, as opposed to agent focus which shares its affix set with intransitive verbs. This distinction can be seen in Table 2 for 'abilitative'⁵ (ABL) and 'imperative' (IMP) moods. In addition to the morphological difference, the abilitative affixes are 'added onto' the four transitive indicative (IND) forms, but supplant the indicative affixes for intransitives and agent focus. Note that ABL and IND-P forms are mutually exclusive. Moods other than the two indicative ones are not employed in the examples in this paper.

⁵ 'Abilitative' mood is the label given to the affixation which communicates that the action of the verb was able to be accomplished, or indeed was accomplished, whether or not the action was intended by the agent.

3.3 The affix *ni-/in-*⁶

The verbs in the examples above are the basic indicative mood, that is, they are inflected only for focus. When the basic indicative PF or RF is used, an agent is obligatory, and the agent must be an ergative pronoun which is phonologically attached to the verb (the bound pronoun usually occurs with IF as well, but it is not obligatory). The basic indicative plus ergative pronoun is the preferred construction when the agent is first or second person, and is used when the speaker is referring to something in the immediate situational context. When the speaker is referring to something outside of the immediate context, i.e. relating a past event or speaking of a future or hypothetical event, he is more likely to use an affixed form of the verb. The most common affix used is *ni-/in-*. This has been labeled IND-P, because as will be seen, this affix is passive-like if not always indicative of a true passive. When the agent is a third person pronoun, the verb affixed with *ni-/in-* is preferred to the unaffixed form but not required. When the agent is a full noun phrase, however, this affix is required.⁷ When the verb is affixed with *ni-/in-*, the agent, whether pronoun or noun, must be marked with the ergative particle *e'/ni* (reflecting dialectal variation). The following examples illustrate this usage. The affix is glossed as 'P'.

- (9) *Bay na bowa-ku palauk-in pe' ma-iya.*
 PST CP bring.PF-1SG.ERG viand-the there OBL-3SG
 'I have already taken the viand there to her.'
- (10) *Bay na b-in-owa e'-na palauk-in pe' ma-iya.*
 PST CP P-bring.PF ERG-3SG.ERG viand-the there OBL-3SG
 'He has already taken the viand there to her.'
- (11) *Bay na b-in-owa e' si Inah palauk-in pe' ma-iya.*
 PST CP P-bring.PF ERG PM mother viand-the there OBL-3SG
 'Mother has already taken the viand there to her.'
- (12) **Bay na bowa e' si Inah palauk-in pe' ma-iya.*
- (13) *Bay na b-in-owa si Inah pe' ni doktor.*
 PST CP P-bring.PF PM mother there to doctor
 'Mother has already been taken there to the doctor.'
- (14) *Bay na iya b-in-owa-han e' si Inah palauk.*
 PST CP 3SG.ABS P-bring-RF ERG PM mother viand
 'Mother has already taken her some viand.'
 'She has already been taken some viand by Mother.'

The affix *ni-/in-* often functions as a passive. This is clear in sentences like (13) which do not contain an agent. When an agent occurs, as in (14), it is less clear whether the sentence should be considered transitive or passive. It cannot be that all *ni-/in-* constructions are true passives, because these constructions (and similar ones with *ta-*) are the only transitive options

⁶ *ni-* occurs regularly with vowel initial roots or those that begin with /l/ or /h/. Some speakers prefer *ni-* as well with /d/ initial roots. The infix occurs regularly elsewhere.

⁷ Rather, some inflection is required on the verb. It does not need to be this affix. Verbal affixation is required whenever the ergative *e'/ni* is used, whether the agent is a noun or pronoun. However, for the purposes of this paper, affixations other than *ni-/in-* which do not affect focus will be ignored.

when the agent is a full noun phrase. Word order may be involved in deciding whether a particular sentence is transitive or passive. This is an area for further study.

3.4 Some complexities in the semantics of focus

There is considerable complexity in the 'assignment' of grammatical focus to express the various semantic roles, which is not captured in the Table 2 above. Two of these complexities will be discussed here.

3.4.1 The function of paN-

Below are examples using the root *isi*, which is both the noun 'contents' or 'flesh', and the verb 'to put in' or 'to fill with'. Again, the 'subject' is in bold type.

- (15) *Mani' itu subay ni-isi ni-deyom garapun.*
 peanuts this must (in)-PF.put.in to-inside jar
 'These **peanuts** should be put in a jar.'
- (16) *Isi-han-ku mani' garapun itu.*
 put.in-RF-1SG peanuts jar this
 'I'll fill **this jar** with peanuts.'
- (17) *Isi-han-ta kaa mani' ma-deyom garapun itu?*
 put.in-RF-1DU 2SG.ABS peanuts OBL-inside jar this
 'Shall I put some peanuts in this jar for **you**?'
- (18) *Mainggahan na garapun-in? Bay na pang-isi-han-ku mani'.*
 where CP jar-the PST CP LF-put.in-LF-1SG peanuts
 'Where is that **jar**?' Answer: 'I've already filled it with peanuts.'
- (19) *Duma'in tangan-ku ya bay pang-isi-ku mani'*
 not hand-1SG.GEN DET PST IF-put.in-1SG.ERG peanuts
ni-deyom garapun sumaguwa' suru'.
 to-inside jar but spoon
 'What I'll use to put the peanuts in the jar is not my **hand** but a **spoon**.'
- (20) *Ya pang-isi-ku ma garapun inaan mani'.*
 DET IF-put.in-1SG OBL jar that peanuts
 '**Peanuts** is what I put in that jar.'
- (21) *A-niya' itu batu-batu pang-isi-ku ma lowang pag-la-labay-an-in.*
 ST-EX this small.rock IF-put.in-1SG OBL hole LF-RD-pass.by-LF-the
 'I have here **gravel** with which I will fill the hole in the path.'

The same NP may be the 'subject' of different focus orientations, and still retain the same role in the situational context. Examples (15)–(20) are all about the same event, putting peanuts into a jar. Strictly speaking, 'peanuts' remains the patient in all, for it is that which changes location, but in (20), it is coded for a different role, that of instrument. This sentence is said to be in answer to the question, "What did you put in the jar?" Perhaps the jar was not originally intended for peanuts, but on this occasion, that which was used to fill it was peanuts. In this sense, then, 'peanuts' is a type of instrument because it is used to accomplish

the action of filling. It is what is 'used for' the patient. Thus, the patient-like referent in a particular situation may be expressed as the 'subject' of either a PF or an IF verb. The former expresses what happens to the patient, as in (15); the latter specifies exactly what constitutes the patient and contains an element of contrast (20).

This second function of IF is very common. It is further illustrated in (21) where gravel is used (IF) to fill a hole. The 'subject' of an IF verb, then, may either be the true instrument, i.e. the implement used to perform the action as in (19), or the item(s) used as the patient in accomplishing it. In a very small class of verbs, the IF has completely taken over the function of PF. Among this class are *ba:* 'to tell' and *buwan* 'to give'. The verbal root never occurs by itself, nor accompanied only by the affix *ni/-in-*. Since IF affixation of these verbs functions as PF, there is no true instrument focus. See examples (22) to (24) below.

(22) **Buwan-ku *Binuwan e'-ku relas-ku ma-iya.*

(23) *Pamuwan-ku relas-ku ma-iya.*
 PF.give-1SG.ERG watch-1SG.GEN OBL-3SG
 'I'll give my watch to him.'

(24) *Ya bay p-in-amuwan e'-na ma-aku tinggal dakayu' pilak.*
 DET PST PF-P.give ERG-3SG.ERG OBL-1SG only one peso
 'The only thing he gave me was one peso.'

Just as instrument focus may also choose the 'patient' as its subject, so location focus may also choose the goal. In examples (16) and (18) above, the NP *garapun* 'jar' is the subject of a RF verb in the first, and of a LF verb in the second. The distinction between the two is similar to that seen in IF. In the first, the jar is the recipient of the action. In the second, it is being used as a container for peanuts as opposed to a container for something else.

3.4.2 -an verbs

The class of verbs labelled here '-an verbs' are those in which the RF affixation has taken over the function of PF. As with some IF verbs, neither the unaffixed verb root nor the root affixed with *ni/-in-* alone occurs. Most '-an verbs' belong to one of three subclasses, all of which may be related semantically. One subclass comprises verbs which express motion away from the agent. Included here are 'remove', 'throw' and 'throw away'. The 'focussed NP' or 'subject' is that which is moved away from the agent or other point of reference. The second subclass comprises verbs whose 'subject' is decreased or increased in some way. Included here are verbs such as 'shorten' and 'added-to'. Not included are derived causatives. The third subclass includes most verbs that deal with cleaning. The 'subject' is that which is laundered, scrubbed and hung out to dry. It is clear that the 'subjects' of these verbs are true patients, all having undergone a process of one sort or another. But it may be that they are also being viewed as the end point of the action. Examples using *kose* 'to wash' follow:

(25) *Bay na kose'-an-nu la:y-in?*
 PST CP wash-PF-2SG.ERG plate-the
 'Have you washed **that plate**?'

(26) *Bay na k-in-ose'-an la:y-u.*
 PST CP P-wash-PF plate-that
 '**That plate** has already been washed.'

- (27) *Kose'-an-ta kaa la:y pamanganan-nu.*
wash-PF-1DU.ERG 2SG.ABS plate LF.eat-2SG.ERG
'I'll wash for you a **plate for you to eat from.**'
- (28) *Bay na kaa ka-kose'-an-ku la:y pamanganan-nu.*
PST CP 2SG.ABS PF.ABL-wash-1SG.ERG plate LF.eat-2SG.ERG
'I've already washed a **plate for you to eat from.**'
- (29) *Itu na semmek panapu la.y.*
here CP cloth IF.wipe plate
'Here is the **cloth** for wiping dishes.'
- (30) *Angose' kita tangan-ta bang song amangan.*
AF.wash 1DU.ABS hand-1DU.ERG when soon AF.eat
'**We** should wash our hands before eating.'

Note that the *-an* suffix is not a derivational affix. It does not create a new stem upon which the focus affixation could be built, since it does not appear on either instrument focus (29) or agent focus (30). Rather, *-an* has assumed the function of patient focus for a certain set of verbs. In (29), the *-an* verb, *kose'* 'wash' has been replaced by *sapu* 'wipe'.

What then of recipient focus? Is RF possible with *-an* verbs? No, because the recipient is not the subject (i.e. 'focussed NP'), but an extra argument which is allowed in the clause as a non-oblique argument. In example (27), the beneficiary *kaa* 'you' is absolutive, but it is not the subject of *kose'an*. If the sentence is passivised, i.e. the agent is deleted and the verb is affixed with *-in-*, *la:y* 'plate' is always the subject as in (26). In (28), the passive form was not allowed with an absolutive beneficiary. The abilitative mood was substituted, but here as well, 'plate' is the subject. It may be an inelegant analysis, but it appears that in limited contexts, *-an* verbs will allow an 'extra' absolutive argument, namely the recipient/ beneficiary, into the clause, but this argument will never be the subject, that is, it will never truly be 'focussed' by the verbal affixation.

When the recipient/beneficiary is a full noun phrase, the preference seems to be to topicalise it, that is, remove it entirely from the clause, as in (31), although (32) is also possible. Topicalisation or fronting is a common strategy in Bangingi' for relieving the load of too many arguments in a clause. It seems that three arguments are the maximum that may follow a verb comfortably, and one or two arguments is the preference.

- (31) *Si Inah, bay b-in-illi-han e' si Nanah taumpa'.*
PM mother PST P-buy-RF ERG PM Nanah shoes
'**Mother**, (she) has been bought shoes by Nanah.'
- (32) *Binilli-han e' si Nanah si Inah taumpa'.*
(in).buy-RF ERG PM Nanah PM mother shoes
'Nana will buy **mother** some shoes.'

4 Intransitive and stative affixes

This section will give only a brief overview of intransitive and stative affixes, designed for comparative purposes only. A full analysis would need to take into account the complex and fascinating study of verbal semantics in the Sama languages. This analysis still remains to be done for Bangingi'. For one approach, the reader is referred to Walton (1986).

There are three active indicative intransitive affixes in Bangingi': *aN-*, *pa-*, *-um-*, and for some roots, *a-*. The verbs which employ the first two are mutually exclusive.⁸ The first subset includes verbs such as *anengko* 'sit' and *anengge* 'stand'. The second is a much larger class and includes motion verbs such as *palangngan* 'walk' and *pasakat* 'go up', and position verbs such as *pakkom* 'lying face down' as well as many others. The third affix *-um-* occurs with some of the motion verbs, e.g. *lumangngan* 'walk' and *sumakat* 'go up', and some of the '*a*-verbs' that have a more intransitive sense such as *atacca/tumakka* 'arrive' and *ahabba/humabba* 'topple over'.⁹

The frequency of the *-um-* affix varies considerably with the regional dialect and the age of the speaker. It seems to be more evident in the speech of the older generation, but it is not frequent in any case. In some areas its use is limited to a very few roots in a limited context. Its use is so limited that I have not been able to determine a clear difference in meaning from the same verb affixed with *pa-* or *a-*.¹⁰ It may be a reflex of PAn **-um-* whose use has been restricted to intransitive verbs. At one time it may have served as both an intransitive and agent focus affix as *aN-* does now. However, it may also have come into Bangingi' through convergence with Tausug which employs *-um-* both as an agent focus affix (see Hassan et al., 1994:19) and as an intransitive affix.

The affix *a-* may more properly be described as stative than as intransitive and has been (tentatively) so glossed in this paper. It occurs on adjective-like roots (*a-hap* 'good', *a-haggur* 'cold') and stative-like roots (*a-tuli* 'be asleep', *a-bati* 'be awake', *a-niya* 'the existential 'there is') as well as the more intransitive-like roots mentioned above. In addition, many roots which are affixed by *a-* behave much like transitive verbs with agent and recipient¹¹ focus counterparts, among them are *lasa* 'feel love for', *baya* 'feel desire for' and *ta'u* 'know how/have knowledge of'. The recipient focus of these verbs may be further affixed with *ni-/in-*. Examples using *lasa* follow. See (42) below for *baya*' and *tuli*, and (21) above for the existential *aniya*'.

- (33) *A-lasa aku ma anak-ku.*
ST-love 1SG.ABS OBL child-1SG.GEN
'I love my child.' (lit. 'I feel affection/love towards my child')
- (34) *Ka-lasa-han-ku anak-ku.*
RF-love-RF-1SG.ERG child-1SG.GEN
'I love my child.'
- (35) *K-in-a-lasa-han iya e' Inah-na.*
RF-P-love-RF 3SG.ABS ERG mother-3SG.GEN
'He is loved by his mother.'

⁸ The affix *pa-* does occur with roots which form an intransitive verb with *aN-*, but with these roots, *pa-* is causative and derives a transitive verb.

⁹ *Pa-* may also occur with many of these roots. Depending on the root and other affixation, *pa-* may be either causative or intransitive.

¹⁰ In Yakan, *-um-* indicates less volitional intransitive action (Dietlinde Behrens, pers. comm.).

¹¹ The recipient focus affix for these roots is *ka--an*. In one sense it is abilitative mood (see Table 2), but in that it can be further affixed with *-in-*, it is not (normally ABL and IND-P are mutually exclusive). Presumably, *ka--an* for these roots may be affixed with *-in-* because there is no other (basic) indicative form.

A few stative/adjectival roots are affixed by *ma-* instead of *a-*. These are *maasin* 'salty', *mahaddon* 'hungry' and *maase* 'feel compassion for'. These occurrences of *ma-* may also be due to the influence of Tausug which employs *ma-* as an adjectival/stative affix.

It is frequently difficult to classify Sama verb roots as stative, intransitive or transitive. Affixation and use in a sentence determines the classification, as below:

- (36) *A-tambol gawang-in.*
ST-close door-the
'The door is closed.'
- (37) *A-tambol gawang-in e' baliyu.*
ST-close door-the ERG wind
'The door was closed by the wind.'
- (38) *Gawang-in, pa-tambol baran-na.*
door-the INT-close body-3SG.GEN
'The door, it closed by itself.'
- (39) *Bay na t-in-ambol e' si Nanah gawang-in.*
PST CP P-close.PF ERG PM Nanah door-the
'Nanah has already closed the door.'
- (40) *A-haggut tood baliyu-u.*
ST-cold very wind-that
'That wind is very cold.'
- (41) *Bang aku a-tawwa' e' baliyu, ni-haggut aku.*
when 1SG.ABS ST-strike ERG wind P-cold 1SG.ABS
'When I am struck by the wind, I feel cold.'
- (42) *K-in-aruu' tooran aku. A-baya' aku a-tuli.*
P-sleepy very 1SG.ABS ST-desire 1SG.ABS ST-sleep
'I'm very sleepy. I want to sleep.'

Likewise, it is difficult to assign one meaning to the various verbal affixes as has been seen with *a-* in the preceding examples. The affix *ni-/in-* also can be variously transitive, passive or stative-like as with *nihaggut* in (41) and *kinaru'* in (42) above. Yet a common semantic component can be discerned. The subject in every case has been affected by the state or action of the verb. Note that in (41) the *in-* affix indicates that the coldness is a temporary state that has affected the subject, whereas in (40) the *a-* indicates that the coldness is an intrinsic characteristic of the subject.

As a non-native speaker of this language who is still learning to speak it properly after many years, I am fascinated by the semantic range of the affixes (both verbal and nominal) and their adaptability to the semantic requirements of the roots. Of course, the possibility of fine semantic distinctions in human language is nearly endless, as we have learned in the plethora of semantic roles that have been proposed at one time or another. This forces languages to broaden the semantic range of the limited number of grammatical constructions. Sama Bangingi' is no exception. But the particular way in which the Bangingi' speak of the world around them will hopefully contribute much to the understanding of the semantics of focus in Austronesian languages.

References

- Hassan, Irene U., Seymour A. Ashley and Mary L. Ashley, 1994, *Tausug–English dictionary: kabtangan iban maana* (2nd edition), Sulu Studies 6. Jolo: Notre Dame of Jolo College, Manila: Summer Institute of Linguistics.
- Hopper, Paul J, 1983, Ergative, passive and active in Malay narrative. In Flora Klein-Andreu, ed. *Discourse perspectives on syntax*, 67-88. New York: Academic Press.
- Pallesen, A. Kemp, 1985, *Culture contact and language convergence*. Manila: Linguistic Society of the Philippines.
- Walton, Charles, 1986, *Sama verbal semantics: classification, derivation and inflection*. Manila: Linguistic Society of the Philippines.

Voice and role in two Philippine languages

WALTER SPITZ

1 Introduction

Perhaps inevitably, most linguists' discussions of voice and role reflect highly Eurocentric models of language. Under the sway of Indo-European grammar, such discussions often tacitly assume the universal applicability and descriptive adequacy of a binary 'active'/'passive' distinction (for voice) and the participation of a universal set of 'cases' (or roles), concepts which can prove problematical and even misleading when applied to languages from other families. In particular, the Philippine languages offer an interesting challenge to the noun-oriented, reocentric priorities of Western metalinguistic discourses. The shortcomings of these priorities have long been evident to Austronesianists. Like other Philippine languages, Hiligaynon (Central Visayan) and Yogad (Northern Cordilleran)¹ employ an elaborate system of verbal voice markers which, eluding the active/passive binarism of English, highlight one of four possible event phases (i.e. incept, middle, goal, limit), achieving various epiphenomenal 'role' effects which are often reminiscent of Indo-European cases. However, these two languages exhibit a robust contrast in the degree to which voice is mixed with the content of role. With the voice affixes in play, the two nuclear roles of Hiligaynon are consistently motile/inert (or dynamic), while Yogad roles are more consistently eruptive/post-eruptive (or relational) in character. In contrast to English voice, voice in these languages is markedly more sensitive to verbal event semantics than to

¹ Hiligaynon, which is spoken primarily on the islands of Negros Occidental and Panay, has been discussed at length in Métrida and Aparicio (1894), Ruiz (1968), Wolfenden (1971, 1975), Zorc (1977), and Spitz (1997). Yogad, spoken in the Cagayan Valley of Luzon, has been largely overlooked by linguists but is discussed in Healey (1958), in a primer by Healey and Healey (1956), in Spitz (1997), and most recently and thoroughly, in Davis et al. (1998). Most of the Hiligaynon data were provided by Ms Concha Ponce, of Ilog, Negros Occidental, and by the Rev. Ernest Dagohoy, of San José de Buenavista, Panay. The Yogad data were provided by Dr Angel Mesa of Echague (Isabela Province), Luzon. Most data-gathering was accomplished in a series of weekly sessions conducted on the Rice University campus. All three consultants currently reside in Houston, Texas. I heartily thank them for their vital participation in the current project. I also thank Drs Philip W. Davis and David Zorc for their invaluable comments and suggestions. All errors are my own.

topicality, transitivity, or other noun-heavy semantics. Seen side by side, these Philippine languages throw into relief the complementary semantics of voice and role.

In both Hiligaynon and Yagad, the propositional nucleus of clauses, often characterised in terms of a putative 'VSO' configuration, hosts a maximum of two roles, i.e. the 'S' and 'O'. The 'VSO' characterisation of the nucleus is not entirely accurate but rather reflects the well-recognised cross-linguistic tendency for 'agents' (often construed as 'subjects') to be highly topical — i.e. for people to talk about active things more than about inactive ones (see Givón 1979:336-337). More consistently, the 'V' position marks rheme, semantically the most dynamic portion of the nucleus, which tends to be verb-like but which also hosts participants which are 'new' to a given role, including interrogative pronouns. Instead of role *per se*, the 'S'-position hosts the participant which is the most highly topical in the sense of being expected or established in the discursive context, and the 'O'-position hosts a less well expected one. Other particulars, their role status even less well specified and quite variable, may occur in the periphery of the proposition, beyond the reach of voice. Rather than relying upon nominal case markings to specify the role which nuclear participants play in events, these languages employ verbal affixes which distinguish a number of different voices which an active/passive dichotomy is too crude to reflect. Each voice (or combination of voices) selects a particular phase for focus, which is complementarily marked by a nominalising determiner. Each voice affix also indicates aspect (either realis or irrealis). In the Hiligaynon instances listed in (1), each verbal voice affix selects one of four possible phases of each event, while the determiner *ang* focusses the nominalised particular which is operative at the selected phase:²

- (1)a. *Nag-sulát ang babáye sang libro.*
 NAG-write FOC woman UNF book
 'The woman wrote a book.'
- b. *Gin-sulát sang babáye ang libro.*
 'The woman wrote the book.'
- c. *I-sulát sang babáye ang lápís.*
 'The woman will write with the pencil.'
- d. *Sulat-án sang babáye ang papél.*
 'The woman will write on the paper.'

In (1a), *nag-* marks the (motile) incept of the event, i.e. the 'woman' who effects the writing of the book; in (1b), *gin-* selects the (inert) goal of the writing, i.e. the 'book'; in (1c), *i-* marks the (inert) middle phase, here the means of writing, i.e. the 'pencil'; and in (1d), *-an* highlights the marginally involved (inert) limit of the writing, beyond which the effects cannot extend, i.e. the 'paper' written on. The determiner *sang* (or *ni* for proper nouns) marks participants which are nuclear but unfocussed, i.e. not selected by voice.

Although each voice may initially seem to be highlighting 'agent', 'patient', 'instrument', and 'location' respectively, further examination problematises this assumption. In (2), *-an* may seem to select a range of roles:

² The following abbreviations appear in the present text: 1,2,3 first, second, third person; DU dual; FOC focussed; IMP imperfective; LINK linkage, ligature; OBL oblique; PL plural; RHM rheme; SG singular; UNF unfocussed.

- (2)a. *Sulat-án mo ang lamésa.*
 write-AN 2SG.UNF FOC table
 'Write at **the table.**'
- b. *Pabokal-án mo ang manók.*
 boil-AN 2SG.UNF FOC chicken
 'Boil **the chicken.**'
- c. *Tabakó-an mo ang abáno.*
 smoke-AN 2SG.UNF FOC cigar
 'Smoke **part of the cigar/one of the cigars.**'
- d. *Himo-án mo ang báta'.*
 make-AN 2SG.UNF FOC child
 'Make it/something **for the child.**'

In (1d) and (2a), *-an* seems to highlight 'the paper' and 'the table' as the 'locations' of the writing, the loci which are 'written on or at'. In (2b), *-an* seems to select the 'chicken' as a 'patient', the 'direct object' of the 'boiling'; likewise, (2c) focusses the 'cigar(s)' as 'patient', but this time as a partially affected quantity, i.e. as something 'smoked (away) at'. Finally, (2d) focusses the 'child' as an apparent 'benefactee', the one that something is 'made for'. Obviously, framing the 'solution' to *-an* in terms of 'role' (i.e. 'location', 'patient', and 'benefactee') will be ultimately unsatisfactory. The consistency of *-an* becomes clearer if we compare (2b) with the goal-oriented form *-on* in (3):

- (3) *Pabokal-ón mo ang túbig.*
 boil-ON 2SG.UNF FOC water
 'Boil the water.'

Here, *-on* selects the 'water' as the very 'means and substance' of the boiling. In contrast, the 'chicken' focussed by *-an* in (2b) participates 'off to one side' of the process. The 'peripheral involvement' semantics seen with *-an* resonates with the 'partial consumption' of the 'cigar(s)' in (2c) and with the likewise marginal involvement of the 'table' and 'paper' in the writing in (1d) and (2a) and with the benefactee 'child' in (2d). Though nuclear to the proposition, these *-an-* focussed participants have in common a peripheral, limited involvement in the semantic exhaustion of each event; hence my use of the term limit for this voice. The limiting semantics of Hiligaynon *-an* is essentially the same as that found with its cognates in Tagalog and Yogan.

The present study attempts to refine understanding of voice and role in a way that accommodates the Philippine languages as closely as possible. In the process, a robust systematic distinction in event construction emerges: namely, the differing degrees to which role and voice may interpenetrate.

2 The semantics of incept: motility vs. eruption

2.1 The motile semantics of Hiligaynon *nag-*

The Hiligaynon verbal voice affixes are listed in Table 1. My principal emphasis will be on *nag-*, *gin-*, and *i-* in Hiligaynon and on their cognates in Yogan, all of which will be briefly contrasted with the other affixes.

Table 1: Hiligaynon inflectional voice affixes

	Realis	Irrealis
Motile	<i>nag-</i> <i>naka-</i> <i>nang-</i>	(-um-) <i>mag-</i> <i>maka-</i> <i>mang-</i>
Inert	-in- gin- na-	i- -on ma- -an

In Hiligaynon, the durative affix *nag-* defines the inceptive phase of the event, which is occupied by the relatively motile role of the propositional nucleus. The common practice in Philippinist literature of labelling *nag-* (and others) as a marker of 'agent focus' is potentially misleading, though *nag-* does focus agent-like roles in (1a) and in (4)–(5):

- (4) *Nag-pínta ang pintór sang baláy.*
NAG-paint FOC painter UNF house
'The painter painted a house.'
- (5) *Nag-langóy ang báboy sang subá'.*
NAG-swim FOC pig UNF river
'The pig swam a river.'

However, such 'agentive' content is not consistently involved. If a 'prototypical' agent provides 'initiative', 'intention', and 'control' (preferably extending to a patient), (1a) and (4) obviously involve more or less 'prototypical' agent/patient contrasts, with *nag-* selecting 'agents' (i.e. the 'woman' and the 'painter') for focus by *ang*, with the unfocused, *sang*-marked 'patients' being respectively the effected 'book' and the affected 'house'. Likewise, some may see *ang báboy* 'the pig' in (5) as less prototypically agentive since *sang subá'* is relatively unaffected.

The agentive effects derive holistically from semantics peculiar to events such as 'writing', 'painting', and 'swimming' rather than from the grammar per se. Such events, susceptible to human agenthood, intentionality, control and the like I broadly term accidental, after Davis (1995) and Davis et al. (1998). With more essential events, such as 'getting fat', 'melting', and 'getting hot', which are less dependent on (human) intervention and often more spatiotemporally diffuse in Aktionsart, agenthood may often be unthinkable, and *nag-* marks the incept without implying the participation of an agent. In the 'middle voice' events of (6)–(8), the focussed incepts are clearly not agentive:

- (6) *Nag-támbok ang propesor sang inúm.*
NAG-fat FOC professor UNF drinking
'The professor got fat from drinking.'
- (7) *Nag-túnaw ang kalámay.*
NAG-melt FOC sugar
'The sugar melted.'

- (8) *Nag-ínit ang túbig.*
 NAG-heat FOC water
 'The water heated.'

In (6)–(8), *nag-* focusses *ang propesor*, *ang kalámay*, and *ang túbig* as motile incepts, the 'origins' — but not the 'causes' — of the processes named. The 'S'-role is not 'agentive' but simply motile, while any 'O'-role is relatively inert rather than specifically 'patientive', as shown by the relationship between the bibulous professor and his/her 'drinking' in (6). With accidental events (e.g. 'writing', 'painting', and 'swimming'), i.e. those which are kinetic, susceptible to the participation of (preferably human) intervention, any 'agentive' (etc.) semantics derives from an understanding of the dimension of practical experience rather than from the imperatives of grammar. In short, Hiligaynon *nag-* may accord with the agentive without specifying it, as Ruiz (1968:22) has recognised. The Hiligaynon voices achieve their 'role' effects by modulating variations on the nuclear role contrast of motility/inertness, a contrast that is essentially more dynamic than manipulative.

Hiligaynon *nag-* is felicitous only where a well defined trajectory from motile incept to inert exhaustion can be inferred. It is not felicitous in events with diffuse role definition, i.e. those with blurry agent/patient distinctions or with an inefficient or distributed performativity. In particular, *nag-* does not happily co-occur with experiential (subset of accidental) events of cognition, body function, and happenstance, e.g. *kilála* 'know (somebody)', *hibaló* 'know (that...)', *lípay* 'happy', *subó* 'sad', *táwo* 'be born', *gútom* 'hungry', *úhaw* 'thirsty', *busóg* 'full', and *dúla* 'lose'. Experiences (or 'passions' or 'affects') typically involve motile/inert contrasts reflect the play of circumstance as much as the operations of any single participant. The motility in such events is too diffuse (being both agent-like and patient-like) to be compatible with *nag-*. Such co-occurrence restrictions on *nag-* with experiential roots are not simply a matter of matching (or not matching) certain lexemes with certain morphemes but rather of recognising the dynamics of events as wholes. To illustrate, the roots *patáy* 'die/kill' and *kíta* 'see', which often name experiences, can co-occur with *nag-* if the event is distributed among participants in an efficiently 'transitive' or 'reciprocal' fashion which construes the 'S'-role as a well defined 'origin' which extends to a distinct 'consequence':

- (9)a. **Nag-patáy ang ákon íloy.*
 NAG-die/kill FOC 1SG.UNF mother
- b. *Nag-patáy ang ákon íloy sang lamók.*
 NAG-die/kill FOC 1SG.UNF mother UNF mosquito
 'My mother killed a mosquito.'
- (10)a. **Nag-kíta' ang idó' sang kuríng.*
 NAG-see FOC dog UNF cat
- b. *Nag-kíta' kamí.*
 'We saw/visited each other.'

The 'experiential' readings of these roots (i.e. 'dying' and 'seeing') are semantically too diffuse in performance for *nag-*; hence the rejection of (9a) and (10a). However, if these roots are interpreted as the actions 'killing' and 'visiting', as in the acceptable (b)-sentences, the motile roles are sufficiently efficient or focussed to be compatible with *nag-*, which accords with incepts which are semantically highly focussed, as in (11):

(11)a. *Nag-patáy ang polís sang kriminál.*
 NAG-kill FOC police UNF criminal
 'The policeman killed a criminal.'

b. **Nag-patáy ang hubón sang kriminál.*
 NAG-kill FOC crowd UNF criminal

In (11b), *ang hubón* 'the crowd' is grammatically focussed with the determiner *ang*; however, this participant is semantically too diffuse to effectively fill the motile role, in contrast to *ang polís* in (11a). It is the degree of such semantic focus/diffuseness which *nag-* and other voice forms define.

Issues of semantic focus interpenetrate with those of control. Experiences can be divided roughly into low-control and no-control categories, their focussed participants being more affected than effective. In Hiligaynon, the latter include 'dying' and 'being hungry'. Such events, whose focussed roles are more patientive than agentive, require the inert voice *na-*:

(12) *Na-patáy ang ákon íloy.*
 NA-die FOC 1SG.UNF mother
 'My mother died.'

Here, *na-* marks 'dying' as a process that befalls the focussed participant, marking *ang ákon íloy* as the terminal participant rather than the origin of the event, which arises from a more diffusely distributed circumstance (e.g. the human condition).

Other experiential events such as 'seeing' or 'hearing' allow more control or cooperation on the part of the motile role and hence may occur with the motile (or inceptive) potential forms *naka-* and *maka-* in addition to the inert *na-*. While *na-* and *naka-* select different nuclear roles (inert and motile respectively) for focal attention, they both mark a more diffuse performativity than *nag-*, compatible with experiential roots, as in (13):

(13)a. *Na-kíta' sang idó' ang kuríng.*
 NA-see UNF dog FOC cat
 'The dog saw **the cat**.'

b. *Naka-kíta' ang idó' sang kuríng.*
 'The **dog** saw a/the cat.'

Here, *na-* selects *ang kuríng* as the more inert participant, while *naka-* selects *ang idó'* as the more motile one. The diffusive semantics of both the inert form *na-* and the motile form *naka-* suggests that the perceived 'object' impacts the perceiving 'subject' at least as much as the 'subject' does the 'object'. When more high-control, accidental roots such as *sulát* 'write', *naka-* and *maka-* mark a less efficient, abilitative semantics:

(14)a. *Naka-sulát ang babáye sang líbro.*
 NAKA-write FOC woman UNF book
 'The woman **managed** to write a book.'

b. *Maka-sulát ang babáye sang líbro.*
 'The woman **can/is able to** write a book.'

In (14), the ability of the event to exhaust itself efficiently is at issue. This abilitative semantics reflects a heightened aspectual diffuseness which contrasts with the more tightly focussed semantics of *nag-*. With the 'distributive' forms *nang-* and *mang-*, the link between motile and inert roles is maximally inefficient so that the unfocussed inert phase is even more broadly distributed through time or space (see Spitz 1997:126-132). This 'inefficient',

maximally diffusive semantics takes various manifestations, including 'intentional' readings (with roots of sensory perception), as in (15), and plural/serial inert roles (with more accidental, manipulative roots), as in (16):

(15) *Nang-íta' ang babáye sang báyo'.*

NANG-see FOC woman UNF dress

'The woman **looked for** a dress.'

(16)a. *Nang-hátag ang maéstra sang mga libro.*

NANG-give FOC teacher UNF PL book

'The teacher gave out **some books**.'

b. *Nag-hátag ang maéstra sang (*mga) libro (sa estudyánte).*

'The teacher gave a book (to the student).'

In (15), the root *kitá'* combines with *nang-* to yield 'looking for' as opposed to 'seeing' (see (10) and (13)). This 'intention to see' reflects a prolongation of the inert phase which, with 'manipulative' events, yields a serial distribution of affectedness involving multiple inert participants, as seen with the 'books' in (16a). While the sequence *nag-/mag-*, *naka-/maka*, *nang-/mang-* is progressively more diffusive in terms of event performance, the realis form *-um-* is maximally focussed. The incept does not extend to a fully realised exhaustion. Hence while *Nagbása akó sang libro* and *Bumasa akó sang libro* may be identically glossed 'I read the book', the *nag-* alternative suggests a complete reading and the *-um-* alternative an incomplete, spot reading.

As these alternative inceptive/motile affixes show, Hiligaynon grammar is markedly more sensitive to the semantic dimensions of motility and focus than to transitivity *per se*. Although 'transitive' clauses abound, these reflect the interaction of the morphosyntax with specific accidental event semantics. Such clauses as (4)–(6) illustrate that a given morphosyntactic configuration may indifferently harbour either 'transitive' and 'intransitive' semantics. In other words, there is no grammar of transitivity; the morphosyntax is merely complicit in constructing transitivity without specifying it.

The grammar's insensitivity to transitivity as such also appears in the use of oblique prepositional forms to mark inert participants which are well defined (because physically or discursively 'at hand') but unfocussed. With *nag-* highlighting the motile role, inert pronominals and proper forms are marked as obliques, the former with *sa* and the latter with *kay*, forms also applicable to 'locatives' and 'recipients':³

(17)a. *Nag-hámpa' ang maéstra sa íya.*

NAG-strike FOC teacher OBL 3SG.UNF

'The teacher hit **him/her**.'

³ Such 'antipassive' configurations, with the agent in focus and the patient oblique, are of course not uncommon across languages. Spanish (the language of the longtime imperialist governors of the Philippines) has a comparable animacy/definiteness hierarchy, whereby the preposition *a* marks 'empathy-worthy' direct objects, i.e. those which are both human and referential. In Spanish, objective proper names require this preposition, as in *He empujado a Roberto* 'I've pushed Roberto' (but not **He empujado Roberto*). With common nouns, such empathy is more manipulable; both *Estoy buscando a una criada* and *Estoy buscando una criada* may be glossed 'I'm looking for a maid(servant)', the latter implying that I need my room cleaned and that any maid will do, the former implying that I am seeking a particular maid, e.g. the one who stole my watch.

b. *Nag-hámpa' ang maéstra kay Roberto.*
 'The teacher hit **Roberto.**'

c. *Nag-hátag ang maéstra sang bóla*
 NAG-give FOC teacher UNF ball

sa báta' / kay Roberto.

OBL child / OBL Roberto

'The teacher gave a ball **to the child/to Roberto.**'

In (17a) and (17b), the forms *sa* and *kay* render the clauses grammatically intransitive while they remain semantically transitive. Here and elsewhere, the morphosyntax does not reliably indicate transitivity, a situation which is markedly more pronounced in Yogad.

2.2 The eruptive semantics of Yogad *nag-*

While the motile/inert role contrasts of Hiligaynon may seem 'thin', Yogad role is even more minimalistic. The voice affixes of Yogad, listed in Table 2, are superficially similar to those of Hiligaynon, yet their apparent similarity conceals a robustly different construction of voice, one which involves virtually no admixture of role. The 'S'- and 'O'-roles do not consistently imply a trajectory of influence from the 'S' to the 'O'; these nuclear roles more consistently involve a contrast between eruptive and post-eruptive semantics, as elaborated below.

Table 2: Yogad inflectional voice affixes

	Realis	Irrealis
Eruptive	<i>-inum- ~ num-</i> <i>nag-</i> <i>naka-</i> <i>nang-</i> <i>nagg-</i>	<i>-um-</i> <i>mag-</i> <i>maka-</i> <i>mang-</i> <i>magg-</i>
Post-Eruptive	<i>-in-</i> <i>na-</i> <i>ni-</i>	<i>-uhn</i> <i>ma-</i> <i>i-</i> <i>pag-</i> <i>pang-</i> <i>pagg-</i> <i>-an</i>

Yogad voice affixes tend to be less selective of the roots they combine with, i.e. more 'promiscuous', than their Hiligaynon cognates. For instance, Yogad *nag-* may felicitously co-occur with certain diffuse 'experiential' events such as 'getting hungry'. However, in such an instance, the focussed role is construed as the origin for the event, where the event 'erupts', as in (18):

- (18) *Nab-bisín yu táwlay.*⁴
 NAG-hungry FOC person
 'The man went hungry.'

Sentence (18) describes the beginning of a fast; the 'man' goes hungry deliberately. The 'volitional' reading reflects the eruptive semantics of *nag-* (which selects *yu táwlay* as the stage from which the hunger erupts) in combination with commonsense recognition of the potential for human intervention in effecting hunger. Other experiential events, e.g. passively 'feeling hungry' (as opposed to deliberately 'going hungry'), 'dying' (*tay*), or 'seeing' (*itá*), are aspectually more diffuse and may combine with the affixes *na-* ('dying' accepting only the form) or *naka-* but not *nag-*, much as in Hiligaynon. Again, however, the acceptance of such roots by Yogad *nag-* contrasts with the greater restrictiveness of Hiligaynon *nag-*, which is more consistently geared toward marking motility. In (19), *nag-* selects *yu táwlay* 'the man' as the locus where the 'hair' erupts:

- (19) *Nab-búk yu táwlay.*
 NAG-hair FOC person
 'The man grew hair/became hairy.'

While Yogad *nag-* may mark 'agent' or some other motile role, it need not do so, as illustrated further in (20)–(21):

- (20) *Nat-tuláng yu ikán.*
 NAG-bone FOC fish
 'The fish was/turned out boney.'
- (21) *Nab-binaláy yu profesór.*
 NAG-house FOC professor
 'The professor has a house.'

In (20)–(21), nothing happens; the 'fish' does not suddenly sprout 'bones', nor the 'professor' a 'house'. Rather, *nag-* selects the *yu*-marked nominal elements (i.e. the 'professor' and the 'fish') as the eruptive loci of the verbalised elements; the 'bone' and 'house' manifest themselves with respect to the 'fish' and 'professor'.

Elsewhere, *nag-* may allow 'mediopassive' readings. In (22), *nag-* focusses alternately a semantically agentive 'woman' and a patientive 'paper' (see Davis et al. 1998:156–165):

- (22)a. *Nap-pissáy yu babáy.*
 NAG-tear FOC woman
 'The woman tore something.'
- b. *Nap-pissáy yu papél.*
 'The paper tore (by itself).'

Yogad grammar is even more indifferent to semantic transitivity than that of Hiligaynon. While the voices of Hiligaynon maintain a dynamic motile/inert contrast in the propositional nucleus, such a contrast is often lacking in Yogad clauses. Beginning from the inherently well defined eruptive 'S'-role, the roles are increasingly vague in their definition vis-à-vis their integration into the event. The determiner *tu* (or *tu ku...* for pronominal and proper-nominal

⁴ Note that the final *-g* of *nag-* (and of *mag-* and *pag-*) assimilates completely to any subsequent initial consonant. Such morphotactic assimilation is more widespread in Yogad than in Hiligaynon.

participants) marks all successive unfocussed post-eruptive participants, regardless of their propositional nuclear/peripheral status or of their performative role in the event, as seen in (23)–(24):

- (23) *Nab-biláng yu méstru tu binaláy.*
 NAG-count FOC teacher UNF house
 ‘The teacher counted **in the house.**’
 ‘The teacher counted **houses.**’
- (24) *Nad-digút yu babáy tu danúm.*
 NAG-bathe FOC woman UNF water/river
 ‘The woman bathed **in the river.**’
 ‘The woman bathed **with water.**’

In (23), the ‘house(s)’ can be interpreted as either ‘patient’ (if the nucleus contains two roles) or ‘location’ (if it contains only one). In (24), the ‘water’ may be either ‘location’ or ‘instrument’. The language does not further specify the precise performativity of each *tu*-marked post-eruptive particular. Outside of the nucleus, beyond the reach of voice, roles are increasingly ill defined by the grammar, requiring commonsense integration of such particulars into the event without the aid of prepositions like the semi-deictic Hiligaynon forms *sa* and *kay*. The seemingly lax definition of more peripheral roles may be further illustrated in events which clearly contain two nuclear roles:

- (25) *Nap-pórba yu babáy tu sinnún tu Kmart.*
 NAG-try FOC woman UNF clothes UNF Kmart
 ‘The woman tried on clothes **at Kmart.**’
- (26) *Nak-kánna yu estudyánte tu méstru tu batú.*
 NAG-hit FOC student UNF teacher UNF rock
 ‘The student hit a teacher **with a rock.**’

In both (25) and (26), *tu* marks two successive post-eruptive participants. While the nuclear participants (i.e. *tu sinnún* and *tu méstru*) are patient-like, those occurring outside the nucleus, where relations are less well defined, may be interpreted as ‘locatives’ (such as *tu Kmart*) or ‘instruments’ (such as *tu batú*); the morphosyntax does not make a distinction, and it would be presumptuous for linguistic analysts to do so. The relative indifference of Yogad voice to role — and particularly to dynamics — is a pervasive condition throughout the language, as is evident from the post-eruptive voices discussed in §3 and §4.

Yogad *nag-* is supplemented by other motile/inceptive forms which mark a performativity that is less tightly focussed. The forms *naka-* and *maka-* mark abilitative/potential events, as in Hiligaynon. The forms *nang-* and *mang-* suggest a more diffuse performance of event (whether transitive or not) than *nag-* and *mag-*, with a correspondingly greater spatiotemporal distance between the eruption and exhaustion, as seen in (27):

- (27)a. *Nag-immugúd kan tu atú-m.*
 NAG-care 1SG.FOC UNF dog-2SG.UNF
 ‘I took care of your dog.’
- b. *Nang-immugúd kan tu atú-m.*
 ‘I took care of your dog.’

In (27a), *nag-* suggests a spontaneous, short-lived act of kindness, while *nang-* in (27b) implies a prolonged process, e.g. my caring for ‘your.dog’ during your vacation. Both forms

allow inert participants to be transitively affected, the instance with *nag-* being the more punctual. When a post-eruptive participant is not specified, the difference between the two forms may incidentally invoke 'transitivity'. In (28a), *nag-* terminates efficiently in the participant where it erupts, while *nang-* in (28b) requires the event to exhaust itself less efficiently, in a more distant locus (see Davis et al. 1998:176):

(28)a. *Nag-usíp kan.*
NAG-haircut 1SG.FOC
'I got a haircut.'
*'I cut someone's hair.'

b. *Nang-usíp kan.*
'I cut someone's hair.'
*'I got a haircut.'

Note that the contrast between *nag-* and *nang-* is not primarily one of transitivity, given such instances as (27). Interestingly, Hopper and Thompson's (1980) classic analysis treats both 'punctuality' and 'affectedness of object' as complementary 'features' of transitivity. In such a scenario, one might expect the more 'punctual' (i.e. focussed) form *nag-* to be consistently more 'transitive' (i.e. to require an 'object'), but this is not the case in (28a); rather, it is the 'nonpunctual' (or more diffusive) form *nang-* which takes an implicit 'object' in (28b). Obviously, *nag-* and *nang-* (like other voice forms) are not primarily attuned to 'transitivity'; rather, relative 'punctuality' (or, in our terminology, focus) of performativity is more to the mark. The relatively unproductive forms *nagg-* and *magg-* are even more diffusive, as in (29):

(29)a. *Mag-ítá yu atú tu pasyénte.*
MAG-see FOC dog UNF patient
'The dog is **seeing** for a patient.' (i.e. as a seeing-eye dog)

b. *Magg-ítá yu atú tu pasyénte.*
'The dog is **searching** for a patient.' (e.g. pursuing an escapee)

Thus, the series *nag-/mag-*, *naka-/maka-*, *nang/mang*, *nagg-/magg-* marks a progression of increasingly diffusive semantics. In the other direction, the form *-um-* (realis counterpart *-inum-*) is even more tightly focussed than *nag-*, marking eruptions which are partially initiated but not complete, involving very limited follow-up, as in (30) (see Davis et al. 1998:169-173):

(30)a. *Nab-bibbíd kan tu nobéla.*
NAG-read 1SG.FOC UNF novel
'I **read** a novel.'

b. *B-inum-ibbíd kan tu nobéla.*
'I **convinced myself to read** a novel.'

In (30a), the reading has been completed, whereas in (30b) it has not even been properly initiated but merely intended. This 'intentional' reading of *Yogad -um-* is less active than the 'partial' or 'spot' reading observed with the Hiligaynon *-um-*. With other roots, *-um-* may mark events that are reversible, without consequences, or simply 'in the nature of things' (see Davis et al. 1998:169). Just as the limiting form *-an* (illustrated in (2)) marks 'patients' that are only partially affected, *-um-* marks 'agents' which are only partially effective. Given the

limited effectiveness/affectedness marked by each form, *-um-* and *-an* may be thought of as analogues to each other which differ primarily in the event phase that each selects.⁵

In both languages, *nag-* selects a given participant as the most efficiently effective launching point for an event. The motility/eruption differential between Hiligaynon and Yogad, summarised for *nag-* in Table 3, is pervasive among all of the most highly focussed voices, as discussed further in §3 and §4.

Table 3: Hiligaynon and Yogad *nag-* compared

Hiligaynon <i>nag-</i>	Yogad <i>nag-</i>
Motile, dynamic	Eruptive, relational
More selective (of actions, not experiences)	More promiscuous (accepting experiential roots)
More control-sensitive	Less control-sensitive

3 The semantics of exhaustion: goal vs. pervasiveness

3.1 The goal semantics of Hiligaynon *gin-*

While *nag-* and other incept-selective voices highlight the inception of events, the inert voices highlight their 'down side'. A preliminary differentiation between motile and inert voices is illustrated in (31):

(31)a. *Nag-túnaw ang kalámay.*
 NAG-melt FOC sugar
 'The sugar **melted**.'

b. *Gin-túnaw ang kalámay.*
 'The sugar **was melted**.'

Sentence (31a), with *nag-*, frames the 'sugar' as the inceptive locus of the 'melting'. In (31b), the *gin-*-focussed 'sugar' occupies the endpoint of the process. Although *gin-* might seem to be marking 'passive', this characterisation is potentially confusing since 'passive' has acquired certain baggage that would be contraband in the present linguistic territory. In numerous languages, including English, 'passive' normally involves 'object promotion' and/or 'subject demotion', but *gin-* does not require it. In (32), *sang báta* 'the child', being the most expected participant, occupies the continuous topic 'S'-position, even while the focus of attention falls on *ang kalámay* 'the sugar':

⁵ Zorc (1977), labels *-um-* the active punctual; *nag-*, the active durative; *naka-/maka-*, the active potential; *nang-/mang-*, the active distributive; *gin-*, the passive instrumental/durative; *na-/ma-*, the passive potential; *i-*, the instrumental punctual and *-an*, the local. I have supplemented this terminology for a number of reasons, most importantly: (a) to delineate more forcefully the phase orientation of voice, which transcends the 'active/passive' binarism of most metalinguistic discussions, (b) to more elegantly express the interrelationship of forms at a given phase along a continuum of focus/diffuseness, and (c) to enhance the consistency of labelling (since, for example, *na-* may mark 'objects' of sensory perception as opposed to merely potential objects, and *-an* may mark non-locative participants).

- (32) *Gin-túnaw sang báta' ang kalámay.*
GIN-melt UNF child FOC sugar
'The child melted the sugar.'

However, inert participants may be topicalised by occurrence in the 'S'-position:

- (33) *Gin-túnaw ang kalámay, sang báta.'*⁶
'The sugar was melted by a child.'

Here, the *gin*-focussed *ang kalámay* is the most topical (or 'expected') particular, and the sentence as a whole conforms more completely to a 'passive' configuration, though even here the motile role is not made oblique, as its English counterpart would be, but is simply less topical. The independent manipulability of topic (signalled by the position immediately after the rheme/verb), nominal focus (signalled by the determiners *ang, sang*, etc.), and role (signalled by correlative marking by the determiners and voice affixes) underscores the difficulty of applying such Indo-European categories as 'active'/'passive' and even 'subject'/'object' to the Philippine languages (see Schachter 1976).

Another objection to a 'passive' characterisation of *gin*- is the sheer inadequacy of an 'active'/'passive' contrast to account for the aspectual variation in the voices encountered in the Philippine languages, which, as already discussed, marks a cline of focus/diffuseness. In all of its applications, *gin*- implies a punctuality of execution which contrasts with the relative diffuseness signalled by *na*-, as seen in (34)–(36):

- (34)a. *Gin-gutóm ang idó' kahápon.*
GIN-hungry FOC dog yesterday
'The dog got hungry yesterday.'

- b. *Na-gutóm ang idó' kahápon.*
'The dog was hungry yesterday.'

- (35)a. *Gin-káon ang bugás.*
GIN-eat FOC rice
'The rice was eaten.'

- b. *Na-káon ang bugás.*
'The rice managed to get eaten.'

- (36)a. *Gin-anó ang idó' ?*
GIN-what FOC dog
'What was done to the dog?'

- b. *Na-anó ang idó' ?*
'What happened to the dog?'

In (34)–(36), in contrast to the (b)-sentences, with *na*-, the *gin*-marked (a)-sentences suggest events whose unspoken incepts are relatively well definable so that each event is focussed in its performativity. Sentence (34a), with the experiential root *gutóm*, indicates that agent/patient contrasts *per se* are not at stake. Nobody made the 'dog' hungry; it simply got hungry. Thus, no agent need be involved, though one may be, as implied in (35b). In all the (b)-sentences, with *na*-, the performance of the event is more diffuse and open to happenstance

⁶ Without the pause between *ang kalámay* and *sang báta'*, here indicated by a comma, the sequence *ang kalámay sang báta'* reads as the single constituent 'the sugar of the child'.

than in the (a)-sentences, with *gin-*, whose efficiently punctual semantics is consonant with agent.

As discussed in relation to *nag-* earlier, the inert voices *gin-*, *na-*, and *ma-* differ in terms of focus/diffuseness of the phases they select, each voice in the sequence being more diffusive than the preceding one. While *gin-* marks the goal of a trajectory stemming from a well defined origin, *na-* merely acknowledges that an ill-defined crux or turning point has been passed, marking a more diffuse playing out of the event. Even more diffusive is *ma-*, which is not merely the irrealis counterpart of *na-*. One indication of this diffusiveness is the variability of the placement of nominal focus, since *ang* may select either the motile role (indicating an action or process so imminent as to seem inevitable) or on the inert one (indicating the potential for a process or action to occur), as illustrated in (37a) and (37b) respectively:

- (37)a. *Ma-sulát ang babáye sang sulát.*
 MA-write FOC woman UNF letter
 'The woman **is about to** write a letter.'
- b. *Ma-sulát sang babáye ang sulát.*
 'The woman **can/is able to** write the letter.'

In either case, the *ang*-focussed participant comes into play at a point where the event is seemingly inevitable and beyond intentionality. Another facet of the semantic diffusiveness of *ma-* is its applicability to essential qualities that are variable, as in (38):

- (38)a. *Ma-talóm ang kotsílyo.*
 MA-sharp FOC knife
 'The knife is sharp.'
- b. *Talóm ang kotsílyo.*
 'The knife is sharp.'
 'Knives are sharp.'

Sentence (38a) describes the current condition of a particular knife, whose sharpness may vary from time to time depending on its recent treatment. Sentence (38b) functions as either a description of a knife which never gets dull or as a categorial statement about knives in general; variability is pointedly ignored. The variability in performative focus seen in the sequence *gin-*, *na-*, *ma-* offers an inert-voice counterpart to the variability already encountered among the motile voices, a variability which transcends the typical 'passive' characterisation.

In this context of diminishing degrees of focus, *gin-* (like its irrealis counterpart *-on*) most sharply defines participants which are most fully enveloped by their events. These fall within a range of affectees, as seen in (39):

- (39)a. *Gin-tiró sang babáye ang pusíl sa makáwat.*
 GIN-shoot UNF woman FOC pistol OBL thief
 'The woman shot **the pistol** at the thief.'
- b. *Gin-tiró sang babáye ang bála sa makáwat.*
 'The woman shot **the bullet** at the thief.'
- c. *Gin-tiró sang babáye ang makáwat.*
 'The woman shot **the thief**.'

In (39), *gin-* highlights participants which operate at different points in the progress of the event of 'shooting', i.e. the 'pistol' used to fire the bullet, the 'bullet' ejected from the pistol, or the 'thief' who 'catches' the bullet. In any case, however, *gin-* highlights the efficiently attained goal of the process. The relatively focussed semantics of this voice precludes its felicitous combination (minus other affixes) with such qualitative, diffusive (i.e. more essential) roots as *dakó* 'big', *támbok* 'fat', *putí* 'white', *dyútay* 'little/few', and *dóktor*. But, as with *gutóm* 'hungry' in (34a), *gin-* may mark seemingly gradual performances if a reasonably well defined motile-inert relation can be understood, as in a standard folk explanation for a neonate's unexpectedly dark pigmentation:

- (40) *Gin-panámkon ang báta' sa úling.*
 GIN-crave FOC child OBL charcoal
 'The child came out dark because of its mother's craving for charcoal.'
 (lit. 'The child was craved unto charcoal.')

The goal semantics of *gin-* contrasts interestingly with the more punctual semantics of *-in-*. According to Wolfenden (1971:118), *-in-* "seems to be a free variant of *gin-* and can usually be substituted for it with no change of meaning." Indeed, both forms may select the inert termini of events, as in:

- (41)a. *Gin-kagát sang kuríng ang idó'.*
 GIN-bite UNF cat FOC dog
 'The cat bit **the dog**.'
- b. *K-in-agát sang kuríng ang idó'.*
 'The cat bit the dog.'

However, equivalence does not imply identity; the identical English glosses in (41) gloss over an important semantic difference; *-in-* does not define a simple, clear trajectory from the event incept to the event exhaustion. Sentence (41b) suggests a casual or incidental occurrence; e.g. the 'cat' may have accidentally nipped the 'dog' during play. The action described by sentence (41a) is more intense: the 'cat' behaved in more deliberate fashion ('you could see it coming') and bit more deeply. Thus, (41a), but not (41b), might felicitously end a narrative (David Zorc, pers. comm.). The lack of direction associated with *-in-* is ultimately irreducible to either role or aspect. In addition to such inflectional, voice-related applications as seen in (41b), *-in-* appears in derivational contexts:

- (42)a. *Nag-gowá' ang túbig.*
 NAG-out FOC water
 'The water ran out.'
- b. *Nag-g-in-ówa' ang túbig.*
 'The water ran out (**in all directions**).'
- c. *Mag-a-g-in-ówa' ang túbig.*
 MAG-IMP-out-IN-out FOC water
 'The water will run out (**in all directions**).'

In (42a), the 'water' seems to have run out through a single hole or leak. In (42b) and (42c), *-in-* points to an ill defined exhaustion following the motile incept marked by *nag-* and *mag-*. Such effusive semantics are also notable in such nominalisations as *tinúig* 'many years' (cf. *túig* 'year') and *pinolítika* 'political ways/modes of engagement' (cf. *polítika* 'politics').

3.2 The pervasive semantics of Yogad *-in-*

Like Hiligaynon *gin-*, Yogad *-in-* marks an efficiently focussed exhaustion. This highly focussed completeness contrasts with the more diffusive semantics of *na-*:

- (43)a. *G-in-afút nu babáy yu mammánok.*
 catch-IN-catch UNF woman FOC bird
 'The woman **caught** the bird.'
- b. *Na-gafút nu babáy yu mammánok.*
 'The woman **managed to catch** the bird.'
- (44)a. *T-in-akáw nu dóktor yu kwártu.*
 steal-IN-steal UNF doctor FOC money
 'The doctor **stole** the money.'
- b. *Na-takáw nu dóktor yu kwártu.*
 'The doctor **managed to steal** the money.'

Note that the determiner *nu* marks eruptive participants that are unfocussed, just as *tu* marks post-eruptive ones. In (43)–(44), the (a)-sentences name more punctual events than the (b)-sentences. In (43a), the 'woman' seems to have caught the 'bird' directly, in her hands, while (43b) implies that she did so indirectly, through a trap, or that she experienced some difficulty. Likewise, (44a) might imply an efficient 'hit', while (44b) suggests that the doctor might have bungled through the job or encountered complications.

Unlike Hiligaynon *gin-*, Yogad *-in-* can co-occur with roots which may in other contexts name qualities, e.g. *dakál* 'big', *fulláw* 'white', and *bágu* 'new', a combination which may apply to efficient processes involving agent/patient contrasts:

- (45) *D-in-akál nu pintór yu letrátu.*
 big-IN-big UNF painter FOC picture
 'The painter enlarged the picture.'
- (46) *F-in-ulláw nu dentísta yu ngipán-ku.*
 white-IN-white UNF dentist FOC tooth-1SG.UNF
 'The dentist whitened my teeth.'
- (47) *B-in-águ nu estudyánte yu ugáli-na.*
 new-IN-new UNF student FOC trait-3SG.UNF
 'The student changed his/her habits.'

Although *-in-* seems to be marking (inert) role, this effect is incidental. Unlike Hiligaynon *gin-*, Yogad *-in-* is concerned with outcomes at the expense of processes and, like its Hiligaynon cognate *-in-*, may occur in events lacking a role contrast, as in the following (see Davis et al. 1998:200-250):

- (48) *S-in-íri yu presidénte.*
 lie-IN-lie FOC president
 'The president is **full of lies**.'
- (49) *T-in-úppak yu arinóla.*
 spit-IN-spit FOC bedpan
 'There's **plenty of spit** in the bedpan.'

In (48)--(49), the focussed 'president' and 'bedpan' are suffused with 'lies' and 'spit' respectively. The pervasive semantics of *-in-*, which never accommodates imperfective readings in the manner of its Hiligaynon cognate, signals a final state of merger beyond the dynamics of merging, a perfective 'closure'. This perfectivity (seen also in the derivation of *-inum-* from *-um-*) is consonant with a disinterest in a directed 'flow' or 'transitivity' *per se*. Even on a formal level, pervasive involvement seems to be manifest in the infixation of this morpheme, with its 'disruption' of morpheme boundaries. The pervasive semantics of *-in-* is echoed more feebly in the irrealis/imperfective form *-uhn* (cognate with Hiligaynon *-on*), which is far less productive (see Davis et al. 1998:263-277). The high degree of aspectual focus seen with Yogad *-in-* contrasts with the increasingly greater diffuseness of *na-* and *ma-*, much as in Hiligaynon. In addition to the semantics already described in §3.1 for its Hiligaynon cognate, Yogad *ma-* also marks processes and conditions which are repetitive or graded, as in (50)--(51) (see Davis et al. 1998:217-230):

- (50) *Ma-dáfung kan.*
 MA-meet 1SG.FOC
 'I have a lot of greeters.'
 'I constantly greet people.'
 'I will be met.'
- (51)a. *Ma-lasáng yu kòrsonsílyu-ku.*
 MA-red FOC underwear-1SG.UNF
 'My underwear is reddish.'
- cf. b. *Lasáng yu kòrsonsílyu-ku.*
 'My underwear is red.'

To sum up, with all voices including *-in-*, any agent/patient dynamic is an epiphenomenon of specific event semantics as modulated by the morphosyntax. In contrast to Hiligaynon *gin-*, Yogad *-in-* more consistently specifies relations, not dynamics, as summarised in Table 4 (see Davis et al. 1998:141-142).

Table 4: Hiligaynon *gin-* and Yogad *-in-* compared

Hiligaynon <i>gin-</i>	Yogad <i>-in-</i>
Inert	Post-eruptive
Patient, goal	Patient, goal
Experiencer	Experiencer
More selective (of actions, accidental roots)	More promiscuous (accepting durative, essential roots)
Bounded interaction	Pervasive interaction

4 The semantics of the middle phase: dynamics vs. relations

4.1 The middle semantics of Hiligaynon *i-*

Perhaps the clearest demonstration of the distinction between the Hiligaynon emphasis on dynamics and the Yogad emphasis on relations is to be found among the middle voices. In each language, the affix *i-* selects the middle phase of each event for focus, a phase past the

incept yet prior to the exhaustion of the event. In Hiligaynon, the *i*-focussed participants are consistently translative, i.e. moved by a prior motile force toward an unspecified goal:

- (52) *I-táhi' sang babáye ang dágom.*
 I-sew UNF woman FOC needle
 'The woman will sew **with the needle.**'
- (53) *I-lígid sang táwo ang úlo sa íya asáwa.*
 I-roll UNF person FOC head OBL 3SG.UNF wife
 'The man will roll **the head** to his wife.'
- (54) *I-bú'bo' sang dóktor ang kapé.*
 I-pour UNF doctor FOC coffee
 'The doctor will pour **out the coffee.**'

In (52), *i*- seems to be 'instrumental'. However, (53) and (54) seem to mark 'patients'; from the perspective of English, the morphosyntax conflates two different roles. The consistency of *i*- lies in its selection of the middle phase of the event, where the process is under way but has not yet exhausted itself. The 'head' is not merely 'rolled' but 'rolled to' somebody; the 'coffee' is not merely 'poured' but 'poured out' to another location. In other words, *i*- selects a phase that is one remove from the exhaustive phase (or 'final resting place') selected by *gin*-, *na*-, and *ma*-. This middle semantics is commensurate with the irrealis aspect (its realis counterpart, *gin*-, selects both middle and exhaustive phases). The middle phase marked by *i*- selects any particular which is 'sold', 'given', 'thrown', or otherwise transferred, but it is incompatible with more essential events for which 'translative' readings would be irrelevant, e.g. 'being hungry', 'being white', 'being tired', 'growing', or 'dying'. Hiligaynon *i*- prefers kinetic or accidental events involving a relatively inert participant responding to a more motile force in a way that is consistent with the essentially dynamic character of voice in Hiligaynon.

4.2 The middle semantics of Yogad *i*- and *pag*-

Like its Hiligaynon cognate, the Yogad middle voice *i*- (and likewise its realis counterpart *ni*-) can highlight translative participants, as in (55) and (56):

- (55) *I-kánna-mo yu bátu tu dínding.*
 I-hit-2SG.UNF FOC rock UNF wall
 'Hit **the rock against** a wall.'
- (56) *I-taráng nu profesór yu sombréru tu dínding.*
 I-hang UNF professor FOC hat UNF wall
 'The professor will hang **the hat** on a wall.'

In (55)–(56), a (human) agent 'moves' the focussed participants (i.e. the 'rock' and 'hat') toward an unspecified goal. However, with more experiential events such as 'growing big' or 'dying', Yogad *i*- may indicate conditions which occasion the eruption of the event, in violation of the more expected flow of influence from the 'S' to the 'O':

- (57) *I-dakál nu anak yu paginúm tu gatták.*
 I-big UNF child FOC drinking UNF milk
 'The child will grow big **by drinking milk.**'

- (58) *I-patáy nu kolák-ku yu awán tu asikásu.*
 I-die UNF friend-1SG.UNF FOC absence UNF concern
 'My friend will die **through neglect.**'

In (57)–(58), *nu anák* 'the child' and *nu kolákku* 'my friend' occupy the event phase where the 'growing big' and 'dying' erupt. The *i*-focussed particulars *yu paginúm tu gatták* 'the drinking of milk' and *yu awán tu asikásu* 'the absence of concern' are conditions operating in close proximity to each eruptive participant. (Once again, the grammar does not mark a consistent trajectory from a motile 'S' to an inert 'O'.) In contrast to its Hiligaynon cognate, Yogad *i*- often functions independently of any semantics of intension or extension.⁷ The relational character of this semantics can be more clearly appreciated when *i*-, marking the proximate middle, is contrasted with *pag*-, which, highlighting the distal middle, selects particulars which are operative at a greater spatiotemporal remove from the inceptive and exhaustive event termini.⁸ In (59)–(60), *i*- selects the 'bad workmanship' and the 'fruit' as conditions residing imminently in the very loci where they erupt (e.g. the 'fruit' being the very embodiment of the tree's fruitfulness), while *pag*- selects the 'cold (weather)' and the 'use of fertiliser' as conditions external to the eruptive particulars:

- (59)a. *I-bákka nu bintána yu marál ya pàkkrabáho.*
 I-shatter UNF window FOC bad LINK workmanship
 'The window will shatter **due to bad workmanship.**'
- b. *Pab-bákka nu bintána yu malábat.*
 'The window will shatter **from the cold.**'
- (60)a. *I-bungá nu kayú yu frútas.*
 I-fruitful UNF tree FOC fruit
 'The tree is bountiful **through its fruit.**'
- b. *Pab-bungá nu kayú yu pangiyúsa tu abóno.*
 'The tree is bountiful **due to the use of fertiliser.**'

The alignments of voices and participants could not sensibly be reversed in these instances; e.g. *i*- in (59) could not select *malábat*. This 'internal/external' contrast reflects a differential in spatiotemporal distance between the focussed participants and the corresponding eruptive ones; *pag*- consistently marks a greater detour from the straight trajectory of eruption-exhaustion than *i*-. Elsewhere, the *i*-/*pag*- contrast may be predominantly aspectual, as in (61)–(62):

⁷ Hiligaynon *i*- may, however, mark the means of an event's accomplishment in nominal phrases, as in (i) (David Zorc, pers. comm.):

(i) *Ang sigarilyo ang i-patáy sa ímo.*
 FOC cigarette FOC I-kill OBL 2SG.UNF
 'Cigarettes will be the thing that kills you.'

Verbal phrase occurrences of Hiligaynon *i*- seem, however, to be consistently translative. Essential roots naming qualities require *pa*- to co-occur with *i*-:

(ii) *I-pa-dakó' sang báta' ang pag-inúm sing gátas.*
 I-PA-big UNF child FOC PAG-drink UNF milk
 'The child will get big through the drinking of milk.'

In general, 'experiential' roots in Visayan (e.g. 'getting big', 'getting sick', 'dying') seem to prefer the form *na*- to *i*- (David Zorc, pers. comm.).

⁸ The Hiligaynon form *pag*- does not mark voice in finite clauses but rather participials, as in *paginúm sang serbésa* 'the drinking of beer'.

- (61)a. *I-takít nu estudyánte yu siffún.*
 I-ill UNF student FOC cold
 'The student **is getting** ill from a cold.'
- b. *Pat-takít nu estudyánte yu siffún.*
 'The student **will get** ill from a cold.'
- (62)a. *I-darál nu kárne yu patú ya tyémpu.*
 I-destroy UNF meat FOC hot LINK weather
 'Meat **spoils** due to the heat of the weather.'
- b. *Pad-darál nu kárne yu patú ya tyémpu.*
 'The meat **will spoil** due to the heat of the weather.'

In the *i*-marked (a)-sentences, the exhaustion is more imminent than in the *pag*-marked (b)-sentences. In (61a), the student's illness is already underway, in progress; in (61b), it is more prospective and perhaps preventible if certain actions are taken. The sentences of (62) offer yet another variation on the proximate/distal contrast; (62a), with *i*-, is read as a categorial statement about meat, i.e. one which is always imminently 'in force', while (62b), employing *pag*-, is a prediction applied to a specific instance.

The various readings of these voices flow from a commonsense understanding of the inherent dynamics of given events and the possible and/or appropriate articulation of the participants within them. With *pag*-, any role interpretation differs drastically according to whether the focussed participant is human or nonhuman, as in (63):

- (63)a. *Pag-gáku nu méstru yu atáwa-na tu abóbo.*
 PAG-cook UNF teacher FOC wife-3SG.UNF UNF adobo
 'The teacher will cook adobo **for/in lieu of his wife.**'
- b. *Pag-gáku nu méstru yu oven tu adóbo.*
 'The teacher will **use the oven** to cook adobo.'

In (63a), the focussed human participant *yu atáwana* 'his wife' names an inactive 'benefactive' participant for whom the eruptive 'S'-role (*nu méstru*) serves as 'proxy'. In (63b), *pag*- selects the nonhuman participant *yu oven* as the means or 'instrument' by which the cooking is effected. In both 'proxy' and 'instrument' readings, the eruptive participant is more active than the focussed post-eruptive one, which enters the event at point well removed from the actual inception or exhaustion. While Hiligaynon voice specifies the dynamics of motility/inertness, Yogad specifies the relations of eruption/post-eruption. Where Hiligaynon drives, Yogad drifts.

With the compound form *pinag*- (a perfective combination of *pag*- and *-in*-), a familiar indifference to the dynamics of performativity appears, as exemplified in (64):

- (64) *P-in-ag-gatáng nu méstru yu atáwa-na tu adóbo.*
 'The teacher **had his wife cook** adobo.'
 'The teacher cooked adobo **for/in lieu of his wife.**'

In (64), the combination of *pag*- with *-in*- allows both 'causative' and 'benefactive' readings of *yu atáwana*. As seen in (48)–(49), *-in*- does not specify 'role' but instead contributes a perfectivity that manifests itself beyond dynamics. In combination with other, less focussing voices, *pag*- favours 'causative' readings.

In marking graded digressions from the eruptive-exhaustive trajectory of events, the Yogad middle voices *i-*, *ni-*, and *pag-* isolate well defined, focussed particulars. The high degree of focus or definition marked by these voices contrasts, in a now-familiar pattern, with the more diffusive semantics seen in two alternatives to *pag-*, i.e. *pang-* and *pagg-*. The affix *pang-* selects particulars which are not isolable as participants but which emerge as artefacts of the event's playing out, in contrast to *pag-*, which as a verbal inflection selects more highly 'institutionalised' particulars (see Davis 1996):

- (65)a. *Pag-afút nu presidénte yu kwártu.*
 PAG-win UNF president FOC money
 'The president will **use money** to win.'
- b. *Pang-afút nu presidénte yu mapí ya pag-gobyérno.*
 'The president will win **through his record of good governing.**'

In (65a), *pag-* selects *yu kwártu* 'the money' as the 'means of winning', a particular which is more focussed or institutionalised than *yu mapí ya paggobyérno* '(his) record of good governing', selected by *pang-* in (65b), which is more broadly distributed in time and/or space. A reversal of *pag-* and *pang-* in the examples of (65) would not be sensible. Similarly, the modestly productive form *pagg-* names particulars which are even more diffuse, in a manner analogous to *nagg-/magg-* (Philip W. Davis, pers. comm.):

- (66)a. *Pag-itá-ku yu anteóhos.*
 PAG-see-1SG.UNF FOC glasses
 'I'll use the glasses to **see.**'
- b. *Pagg-itá-ku yu estudyánte tu binaláy.*
 PAGG-see-1SG.UNF FOC student UNF house
 'I'll **look for** a house **for the student.**'
- (67)a. *Pab-bannád nu dentísta yu anaesthesia tu ngipán.*
 PAG-numb UNF dentist FOC anaesthesia UNF tooth
 'The dentist uses **anaesthesia** to numb the tooth.'
- b. *Yu pabba-bannád nu dentísta ay mapí.*
 FOC PAGG-numb UNF dentist RHM good
 'The **way** the dentist numbs **is good.**'

In its 'benefactive' and 'manner' readings in (66)–(67), *pagg-* marks spatiotemporally diffuse performances, while *pag-* marks more focussed ones.

To recover our main theme, the Yogad post-eruptive irrealis voices *-an*, *i-*, and *pag-*, and the compound form *ipag-* mark a scale of increasing divergence from the direct trajectory of progress of an event from its incept to its exhaustion, with *-an* marking the most direct path and *ipag-* the least direct one (see Davis et al. 1998:136–148):

- (68)a. *Pínta-n nu estudyánte yu binaláy.*
 paint-AN UNF student FOC house
 The student will paint **the house.**'
- b. *I-pínta nu estudyánte yu lasáng ya pínta tu binaláy.*
 'The student will paint the red paint onto a house.'

- c. *Pap-pínta nu estudyánte yu brótya tu binaláy.*
‘The student will **use the brush** to paint a house.’
- d. *I-pap-pínta nu estudyánte yu propesór tu binaláy.*
‘The student will paint a **house for the professor.**’

In (68a), *-an* focusses ‘the house’ as the final affectee of the ‘painting’, the limit of the event reached via the most direct route. In (68b), *i-* highlights ‘the red paint’ as a participant which is one step away from this limit. Sentence (68c), with *pag-*, selects ‘the brush’ as filling the role which is more removed. Finally, the complex form *ipag-* selects ‘the professor’ as the ‘benefactee’, i.e. as the post-eruptive role farthest removed from the process of execution. The ‘professor’ functions as an Aristotelian final cause, being both occasion and goal of the entire event process. The essentially relational character of the Yogad voices contrasts robustly with the more dynamic voices of Hiligaynon, as seen with respect to the middle form *i-*, as summarised in Table 5.

Table 5: Hiligaynon and Yogad *i-* compared

Hiligaynon <i>i-</i>	Yogad <i>i-</i>
Inert	Post-eruptive
Instrument	Instrument
Translative patient	Translative patient
More selective (of accidental roots)	More promiscuous (occurring with essential roots)
	Inherent, proximate cause/condition

5 Concluding connections

According to Gregory Bateson, ‘THE DIVISION OF THE PERCEIVED UNIVERSE INTO PARTS AND WHOLE IS CONVENIENT AND MAY BE NECESSARY, BUT NO NECESSITY DETERMINES HOW IT SHOULD BE DONE’ (Bateson 1980:42, capitals in the original). Decades earlier, Saussure, forging the modern ‘linguistic consciousness’, had observed that ‘no one object of linguistic study emerges of its own accord’ (Saussure 1983:9). In focussing the current discussion of voice on the verbal affixes, I have tried to follow the priorities of these languages as indicated by their morphosyntax, to examine what seems most interesting and informative from an Indo-European perspective. One result has been a refinement of our understanding of ‘voice’.

Davis (1995) and Davis et al. (1998:192-196) note that the term ‘voice’ has been applied to several disparate phenomena, each emphasising a different aspect of event-participant relations. One usage highlights the relative effectiveness/affectedness of participants in their events, taking into account the relative humanness (or motility) of given participants (see the top tier of Figure 1, adapted from Davis et al. 1998:191). As noted in Davis et al. (1998:192), discussions of ‘transitivity’ (see Hopper & Thompson 1980), and ‘middle’ and ‘mediopassive’ voice (see Kemmer 1993) have typically emphasised this sense of ‘voice’.

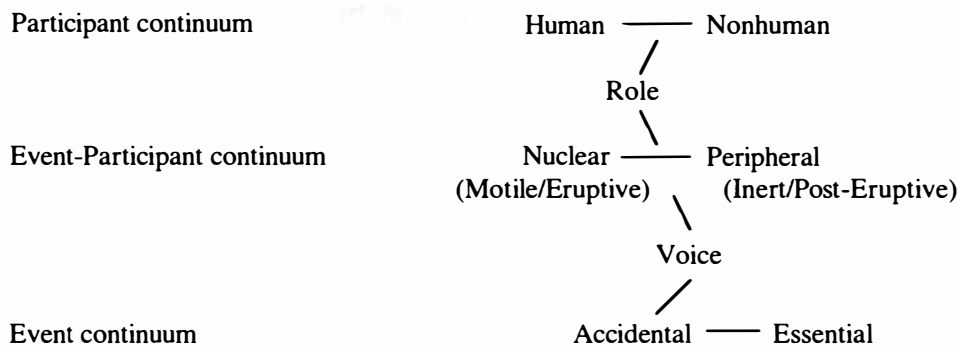


Figure 1: Relations of voice, role, event and participant

A second common understanding of 'voice', as discussed by Davis et al. (1998:192), addresses the arrangement of participants into a propositional nucleus and periphery, whereby participants are related to events via role, as seen in discussions of 'active', 'passive', 'antipassive' and so on (the linkage of the top and middle tiers in Figure 1). This tradition is exemplified in Foley and Van Valin (1984) (in terms of 'core' arguments/participants). As we have seen, however, the Philippine languages do not greatly elaborate such nominal/participant relations. None of the voice affixes requires object (or patient) promotion or subject (agent) demotion, and it is impossible to establish either the motile/eruptive or inert/post-eruptive forms as most 'basic'; and at any rate the voices elude a simple binary opposition, given the full range of their semantics. An 'ergative/absolute' characterisation is likewise uninformative about the specific semantic dimensions of these particular languages, as is implied by the very existence of the differences discussed above. (In this context, *Yogad* is remarkable for its lack of an antipassive.) It is unfortunate that the widespread quest for universals typically discourages the recognition and examination of such interesting interlinguistic differences.

Following Davis (1995) and Davis et al. (1998), the present discussion offers a third take on 'voice', one which highlights verbal event semantics, in particular the configuration of event content by the verbal affixes (the connection between the bottom tier and the middle one). In this context, event phases are selected, with varying degrees of focus, and with minimal attention given to participant semantics. The present encounter of established metalinguistic terminologies with linguistic 'novelty' has thus prompted a simultaneous 'deterritorialisation' and 'reterritorialisation' of the categories 'voice' and 'role'. With their relatively heavy emphasis on verbal voice phenomena and correspondingly light emphasis on nominals role semantics, the Philippine languages are generally well attuned to a way of speaking and thinking that the physicist David Bohm (1983:30-31) characterised as the 'rheomode', wherein 'movement is to be taken as primary in our thinking by allowing the verb rather than the noun to play a primary role.' This mode contrasts with what I may call the 'reomode' of Indo-European, whose grammatical emphasis on the noun conditions our understanding of events as the product of transitive relations between subjects and objects, of 'things doing things to things'.

A predictable characteristic of languages cast in the rheomode is a general grammatical indifference to control and transitivity (i.e. 'effectiveness' and 'affectedness'). Although the morphosyntax of each language does indeed recognise degrees of 'control' and 'transitivity', it does so incidentally, in terms of spatiotemporal focus/diffuseness. As discussed throughout, the grammatical indifference to control and transitivity *per se* is more marked in

Yogad, whose eruptive/post-eruptive role contrast is 'thinner' than the more dynamic and 'assertive' motile/inert contrast of Hiligaynon.

Along with the relatively low priority given to control/transitivity seen thus far, one might expect to encounter a corresponding degree of laxity in morphological causativity, and indeed, the morphological *pa-* 'causative' constructions, discussed in detail in Spitz (1997), are 'weak' or 'permissive' rather than 'causative' *per se*. In each language, the form *pa-* effects a displacement of performance from the 'S'-role to a later, possibly peripheral participant, as in sentence (69) from Yogad:

- (69) *Nap-pa-kánna yu babáy tu táwlay tu estudyánte.*
 NAG-PA-hit FOC woman UNF person UNF student
 'The woman **had/let** a man be hit by a student.'

In (69), the 'woman' may delegate the hitting or merely allow it or fail to prevent it. The highly epiphenomenal nature of such 'causation' is especially obvious when *pa-* occurs without other voice inflections, as in (70)–(71), exemplifying each language:

- (70) *Pa-bérde ang hílámón.* (Hiligaynon)
 PA-green FOC grass
 'The grass **is turning** green.'
- (71) *Pa-sándig yu kayú.* (Yogad)
 PA-lean FOC tree
 'The tree **tends to** lean.'

In each instance, *pa-* indexes a displacement of some participant from one condition or locus to another without any agentive motivation or control; each event just happens. Yet even here, an interlinguistic contrast between motility and eruption is observable. Unlike Yogad, Hiligaynon may enhance the displaced motility through 'reduplication':

- (72)a. *Gin-pa-ági sang dóktor ang kàrabáw sa subá'.*
 GIN-PA-pass UNF doctor FOC buffalo OBL river
 'The doctor **let** the buffalo cross the river.'
- b. *Gin-pa-pa-ági sang dóktor ang kàrabáw sa subá'.*
 'The doctor **saw to it that** the buffalo crossed the river.'

In (72a), the single occurrence of *-pa-* displaces the performance of 'passing' from the 'doctor' to the 'buffalo'; the doctor may have simply allowed the buffalo to pass. In the reduplicated version of (72b), the doctor has done something active, e.g. waved his arms, to encourage the buffalo to pass. While Yogad disallows such 'dynamic' or 'assertive' reduplication of *pa-* for intensification, it does exploit the distal middle semantics of *pag-* to achieve a more direct (and hence more 'intense') causal relation between the nuclear roles, as in the first gloss of (64) above.

The 'assertiveness' differential between the motile and the eruptive is, as we have noted, also reflected in the construction of participants. Hiligaynon inert roles, including obliques, are contrastively marked with either *sa* (for nonce-defined or pronominal participants) or *kay* (for particulars which are 'emancipated' from the immediate context of speech, including proper nominals); see (17). Furthermore, the Hiligaynon unfocussed pronouns alternate between pre-positional unbound forms (which, as genitives, precede their constituents) and post-positional clitics, as summarised in Table 6.

Table 6: Hiligaynon pronouns

	Focussed	Unfocussed	
		Prepositional	Postpositional
1SG	<i>akó</i>	<i>ákon</i>	<i>-ko/-nákon</i>
2SG	<i>ikáw/ka</i>	<i>ímo</i>	<i>-mo/-nímo</i>
3SG	<i>s(í)ya</i>	<i>íya</i>	<i>-níya</i>
1PL.INC	<i>kitá</i>	<i>áton</i>	<i>-ta/-náton</i>
1PL.EXC	<i>kamí</i>	<i>ámon</i>	<i>-námon</i>
2PL	<i>kamó</i>	<i>ínyo</i>	<i>-nínyo</i>
3PL	<i>silá</i>	<i>íla</i>	<i>-níla</i>

Focussed forms always name full participants; and the unfocussed forms, either full participants or genitives. Unfocussed prepositional forms may stand alone as definite inert participants after *sa*, or they may name 'assertive', 'emphatic', or 'contrastive' genitives:

- (73)a. *Gin-hátag-ko ang áwto-nákon sa ímo.*
GIN-give-1SG.UNF FOC car-1SG.UNF OBL 2SG.UNF
'I gave my car to you.'
- b. *Iní íya síya, indí' ímo.*
this.FOC 3SG.UNF chair not 2SG.UNF
'This is her chair, not yours.'

Among the unfocussed postpositional forms, the first and second person singular and the first person plural inclusive further differentiate between more emphatic forms (i.e. the bisyllabic n-initial ones) and less emphatic, 'default' forms (i.e. the monosyllabic ones).

Yogad morphology offers even fewer nominal/participant distinctions. Yogad merely employs successive instances of *tu* (or *tu ku...* for both proper or pronominal forms) to mark any unfocussed post-eruptive participants. The pronouns are simply either focussed or unfocussed, the latter forms being clitics whose initial consonants may assimilate to any preceding consonant, as with the free variants *nonót-ra* and *nonó-da* ('their minds'). In Table 7, the unfocussed second person singular and third person plural forms *-nu* and *-da* follow consonants; and *-m* and *-ra*, vowels.

Table 7: Yogad pronouns

	Focussed	Unfocussed
1SG	<i>kan</i>	<i>-ku</i>
2SG	<i>ka</i>	<i>-nu/-m</i>
3SG	<i>(ya baggí-na)</i> ⁹	<i>-na</i>
1DU.INC	<i>kitá</i>	<i>-ta</i>
1PL.INC	<i>kitám</i>	<i>-tam</i>
1PL.EXC	<i>kamí</i>	<i>-mi</i>
2PL	<i>kam</i>	<i>-maw</i>
3PL	<i>sirá</i>	<i>-ra/-da</i>

⁹ There is no proper third person singular focussed pronoun in Yogad. The phrase 'his/her body' serves in its stead.

Thus, even in their morphotactics, Yogad forms (i.e. pronouns and the voices *nag-*, *mag-*, and *pag-*) exhibit a greater degree of continuity, indistinctness, or 'blending in' than their more role-heavy, 'assertive' Hiligaynon cognates, an assimilation that 'rimes' with the extremely minimal role contrast in this language. Seen side by side, these two languages revalue the complementary participation of voice and role in the construction of events, challenging researchers to rethink these categories in the face of alternative virtualities.

References

- Bateson, Gregory, 1980, *Mind and nature: a necessary unity*. New York: Bantam Books.
- Bohm, David, 1983, *Wholeness and the implicate order*. London and New York: Ark Paperbacks.
- Davis, Philip W., 1995, Dimensions of voice. Paper presented at the International Conference on Functional Approaches to Language, 24th July, Albuquerque, NM.
- 1996, Transitivity in a Philippine language. Paper presented at the 25th LACUS Forum, Sep. 26, Toronto, Ontario.
- Davis, Philip W., John W. Baker, Walter L. Spitz and Mihyun Baek, 1998, *The grammar of Yogad: a functional explanation*. Munich: Lincom Europa.
- Foley, William and Robert Van Valin, 1984, *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Givón, Talmy, 1979, *On understanding grammar*. New York: Academic Press.
- Healey, Alan, 1958, *Notes on Yogad. Studies in Philippine Linguistics (Oceanic Linguistics 3:77-82)*. Sydney: University of Sydney.
- Healey, Alan and Phyllis Healey, 1956, *First primer: Yogad*. Manila: Summer Institute of Linguistics.
- Hopper, Paul J. and Sandra A. Thompson, 1980, Transitivity in grammar and discourse. *Language* 56:251-299.
- Kemmer, Suzanne, 1993, *The middle voice (Typological Studies in Language 23)*. Amsterdam: John Benjamins.
- Méntrida, Alonso and M.R.P. José Aparicio, 1894, *Arte de la lengua Bisaya-Hiligayna de la isla de Panay*. Manila: Tamborong.
- Ruiz, Macario B., 1968, *The behavior of Hiligaynon verb roots*. Iloilo City: Central Philippine University Research Center.
- Saussure, Ferdinand de., 1983, *Course in general linguistics*, trans. by Roy Harris. LaSalle, IL: Open Court Classics.
- Schachter, Paul, 1976, The subject in Philippine languages: topic, actor, actor-topic, or none of the above? In Charles N. Li, ed. *Subject and topic*, 491-518. New York: Academic Press.
- Spitz, Walter L., 1997, Lost causes: morphological causative constructions in two Philippine languages. PhD dissertation, Rice University.
- Wolfenden, Elmer, 1971, *Hiligaynon reference grammar*. Honolulu: University of Hawaii Press.
- 1975, *A description of Hiligaynon syntax*. Norman, OK: Summer Institute of Linguistics.
- Zorc, David Paul, 1977, *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. Canberra: Pacific Linguistics.

The position of Chamorro and Palauan in the Austronesian family tree: evidence from verb morphosyntax

ERIK ZOBEL

1 Introduction¹

Palauan and Chamorro, spoken in Micronesia on the Palau (Belau) and Mariana Islands, respectively, have long been recognised as outliers in the Pacific region, with stronger ties to the languages of the Philippines and Indonesia than to neighbouring languages. In his ground-breaking monograph, Dempwolff (1934–38) divided the Austronesian language family into an ‘Indonesian’ and a ‘Melanesian’ subgroup. It has become apparent that according to this scheme, Palauan and Chamorro must be included in the ‘Indonesian’ subgroup, since they do not share the innovations characterising the ‘Melanesian’ subgroup (which latter under the label ‘Oceanic’ has remained firmly established as a well-defined subgroup of the Austronesian family).

In Dyen’s lexicostatistical classification of the Austronesian languages (1965), Chamorro and Palauan are isolates of the ‘Malayo-Polynesian’ linkage, coordinate to subgroups of relatively high order. In spite of the problematic nature of lexicostatistics (Dyen’s classification fails to recognise well-established subgroups such as Oceanic), it illustrates the isolated character of Palauan and Chamorro with regard to their lexicon.

Blust (1977) proposed a classification of the Austronesian languages which up to now has gained wide acceptance (Figure 1). It contains two nodes relevant to the discussion here: the Malayo-Polynesian (MP) subgroup, based on phonological, lexical and grammatical innovations (Blust 1995); and the Central-Eastern Malayo-Polynesian (CEMP) subgroup (chiefly based on lexical innovations; Blust 1993), which includes the Central Malayo-Polynesian (CMP) and the Eastern Malayo-Polynesian (EMP) group. The latter contains the

1. I am indebted to Bernd Nothofer, Malcolm Ross and Sander Adelaar for their comments on earlier drafts of this paper.

Oceanic languages. MP languages that are not included in CEMP were grouped together by Blust in a Western MP (WMP) group.

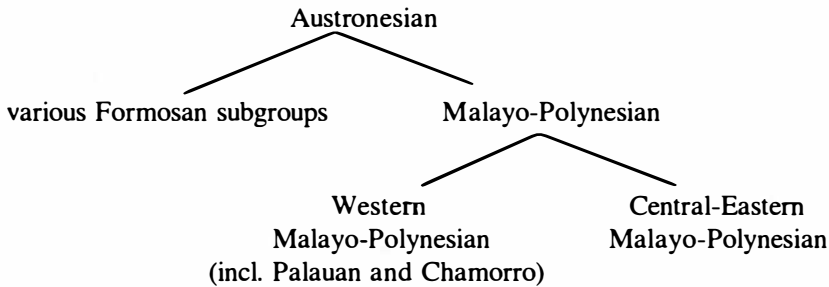


Figure 1: Austronesian family tree (following Blust)

According to this classification, Chamorro and Palauan must be included in the WMP subgroup, since they are clearly MP languages, and do not share the innovations defining CEMP. The next task is to establish closer ties between these languages and other WMP languages.

The phonological histories of the two languages give no clue apart from showing that they are non-Oceanic Malayo-Polynesian languages. The sound changes of both are either found in many other WMP languages (Chamorro: merger of **e* and **u*, Palauan: merger of **D* and **Z*, **ñ* and **n*) or are unique (Chamorro: merger of **D* and **k*, **j* and **q*; Palauan: merger of **j* and **R*, vocalisation of **l* and **p*). There are no phonological innovations common to both languages, apart from trivial ones (loss of **S*, stress on the PMP penultimate). Unlike Chamorro, Palauan shares with neighbouring Yapese and many Nuclear Micronesian languages the loss of final vowels, which is an areal feature in that part of the Pacific.

There have also been attempts to use grammatical evidence to establish the closer affiliations of Palauan and Chamorro. Pätzold (1968) demonstrated that many Palauan affixes (verbal and nominal) are shared with languages of the Philippines and Sulawesi, but this just proves the conservatism of Palauan in this respect.

For Chamorro, Topping (1973) claimed on the basis of its verbal system, that it should be grouped with Philippine languages such as Tagalog or Ilokano. His argument rests mainly on his focus analysis of the Chamorro verbal system, which is not fully appropriate, as I will show later. Starosta and Pagotto (1991) compared the Chamorro verbal system with Formosan and focus-preserving Malayo-Polynesian languages. They note the divergent character of Chamorro, which has led them to state that Chamorro is an early offshoot from PAN. However, if Chamorro is compared with languages further south, it can be seen that this divergence results from innovations which are not peculiar to Chamorro, but are shared by most languages of Indonesia and Oceania.

In this paper, I will use evidence from verbal morphosyntax to propose a modified subgrouping of the Malayo-Polynesian languages based on exclusively shared innovations, and establish the position of Chamorro and Palauan within this subgrouping.² I will discuss the affixation of verbs in the focus system, and in derivations from this system, and the interplay of these verbal affixes with the pronoun sets.

² The subgrouping hypothesis proposed here is elaborated in my dissertation (in progress).

A note on terminology: here, antipassive and passive are defined solely as syntactic surface categories. Both are syntactically intransitive, with the agent (A) or the object (O), respectively, as subject (S), and the other participant absent, incorporated or placed in an oblique relation. This presupposes that transitive and intransitive constructions are clearly distinguishable (e.g. in person marking), which is the case for Palauan and Chamorro.³ The basic transitive construction of both ergative and nominative languages is called 'active'. Focus languages are treated here as neither ergative nor nominative, since it is not clear whether actor focus (AF) or non-actor focus (non-AF) should be the 'basic' transitive form.

2 Reconstruction of the PMP verbal system

Most Formosan and many MP languages have a verbal system usually dubbed a 'Philippine-type' focus system. In line with Blust's subgrouping, it is safe to assume that this verbal system is inherited from PAN. Ross (1995) has reconstructed the focus system of PAN by concentrating on Formosan evidence, supported by additional data from MP languages. His reconstruction is summarised in Table 1.

Table 1: PAN focus system

	Past	Non-past	Atemporal	Projective
AF	<i>⟨um⟩</i>	<i>⟨um⟩</i>	∅	<i>⟨um⟩ -a</i>
UF	<i>⟨in⟩</i>	<i>-en</i>	<i>-u</i>	<i>-aw</i>
LF	<i>⟨in⟩ -an</i>	<i>-an</i>	<i>-i</i>	<i>-ay</i>

A reduplicated form of the non-past served as a progressive. There was also a 'stative passive' expressed by **ma-/*ka-*. It probably occurred in both non-actor focuses forming the paradigm found in Table 2.

Table 2: PAN stative passive

	Past	Non-past	Atemporal	Projective
UF'	<i>mina-</i>	<i>ma-</i>	<i>ka-</i>	?
LF'	<i>mina- -an</i>	<i>ma- -an</i>	<i>ka- -i</i>	?

The PMP verbal system did not differ much from the PAN verbal system (Table 3). The main innovations are the emergence of a fourth focus and the stem extensions **paN-* and **paR-*. The fourth focus using **Si-* is also found in many Formosan languages (e.g. Paiwan, Atayal, Bunun); the prefix **Si-* probably already existed at the PAN level as a noun forming affix (Ross 1995). The stative passive remained as in PAN.

³ With this definition, it will be seen below that nominative languages can have an antipassive (e.g. Palauan), just as ergative languages can have a passive (e.g. Chamorro).

Table 3: The PMP verb system

	Past	Non-past	Atemporal	Projective
AF	<i>⟨um⟩</i> <i>minaR-</i> <i>minaN-</i>	<i>⟨um⟩</i> <i>maR-</i> <i>maN-</i>	\emptyset <i>paR-</i> <i>paN-</i>	<i>⟨um⟩ -a</i> <i>maR- -a</i> <i>maN- -a</i>
UF	<i>⟨in⟩</i>	<i>-en</i>	\emptyset / <i>-u</i> (?)	<i>-a/aw</i> (?)
LF	<i>⟨in⟩ -an</i>	<i>-an</i>	<i>-i</i>	<i>-ay</i>
IF	<i>ini-/i- ⟨in⟩</i>	<i>i-</i>	<i>-an</i>	?
UF'	<i>mina-</i>	<i>ma-</i>	<i>ka-</i>	?
LF'	<i>mina- -an</i>	<i>ma- -an</i>	<i>ka- i</i>	?
IF'	<i>minai-</i>	<i>mai-</i>	<i>ka- -an</i>	?

The non-atemporal forms **maN-*, **maR-* and **ma-* are portmanteaus of **⟨um⟩* plus **paN-*, **paR-* and **ka-* with **um*. In the following discussion, I will call all affixes containing **⟨um⟩* *M-affixes* (including *⟨um⟩* itself), as opposed to *base affixes* (all non-AF affixes, and AF atemporal).

The stem extensions **paR-* and **paN-* are a characteristic of the Malayo-Polynesian languages: **paR-* is also found in Formosan languages, but restricted to forming reciprocals, while **paN-* with nasal substitution is an innovation particular to the Malayo-Polynesian group. Although it is difficult to establish the original function of the stem extensions, it can be roughly extrapolated from the modern daughter languages that in PMP **paR-* had a durative and reflexive/reciprocal function, while **paN-* had a distributive function, describing an action involving plural agents or objects. Both functions are transitivity-reducing, so it is not surprising that they are found mainly in AF (see below on the pragmatics of AF).

In many Malayo-Polynesian languages, the use of the stem extensions in non-AF has been limited to focussing circumstantial participants, such as location (LF), reason and instrument (IF) and occasionally time and manner, whereas non-AF without stem extensions focuses core roles, such as undergoer and goal. These 'circumstantial' focus forms (Table 4) are used mainly in cleft or equational constructions, which employ only the past and non-past tenses.

Table 4: Post-PMP 'circumstantial' focus forms

	Past	Non-past	Atemporal	Projective
LF''	<i>pinaR- -an</i> <i>pinaN- -an</i>	<i>paR- -an</i> <i>paN- -an</i>	<i>(paR- -i)</i> <i>(paN- -i)</i>	<i>(paR- -ay)</i> <i>(paN- -ay)</i>
IF''	<i>ipinaR-</i> <i>ipinaN-</i>	<i>ipaR-</i> <i>ipaN-</i>	<i>(paR- -an)</i> <i>(paN- -an)</i>	(?) (?)

In AF, the choice of *⟨um⟩*, *maR-* and *maN-* has become lexicalised in many Malayo-Polynesian focus languages, although some languages still allow all three forms with one verb.

No Malayo-Polynesian language has retained the system of Table 3 completely. In particular, the atemporal and projective non-AF forms have been conflated into a single category in all Malayo-Polynesian languages. Yet both sets have to be reconstructed 'from the top down', since they can be posited for PAN and reflexes of both are found in modern Malayo-Polynesian languages.

Noun case marking in PMP was as in modern Philippine languages, with nominative (marking the 'focussed' noun phrase), genitive (possessive, unfocused agent) and oblique (unfocused non-agent). The corresponding pronoun sets of PMP can be reconstructed as in Table 5.

Table 5: PMP pronoun sets

	Nominative	Genitive	Oblique
1SG	<i>(i-)aku</i>	<i>-(ng)ku</i>	<i>aken</i>
2SG	<i>(i-)kau</i>	<i>-mu/-nu/-u</i>	<i>iu(n)</i>
3SG	<i>sia</i>	<i>-nia</i>	<i>(ia ?)</i>
1PL.INC	<i>(i-)kita</i>	<i>-(n)ta</i>	<i>aten</i>
1PLEXC	<i>(i-)kami</i>	<i>-mami/-nami</i>	<i>amen</i>
2PL	<i>(i-)kamu(yu)</i>	<i>-muyu</i>	<i>imuyu(n)</i>
3PL	<i>siDa</i>	<i>-niDa</i>	<i>(iDa ?)</i>

In verb-initial sentences, genitive and nominative pronouns immediately follow the verb in that order (ignoring particles that also immediately follow the verb). If the sentence begins with a negative, adverb or any other member of a class of 'preverbs', genitive and nominative pronouns are fronted to immediately follow the preverb. In such constructions, the verb usually is in the atemporal. Examples (1) and (2) from Cebuano illustrate the fronting of pronouns.⁴

- (1) *Gi-tawg-an nako siya.*
 PST-call-LF 1SG.GEN 3SG.NOM
 'I called him.'
- (2) *Wa nako siya tawg-i.*
 NEG 1SG.GEN 3SG.NOM call-ATEMP.LF
 'I didn't call him.'

It can be assumed that the pragmatics of focus choice in PMP functioned as in modern Philippine languages. Focus selection is triggered by syntactic or pragmatic criteria. Syntactic criteria involve cases where the verb is nominalised, as in relative, existential and cleft clauses and in most WH-questions. Here, the verb must take the focus corresponding to the function of the highlighted NP. If the verb is not nominalised, focus is triggered by pragmatic criteria: roughly speaking, if the object NP is definite/referential and totally affected by the action,

⁴ Abbreviations used are: ACT active participle, AF actor focus, ANTI antipassive, APPL applicative, ART article, ASP aspect marker, ATMP atemporal, CONJ conjunction, GEN genitive, GER gerundive, HUM human, IF instrument focus, INTR intransitive, IRR irrealis, LF locative focus, LOC locative, NEG negative, NOM nominative, OBJ object, OBL oblique, PASS passive, PL plural, PERS personal article, POSS possessive, PST past, REAL realis, RED reduplication, RES resultative, SG singular, SUB subjunctive, UF undergoer focus.

this NP triggers the corresponding non-AF as in example (3) from Tagalog (if there is more than one non-agent core NP, case hierarchy determines focus selection). Zero anaphora for focussed non-agents is very common in most focus languages and was certainly a feature of PMP.

- (3) *D-in-alaw ko siya.*
 UF.PST-invite 1SG.GEN 3SG.NOM
 'I visited him.'
 (non-AF, definite object)

If the object NP is indefinite, or definite but partially affected, AF is selected,⁵ as in the Tagalog sentence (4).

- (4) *D-um-alaw ako ng mga kaibigan.*
 AF.PST-invite 1SG.NOM GEN PL friends
 'I visited some friends.'
 (AF, indefinite object)

The syntactic trigger always overrules the pragmatic trigger, as shown in example (5), from Tagalog: on pragmatic grounds, non-AF would be selected, but since the agent is highlighted in a construction that requires nominalisation of the verb, AF is chosen.

- (5) *Sino ang d-um-alaw sa kanya?*
 who NOM AF.PST-invite OBL 3SG.OBL
 'Who visited him?'
 (AF, definite object, but agent is questioned)

3 Grammatical sketch Of Chamorro

This short sketch is largely based on the descriptions by Topping (1973) and Cooreman (1987). Additional material is from Costenoble (1940).

In Topping (1973), Chamorro is described as having a focus system of the Philippine type. I will show below that the notion of focus (in the sense employed here) is not really applicable to Chamorro. Following Cooreman (1987), the Chamorro verbal system is better described as a split-ergative system. The ergativity-split is conditioned by mood: Chamorro distinguishes realis and irrealis mood: in realis there is ergative pronoun marking, while in irrealis there is nominative marking.

3.1 Chamorro pronoun sets and verbal morphosyntax

3.1.1 Pronoun sets

There are four pronoun sets in Chamorro (Table 6) with the following functions:

- Set A has two slightly different subsets depending on the mood of the sentence:
 - the agentive Set A1 marking A in realis mood;
 - the nominative Set A2 marking S and A in irrealis mood;
- Set B (absolute) pronouns mark S in realis mood and O in both moods;

⁵ Note that definiteness of the agent is not decisive for the selection of AF.

- the possessive set functions as possessor and A in certain nominalisations;
- free pronouns occur elsewhere.

Table 6: Chamorro pronoun sets

	Set B absolute	Set A1: agentive	Set A2:IRR nominative	Possessive	Free
1SG	<i>yo'</i>	<i>hu</i>	<i>(bai) hu</i>	<i>-hu/ku</i>	<i>guahu</i>
2SG	<i>hao</i>	<i>ɯn</i>	<i>ɯn</i>	<i>-mu</i>	<i>hago</i>
3SG	<i>gue'</i>	<i>ha</i>	<i>u⁶</i>	<i>-ña</i>	<i>guiya</i>
1PL.IN	<i>hit</i>	<i>ta</i>	<i>(u) ta</i>	<i>-ta</i>	<i>hita</i>
1PLEX	<i>ham</i>	<i>in</i>	<i>(bai) in</i>	<i>-mami</i>	<i>hami</i>
2PL	<i>hamyo</i>	<i>en</i>	<i>en</i>	<i>-miyu</i>	<i>hamyo</i>
3PL	<i>siha</i>	<i>[ma]⁷</i>	<i>uha/w/[uma]</i>	<i>-ñiha</i>	<i>siha</i>

The Sets A1 and A2 only differ in the third person singular and plural, and by the use of optional irrealis markers in some forms of Set A2. Both subsets of Set A are derived from the PMP genitive set (which is also directly continued in the possessive set), while Set B and the free set reflect the PMP nominative set.

The fact that the intransitive subject can be expressed by two pronoun sets makes it possible to distinguish clearly between syntactically transitive and intransitive constructions.

3.1.2 Intransitive verbs

Intransitive verbs can be divided into three classes depending on whether they take *<um>*, *ma(N)-* or \emptyset as singular realis affixes. Plural pronouns with singular verb forms have dual meaning. Intransitive affixes are given in Table 7.

Table 7: Intransitive verbal affixes in Chamorro

Realis singular	Irrealis singular	Realis plural	Irrealis plural
<i><um></i>	\emptyset	<i>maN-</i>	<i>faN-</i>
<i>ma(N)-</i>	<i>fa(N)-</i>	<i>manma(N)-</i>	<i>fanma(N)-</i>
\emptyset	\emptyset	<i>maN-</i>	<i>faN-</i>

The following pair illustrates the use of Set A and B pronouns in irrealis (6) and realis (7) mood:

- (6) Irrealis:
Para ta hanao.
 IRR 1PL.INC.A go
 'We will go.'

⁶ Actually, *u* is an irrealis marker that is obligatory in the third person and optional in the first person plural inclusive; *bai* is the irrealis marker for the first person (exclusive).

⁷ *ma-* is historically — and probably also synchronically — identical to the passive prefix *ma-*.

- (7) Realis:
Humanao hit.
 REAL.SG-go 1PL.INC.B
 'We went.'

Note that the irrealis forms that take Set A pronouns do not use *M*-affixes.

3.1.3 Transitive verbs

In the basic active construction, a transitive verb base is preceded by a Set A pronoun marking the agent, while the object is either a pronoun of Set B, as in (8) and (9), or a definite noun, as in (10). As illustrated in examples (8) and (9), active transitive forms do not change with mood, except for the slight difference in pronoun marking of the agent with Sets A1 (realis) and A2 (irrealis).

- (8) *Hu-li'e' gue'.*
 1SG.A-see 3SG.B
 'I saw him.'
- (9) *Para bai-hu-li'e' gue'.*
 IRR IRR-1SG.A-see 3SG.B
 'I will see him.'
- (10) *Hu-li'e' i lepblo.*
 1SG.A-see ART book
 'I saw the book.'

If the agent has to be extracted, as in cleft-, WH-clauses, relative clauses and equi-NP deletion, the agentive Set A pronoun is replaced by the infix $\langle um_2 \rangle$.⁸ The object is still represented by a Set B pronoun, as in (11).

- (11) *Hayi lum₂i'e' gue'?*
 who ACT-see 3SG.B
 'Who saw him?'

Topping has described $\langle um_2 \rangle$ as an actor focus affix. However, the latter sentence shows that the concept of focus — at least in the Philippine-type sense — is inapplicable to Chamorro, since transitive verbs with $\langle um_2 \rangle$ are Janus-faced in an odd way: to the left, the agent is highlighted, which would require AF in Philippine-type focus languages; to the right, the object is marked by an absolutive Set B pronoun, which corresponds to non-AF in focus languages. Avoiding the term focus, I will call forms with $\langle um_2 \rangle$ 'active participles', following Esser (1927) in his description of Mori.⁹ For intransitive verbs the participle is identical to the realis form of the verb.

⁸ This infix $\langle um_2 \rangle$ is homophonous with the intransitive infix $\langle um \rangle$ but not identical, since the latter only occurs with singular subjects, while the former is indifferent to number.

⁹ The term participle is employed, since $\langle um_2 \rangle$ replaces Set A person markers, which in the literature are often referred to as conjugation markers.

The affix *maN-/faN-* is employed to form an antipassive from transitive verbs,¹⁰ i.e. a form that is syntactically intransitive with the underlying agent as subject. The antipassive is mainly used if the object is indefinite; the latter can be left unexpressed or expressed by an unmarked noun, as in (12) and (13). Example (12) is in realis mood, and the agent/subject is represented by a Set B pronoun, while in (13), it is represented by a pronoun of Set A preceding the irrealis variant of the antipassive prefix. The intransitive nature of the antipassive can be seen in the use of either Set A or Set B pronouns for the agent/subject, depending on mood, in the same manner as in examples (6) and (7) above.

- (12) *Man-(t)aitai yo' lepblo.*
 ANTI.REAL-read 1SG.B book
 'I read a book.'
- (13) *Para bai-hu-fan-(t)aitai lepblo.*
 IRR IRR-1SG.A-ANTI.IRR-read book
 'I will read a book.'

With certain transitive verbs, the antipassive can also occur with definite objects, e.g. if the object is partially affected. The object then has an oblique or locative case marker, as in example (14). The active counterpart of (14) is sentence (15).

- (14) *Mam-(p)atek hao gi ga'lagu.*
 ANTI-kick 2SG.B LOC dog
 'You kicked at the dog.'
- (15) *Un-patek i ga'lagu.*
 2SG.A-kick ART dog
 'You kicked the dog.'

There are two passive affixes, *in>* and *ma-*, which are used in both realis and irrealis. Verbs with *in>* and *ma-* are syntactically intransitive, with the patient as subject. The intransitive nature of the passive is apparent in example (16), where the underlying patient is marked by a Set A pronoun, since it is the subject of an irrealis sentence. The agent, if present, is normally marked as oblique case, as in (17), although in cleft and similar constructions the agent can be marked as a possessor, as illustrated in (18).

- (16) *Ti un-hin>engge.*
 NEG 2SG.A-PASS-believe
 'You won't be believed.' (= 'He won't believe you.')
- (17) *Lini'e' si Maria as Pedro.*
 PASS-see PERS Maria OBL.PERS Pedro
 'Maria was seen by Pedro.'
- (18) *Hafa lini'e'-ña si Maria?*
 What PASS-see-3SG.POSS PERS Maria
 'What did Maria see?'

¹⁰ The infix *um₂* instead of *maN-* occurs with at least two verbs, namely *gimen* ('drink') and the suppletive *chocho* ('eat' – the corresponding active base is *kanno*').

- (19) *Ma-li'e' i palao'an.*
 PASS-see ART woman
 'The woman was seen.'

The *ma*-passive is used when the agent is unidentified, as in example (19), or third person plural. Otherwise, the choice between active and *in*-passive is dependent on rather complex discourse factors (Cooreman 1987). The transitive affixes are summarised in Table 8.

Table 8: Transitive verbal affixation in Chamorro

	Realis	Irrealis	Participle
Active	\emptyset	\emptyset	$\langle um_2 \rangle$
Antipassive (SG)	<i>maN-</i>	<i>faN-</i>	
<i>in</i> -passive	$\langle in \rangle$	(in)	
<i>ma</i> -passive	<i>ma-</i>	$(ma-)$	

The applicative affix *-i* (variant form: *-iyi*) has several functions, one of its main functions being the derivation of transitive stems from intransitive verbs and also sometimes from nouns. If suffixed to transitive verbs, it has benefactive function: the beneficiary then becomes the direct object, while the original direct object is put into oblique case. The suffix *-i* is not a focus affix since it can combine with all the above mentioned transitive affixes, giving *maN-i*, $\langle um_2 \rangle -i$, $\langle in \rangle -i$, and *ma-i*. Below I give examples for verbs suffixed with *-i*:

- hanagu-i* 'go to' (< *hanao* 'go')
apas-i 'pay' (< *apas* 'wage')
tugi'-i 'write to/for' (< *tuge* 'write something').

Examples (20) and (21) illustrate the use of *-i* in a basic active clause and with the active participle affix $\langle um_2 \rangle$, respectively.

- (20) *Hu-tugi'-i si Pedro ni katta.*
 1SG.A-write-APPL PERS Pedro OBL letter
 'I wrote the letter to Pedro.'
- (21) *Hu-konfotme kum₂uentus-i i ma'gas.*
 1SG.A-agree ACT-talk-APPL ART boss
 'I agree to talk to the boss.'

Transitive verbs can take the suffix *-(y)on* to form a stative verb 'capable of being X-ed'. (Occasionally, intransitive verbs can also take this suffix to express 'capable of X'.) Some examples include:

- atan-on* 'nice to look at' (< *atan* 'look at')
honggi-yon 'credible' (< *honggi* 'believe')
falagu-yon 'capable of running' (< *malagu* 'run').

The circumfix *faN-* *-(y)an* forms location nouns from verbs, for example:

- fañochoyan* 'eating place' (< *chocho* 'eat').

3.2 Historical derivation of the Chamorro verbal system

Table 9 gives an overview of how the PMP focus affixes of Table 3 are reflected in Chamorro:

Table 9

PMP	Chamorro		
AF:			
Past	* <i>minaN-</i>	(no reflex)	
Non-past	* <i>maN-</i>	> <i>maN-</i>	antipassive, realis mood
	* <i>um</i>	> <i>um₂</i>	active participle
Atemporal	* <i>paN-</i>	> <i>faN-</i>	antipassive, irrealis mood
Non-AF:			
Past	* <i>in</i>	> <i>in</i>	passive
	* <i>in</i> - <i>an</i>	> (<i>in</i> - <i>i</i>)	passive
Non-past	* <i>-en</i>	> <i>-(y)on</i>	(derives stative verbs)
	* <i>paN-</i> - <i>an</i>	> <i>faN-</i> <i>-(y)an</i>	(derives nouns)
Atemporal	* \emptyset	> \emptyset	active, realis/irrealis mood
	* <i>-i</i>	> <i>-i</i>	active, realis/irrealis mood

The following developments from PMP to Chamorro should be noted:

- PMP non-AF has become the Chamorro active, while AF developed into an antipassive, with the notable exception of PMP **um*, which has turned into the active participle, combining features of AF and non-AF.
- In the antipassive, the past/non-past distinction is lost, the non-past becoming the general realis form; the atemporal has become irrealis mood.
- The non-AF atemporal has become the general tense form of the active, while its past form has become a passive; the non-AF non-past is retained residually in lexical derivations.
- Further, there are two innovations that are not seen in Table 9:
- The PMP LF atemporal suffix **-i* has become the Chamorro applicative suffix *-i* which can combine with affixes that are derived from the PMP AF, *maN-*, *um₂*, see example (21).
- In the PMP atemporal, a pronominal agent with AF verbs is of the nominative set, while in non-AF it is of the genitive set, and both are fronted to preverbal position if the clause is opened by a negator (or any other preverbal modifier that requires the atemporal form of the verb). This is exemplified in sentences (22) in AF and (23) in non-AF from Waray-Waray (Central Philippines), which has retained the pattern assumed here for PMP.

- (22) *Waray pa ako kaon.*
 NEG yet 1SG.NOM eat
 'I haven't eaten yet.'

- (23) *Waray ko kaun-a.*
 NEG 1SG.GEN eat-ATEMP.UF
 'I haven't eaten [it].'

In contrast, sentences (24) and (25) show that in Chamorro, both in antipassive (< PMP AF) and in active voice (< PMP non-AF), the irrealis is preceded by Set A pronouns, which are derived from the PMP genitive set. Note that the verb forms which combine with these preposed pronouns employ *base affixes* (including \emptyset).

- (24) *Bai-hu-fa[n]-taitai.*
 IRR-1SG.A-ANTI.IRR-read
 'I will read (something).'
- (25) *Bai-hu-taitai i lepblo.*
 IRR-1SG.A-read ART book
 'I will read the book.'

Functionally, selection of active/antipassive in Chamorro corresponds to selection of non-AF vs. AF based on pragmatic criteria in Philippine-type focus languages. Significantly, the syntactic criteria for focus selection in 'focus' languages are *not* relevant for the selection of active/antipassive in Chamorro. This is apparent in the use of the active participle with $\langle um \rangle$. This is used in situations where Philippine-type languages have syntactically conditioned AF, as exemplified in the Tagalog sentence given above in example (5).

4 Grammatical sketch of Palauan

The following sketch mainly draws from two sources: Josephs (1975), which contains a host of sample sentences, although the analysis of data is inadequate at times; and Lemaréchal (1991), who has reinterpreted a good deal of the former's analysis in a much clearer way (see the appraisal by Josephs 1994). Additional information is taken from Pätzold (1968).

In order to identify PMP morphemes that have been retained in Palauan, note the following sound changes:

- unstressed vowels become ∂ (*e* in Palauan orthography) or \emptyset ;
- pre-stress **pa-* > *o-*, **pina-* > *ul(e)-*;
- Loss of **R* in clusters, as in **maR-* > *me-*, **paR-* > *o-*;
- **n* > *l*, as in **in* > $\langle i \rangle l$.
- The infix $\langle m \rangle$ (< **um*) is often realised as a back semivowel or as backing of the stem vowel.

4.1 Palauan pronoun sets and verbal morphosyntax

Palauan has a nominative pronominal agreement system. There are five sets of pronouns or pronominal affixes, which are given in Table 10.

Table 10: Palauan pronoun sets

	Free	NOM I	NOM II	OBJ	POSS
1SG	<i>ngak</i>	<i>ak</i>	<i>k-</i>	<i>-ak</i>	<i>-k</i>
2SG	<i>kau</i>	<i>ke</i>	<i>òm-</i>	<i>-au</i>	<i>-m</i>
3SG	<i>ngii</i>	<i>ng</i>	<i>l-</i>	<i>-ii</i>	<i>-l</i>
1PL.IN	<i>kid</i>	<i>kede</i>	<i>d-</i>	<i>-id</i>	<i>-d</i>
1PL.EX	<i>kam</i>	<i>aki</i>	<i>kim-</i>	<i>-am</i>	<i>-(m)am</i>
2PL	<i>kemiu</i>	<i>kom</i>	<i>òm-</i>	<i>-emiu</i>	<i>-(m)iu</i>
3PL.HUM	<i>tir</i>	<i>te</i>	<i>l-</i>	<i>-terir</i>	<i>-rir</i>
3PL.NON-HUM				\emptyset	

The free, NOM I and object sets are clearly derived from the PMP nominative set, while the NOM II set is related to Chamorro Set A, and together with the possessive set is derived from the PMP genitive set.

The choice between the two nominative pronoun sets depends largely on syntactic criteria: if the verb is clause-initial, the first set is used. An exception to this are imperative sentences, where the second set is employed. If the verb is preceded by a subject (= S, A) NP (or a part of it), then there is no nominative pronoun; if it is preceded by any other constituent (object, adverbial etc.), the second set is used. The second set is also obligatory after certain conjunctions.

Many verb forms alternate depending on whether they are preceded by a NOM I pronoun (or a nominative NP) or by a NOM II pronoun. For convenience, I will call the first verb form indicative and the latter subjunctive. Subjunctive forms never occur without a preceding NOM II pronoun; they are also never found in sentence initial position, except in imperative sentences.

The following examples illustrate the correlation between nominative pronoun, verb affix and word order. Example (26) is a verb initial sentence with a NOM I pronoun preceding an indicative verb. Examples (27) and (28) are rearranged versions of (26): in (27), the subject precedes the verb, in which case there is no nominative pronoun, while in (28), the object precedes the verb, which therefore has to be in its subjunctive form preceded by a NOM II pronoun.

(26) *Ng-meng-(ch)uiu er a hong a Droteo.*
 3SGI-ANTI-read OBL ART book ART Droteo
 'Droteo is reading the book.'

(27) *A Droteo a meng-(ch)uiu er a hong.*
 ART Droteo ART ANTI-read OBL ART book
 'Droteo is reading the book.'

(28) *A hong l-ong-(ch)uiu er ngii a Droteo.*
 ART book 3SGII-ANTI-read OBL 3SG.FREE ART Droteo
 'As for the book, Droteo is reading it.'

Example (29) illustrates the exceptional sentence-initial position of the subjunctive with a Nom II pronoun in an imperative sentence. The non-imperative counterpart of (29) is (30).

- (29) *D-o-rael!*
1PL.INCII-INTR-go
'Let's go!'
- (30) *Kede-me-rael.*
1PL.INCI-INTR-go
'We go.'

All transitive verbs and many intransitive verbs have two tense forms, past and non-past, and distinct forms for indicative and subjunctive. Transitive verbs can occur in active voice, antipassive voice and three forms of passive voice. Transitive affixes of Palauan are given in Table 11.

Table 11: Overview of basic verbal affixes in Palauan

	Indicative		Subjunctive	
	Non-Past	Past	Non-Past	Past
Intransitive	<i><m></i> <i>me-</i>	<i><l></i> <i>mil-</i>	\emptyset <i>o-</i>	<i><l></i> <i>ul-</i>
Transitive				
Active	<i><m₂</i>	<i><l₂</i>	\emptyset	<i><(i)l₂</i>
Antipassive	<i>meN-</i>	<i>mileN-</i>	<i>oN-</i>	<i>uleN-</i>
Passive	<i>me-</i>	<i>mil-</i>	<i>(me-)</i>	<i>(mil-)</i>
Resultative	<i><l>, <l> -el</i>			
Gerundive	<i>-el, -all</i>			

In the active, an object pronoun is obligatory and agrees with an overt object NP, as exemplified in sentence (31). By definition, a verb in the antipassive cannot take an object suffix; if the object is a pronoun or a definite NP, it takes the oblique marker *er*.¹¹ Examples (32) and (33) are antipassive sentences with an indefinite and a definite object respectively. Both active and antipassive have distinct forms for past and non-past, and indicative and subjunctive.

- (31) *Ak-kilisii* *a* *kiokl.*
*Ak-<il<sub>2 *kios-l*
1SGI-ACT-dig-3SG.OBJ ART dig-GER (=hole)
'I (completely) dug the hole.'</sub>*
- (32) *Ak-milengiis* *a* *kiokl.*
Ak-mileN-kios
1SGI-ANTI.PST-dig ART hole
'I was digging holes.'

¹¹ Josephs describes the difference between what is called active and antipassive here as an aspectual distinction between perfective and imperfective aspect. My interpretation follows the analysis of Lemaréchal, although with a different terminology.

- (33) *Ak-milengiis er a kiokl.*
Ak-mileN-kios
 1SGI-ANTI.PST-dig OBL ART hole
 'I was digging the hole.'

There are three passive forms: a verbal passive (*me-*) with two tense forms, but not distinguishing between indicative and subjunctive (34); and the resultative (35) and gerundive (36) passives, which are better regarded as derivations outside of the transitive voice paradigm, as they are often employed as nouns. All passives generally do not allow the explicit occurrence of the agent.¹²

- (34) *A blai a mil-seseb.*
 ART house ART PASS.PST-burn
 'The house was burnt.'
- (35) *A ulaol a ngɔatech.*
 ART floor ART RES-clean
 'The floor is cleaned.'
- (36) *A ulaol a ngetach-el.*
 ART floor ART clean-GER
 'The floor has to be cleaned.'

4.2 Historical derivation of the Palauan verbal system

Table 12 gives an overview of how the PMP focus affixes are reflected in Palauan:

Table 12: PMP focus affixes and their Palauan reflexes

PMP	Palauan			
AF:				
Past	<i>*minaN-</i>	>	<i>mileN-</i>	antipassive, past indicative
	<i>*amin></i>	>	<i>il₂</i>	active, past indicative
Non-past	<i>*maN-</i>	>	<i>meN-</i>	antipassive, non-past indicative
	<i>*um></i>	>	<i>m₂</i>	active, non-past indicative
Atemporal	<i>*paN-</i>	>	<i>oN-</i>	antipassive, non-past subjunctive
Non-AF:				
Past	<i>*in></i>	>	<i>⟨i⟩l₂</i>	active, past subjunctive
	<i>*in></i>	>	<i>ɔ</i>	(derives resultative)
	<i>*in> -an</i>	>	<i>ɔ -el</i>	(derives resultative)
	<i>*pinaN- (-an),</i> <i>*(i)pinaN-</i>	>	<i>uleN-</i>	antipassive, past subjunctive?
Non-past	<i>*-en, *-an</i>	>	<i>-el</i>	(derives gerundive)
Atemporal	<i>*∅</i>	>	<i>∅</i>	active, non-past subjunctive
	<i>*-i</i>		(no reflex)	

¹² Josephs gives 'awkward' examples of the verbal passive with an agent carrying the oblique marker 'er' which he suspects to be based on an English model.

The following developments from PMP to Palauan should be noted:

- Palauan has nominative agreement, with nominative pronouns always prefixed to the verb. The PMP nominative pronouns in post-verbal position are only preserved with object marking function.
- As in Chamorro, PMP non-AF has become active voice in Palauan, while AF developed into an antipassive, although there is 'cross-over', i.e. some active forms have an AF origin and some antipassive forms are derived from a non-AF source. Again, we find the unusual use of *m₂/il₂* with a following object suffix derived from the nominative set, combining features of AF and non-AF.
- The PMP tense distinction is preserved in Palauan, the only exception being the non-AF non-past, which has become a derivational affix. Its function has been taken over by the non-AF atemporal.
- Among the PMP non-actor focuses, only UF is preserved in the transitive paradigm of Palauan. The past subjunctive forms of antipassive voice are probably derived from LF and IF, since its function corresponds to the function of PMP 'circumstantial' non-AF focus forms, although there is no trace of the characteristic focus affixes.
- The indicative uses affixes derived from PMP *M*-affixes, while subjunctive forms are derived from base affixes. The Nom II pronoun set is equivalent to Set A in Chamorro, and like its Chamorro counterpart, is incompatible with *M*-affixes.

What has been said above about the functional correspondence between non-AF vs. AF in PMP and active vs. antipassive in Chamorro, also holds for Palauan.

4.3 The *-akl* suffix

Unlike in Chamorro, there is no productive applicative suffix. However, as noted by Pätzold (1968), some verbs seem to contain a fossilised affix *-akl*. Combined with the gerundive suffix *-(e)l*, this gives *-ekill* pointing to a synchronic deep form */-akil/* (< **-akin*). There are a few pairs of verbs where the root occurs both with and without */-akil/* (*/-okil/* in one instance):

<i>techolb</i>	'wash, baptize'
<i>techelbaki</i>	'dive into'
<i>toir, tir</i>	'chase'
<i>tiraki</i>	'follow, obey'
<i>iub, ibng</i>	'sneak out, avoid'
<i>ibngokil</i>	'sneak out, avoid'
<i>-renges</i>	'hear'
<i>beko/de/rengesaki</i>	'having sharp hearing' (the prefix <i>beko-</i> , <i>beke-</i> means 'good at doing something'; the additional <i>de-</i> cannot be explained)

It is very hard to extrapolate the original function of */-akil/* from these few examples. There are more examples where however the semantic distance of the pair is too great to exclude mere coincidence:

<i>dibechakl</i>	'cross'
<i>dibech</i>	'invent'

Most verbs that seem to bear a suffix /-akil/ do not occur without it, i.e. there is no way of showing that it is not an integral part of the stem:

<i>bedechakl</i>	'throw down, drop'
<i>ngeriakl</i>	'move forward'
<i>techemakl</i>	'stuff'
<i>techerakl</i>	'pick up with a hook'

5 Chamorro and Palauan innovations and their occurrence in other Malayo-Polynesian languages

If the innovations described in §3.2 and §4.2 are compared, it can be seen that Chamorro and Palauan share the following innovations if compared with PMP:

1. The pronoun set A or NOM II (derived from, but distinct from the possessive set), which occurs before verbs with base affixes;
2. the syntactic and semantic differentiation of $\langle um_2 \rangle / \langle m_2 \rangle$ and *maN-/meN-*, with * $\langle um \rangle$ taking over functions associated with non-AF.

Chamorro also has the following innovations not found in Palauan:

3. the circumfixes *maN-* *-i* and $\langle um_2 \rangle$ *-i*, combining PMP AF and non-AF (atemporal LF) affixes;
4. the loss of the tense (aspect) distinction involving **in*. The infix *in* has become a passive marker.

Apart from being a nominative language, Palauan seems to reflect one important innovation not found in Chamorro:

5. the suffix /-akil/.

5.1 The Set A pronouns

This pronoun Set A is found in many other Malayo-Polynesian languages: Sumatran languages (including languages of the Barrier Islands), Malay, Embaloh, Old Javanese, Sulawesi languages (excluding the focus languages of the North), and CEMP languages. Many of these languages have defective sets, e.g. Batak, Malay, Kaili. Set A is not found in the Philippines, Northern Sulawesi and Borneo (except for Malayic and Tamanic), nor in Sundanese and Balinese.

The most innovative feature of Set A pronouns is that they are placed before the verb, i.e. they are proclitics or — in most languages — prefixes with a fixed position, unlike the nominative or genitive sets in PMP or PAn, which are enclitics that are usually subject to raising.

Three types of languages can be distinguished according to the function of the Set A pronouns:

- (I) Set A pronouns have strictly agentive function, i.e. they are only used as agent markers with transitive verbs, as in Batak, Malay, Lampung, Embaloh, Kaili, Saluan, Totoli, Mandar.
- (II) Set A pronouns occur in a split-ergative system, i.e. generally marking the agent with transitive verbs, but also marking the subject with intransitive verbs¹³ in certain constructions, as in Chamorro, Buginese, Mori, Pamona, Nias.
- (III) Set A pronouns are part of a nominative agreement system, not discriminating between transitive and intransitive verbs, as in Palauan, Muna-Buton languages, and most CEMP languages.

In languages of type I, Set A pronouns co-occur with transitive verb forms derived from the PMP non-AF atemporal. This is also the case in languages of types II and III; here, additionally, intransitive verbs (including derived forms) taking Set A pronouns occur in a form derived from the PMP AF or intransitive atemporal. Usually, Set A pronouns are not compatible with M-affixes.¹⁴

Examples of Set A pronouns in languages of type I:

Karo Batak (Woollams 1996):

- (37) *Ku-guas takal-na.*
 1SG-thump head-3
 'I clobbered him on the head.'

Mandar:¹⁵

- (38) *U-issam-mi.*
 1SG-know-ASP.3
 'I already know.'

For languages of the types II and III, I will restrict myself to giving examples of their occurrence with intransitive/AF verb forms, especially when extended with **paR-/paN-*, to illustrate that these are a continuation of the PMP atemporal. Nias, Bugis and Bungku are western MP languages, while Kambera and Buli represent CMP and EMP, respectively.

Chamorro:

- (39) *Para bai-hu-fa-lagu.*
 IRR IRR-1SG-fa-run.
 'I will run.' (fa- < **paR*, cf. *ma-lagu*)

Palauan:

- (40) *D-o-rael!*
 1PL.INCII-o-go
 'Let's go!' (o- < **paR-*, cf. *me-rael*)

¹³ Including derived intransitive forms of transitive verbs.

¹⁴ Exceptions to this are found in languages which have completely lost the AF atemporal in favour of M-forms, even in imperative function, e.g. in Toraja or Banggai.

¹⁵ Bugis, Mandar, Bungku and Pitu Ulunna Salu data are from my own fieldnotes.

Nias (Sundermann 1913):

- (41) *Mi-o-fanö!*
 2PL-*o*-go
 'Go away (PL)!' (o- < *(p)aR-, cf. *mo-fanö*)

Bugis:

- (42) *Aja' mu-ac-cue:!*
 Don't 2-aC-follow!
 'Don't follow!' (aC- < *(p)aR-, cf. *mac-cue:*)

Bungku:

- (43) *Nahina-po ku-pong-kaa.*
 not-yet 1SG-poN-eat
 'I haven't eaten yet.' (poN- < *paN-, cf. *mongkaa*)

Kambera (Klamer 1994):

- (44) *Nggiki hi u-pa-taru?*
 why CONJ 2SG-pa-watch
 'Why are you watching?' (pa- < *paR-)

Buli (Maan 1951):

- (45) *...fare d-fa-pun-pun.*
 CONJ 3PL-fa-RED-hit
 '...and they hit each other.' (fa- < *paR-)

This use of Set A pronouns with AF/intransitive verb forms is not found in languages of type I. But it has to be noted that in these languages, verb forms derived from the PMP AF or intransitive atemporal are only used as imperatives (Mandar, Totoli, Saluan), or are not reflected at all (Batak, Malay). Many languages of this type have a defective set of preposed pronouns (Totoli, Saluan, Kaili, Embaloh, Batak).

The pronoun Set A could be either taken as a common innovation that occurred in a meso-language from which all above-mentioned languages have derived, or as an independent parallel innovation. The latter view is proposed by Himmelmann (1996) and Wolff (1996), who regard the defective sets as incipient stages to a full paradigm. In contrast, van den Berg (1996) reconstructs a full set for the parent language of at least some languages discussed here (Proto Celebic), but only for transitive verb forms: the extended use of set A pronouns with intransitive or AF verbs in languages of types II and III (both types being represented in his Celebic group) he regards as a later development.

Here I propose that Set A is a common innovation of all the languages in which it occurs, and that it originally was used with *both* non-AF and AF (and intransitive) atemporal verb forms, although the use with AF forms was more limited than with non-AF forms. Languages of types II and III offer ample evidence of this, especially since they include the isolated Chamorro and Palauan languages. Only in a later, and in most cases independent parallel development, have languages belonging to type I restricted the use of AF and intransitive atemporal to functions where there is usually no person marking (e.g. imperative), or have lost this atemporal completely, leading to the restriction of Set A pronouns to transitive forms derived from the PMP non-AF.

5.2 Antipassive *maN-* versus active participle *⟨um⟩*

At the end of §2 I discussed the pragmatics of AF in focus languages. Where pragmatic criteria require AF in PMP, Chamorro and Palauan have the antipassive, formed by a reflex of PMP **maN-/paN-*, while in those cases where AF is conditioned by syntactic criteria, a continuation of PMP **⟨um⟩* is used (here marked *⟨um⟩*). The objects of verbs carrying this *⟨um⟩*, especially pronominal objects, have a case form derived from the PMP nominative.

The syntactic and pragmatic differentiation of **maN-* and **⟨um⟩* and the particular object marking after **⟨um⟩* are a distinctive innovation that is found in only a few areas, which however have a widely scattered distribution: the South Sulawesi,¹⁶ Bungku-Tolaki and Muna-Buton groups of Sulawesi, older Toba Batak, Nias and Enggano, and probably Old Javanese and Old Balinese.¹⁷ In the following examples, (46) and (47) are equivalent to the Chamorro sentence (11), while the Nias sentences (48) and (49) parallel examples (31) and (32) from Palauan.

Pitu Ulunna Salu (South Sulawesi):

- (46) *Menna mu-hambi-ko?*
 who *mu*-hit-2SG
 'Who hit you?' (cf. *ku-hambi-ko* 'I hit you.')

Bungku:

- (47) *Inai 'umala-o?*
 who *⟨um⟩*-take-3SG
 'Who took it?' (cf. *ku-ala-o* 'I took it.')

Nias (southern dialect, Sundermann 1913):

- (48) *Gu-t⟨um⟩agu(-ya).*
 1SG.IRR-*⟨um⟩*-sew(-3SG)
 'I will sew it.' (definite object)
- (49) *Gu-man-(t)agu.*
 1SG.IRR-*maN*-sew
 'I will sew.' (indefinite object)

Note that the above mentioned languages (with the exception of Old Balinese) also display innovation 1. However, other languages of that group do not have a reflex of **⟨um⟩* in the transitive paradigm, but generally use a reflex of **maN-* (occasionally **maR-*). In some cases, this can be shown to be a later development:

- modern Javanese generally has *N-* (< **maN-*) where Old Javanese used *(m)aN-*, *(m)a-* (< **maR-*) and *⟨um⟩* (Kern 1918–20; Zoetmulder & Poedjawijatna 1961);
- in modern spoken Toba Batak, *⟨um⟩* is replaced by *maN-* (van der Tuuk 1971);
- Embaloh, shown to be closely related to Buginese, generally uses *maN-* (Adelaar 1994, 1995), with *m-* (< **⟨um⟩*) only occurring before vowels.

¹⁶ In Proto South Sulawesi, **⟨um⟩* is reflected as a prefix: **um-/mu-*.

¹⁷ In some of these languages, a reflex of **maR-* occurs next to **maN-* in an antipassive function, the choice of which is lexically determined.

Based on these three examples, it is probable that the **ma[N/R]-* vs. **um* contrast was also lost in many other languages in favour of **maN-*.

5.3 The affix combinations *maN-* *-i* and *um* *-i*

This innovation can be described as a symmetrisation of the focus system. As described above, the PMP focus system is asymmetrical since two or three non-actor focuses are matched by one actor focus. In Chamorro, the non-AF suffix *-i* can co-occur with the AF affixes *um* and *maN-*. Thus, Chamorro *-i* has become an 'applicative' affix.

Innovation 3 is found in all languages that have the first innovation, provided they have a reflex of the affixes **maN-* and **um* (which is not the case for most CEMP languages) and the atemporal non-AF suffixes (which is not the case in Palauan). It is further found in Balinese, Madurese and Sundanese. The symmetrisation is not restricted to former LF; the AF affixes can also combine with the PMP IF atemporal **-an*, giving **maN-* *-an* and *um* *-an* (e.g. Totoli, South Sulawesi, Selako). Note that in all languages (except Sundanese), this innovation involves the atemporal form of PMP non-AF.

5.4 Loss of past/non-past tense distinction

This innovation is found in almost all languages in which the former three innovations have taken place. An exception to this are Palauan and a few languages in a small stretch of Central Sulawesi, namely Saluan, Balantak, Kaili, and the Tomini-Tolitoli languages. These languages have retained the original PMP tense distinction.

In languages that have lost the past/non-past distinction, it is the non-past form that has been lost in AF. In non-AF, some of these languages retain all tense forms, but the atemporal has become the general form, while past and non-past forms have been relegated to specialised meanings. This is the case for Chamorro (see §3.2), Toba Batak and Buginese.

5.5 Palauan /-akil/

This fossilised suffix is probably related to the widespread applicative suffix **-aken*, although the vowel in the underlying final syllable presents a problem, since Palauan /i/ is not a regular reflex of PAn **e*. However, Sirk (1996) pointed out that there is much variation in reflexes of **-aken*, and the Palauan form falls well within this variation.

Reflexes of **-aken* occur in most languages that have at least one of the above innovations, and functionally, it has taken over the role of the atemporal IF suffix **-an*. It is not found in any language that clearly did not participate in innovation 3, i.e. its introduction must have post-dated the symmetrisation of the focus system. From its distribution, **-aken* can not be reconstructed for PMP as a suffix, but it is likely that it represents the capture of an oblique-case marker **(a)ken* common in the languages of the Philippines, which also survives as the Malay preposition *akan*.

In some language groups on Sulawesi, there is evidence that this capture occurred gradually: here, in some functions, **-aken* already occurs as an inseparable affix, while in others, **aken* still betrays its originally prepositional nature (Mead 1998). The replacement of inherited IF atemporal **-an* by **(-)aken* also must have occurred gradually, with both

morphemes occurring side by side at some stage having related but distinct functions. This is witnessed by the Malayic subgroup, where some languages have a reflex of **-an* (Kendayan, Selako, some Minangkabau dialects), while others have **-aken* (Malay, Serawai, Banjar). In Muna, both reflexes of **-an* and **-aken* are found: **-an* combines with second and third person pronouns, while a continuation of **-aken* is used with nouns.

Functionally, the relation of the Palauan suffix /-akil/ with **-aken* is unclear, since there is little agreement between them. The major functions of **-aken* are: causative (competing with the inherited causative prefixes **pa-* and **paka-*), benefactive, and instrumental, none of which can be assigned to the occurrence of /-akil/ in Palauan. One has to bear in mind however that the Palauan suffix is fossilised, and a closer inspection of its occurrences might reveal a connection with the functions of **-aken* listed above. At the present stage, Palauan is at best considered an imperfect witness for the emergence of the suffix **-aken*. It is even possible that Palauan words with /-akil/ reflect very early loans from an Oceanic language (cf. Proto Oceanic **-akini*), such as Yapese.

6 The Nuclear Malayo-Polynesian subgroup

6.1 Reconstruction of Proto NMP

The innovations discussed in §5 are not independent from each other, especially the innovations proposed in §5.1, §5.2 and §5.3, referred to as innovations 1, 2, 3 respectively in the following discussion:

- All languages displaying innovation 2 also have innovation 1 or 3. In turn, all languages that participate in innovation 1 or 3 and which still make use of the infix ⟨*um*⟩ in the transitive paradigm, also share innovation 2.¹⁸
- No language that shares innovation 1 has evidence that it did not participate in innovation 3, i.e. if a language has Set A pronouns, and has preserved both the AF affixes **maN-* or ⟨*um*⟩ and the non-AF suffixes **-i* or **-an*, it will also make use of the innovative affix combination(s). However, a few languages that display innovation 3 do not share innovation 1, such as Balinese, Sundanese and Madurese.¹⁹
- Innovations 1 and 3 are also functionally interdependent: the use of the atemporal non-AF suffixes in innovation 3 presupposes that in non-AF, atemporal forms are more frequent than the past and non-past forms, because the atemporal has taken over some of the functions of the past and non-past tenses. This is certainly the case in languages which share innovation 1.²⁰

This suggests that innovations 1, 2 and 3 occurred together in a common meso-language which was a daughter language of Proto Malayo-Polynesian and from which all languages

¹⁸ An exception is Acehnese, which has proclitic pronouns that appear to be related to the Set A pronouns, but uses ⟨*eum*⟩ (from **um*) in *de*-transitivising function, which is quite the opposite of innovation 2.

¹⁹ In Sundanese and Madurese, this might be due to the fact that pronouns have largely been relexified with nouns. This argument however does not hold for Balinese, where we still find the original PMP pronouns in Old Balinese and modern Bali Aga dialects, without any evidence for Set A pronouns.

²⁰ This argument however cannot be applied to Balinese and Madurese.

mentioned above derived. I will call this subgroup *Nuclear Malayo-Polynesian* (NMP), as it contains both Malay and the Polynesian family, and the meso-language *Proto Nuclear Malayo-Polynesian* (PNMP). (See Sirk 1978, 1996 for earlier attempts to use innovations 1, 3, and 5 for reconstruction and subgrouping.)

Innovation 4 (§5.4) is found in almost all NMP languages, except for those mentioned in §5.4. In one case, two closely related languages are separated by this innovation, namely Kaili (which has retained the tense distinction) and Pamona (where the distinction is lost). This shows that loss of tense is probably a drift-like phenomenon in the NMP subgroup.

Innovation 5 (§5.5) also postdates PNMP, as it is only found in languages that have innovations 1, 2 and 3, although not in all of them. As illustrated in §5.5, this innovation involved the gradual capture of the preposition **aken*, eventually replacing the atemporal IF suffix **-an*. We can assume that like innovation 4, this capture is the result of drift within the NMP subgroup.

The reconstruction of the PNMP system for transitive verbs given in Table 13 accommodates the evidence given by its daughter languages.

Table 13: The PNMP verb system

	Past	Non-past	Atemporal ²¹
Actor focus	<i>minaR-</i> <i>minaN-</i> (<i>umin</i>)	<i>maR-</i> <i>maN-</i> (<i>um</i>)	<i>paR-</i> <i>paN-</i> (\emptyset)
Actor participle	(<i>umin</i>)	(<i>um</i>)	
Patient focus	(<i>in</i>)		
Gerundive		-en	
Actor focus	<i>minaR- -i</i> <i>minaN- -i</i> (<i>umin</i>) -i	<i>maR- -i</i> <i>maN- -i</i> (<i>um</i>) -i	<i>paR- -i</i> <i>paN- -i</i> (-i)
Actor participle	(<i>umin</i>) -i	(<i>um</i>) -i	
Patient focus	(<i>in</i>) -an	-i	-i
Gerundive		-an	
Actor focus	<i>minaR- -an</i> <i>minaN- -an</i> (<i>umin</i>) -an	<i>maR- -an</i> <i>maN- -an</i> (<i>um</i>) -an	<i>paR- -an</i> <i>paN- -an</i> (-an)
Actor participle	(<i>umin</i>) -an	(<i>um</i>) -an	
Patient focus	(<i>i-</i>) (<i>in</i>)	-an	-an
Gerundive		<i>i-</i>	

This system differs from the PMP system in Table 3 in the following aspects:

- For each patient focus (\emptyset , -i, -an) there is a corresponding actor focus form; this is the result of innovation 3.

²¹ There was still a projective in PNMP, since it is found e.g. in Old Javanese; it is however not relevant for the discussion here.

- The patient focus non-past is formally identical to the atemporal; the PMP non-past has acquired a gerundive meaning.
- The active participle belongs to patient focus, since it is followed by a patient in nominative case; to the left however, it highlights the agent (innovation 2).
- Pronominal case marking has been reshaped in the atemporal: in AF, a fronted pronoun in pivot function in PNMP is of the genitive set, not of the nominative set as in PMP, while in non-AF, only the genitive pronoun (agent) is fronted, while the nominative pronoun (object) is not fronted, schematically:

	PMP:	PNMP:
AF:	PRV <i>aku</i> V	PRV <i>ku</i> V
UF:	PRV <i>ku sia</i> V	PRV <i>ku</i> V <i>sia</i>
LF:	PRV <i>ku sia</i> V- <i>i</i>	PRV <i>ku</i> V- <i>i</i> <i>sia</i>
IF:	PRV <i>ku sia</i> V- <i>an</i>	PRV <i>ku</i> V- <i>an</i> <i>sia</i>

Thus, in clauses with a verb in atemporal aspect, a pronominal agent is always of the genitive set, whether in AF or non-AF. Before the breakup of PNMP, these pronouns must have shifted their position from enclitic on the preverb to proclitic on the verb; because of their new position, they evolved into a set of their own, distinct from the genitive set, in all NMP daughter languages (innovation 1).

The latter innovation was probably the starting point for the development of the remaining innovations. They probably took place in the following order (see Wolff 1996 and Sirk 1996 explaining the emergence of innovations 1 and 3 in a similar way):

- With the development of pronoun set A, AF and UF atemporal become formally identical (although still differing with regard to the case marking of accompanying NPs). This leads to further symmetrisations of the focus system:
- In AF, **um* (or *M*-affixes) and Set A pronouns are in complementary distribution: since **um* occurs if the agent is in fronted position, this use is extended to non-AF, probably first in UF (*Q* refers to the object NP in nominative case):

	Atemporal	Agent fronted
AF:	<i>ku</i> -V	A <i>um</i> -V
UF:	<i>ku</i> -V <i>Q</i>	A <i>um</i> -V <i>Q</i> .

Thus, in UF the patient is in nominative case even if A is in preverbal focus position and the verb takes **um*, which is a major departure from the original PMP system where patients (especially pronominal patients) of verbs with **um* are in oblique case (innovation 2).

- Later this use is extended to the other non-AFs:

	Atemporal	Agent fronted
LF:	<i>ku</i> -V- <i>i</i> <i>Q</i>	A <i>um</i> -V- <i>i</i> <i>Q</i> .
IF:	<i>ku</i> -V- <i>an</i> <i>Q</i>	A <i>um</i> -V- <i>an</i> <i>Q</i> .

Eventually, **i* and **an* also combine with **maN*- and **maR*-.²² Thus **i* and **an* become applicative suffixes, independent of focus. This co-occurrence of AF and

²² This is a post-PNMP development, as in some NMP languages there is a constraint on the use of **i*-*an* with **maN/RJ*-, e.g. in Buginese and Chamorro.

non-AF affixes leads to symmetrisation of the PMP focus system, where originally one AF contrasts with three non-AFs (innovation 3).

Two widespread phenomena are post-PNMP drifts:

- (d) The loss of the PMP tense distinction: former AF past forms are lost completely, while the former non-AF past tense forms **in* and **in* -*an* acquire passive function without temporal connotation (innovation 4).
- (e) The emergence of the applicative suffix **-aken* (innovation 5).

6.2 From PNMP to Chamorro

Chamorro has retained the PNMP system quite faithfully. However, some of the development described in §3.2 are post-PNMP innovations:

- (i) PMP/PNMP Instrument Focus is lost completely;
- (ii) **in* -*an* has been replaced by *in* -*i*;
- (iii) the tense distinction between past and non-past is lost: with intransitive verbs and in AF, only the non-past form survived as the general form in realis mood, while in non-AF, the non-past form (= the PMP atemporal) becomes the general form for both moods, and the past form becomes a passive;
- (iv) the PMP stative passive **ma* -/*ka* - is generalised as *ma* -.
- (v) the PNMP atemporal is used for irrealis mood.

These developments, especially (ii) and (iii), have also occurred in many other NMP languages, but most probably as a result of drift (loss of tense, see §6.1) or paradigmatic leveling (emergence of *in* -*i*).²³

Incidentally, the morphosyntax of Chamorro as described in §3 is almost identical to that found in Mori (Esser 1927; Bارسel 1994).²⁴

6.3 From PNMP to Palauan

The case of Palauan is a little more complicated. The most significant innovation is the development of nominative agreement. Although this probably happened under the influence of Yapese and western Trukic isolects, with which Palauan also shares other areal features, the constructions that are involved already existed in PNMP.

Indicative clauses with Nom I pronouns are derived from PNMP clauses with preposed subject. For that reason, the indicative active contains the PNMP active participle infix **um*₂. Intransitive constructions with a subject pronoun following the verb fell into disuse. Cf. the following examples with the etyma **maR-zalan* 'to walk' and **tanem* 'to plant':

²³ In eastern Central Sulawesi, Saluan has retained *in* -*an*, while the related Balantak has innovative *ni* -*i*.

²⁴ The similarity is a formal one: these two languages differ quite strongly in the pragmatics of the use of the passive and of the atemporal forms of intransitive verbs.

PNMP	Pre-Palauan, with fronting of subject/agent	Palauan
* <i>MaR-zalan akuf.</i> (INTR)	* <i>Aku maR-zalan.</i>	<i>Ak-me-rael.</i> (INTR)
* <i>MaN-(t)anem akuf.</i> (AF)	* <i>Aku maN-(t)anem.</i>	<i>Ak-mel-(d)alem.</i> (ANTI)
* <i>Ku-tanem iaf.</i> (non-AF)	* <i>Aku t_{um}anem ia.</i>	<i>Ak-d_olem-ii.</i> (ACT)

† *aku* and *ia* represent PMP nominative pronouns.

More complex is the origin of the subjunctive in Palauan. As has been illustrated in Table 12, it is partially derived from the original atemporal, partially from non-AF forms. The original atemporal is retained in imperative and negative clauses, and after certain conjunctions; all these only use non-past subjunctive forms, as illustrated in (50) and (51).

(50) *D-o-rael!* (< **ia-paR-Zalan!*)
1PL.INCII-SUB-go
'Let's go!'

(51) *A le-me a chull...*
if 3SGII-come ART rain
'If it rains...' (lit. 'If the rain comes...')

In the case of fronting of non-subject constituents, the subjunctive is derived from non-AF forms of PMP, which can be seen from the fact that there is a past subjunctive form. Unlike in other NMP languages, *all* genitive pronouns occurring with non-AF verbs have been fronted, not only with atemporal forms, but also with past forms, as in (52).

(52) *A ngikel a le-k_{il}a a bilis.*
ART fish ART 3SGII-ACT.PST-eat ART dog
'The fish were eaten up by the dog'

Here, *le-k_{il}a* is derived from PMP **k_{in}an-nia*, with raising of the genitive pronoun. This is not found in other NMP languages, where the genitive pronoun has become fixed in the position *following* verbs in non-AF past tense (or in the passive that has developed from it).²⁵

As in Chamorro, the PMP stative passive **ma-/ka-* is preserved in Palauan as the verbal passive prefix *me-*, without, however, retaining the atemporal form **ka-*.

6.4 Scope and position of Nuclear Malayo-Polynesian within the Malayo-Polynesian family

If one takes the innovations discussed in §6.1 as diagnostic evidence, the NMP subgroup includes the languages of the CEMP group, Chamorro and Palauan, and most WMP languages of Malaysia and Indonesia. Not included in the NMP group are the following WMP languages: the languages of the Philippines, the three Northern Sulawesi groups

²⁵ Compare the following examples from NMP languages: Kaili *ni-kande-ku*, Saluan *k_{in}aan-ku* 'I ate it', Nias *ni-rongo-mi* 'what you (PL) heard', as opposed to *ku-kande*, *ku-kaan* 'I will eat it' and *mi-rongo* 'you hear'.

(Gorontalo-Mongondic, Minahasan, Sangiric), the Sama-Bajau languages, Malagasy, and all languages of Borneo with the exception of the Malayic and Tamanic groups.

Blust (1999) has pointed out that his Western group of Malayo-Polynesian (WMP) is not to be understood as a subgroup tied together by exclusively shared innovations, but just as an umbrella term for all MP languages not included in the Central-Eastern subgroup. Here, I have shown that indeed this WMP group has to be broken up by adding a node to his tree, as shown in Figure 2.

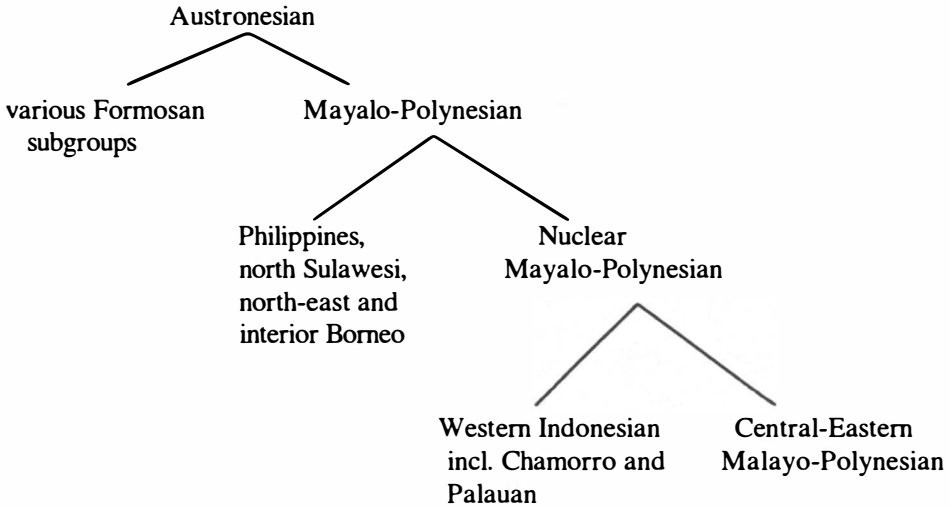


Figure 2: Modified Austronesian family tree

But in turn, it has to be emphasised that the upper branches leaving the MP and NMP nodes are not to be taken as well-defined subgroups. There may be more than one subgroup of MP coordinate to NMP, just as 'Western Indonesian' might contain several subgroups coordinate to CEMP.

In the introduction I mentioned that Chamorro and Palauan have no (non-trivial) common phonological innovations. Also the grammatical aspects discussed here do not point to a closer relation of these two languages to each other, compared with the remaining NMP languages. What they share is a certain morphological conservatism (e.g. the retention of the *maN-/um* distinction), and the retention of the passive prefix **ma-*, with loss of the atemporal form **ka-*. Only a few 'Western Indonesian' languages have preserved **ma-/ka-* as a productive morpheme. In most languages it has been replaced by a generalised **ka-* (Javanese, Toba Batak), or the widespread innovation **taR-*, which is also found in CEMP languages. Another feature that is widespread among other NMP languages but not found in Chamorro and Palauan, is the capture of the preposition **(a)ken* as a benefactive and instrumental applicative suffix. Taken together, these points indicate that Chamorro and Palauan are early offshoots from PNMP.

7 The dispersal of MP and NMP: a scenario for early migrations

The family tree proposed by Blust allows some tentative conclusions about the homeland of Austronesian speakers and the way they expanded into the archipelago and Oceania. Most certainly, Taiwan must have the longest history of Austronesian settlement; from there speakers of PMP moved south to the Philippines and further to Sulawesi, Borneo, and the Sunda islands. Speakers of Proto CEMP then broke away to migrate to eastern Indonesia and on to Oceania. These movements can be partly correlated with archeological findings (Bellwood 1985).

The NMP hypothesis allows us to refine this scenario in the following way: From the Philippines, there were movements to Borneo and Sulawesi. The latter island is a good candidate for the center of NMP dispersal, since there we find a large number of morphologically conservative languages with diverse grammatical systems. From Sulawesi, NMP speakers expanded to the Sunda islands, to parts of Borneo (as Malayic and Tamanic speakers), and to the east and south-east. It was probably also from Sulawesi that the speakers of Chamorro and Palauan (or better: Pre-Chamorro and Pre-Palauan) sailed to the northeast to the distant islands of Palau and the Marianas. The early dates of the first settlement of these islands are in accord with the assumption that Chamorro and Palauan are early breakaways from PNMP.

Of course, these speculations do not take into account the possibility that NMP languages were also spoken in other areas, e.g. in the Philippines, and have later been replaced by focus-retaining languages. Thus, Palauan and Chamorro speakers may have departed from an area outside of the present-day NMP area.

Chamorro must have remained in relative isolation for a long time, although a few loanwords from Western MP languages and Oceanic can be detected (Costenoble 1940). At one point during colonial times, the Chamorro population experienced a catastrophic reduction. This and other factors resulted in massive borrowing from Spanish and, to a lesser degree, from Philippine languages. Yet, its basic verbal morphosyntax seems to have remained quite unchanged.

In the case of Palauan, speakers of it had contacts already in pre-colonial times with speakers of Oceanic languages, first with Yapese, then with Trukic. Ross (1996) has shown that Yapese is an early offshoot of Oceanic that has borrowed heavily from Palauan and Nuclear Micronesian languages. In turn, some of the aspects where Palauan deviates from PNMP might be explained by convergence with Yapese.

References

- Adelaar, K. Alexander, 1994, The classification of the Tamanic languages. In Tom Dutton and Darrell Tryon, eds *Language contact and change in the Austronesian world*, 1-42. Berlin: Mouton de Gruyter.
- 1995, Problems of definiteness and ergativity in Embaloh. *Oceanic Linguistics* 34:375-409.
- Barsel, Linda A., 1994, *The verb morphology of Mori, Sulawesi*. Canberra: Pacific Linguistics.
- Bellwood, Peter, 1985, *Prehistory of the Indo-Malaysian archipelago*. Sydney: Academic Press.

- Berg, René van den, 1996, The demise of focus and the spread of conjugated verbs in Sulawesi. In Steinhauer, ed. 1996: 89-114.
- Blust, Robert A., 1977, The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *University of Hawaii Working Papers in Linguistics*, 9/2:1-15.
- 1993, Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32:241-293.
- 1995, The position of the Formosan languages: method and theory in Austronesian comparative linguistics. In Li et al. eds, 1995:585-650.
- 1999, Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun and Paul Jen-kuei Li, eds *Selected papers from the Eight International Conference on Austronesian Linguistics*. Taipei: Institute of Linguistics, Academia Sinica.
- Cooreman, Ann M., 1987, *Transitivity and discourse continuity in Chamorro narratives*. Berlin: Mouton de Gruyter
- Costenoble, H., 1940, *Die Chamoro-Sprache*. 's-Gravenhage: Nijhoff.
- Dempwolff, Otto, 1934-38, *Vergleichende Lautlehre des austronesischen Wortschatzes*. Berlin: Reimer.
- Dyen, Isidore, 1965, *A lexicostatistical classification of the Austronesian languages*. International Journal of American Linguistics Memoir 19.
- Esser, S.J., 1927-33, *Klank- en vormleer van het Morisch*. Verhandelingen van het Koninklijk Bataviaasch Genootschap van Kunsten en Wetenschappen 67, 3/4. Leiden, Bandung.
- Himmelmann, Nikolaus, 1996, Person marking and grammatical relations in Sulawesi. In Steinhauer, ed. 1996:115-136.
- Josephs, Lewis S., 1975, *Palauan reference grammar*. Honolulu: University Press of Hawaii.
- 1994, Review of Lemaréchal (1991). *Oceanic Linguistics* 33:231-256.
- Kern, H., 1918-20, *Spraakkunst van het Oudjavaansch: Verspreide Geschriften VIII + IX*. 's-Gravenhage: Nijhoff.
- Klamer, Marian, 1994, *Kambara: a language of Eastern Indonesia*. Den Haag: Holland Academic Graphics.
- Lemaréchal, Alain, 1991, *Problèmes de sémantique et de syntaxe en Palau*. Paris: Edition du Centre National de la Recherche Scientifique.
- Li, Paul Jen-kuei, Dah-an Ho, Ying-kuei Huang and Cheng-hwa Tsang, eds, 1995, *Austronesian studies relating to Taiwan*. Symposium Series of the Institute of History and Philology, Academia Sinica 4. Taipei: Institute of History and Philology, Academia Sinica.
- Maan, G., 1951, *Proeve van een Bulische Spraakkunst*. 's-Gravenhage: Martinus Nijhoff.
- Mead, David, 1998, *Proto-Bungku-Tolaki: reconstruction of its phonology and aspects of its morphosyntax*. PhD dissertation, Rice University, Houston.
- Nothofer, Bernd, ed., 1996, *Reconstruction, classification, description: festschrift in honor of Isidore Dyen*. Hamburg: Abera.
- Pätzold, K., 1968, *Die Palau-Sprache und ihre Stellung zu anderen indonesischen Sprachen*. Berlin: Reimer.
- Ross, Malcolm D., 1995, Reconstructing Proto-Austronesian verbal morphology: evidence from Taiwan. In Li et al., eds, 1995:727-791.
- 1996, Is Yapese Oceanic? In Nothofer, ed. 1996:121-166.

- Sirk, Ülo, 1978, Problems of high-level subgrouping in Austronesian. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, 255-273. Canberra: Pacific Linguistics.
- 1996, On the history of transitive verb suffixes in the languages of western Indonesia. In Steinhauer, ed. 1996:191-205.
- Starosta, Stanley and Louise Pagotto, 1991, The grammatical genealogy of Chamorro. In Ray Harlow, ed. *VICAL 2: Western Austronesian and contact languages. Papers from the Fifth International Conference on Austronesian Linguistics*. Auckland: Linguistic Society of New Zealand.
- Steinhauer, Hein, ed., 1996, *Papers in Austronesian linguistics* No.3. Canberra: Pacific Linguistics.
- Sundermann, D. Heinr., 1913, *Niassische Sprachlehre*. 's-Gravenhage: Martinus Nijhoff.
- Topping, Donald, 1973, *Chamorro reference grammar*. Honolulu: University Press of Hawaii.
- Tuuk, H.N. van der, 1971, *A grammar of Toba Batak*. The Hague: Nijhoff.
- Wolff, John U., 1996, The development of the passive verb with pronominal prefix in Western Austronesian languages. In Nothofer, ed. 1996:15-40.
- Woollams, Geoff, 1996, *A grammar of Karo Batak, Sumatra*. Canberra: Pacific Linguistics.
- Zoetmulder, P.J. and I.R. Poedjawijatna, 1961, *Bahasa Parwa*. Jakarta: Obor.

— PART V —

Discussion notes

Final words: the development of the focus system

JOHN U. WOLFF

1 Introduction: reconstruction of Proto Austronesian focus

The last fifty years has seen a geometric increase in the amount of detailed and reliable data available on the Austronesian languages, not only in the form of carefully transcribed and wide-ranging texts in a large number of languages, but also in the form of well-grounded linguistic analyses, which provide answers to the questions which arise in taking the achievements of recent theoretical linguistics into consideration. This has put us in a position to do some serious reconstruction of grammatical features and develop theories which account for the rise of the morphology and syntax that characterise the currently attested languages. We have moved some distance from Dempwolff's dictum to the effect that the modern languages do not present morphological or syntactic data which allow for reconstruction of the grammar of the protolanguage.¹ The foregoing collection of studies is proof of the extent to which our knowledge of the grammatical structures of Austronesian languages has improved. Some of these articles address overtly issues of historical development, and all of them contribute to the endeavor of summarising what is now known about the historical development of forms attested for the modern languages. In this brief essay I will attempt to summarise some of the issues concerning grammatical history which lie at the heart of the grammatical systems of Austronesian languages and which are at stake in the essays of this collection. I will attempt to outline elements connected with these issues which can bring us forward on the path of the total research agenda, the final goal of which

¹ Dempwolff (I. 1934, §1, p.13) wrote:

Diese Sprachen haben keinen derart einheitlichen grammatischen Bau, wie etwa die Semiten- oder die Bantu-Sprachen...Deshalb beschäftigt sich diese Arbeit nur gelegentlich mit grammatischen Übereinstimmungen und beschränkt sich auf die vergleichende Untersuchung von Wörtern, in der Hauptsache auf deren Lautverhältnisse.

'These languages have no unified grammatical structure, as for example the Semitic or Bantu languages have...For this reason this work only deals incidentally with grammatical correspondences and is confined to the comparative investigation of words, most particularly, of their sound correspondences.'

is to reconstruct Proto Austronesian (PAN) syntax and morphology and present a picture of how this developed into the linguistic structures found in the attested languages.

These papers deal with what has been termed 'focus' — that is, the portion of the verbal system which marks the semantic relation between the verb and its argument as the agent, patient, indirectly or partially affected patient or the location of the action, as the patient conveyed, as the instrument, or as the beneficiary of the action. These relations are marked by morphemes affixed to the verbal stem in many Austronesian languages, and this system of affixation has been termed 'focus'. The suitability of this term for the concept has been brought into question by numerous authors including authors of papers in this collection (e.g. Poedjosoedarmo, Himmelmann),² but for better or worse, since this term has been used in many papers in this collection to refer to this feature of the verbal system, I will continue to use it in this essay as well.³ The focus system of verbal affixation is a key element in the grammar of PAN, and an understanding of the focus system will bring us well forward toward a picture of how the structures of the attested languages came into being.

Further, it is important to know that languages with the Philippine type of focus system (§2, below) also distinguish a nominative and a genitive case (as well as a third or even fourth case), marked by separate words (clitics) or by inflected forms (in the case of the pronouns).

To my knowledge, my own work (1973, 1980) was the first attempt to develop a picture of the PAN verbal system by discussing forms from a small number of Austronesian languages located from northern Taiwan to Oceania. This original picture was clarified and expanded by Ross (1995) by taking into account all available data from the Austronesian languages of Taiwan. The resulting system is convincing enough to be quoted in its basic outlines by several of the authors of papers in the volume. Although no one has yet been able to reconstruct with certainty a morpheme of a given shape for every single category, it has been possible to do so for some of them, and in any case the categories are reconstructable with confidence. One version of the categories and the morphemes reconstructed for them is presented in Zobel's contribution to this group of papers (Zobel: Tables 1–4). For convenience of reference I will call a paradigm of this type the 'Philippine-type' system, as the languages of the Philippines all manifest a verbal focus system which is isomorphic or nearly isomorphic with this type, or show what are clearly developments from this system and do not move in any radical way far from it. Verbal systems of a similar type are found south of the Philippines in languages of northern Sulawesi and parts of Kalimantan and also most of the languages north of the Philippines on Taiwan.⁴

² Contributors to this volume are shown by author's name only.

³ Ross (1995) as well as many others refers to this feature as 'pivot' in accordance with terminology adopted for similar phenomena in other language families; Himmelmann, Boutin, Norwood, Donohue refer to it as 'voice'.

⁴ Exceptions are the Rukai languages, which have probably changed the verb system to lose focus. Starosta (1995) believes that Rukai reflects the PAN system and the languages which show the 'Philippine type' system have innovated in common. There is no evidence to support this, and I believe that a careful study of Rukai and its history will demonstrate that the Rukai system is indeed a spin-off from the original focus-type system. In any case, we need not provide a definitive answer to this question here. Should it turn out that Rukai never developed this focus system, then we are dealing in this volume with a PAN which existed after Rukai. Other languages on Taiwan also have moved more than a short distance from the quintessential 'Philippine type' of focus system, e.g. Tsou, which has clearly lost the tense-aspect affixes, Bunun which has developed a series of paraphrastic forms which interact with the inherited tense-aspect affixes, and other languages as well have made other sorts of innovations. We cannot discuss these innovations here. Starosta also speculates that some of these other languages have verbal systems which predate the development of the focus system. No evidence has been adduced to support these speculations.

2 The Philippine type of focus system and developments from it

The hallmark of the Philippine type is that there are there are four focuses: actor (AF), Patient (PF),⁵ Local (LF) and a fourth variously called 'instrumental', 'applicative', or 'benefactive' (usually abbreviated IF). A verb consists of a verb stem plus one and no more than one of these four focus affixes, plus another affix which shows tense, aspect and mode. (The affix may consist of one or two morphemes and the focus and tense affixes may combine into a portmanteau affix.) A verb stem may consist of a root or a root plus a derivational prefix or even several derivational prefixes. The crucial thing to note here is that the verb stem may have only one focus affix at a time in the Philippine-type systems. The languages of the 'Philippine type' are found in a geographically delimitable area. It is the languages south, east, and west of the 'Philippine-type' languages that show differences in structure. These languages, not of the quintessential Philippine type, differ from the Philippine type system in one or more of the following characteristics:

- (a) the verb may contain an AF and additionally one of the other focus affixes, e.g. *Pendau* (Quick, example (44), *nongolia* 'bought for' which consists of *nong-* 'AF-past' (< **paN-* + **-in-* + AF) + *oli* 'buy' + *-a* 'IF' (< **-án* 'IF, atemporal')
- (b) the verb has no tense-aspect distinctions and manifests the 'non-past' affix in AF and the 'atemporal' affix in PF, LF and IF (following Zobel, Tables 1–4)
- (c) the verb stem shows person agent agreement in all or some of its forms. The person agreement marker is typically in the shape of prefixes, but in some languages the person markers are independent words. (In that case, they are most often clitics.) Some languages have person markers which may be either independent words or affixes (e.g. *Sasak* — see *Wouk*).

This third deviation from the Philippine-type is of three sub-types:

- (i) The agent agreement is found only with the PF, LF, and IF.
- (ii) Agent agreement is found with all verb forms, including the AF.
- (iii) There is an additional patient agreement marker for transitive verbs.

A fourth variation on the third deviation is that a few languages are mixed in type: *Chamorro*, for example, shows agent agreement with AF, PF, LF with some verb forms, while some other AF, PF, and LF verb forms do not show agreement with the person of the agent. A fifth variation is that the verb is marked with person agreement only for the first person singular. For other persons and numbers no agreement marker occurs (e.g. in *Tolitoli*, *Wolff* 1996), or the verb has an agent prefix but this is only the third person prefix, or only a first person singular marker plus a third person marker for all the other agents, as for example *Lauje* (see *Himmelmann*).

and a cursory inspection of what is known about them makes it fairly clear that these deviations from the quintessential Philippine type of affixation are in fact later developments. We are only lacking the basic research on the grammatical structure of the languages in question to provide unequivocal confirmation that their verbal systems developed from an earlier 'Philippine'-type. I am convinced that all of the verbal systems of the Austronesian languages on Taiwan descend from a Philippine-type system (as is represented by *Seediq*— see *Holmer*), even *Rukai*.

⁵ Zobel (this volume) refers to this form as 'undergoer focus' (UF). PF is an alternative designation of UF. Otherwise, this article will largely follow Zobel's terminology.

3 Evidence that PAn had focus

The ways the verb fits into the clause — that is, the rules of the basic phrase structure of the clause in Austronesian languages, are remarkably analogous across the board. We can see that the basic building blocks of the clause are the same, or have clearly developed from something that was the same, in the entire area over which the Austronesian languages range (even though individual languages may have undergone developments which brought them far from this basic phrase structure). This is *prima facie* evidence that the Philippine type of focus system or one of the three alternative systems (or six alternatives, if sub-types are counted) is original and the others developed from it, as these focus systems dove-tail closely with the basic phrase structure.⁶ Further, the existence of petrified forms in languages which have lost the basic elements of the focus system guarantee that the focus system characterised PAn.⁷ The question which arises then is: which is the original system and which the secondary, the Philippine type or one of the deviations from the Philippine type?

There is a line of reasoning that can determine which system is original and which a secondary development. This reasoning involves a marshalling of evidence for the manner in which the attested forms came into being. It is in the presentation of this evidence that the papers of this volume make the most significant contribution to moving us forward to our final goal of reconstructing PAn verbal morphology.

3.1 Evidence for the primacy of the Philippine type of focus system

My formulation of the different types of systems in terms of the Philippine type and 'deviations' from it implies that I am presupposing the Philippine type to have temporal primacy — i.e. that the deviating types are developments from the Philippine type of focus system. This question has been addressed specifically (Wolff 1996), and in that article enough evidence was adduced to make it virtually certain that the Philippine system was indeed prior, even though in the absence of support by attested forms (see §3.3 below) the details of the scenario which sketch out how the specific deviating systems developed are in part still speculative. In fact, the papers in this collection that address this question overtly or

⁶ What I am referring to here is a basic clause structure of a predicate and a subject (or topic) and optionally other constituents modifying the clause which may be adverbial expressions, themes, and perhaps play other roles. These basic constituents — the subject or predicate or both — may be modified by (one or more) adjectives or verbs. The basic feature of these constituents is that they may consist of any form; i.e. every form, including the verb with its focus affixes, is free to occur in every construction: predicate, subject, modifier. We cannot go into more detail here, but it is important to point out that Austronesian languages permit the verb to be the subject and the noun phrase which is the argument of the verb to be the predicate (as well as permitting the verb phrase to be the predicate and the noun phrase which is its argument to be the subject); they permit the verb to occur in modifying constructions without in any way changing the morphology of the verb. My contention here is that this characteristic of the syntax of Austronesian languages was also a characteristic of PAn and that the focus systems now attested are consistent with a syntax of this character.

⁷ Although the Oceanic languages have maintained a kind of focus system (see Wolff 1980), only one (or possibly two) of the original morphemes has come down to the modern languages, but much can be seen in the shape of petrified forms which reflect affixes that can be reconstructed. As an example I could quote reconstructed Proto Polynesian forms which reflect affixed forms of the root **kan* 'eat' (quoted from Biggs 1990): **kano* 'flesh' (< **kan-en* the PF non-past 'to be eaten'), **ika* 'fish' (< **is-kan*) IF 'thing to use in eating, put with the staple', **kai* 'food eat' (*kan* [with loss of final **n*] + *-i* the atemporal LF, a productive affix in Oceanic languages), **fanga* 'feed' (< *paN-* + *kan*).

implicitly have accepted (or independently arrived at) the view that the Philippine type is primary. In this collection the paper by Poedjosoedarmo provides crucial evidence from Javanese that the person-agreement morphemes are secondary. Poedjosoedarmo points out that the Old Javanese system shows the characteristics described under deviations (1) and (2):⁸ the verb has the inherited Philippine-type affixes (minus the tense and aspectual forms) and has extended the LF and IF affixes to stems which contain the AF morphemes as well, and, as stated in (2) above, the LF and IF suffixes are the atemporal suffixes. The preposed affixes marking person agreement are absent in Old Javanese, whereas they are clearly present in contemporary Javanese. The origin of these modern Javanese person-agent-agreement affixes is unknown. They appear as a full-blown system in literature in the earliest specimens of modern Javanese from the nineteenth century. Prior to that period Javanese writing used Old Javanese forms with little input from the vernacular spoken by the communities which wrote Old Javanese texts. Hence, there is little by way of attested intermediate forms in Javanese which could shed light on how these forms arose.⁹

Interestingly enough, the history of Malay shows remarkable parallels to the developments in Javanese. The earliest Malay texts (from Srivijaya in the ninth century) show a deviation from the Philippine-type focus system remarkably parallel to that of Old Javanese: there were no person agreement marking morphemes, but the atemporal form of the active with IF and LF affixes occurs in the earliest texts. In summary we can say that Javanese and Malay each attest to two stages of the same language: a later stage in which a preposed agent agreement marker exists with some verb forms and an earlier one in which these markers do not occur — i.e. have not developed.

3.2 Did the deviations from the Philippine type of focus system develop independently in various languages?

These very similar developments in two languages in a period in which they are known to have been in contact were surely not independent. Whether the change originated in Malay and spread to Javanese or originated in Javanese and spread to Malay, or spread to both from a third language, and which language that might have been, is unknown and is a matter for further research. What we can say is that much of this change in the morphological system spread from language to language, a fact which renders the task of finding the analogy which led to the innovation all the more difficult, perhaps impossible. The languages of Bali, Lombok and Sumbawa (all within a subgroup of the western Austronesian languages) offer some clues. Sasak, the most conservative of the group, showing a verbal system close to one from which the Bima and Sumbawa systems developed (Wouk), manifests the following features: preposed pronoun agent markers for some, but not all, PF, LF, and IF verbs and (as opposed to Javanese and Malay and also most of the other languages which show agent prefixes with the PF, LF and IF forms) Sasak also shows preposed agent markers for AF verbs (i.e. has the characteristics designated as (ii) above). Not only that, but the Sasak system

⁸ Actually, the Javanese system is a slight variant on (2) in that although tense and aspect have been lost, much of the modal system of PAn remains, and in fact the modal morphemes in Javanese (both old and new) continue reconstructable PAn affixes.

⁹ There are sharp dialectal differences within Javanese, and it is by no means certain — in fact it is highly unlikely, that standard modern Javanese continues Old Javanese directly. In fact it is clear that some of the modern Javanese dialectal forms lacking in standard modern Javanese continue forms which occur in the Old Javanese texts.

manifests a special characteristic not found in other languages with agent markers: the agent marker may be omitted and the agent expressed by a postposed genitive, as is the case of languages manifesting the Philippine-type focus system. However, Sasak is special in that it does not manifest the temporal affixes of the Philippine-type focus system, the affixes that in other languages must occur on the verbs modified by postposed genitive agent morphemes.¹⁰ However, the factors which motivate a preposed genitive or postposed genitive have not been discovered. It may of course be a matter of subtle nuance which can add little by way of evidence as to the historical origin of each word order. On the other hand, some yet-to-be-discovered facts about these languages may well have something to offer in solving this question.

3.3 Languages which retain the Philippine type of focus affixes and develop the preposed agreement prefix

The languages of Sulawesi, which manifest both postposed and preposed genitive agents with the PF, LF, and IF, use the postposed genitive forms with the reflexes of Philippine-type past tense affixes to refer to the past, and the preposed genitive forms with the inherited atemporal verb form to refer to non-past time, as for example the Kaili-Pamona languages (Meade §3). In an earlier article, I connected this fact (Wolff 1996) with a feature of word order and syntax which characterises the Philippine languages and which can be reconstructed for PAn. This is a word order whereby the genitive pronouns and other short forms of one or two syllables are moved to a position immediately following a modifier of the verb which precedes it if the predicate contains a modifier which precedes the verb. Further, the verb when modified takes neither the past nor the non-Past affix, but rather takes the atemporal affix. To summarise, the word order reconstructable for PAn is as follows:

Verb with Past or non-Past affix + Pronoun Agent
Modifier + Pronoun Agent + Atemporal Verb

I proposed that the preposed pronominal verb forms originated in verb phrases containing PF, LF, and IF verbs having the following make-up: a preposed modifier followed by the agent followed by the atemporal verb. I speculated that this word order with the genitive pronoun before the verb was generalised, spreading to clauses which did not contain a preposed modifier. This is a likely scenario because (1) the verb form which occurs with the preposed agent marker is atemporal and (2) the verb form which occurs after the preposed modifier in the current languages is likewise the atemporal form (or a development from it), and the range of languages in which this feature occurs (that is, in which the genitive is moved to position directly after a preposed modifier) guarantees that this feature is attributable to PAn. The proposition that the word order of the verb phrase which was modified was generalised to unmodified contexts is here termed a 'speculation' because no evidence has turned up for the analogy whereby this generalisation came about. It is here that the Sasak-Balinese

¹⁰ Of course these temporal affixes have lost their temporal reference in languages which have lost tense-aspect distinctions — e.g. Old Javanese *winunuh* 'killed' continues a PAn **binunuq* containing a past infix **-in-* plus the PF affix (which happens to be \emptyset), but in Old Javanese this form has no tense meaning and the infix *-in-* has come to be the marker of PF. The thing to note here is that the postposed genitive form of a noun or pronoun refers to agents with these verbs: *winunuh mami* 'We kill it.'

languages with the preposed and postposed genitive agents might be informative. Further research is needed.

3.4 Preposed markers with the AF

Before we continue with the discussion of the development of the preposed pronominal agent markers we should make an excursus on the matter of the development of these markers with the AF forms. This feature appears not only in Sasak, but also in some languages of Sulawesi, in Chamorro, and in the Oceanic languages (if one takes the so-called 'intransitive verbs' of Oceanic languages to be reflexes of the AF, for the intransitive verbs in many of the Oceanic languages are marked [usually paraphrastically] for person agent). Although the occurrence of the preposed person marker with AF is a feature of widely scattered languages which cannot be within a subgroup, it would be incorrect to assign this to the earliest stages of the development of the preposed person marker: the hypothesis is that the development of the preposed agent with AF took place independently in the Bali-Sumbawa group, in Sulawesi, in Chamorro, and in Oceania. Here, too, the best evidence to prove this hypothesis will come from an account of the analogies which may have led to its rise in each of the various groups in which it is found. Some clues may be given from the Bonggi data presented by Boutin. To summarise in very general terms the relevant aspects of Bonggi: the agent is nominative when it is the argument of an AF verb, and the agent is genitive when it is the argument of a PF, LF, or IF verb, just as is the case for languages of the quintessential Philippine type. If the verb is modified in certain cases (by a past tense modifier, by a conjunction derived from the word for 'place' or by a paraphrastic verb derived from the root *anu*) the agent is always, or may optionally be genitive (depending on the modifier). Thus, for example, Boutin quotes a sentence with an active verb but a genitive agent *ku*:

Bas ku mori nya.
 past I-GEN give-AF him-3SG-NONACTIVE
 'I gave him (something)' (example 74a)¹¹

¹¹ This is a truncated version of the complete facts. It is not only agents of AF which are genitive in these cases, but also the patients of PF, LF, or IF verbs. For example the patient of a PF verb is nominative when not preceded by one of these modifiers, but genitive when modified (optionally or obligatorily, depending on which one). In the following example *ou* 'I' is nominative:

Ou kiohol nya.
 I-NOM was-bitten-PF by-him-GEN
 'I was bitten by him.' (extrapolated from example 72)

But the form for 'I' *ku* is genitive with the preposed modifier *bas*:

Bas ku kiohol nya.
bas I-GEN was-bitten-PF by-him-GEN
 'I was bitten by him.' (example 71)

The term 'preceding modifier' is actually incorrect to describe this construction. This construction in fact consists of a sequence of two forms (two verbs or a stative followed by a verb) and the genitive agent is part of the constituent which contains the first verb or stative rather than being in a constituent with the second verb with which it goes in meaning (of which it is the agent argument). Other languages of the Philippine group manifest analogous phenomena. For example in the following Tagalog sentence, isomorphic with the Bonggi sentence quoted, the agent *niya* is genitive even though the verb of which it is the agent argument is AF:

I have no clues as to the analogy which provided the motivation for the generalisation which permitted the use of the genitive in place of nominative with the active verb, and further work is needed to develop a hypothesis.

4 Batak evidence

Now to return to the question of how the preposed pronominal agent forms developed. Languages outside of Sulawesi which show the Philippine-type focus affixes as well as the verb forms with the preposed agent agreement markers distribute the Philippine-type affixes and the verbs with the preposed agent in a different way from those in Sulawesi. Except for Chamorro and Palauan, none of those outside of the Sulawesi languages have tense.¹² In Javanese and Malay it is a matter of antiquity of form, the older being the Philippine type and the forms with the preposed agent markers being the more modern forms. Toba Batak has developed a preposed agent marker plus the atemporal form of the verb for the PF, LF, and IF,¹³ and also retains the PAn PF affix * \emptyset plus a reflex of the PAn past tense affix *-in-, but in Toba Batak -in- no longer has any tense meaning and has taken on a tenseless PF meaning (as was the case of the Old Javanese and the Old Malay forms reflecting *-in-). The agent of the verb with -in- (if expressed at all) is in most cases a postposed genitive, as is the case in other languages which reflect the Philippine-type focus affixes. The LF and IF of the Philippine-type passive have practically disappeared in Toba Batak. Reflexes of these forms are only attested in petrified remnants. The PF forms of the Philippine-type passive are confined to certain types of contexts — namely, those in which the agent is either not mentioned or not relevant and in embedded constructions — that is, in attributive constructions or in verb phrases meaning ‘the one having had [so-and-so-done to it]’. Karo Batak has developed in a different direction: Karo Batak has combined the two types into a single conjugation, with the preposed pronominal agent prefix used for the first person (singular and plural) and the reflex of the PAn infix *-in- with postposed genitive agents for the other persons (Norwood).¹⁴ In short, the Batak languages provide evidence, as do Malay and Javanese, that the preposed-agent-marked forms are secondary. Interestingly, the Batak languages show a sequence in the development of this type of non-AF focus which is entirely parallel to that manifested in Tolitoli, the first language in northern Sulawesi which shows any steps in this development: namely, in Tolitoli only the first person singular has a person-agreement marker (the other verb forms show Philippine-type affixation). In Karo Batak it is the first person singular and the first person plural which manifest the innovated form, whereas the other verb forms continue the Philippine type. Toba Batak has gone a step further and in that way is remarkably like Lauje (see Himmelmann) which has a first person

Pagkatápos ku ng magbigay sa kanya
 after I-GEN LINKER give-AF to him-DAT
 ‘After I gave something to him.’

- 12 Languages of Sulawesi which have lost tense manifest only the preposed agent-type verb or a development from that (Mead), and if any of the Philippine-type focus forms are retained, it is only in petrified remnants.
- 13 The situation is actually more complicated. Toba Batak, like Tolitoli, has preposed agent forms for the first person only, and for the second and third persons has a prefix *di-*. If there is an agent expressed, it is postposed.
- 14 Currently Karo Batak has passives with first person prefixed and the other persons suffixed, with an optional prefix *i-*, which in older texts appears as *ni-*. The prefix *ni-* is clearly a reflex of PAn *-in-, which, as in Toba Batak, has lost its tense meaning and taken on a voice meaning.

singular marker and then a single prefix for all other persons, singular and plural. These parallel developments, striking as they are, are surely independent and result from natural processes of syntactic change.¹⁵ In short the similarities between the non-AF verb forms in Toba Batak and some of the languages of Sulawesi are surely due to independent developments and no more connected with each other than the numerous cases of change of the past tense affix **-in-* into a marker for PF which occurred independently in many widely separated languages.

5 Chamorro evidence

Chamorro manifests a verbal system which differs in many respects from any which has been discussed so far — that is, it has made innovations on the Philippine-type system in many ways somewhat different from any so far discussed. These features are viewed as innovations (rather than that the features characterising the other languages are the innovations) because there are clear remnants of the original reconstructable features which characterised the Philippine system from which the Chamorro features developed. For one thing Chamorro in its verbal system retains the four foci, AF, PF, LF, and IF, although the PAN affixes which mark these forms have been replaced in a large part, and the inherited morphemes have been given new functions.¹⁶

5.1 **-in-* in Chamorro

First, Chamorro, like Malay, Javanese, Toba Batak and the languages of Sulawesi, retains a reflex of **-in-*, the PAN past-tense marker. In Chamorro, as in those languages *-in-* does not mark the past tense, but marks PF. As in Toba Batak, Old Javanese and Old Malay (and like non-AF forms in the languages with the Philippine focus type), the agent of the verb with reflexes of **-in-* is genitive. The change from past tense to marker of PF is perfectly natural, as pointed out above, and could have occurred independently in Chamorro. In fact, it is very likely to be independent in Chamorro. This is indicated by the fact that the distribution of *-in-* in Chamorro has striking differences from the distribution of the cognate reflexes in Toba Batak and Javanese. Thus, Chamorro *-in-* occurs with PF and LF but not IF (as it does in the Philippine-type systems and in the western Indonesian languages which retain reflexes of

¹⁵ Historically, the scenario whereby Toba Batak person-marking affixes were developed is as follows: a first person (singular and plural inclusive) preposed agent was developed by Toba Batak as well as Karo Batak, by an unknown process (perhaps similar to that outlined for the development of the analogous form in Tolitoli (Wolff 1996)). Toba Batak substituted the prefix *di-* for *-in-*, probably under the influence of Malay, with which it has been in contact for hundreds of years, so that the two variants were in competition, and the verb with *di-* was followed by the agent if the agent was expressed at all. If no agent was expressed the verb with *di-* referred to third person agent. The form with *-in-* came to be confined to the contexts specified above.

¹⁶ Chamorro inherits the non-past PF affix **-en-* in the form of a suffix *-(y)on*, forming a verbal adjective derivation based on transitive verb stems meaning 'capable of being [rooted]', a meaning clearly developed from the meaning we can reconstruct for the PF non-past. Austronesian example is *atanon* 'nice to look at' (< *atan* 'look at') (This Chamorro suffix has further been extended to intransitive verbs, a later development which does not concern us here.) Further, Chamorro inherits the non-past LF affix **-an-* in the form of a suffix *-an* which occurs in conjunction with a prefix *fan-*, forming a nominal derivation based on verb stems meaning 'location of the action', a meaning clearly developed from the meaning we can reconstruct for the PAN LF non-past.

*-in-) and further it does not co-occur with the LF suffix *-an* (as is the case of the other languages) but rather it co-occurs with the atemporal LF suffix, the reflex of **-i*. This shows that the reflex of **-in-* in Chamorro has a morphological status somewhat different from that found in any other language and suggests that the development of *-in-* in Chamorro was quite independent of that of any of the other languages.

5.2 Was the development of the preposed agent agreement marker an independent innovation in Chamorro?

Chamorro also has verbs consisting of preposed agent markers added to the atemporal stem and referring to PF, LF, and IF. Further, in form the first person prefix, at least, is the inherited genitive form. The etymology of the prefixes marking the other persons is unknown.¹⁷ At first sight one would think that the development of this verb form can hardly be independent of the development of analogous forms found in languages of Sulawesi and western Indonesia: the similarities (that is, the preposed genitive pronoun, the atemporal verb form, the reference to PF, LF, and IF) are too striking. Yet a hypothesis that the preposed agent markers characterised the protolanguage is untenable for the reasons discussed above, and a theory that Chamorro developed these forms in common with the languages of Sulawesi does not accord with what we presuppose to be the prehistory of Chamorro. (Admittedly, this is a matter which has not been thoroughly researched.) A third possibility is that Chamorro was influenced by the surrounding Oceanic languages, which also have agent agreement with transitive verbs. This is certainly a possibility, for Chamorro shows influence of the Oceanic languages.¹⁸

There are two other facts which indicate that the development of the preposed person agreement markers in Chamorro was independent of analogous developments in languages of Indonesia: (1) the raw materials which would allow these prefixes to develop are still clearly present in the current language and (2) there are differences in the distribution and meanings of the preposed agent affixes in Chamorro and those of the other languages. To explain further point 1: Chamorro has a rule that certain preposed modifiers of a verb require that the genitive pronoun referring to the agent be affixed to them, and it is possible that the preposed genitive agent marker developed by an analogy which attached the genitive to the verb.¹⁹ To explain further point 2: the preposed genitive agent marker may be attached to several kinds of verbs, although it is most widespread with the reflex of the atemporal verb. It may appear before the reflex of the AF affix, as it may in some of the languages of western Indonesia (where we have argued this occurrence was a later development). But most strikingly and in a way totally different from any other western Austronesian language, these person agreement markers may occur with intransitive verbs, in which case they are usually, but not always,

¹⁷ Zobel (this volume, §3.1.1) states that they developed from genitive forms but gives no evidence supporting this view.

¹⁸ As an example of a grammatical feature which has been borrowed from Oceanic languages we may point to the features of Chamorro similar to the special edible and drinkable possessive classes, found widely in Oceanic languages. Examples are *ga'* 'classifier for pets', *na'* 'classifier for edible things', *ga' ña* 'his pet', *na' mu* 'your food', *i ga' mannok hu* 'my (pet) chicken', *i na' mannok hu* 'my chicken (to eat)'.

¹⁹ For example, *ya* 'like', a preposed modifier which requires a genitive agent after it: *Ya-hu chumochó* 'I like to eat'. This can be compared with *malago'* 'want', a verb which does not require a genitive agent and has a nominative agent *yo'*: *Malago' yo' chumochó* 'I want to eat'.

attached to the reflex of the atemporal verb form.²⁰ In fact, these agreement markers occur even before a verb containing a reflex of **-in-*, in which case the agreement marker refers to the patient (Zobel, example (15)). There is also a strong difference in the meaning of the verb with the preposed agent in Chamorro and in languages of Sulawesi which retain tense. In Sulawesi the PF, IF, and LF forms with the preposed agent agreement marker are non-past. In Chamorro the analogous forms have no tense (although Chamorro distinguishes tense). Strangely enough, the Chamorro agreement markers in intransitive verb forms, which have no parallel in the languages of Sulawesi and western Indonesian, do indeed indicate non-past tense.²¹

There is a further difference between Chamorro and the other languages which show both reflexes of **-in-* and the preposed person agreement marker. The proper noun agent of a Chamorro verb with *-in-* is marked by *as*, whereas a proper noun agent of the Chamorro verb having a preposed agent marker is marked by *si*. In languages of Indonesia which have both the reflexes of **-in-* and the preposed-agent-agreement marker, if the agent is expressed, the agent is genitive with both kinds of verbs.²²

6 A subgroup consisting of the languages which developed personal agreement marking?

The evidence adduced so far cannot provide a definitive answer to the question of whether Chamorro developed the verb forms with the preposed agreement markers independently or together with the languages located further west. If that were the case, the whole question of whether or not the Philippine-type focus system characterised the protolanguage would have to be reconsidered. Little evidence has been adduced from Palauan which can help us reach conclusion on this question (Zobel). Zobel (§6.1) suggests that there was a language 'Proto Nuclear Malayo-Polynesian' which existed after the languages with the quintessential Philippine-type focus system split off from them, and this would be the stage at which these commonalities (that is, the preposed agent agreement marker and the verb containing the AF affix and at the same time LF or IF affixes) developed. This is a speculation which requires a great deal more research if it is to be substantiated. This aspect of the histories of whole slews of languages in Kalimantan, Eastern Indonesia, and Oceania heretofore untouched may

20 The stem with the agent marker is not the reflex of the atemporal verb in all cases. For example, with *maigo* 'sleep', the preposed agent marker is added to *maigo*, not to a reflex of the atemporal form which would be **faigo*: *maigo* 'yo' 'I slept', *para bai humaigo* 'I will sleep'.

21 For example *hufalagu* 'I will run away', which consists of *hu-*, the first-person agentive agreement marker and *falagu*, the verb stem (reflecting the PAN atemporal form), refers to the future. *Malagu* 'ran away', which does not have the first-person marker, is past time.

22 For example the agent is marked by *si* in the following example, which contains a verb with the third person agreement prefix:

<i>hana'i</i>	<i>hao si</i>	<i>-Juan ni</i>	<i>lepblo</i>
was-given-by-him	you	by-John	OBLIQUE-INDEFINITE book
'John gave you a book.'			

However, in the following sentence the agent is marked by *as*:

<i>nina'i</i>	<i>hao ni</i>	<i>lepblo as</i>	<i>Juan</i>
is-given	you	OBLIQUE-INDEFINITE book	by John
'John gives you a book.'			

provide evidence. Most likely, these innovations spread from language to language (Wolff 1996).

7 Verb forms containing AF as well as IF or LF affixes

A final issue which these papers shed light on is the development of the forms containing an AF added to a stem consisting of the atemporal LF or IF verb. The languages of Sulawesi which have these verb forms also have pronominal agent agreement (at least for the first person), and I used this fact to develop a scenario which would explain the development of the verbs containing both AF morphology and the LF or IF morphology (Wolff 1996). As Zobel points out (§6.1), while it is true that languages with preposed person agreement markers also have verbs consisting of AF plus either LF or IF affixes, the converse is not true: there are languages which have verbs containing AF plus either LF or IF affixes, which do not have verbs with proposed personal agreement markers, as for example Old Javanese (Poedjosoearmo). This does not necessarily invalidate the theory that the verbs with AF plus LF or IF affixes originated by the analogies suggested which link the development of verbs with personal markers to the verbs with AF plus LF or IF affixes (Wolff 1996), for each analogy proposed surely occurred only once and the resultant form spread from language to language.

8 Conclusion

In conclusion we may say that these papers have clarified a great deal about the history of the verbal system of the Austronesian languages:

- (a) the Philippine-type focus system is primary in time, and the various deviations from it are secondary;
- (b) person agreement marking on the PF, LF, IF of the transitive verbs developed in languages ranging from western Indonesia through the Pacific, but this feature is in some cases only partly developed — in some cases it is just coming into being;
- (c) the verb forms with preposed person agreement markers have developed further in subsets of the languages which have them; some languages additionally developed these preposed markers with intransitive verbs;
- (d) in some languages verbs with LF and IF affixes also may contain AF affixes; this is a development not independent of (b) in that languages which developed (b) also have undergone development (d);
- (e) patient agreement markers for transitive verbs developed independently in widely separate languages (including some of the Oceanic languages, in addition to the languages discussed here).

References

- Biggs, Bruce, 1990, Pollex: Polynesian reconstructions. Unpublished computer files.
 Dempwolff, Otto, 1934-38. *Vergleichende Lautlehre des austronesischen Wortschatzes*.
 Berlin: Dietrich Reimer, vol.I 1934, vol.II 1937, vol.III 1938.

- Li, Paul Jen-kuei, Dah-an Ho, Ying-kuei Huang and Cheng-hwa Tsang, eds, 1995, *Austronesian studies relating to Taiwan*. Symposium Series of the Institute of History and Philology, Academia Sinica 4. Taipei: Institute of History and Philology, Academia Sinica.
- Ross, Malcolm D., 1995, Reconstructing Proto-Austronesian verbal morphology: evidence from Taiwan. In Li et al., eds, 1995:727-791.
- Starosta, Stanley, 1995, A grammatical subgrouping of Formosan languages. In Li et al., eds, 1995:683-726.
- Van der Tuuk, H.N., 1971, *A grammar of Toba-Batak*. Translated by Jeune Scott-Kemball. Koninklijk Instituut voor Taal-, Land-, en Volkenkunde, Translation Series 13. The Hague: Martinus Nijhoff.
- Wolff, John U., 1973, Verbal inflection in Proto-Austronesian. In A. Gonzalez, ed. *Parangal kay Cecilio Lopez*, 71-91. Quezon City: Linguistic Society of the Philippines.
- 1980, Verbal morphology and verbal sentences in Proto-Austronesian. In P.B. Naylor, ed. *Austronesian studies: papers from the Second Eastern Conference on Austronesian Languages*, 153-167. Ann Arbor: University of Michigan, Center for South and Southeast Asian Studies.
- 1996, The development of the passive verb with pronominal prefix in western Austronesian languages. In Nothofer, Bernd, ed. *Reconstruction, classification, description: festschrift in honor of Isidore Dyen*, 15-40. Hamburg: Abera.

Final words: research themes in the history and typology of western Austronesian languages

MALCOLM ROSS

1 Introduction

My goal in these final words is not to draw conclusions from the chapters in this volume but to identify research themes which they and other recent work suggest. Perhaps other scholars reading this collection will see quite different themes: if so, then this collection of data-oriented pieces dealing with particular languages will be fulfilling its purpose.¹

There are at least three different kinds of research theme on which the chapters may cast light (apart from the contribution that each makes to the study of its particular language(s)). Firstly, they provide more detail for the study of the typology of western Austronesian languages. Secondly, they may contribute to the study of how, for example, the Philippine type became the Indonesian type. And thirdly, they may help us understand more of the culture history of speakers of western Austronesian languages by enabling us to refine our subgrouping and thereby our understanding of their genealogy.

We have distinguished in this volume between Philippine-type and Indonesian-type languages. It seems to me, however, that we need to recognise three types.

There is a sense in which Indonesian-type languages can be seen as 'post-Philippine'; that is, they represent a language type or types that have grown out of the Philippine type. The Indonesian type is perhaps harder to pin down than the Philippine type, simply because it may represent more than one independent set of historical developments. I will reduce the criteria for an Indonesian-type language to one here: it is a language with (at least) two voices, marked by combinations of inherited prefixes (particularly **maN-*) and suffixes (particularly **-i*, **-án/*-[a]kən*). These combinations do not occur in Philippine-type languages. The standard voice matrix for a Malayo-Polynesian Philippine-type language is repeated in Table 1 without some of its contrasts.

¹ Some of the work on which my comments in §3.2 are based was done during a visiting professorship at the University of Frankfurt in the 1998–99 winter semester. I am indebted to the Deutsche Forschungsgesellschaft for financial support and to Bernd Nothofer for facilitating my visit.

Table 1: Proto Malayo-Polynesian verb forms (Ross, Tables 7 and 8²)

	Actor	Patient	Location	Circumstantial
Neutral	<um>√ maN-√	√-ən paN-√-ən	√-an paN-√-an	i-√ i-paN-√
Atemporal	√	√-a	√-i	√-án

If these forms are rearranged in an Indonesian-type matrix as in Table 2, then we get a matrix with two sets of blanks in it:

Table 2: Proto Malayo-Polynesian verb forms rearranged

PMP	Active	Passive
Patient undergoer		
neutral	<um>√ maN-√	√-ən paN-√-ən
atemporal	√	√-a
Location undergoer		
neutral	—	√-an paN-√-an
atemporal	—	√-i
Circumstantial undergoer		
neutral	—	i-√ i-paN-√
atemporal	—	√-án

What has happened in Indonesian-type languages (cf. Table 13 of Ross) is that the two sets of blanks have been filled in with morpheme combinations that do not occur in Philippine-type languages, as shown in Table 3.³

Thus the criterion for an Indonesian-type languages boils down simply to the filling in of these blanks, the process described by Wolff (1996). I have omitted Wolff's other two criteria for an Indonesian-type language (cf. Ross §4.2) simply because there are languages we would probably wish to call 'Indonesian-type' that do not fulfil them. The first is the formation of the passive with actor proclitics (abbreviated GEN=). Although this is shown in Table 3, it is not reflected, for example, in Sundanese or Balinese. The other is the loss of the neutral/perfective distinction, which, as Wolff points out, has not occurred in a number of Indonesian-type languages.

² I refer to the chapters of this volume, including my own contribution to the Overviews section, simply by the name of the author, except in the case of Himmelmann. His contribution to the Overviews section is abbreviated 'Himmelmann I', his contribution on Ratahan and Lauje, 'Himmelmann II'. The abbreviations and conventions used here are the same as those in Ross.

³ In (3) and henceforth, I use GEN for any pronoun set which reflects the PMP genitive pronouns (Ross, Table 3). The probable mechanism by which they became proclitics is discussed by Ross, §4.2.

Table 3: Verb forms in an Indonesian-type language (schematic presentation)⁴

	Active	Passive
Patient undergoer	<um>√ maN-√	GEN=√
Location undergoer	<um>√-i maN-√-i	GEN=√-i
Circumstantial undergoer	maN-√-an	GEN=√-an

If we regard Indonesian-type languages as 'post-Philippine', then there are also languages represented in this volume which we may label 'post-Indonesian'.⁵ In these languages the voice system has broken down: the morphology of voice still occurs, but its function is no longer to mark alternate pivot choices. Again, this language type is probably the outcome of multiple independent developments, so we find various forms of it. I return to the post-Indonesian type in §4.

The languages discussed in the chapters of this volume can be classified as follows:

Philippine-type: Seediq (Holmer), Yogad and Hiligaynon (Spitz), Bonggi (Boutin), Sama Bangingi' (Gault), Southern Sulu Sinama, henceforth SS Sinama (Akamine), Ratahan (Himmelman II)

Indonesian-type: Lauje (Himmelman II), Pendau (Quick), the Bungku-Tolaki and Kaili-Pamona groups (Mead), Tukang Besi (Donohue), Karo Batak (Norwood), Javanese (Poedjosoedarmo), Sasak (?) (Wouk), Chamorro and Palauan (Zobel)

Post-Indonesian: Riau Indonesian and other Malay/Indonesian dialects (Gil), Sasak (?) and Sumbawa (Wouk).

Sasak is shown with a question-mark as both an Indonesian-type and a Post-Indonesian language: the dialect described by Wouk seems to be right on the borderline between the two categories, with the voice system in the process of turning into something else.

The remainder of this chapter falls into three parts corresponding to the three types. Where no source is given for data or analysis from one of these languages, the source is the relevant chapter of this book.

2 Philippine-type languages

Philippine-type languages are represented in this volume by a geographically and genealogically well distributed sample. Only Sama Bangingi' and SS Sinama perhaps belong to the same micro-group. The least represented region here is Borneo, about whose Philippine-type languages (and the transition between them and Indonesian-type languages) we still have little detailed knowledge.

⁴ 'Schematic presentation' means that the forms in the slots may later prove to be incorrect: this is especially the case with regard to circumstantial undergoer *-an.

⁵ 'Post-Indonesian' is not a satisfactory label, particularly because it includes many languages that are located geographically within Indonesia, but I have nothing better to offer at the moment.

Seediq is a very conservative Philippine-type language from northern Taiwan, and Holmer provides the first readily accessible description of its voice system which takes account of modern research (older accounts — Asai 1953 and Pecoraro 1979 — are hard to interpret).⁶ The system he describes is morphologically very similar to that reconstructed for Proto Austronesian, and provides corroborating evidence for the reconstruction of PAN *án- in the circumstantial-voice forms of Table 1 in Ross as an auxiliary. In that table only *án-i and *án-ay were reconstructed, but Seediq appears also to reflect *án-ən and *án-an (as *anun* and *anan*), suggesting that PAN *án- had a wider role.⁷

2.1 The semantics of verbal affixes

A theme which emerges at various points in this volume is that verbal affixes in Philippine-type languages do not solely, or perhaps even principally, mark voice or transitivity. Evidence suggests that 'voice' affixes may mark *Aktionsart* or the more specific semantic role of the pivot participant.

Evidence of *Aktionsart* marking is found in Seediq and in closely related Mayrinax Atayal. Seediq actor-pivot forms are marked unpredictably by either ⟨*m*⟩ or *m-*, but the infix is the more likely choice with dynamic verbs, the prefix with statives. This apparently reflects an earlier situation which remains extant in Mayrinax (Huang 2000), whereby ⟨*um*⟩, *m-*, *ma-*₁ and θ_1 (corresponding atemporal θ) encode dynamic verbs, whilst *ma-*₂ and θ_2 (corresponding atemporal *ka-*) encode statives. (A similar situation is reflected in several Formosan languages; Zeitoun & Huang 2000) The Seediq situation seems to collapse Mayrinax classes into two surface morphological classes. It is clear from Huang's analysis of Mayrinax that the morphology encodes *Aktionsart* rather than transitivity. Holmer also indicates that in Seediq patient voice and locative voice are not simply used to adjust the pivot, but sometimes express *Aktionsart*.

These observations mesh well with the difficulty of describing transitivity in Tagalog (Ross, §2.3) and are echoed by scattered comments in several chapters. Spitz in particular draws attention to the fact that a eurocentric concentration on transitivity and voice misses a large part of what is going on in Yogad and Hiligaynon. Yogad, incidentally, has a Paiwan-like strategy of marking both NPIV and LOC with the phrase marker *tu* (cf. Paiwan *tua*) but GEN with *nu* (Paiwan *nua*), whereas Hiligaynon has a Tagalog-like strategy, marking both GEN and NPIV with *sang* (Tagalog *ng*) but a non-specific patient with *sa* LOC. Far more important to Spitz, however, are the semantic differences between the formally parallel Yogad and Hiligaynon systems, and the fact that in Hiligaynon different affixes alter the *Aktionsart* of intransitives (his examples (31), (34)–(36) and (38)), whilst the same form of a verb can be used with different participants serving as pivot (examples (32), (33), (37) and (39)), i.e. the primary function of these affixes appears not to be transitivity or voice.

Mayrinax verbs encoded as stative are intransitive, but that is a by-product of their stativeness. Verbs encoded as dynamic are either intransitive or transitive, but there is a contrast in semantic role between transitive verbs in ⟨*um*⟩ or *m-* and those in *ma-*₁ or θ_1 : the former are more likely to have a patient undergoer, the latter an undergoer that is less

⁶ There are, however, descriptions of quite closely related Atayal dialects which are more accessible. They include Egerod (1965, 1966, 1969) and L. Huang (1994a,b, 1995, 2001).

⁷ Holmer attributes these forms to analogy, but there is no reason to prefer this origin over inheritance.

affected by the event, e.g. a theme. In Hiligaynon, Spitz notes, the patient voice of certain verbs is marked not by *-on* PV but by *-an*, otherwise LV, or by *i-*, otherwise CV. Spitz' example (2) shows that 'patients' of verbs marked with *-an* are likely to be goals, his (53) and (54) that those marked with *i-* are likely to be themes (i.e. objects moved by the event). Similar examples can be adduced for Tagalog (Schachter & Otones 1972:296-298, 302-305), and Gault's examples (7)–(8) point to a similar phenomenon in Sama Bangingi'. Thus much more study of the meanings of Philippine-type verbal affixes appears to be called for, especially insofar as these have to do with *Aktionsart* or semantic role.

2.2 Changes in verbal affixation in Philippine-type languages

A number of changes in the forms and uses of verbal affixes can be identified during the history of various Philippine-type languages and groups, but in most cases it is not yet possible to pin down exactly where in the genealogical tree (insofar as it is understood) each change occurred, nor to delineate the range of languages in which each change occurred (a notable exception to this generalisation is the analysis of Bisayan innovations by Zorc 1977).

A possible difference between Formosan and other Philippine-type languages concerns intransitives. In Formosan languages, intransitives seem largely to be marked by what are otherwise actor voice affixes; in other Philippine-type languages, at least in the central Philippines, they may be marked by any of the voice-marking affixes, as observed at the end of §2.1.

The PMP reconstructions offered in Ross are incomplete, and there is clearly more to the system than I have reconstructed. In fact, the PMP system seems to have been more elaborated, and therefore more different from the PAn system, than Tables 1 and 7 of Ross imply. One point that emerges is there may have been several PMP circumstantial voice formatives, differing from each other in meaning. Thus we find Ilocano (Rubino 2000) *i-* *-an* 'beneficiary voice', *pag-* and *paN-* 'instrument voice' and *i-* 'theme voice', Tagalog *i-* 'beneficiary/theme voice' and *i-paN-* 'instrument voice', Proto Bisayan (Zorc 1977) **i-* 'instrument voice', Binukid (Post 1992) *ig-* 'instrument voice', Bonggi *pəng-* 'instrument voice', Sama Bangingi' *paN-* 'instrument voice' and *pag-* 'instrument voice, durative', and SS Sinama *paN-* 'instrument voice'. What are we to make of this? We have forms reflecting PMP **i-* 'circumstantial voice', as expected, and then forms reflecting **paN-* and **paR-*, which are unexplained. The existence of Tagalog *i-paN-* suggests that these forms may in fact reflect PMP **i-paN-* 'circumstantial voice, distributive' and **i-paR-* 'circumstantial voice, durative' (cf. Tables 8 and 9 of Ross) with loss of **i-*, and the gloss of Sama Bangingi' *pag-* 'instrument voice, durative' supports this.

Despite the glosses given here, one may wonder why **paN-* and/or **paR-* should have displaced **i-* in some languages. Perhaps Yogad gives us a clue here, as both *i-* and *pag-* function as kinds of circumstantial focus, *i-* taking as pivot a circumstance that is internal to the actor ('The tree is bountiful **through its fruit**'), *pag-* a circumstance external to the actor ('The tree is bountiful **due to the use of fertiliser**') (see Spitz, examples (59)–(60)). We may speculate that the two circumstantial voice forms had subtly different semantics, and that some languages have retained a reflex of one, some a reflex of the other — but obviously more research is needed into a better distributed sample of Philippine-type languages.

2.3 Changes reflected in the voice system of SS Sinama

The three languages from the extreme south-west of the Philippine-type language region, Bonggi, Sama Bangingi' and SS Sinama, have all lost their noun phrase markers and use constituent order to encode grammatical relations. SS Sinama has moved a step or two further towards the Indonesian-type, although it has not taken the criterial step of filling in the 'gaps' in the voice matrix (§1). It has lost the neutral/perfective (or realis/irrealis) distinction, and marks various semantic roles with prepositions.

Like most Philippine-type and many Indonesian-type languages, SS Sinama retains the imbalance between transitive undergoer voices with a specific pivot and a semantically intransitive or semi-transitive actor voice with a non-specific patient (Ross, §2.3). The actor preposition is *leq*, and a normal patient voice clause has the structure in (1), an actor voice clause the structure in (2):

- (1) PV verb + *leq* Actor + Pivot (= Patient)
 (2) AV verb + Pivot (= Actor) + non-specific Patient

However, SS Sinama has innovated what was originally an actor voice with a specific patient, but is now a highly transitive patient voice, giving the structures in (3), where the specific patient is now the pivot:

- (3)a. *leq-N-√* verb + *leq* Actor + specific Patient
 b. *leq* Actor + *N-√* verb + specific Patient

Akamine speculates a little on the history of these forms and on whether the two *leq* forms in (3a) are somehow the same. Historically, I think they are. The historically prior form is the one in (3b). The actor voice verb *N-√* entered into a construction with a preposed actor prepositional phrase (normally unnecessary with an actor voice verb). This seems to have had the effect of increasing transitivity so that the patient was interpreted as specific. The preposed *leq* phrase was probably originally a topicalisation, but it became grammaticised in this position. The next step was that preverbal *leq* came to be interpreted as a marker of transitive voice, such that *leq* could remain in front of the verb while an actor *leq* phrase occurred after it, as in (3a).

These innovations are interesting not only in their own right but because, to judge from analysed data in Adelaar (1992b, 1999), a parallel innovation is reflected in the Western Malayic Dayak languages Kendayan and Salako of western Borneo, where the marker corresponding to SS Sinama *leq* is *di*. The one difference is that in the Salako construction corresponding to (3a), there is no postverbal actor prepositional phrase. The actor prepositional phrase with *di* is either preposed or omitted. Both correspondents of the construction in (3) are used in Salako for a completed event with an undergoer subject, i.e. a highly transitive event.

I have suggested elsewhere (Ross forthcoming) that these changes occurred in early Malayic, and that the construction in (3a), with *di* substituted for *leq*, is the origin of the Indonesian *di-* passive. In that account I suggested that the presence of **N-* in the Salako construction was the result of a later innovation. If the Salako construction arose by the same route as the Sinama construction, then I was wrong, and instead **N-* was originally present, but deleted at some stage in the history of the Malayic passive.

I am assuming that the similar innovations in Bonggi (off the northern tip of Borneo) and Western Malayic Dayak (in the extreme west of Borneo) are not the result of shared

inheritance but of constructional borrowing. It would accordingly be interesting to know whether a similar innovation has occurred elsewhere in the Borneo region. However, I am also assuming that the presence of this construction in Bonggi is not the result of recent borrowing from Malayic, otherwise the verbal prefix would have the form *di-*, as it does in other languages that have apparently borrowed it.

3 Indonesian-type languages

The representation of Philippine-type languages in this volume is skewed in a way that partly reflects recent studies of Indonesian-type languages. The least represented region here is again Borneo, and this reflects how little research has still been done on most of its languages.⁸ The best represented region is Sulawesi (Lauje, Pendau, the Bungku-Tolaki and Kaili-Pamona languages, and Tukang Besi), a result of the recent efflorescence of Sulawesi research.⁹ The presence of contributions on Karo Batak, Javanese and Sasak reflects a growing interest in the languages of Sumatra, Java and their outliers. For some of the more widely spoken languages of this region, there is a good deal of material, but it is not readily accessible to today's linguists.¹⁰ Poedjosoedarmo's chapter on Old and modern Javanese gives us access to information which has long remained out of reach for the linguist who was not a specialist in the area. Wouk's chapter reflects a growing interest in the languages of the islands to the east of Java: there has been a recent spate of work on Balinese and Sasak.¹¹

The reader may wonder why Tukang Besi is included here, as Donohue analyses it as a Philippine-type language. The reason is that Donohue and I are using different criteria: by the criterion given in §1, Tukang Besi is an Indonesian-type language.

3.1 Symmetrical voice and the verb phrase constituent

A number of Indonesian-type languages share two related syntactic features that have only recently attracted much attention. Firstly, these languages have a symmetrical voice system. The term 'symmetrical' can be interpreted in two ways. As I use it here, it refers to a voice system in which both actor and undergoer voices have two core arguments (the actor and the undergoer) by the criteria given in Ross, §2.3.2. That is, both voices are syntactically transitive.

⁸ Recent exceptions are Adelaar (1992b, 1994) and Clayre (1996).

⁹ Detailed descriptive studies include Barsel (1983), van den Berg (1989) and Donohue (1995), supplementing earlier Dutch work. Historical-comparative work has been done by Himmelmann (1996), van den Berg (1996) and particularly (1998). Friberg (1990) provides a collection of texts, and there are now numerous descriptive papers including Martens (1988), Evans (1996), Friberg (1991, 1996), Hanson (2001) and those in Steinhauer ed. (1988) and the *Studies in Sulawesi Linguistics* series (Sneddon ed. 1989, 1991; van den Berg ed. 1994, 1995).

¹⁰ The Batak languages are an exception: we have van der Tuuk (1971), Schachter ed. (1984), Wouk (1986) and Woollams (1996). There is, incidentally, ample published material on Acehnese of north-west Sumatra, but this is not an Indonesian-type language.

¹¹ There are four doctoral dissertations by Balinese-speaking linguists Beratha (1992), Artawa (1994), Pastika (1999) and Arka (1998), as well as one by Clynes (1995), and the language has also attracted interest among syntacticians (e.g. Wechsler & Arka 1998). Other work includes Artawa, Artini and Blake (2001) and Austin (2001), the latter also touching on Sasak and Sumbawa. Austin ed. (1998) is a collection of papers on Sasak, and Wouk (1999) also concerns Sasak.

Secondly, in a more stringent use of the term 'symmetrical', the two voices have mirror-image structure and the non-pivot core argument immediately follows the verb and is inseparable from it (i.e. nothing may intervene between verb and argument). That is, verb + patient form a syntactic constituent (a verb phrase) in an actor-voice clause, and verb + actor form one in a patient-voice clause.

Probably a large majority of Indonesian-type languages have a symmetrical voice system, although a number have both a transitive and an intransitive passive (where the actor is marked by an oblique).¹² The latter include Indonesian, Javanese, Balinese and Sasak. Fewer have a verb phrase constituent, and some have one in only one voice. Languages with a verb phrase in both voices include Pendau, Toba Batak (Emmorey 1984, Schachter 1984: see Ross, example (36)), Karo Batak (Norwood) and Balinese (Artawa 1994, Arka 1998).¹³ Languages with a verb phrase only in the actor voice include Javanese and Tukang Besi. Lauje has a verb phrase only in the patient voice. There are also Indonesian-type languages with other verb phrase structures. The Indonesian transitive passive has actor + verb. The Ngeno-Ngene Sasak data collected by Austin (2001) have patient + oral verb.

Symmetrical voice and the verb phrase constituent are both syntactic features: symmetry depends on transitivity (i.e. the morphosyntactic identification of two core arguments), and the verb phrase constituent is defined by syntactic behaviour. As noted in Ross, §2.3, however, syntactic transitivity is not always matched by semantic transitivity. As in Philippine-type languages, (i) the default voice in many Indonesian-type languages continues to be the transitive passive, and (ii) the actor voice is reserved in independent clauses for non-specific patients and for dependent clauses where the syntax demands the actor voice. Probably a majority of Indonesian-type languages still adheres to these constraints, and it is easier to list languages in which they don't apply than those in which they do. Modern standard Indonesian and Pendau adhere to (i) but not (ii). Sundanese (Hanafi 2001) and Ngeno-Ngene Sasak adhere to neither.

Symmetricality and the constraints just mentioned have clearly been inherited from a Philippine-type ancestor. The origin of the verb phrase constituents is less clear. However, the actor-voice verb + patient phrase and the patient-voice verb + actor phrase seem widely enough distributed to suggest that they too are an inheritance from a Philippine-type ancestor or ancestors. They cannot be attributed to PMP. Rather, it seems likely that in some Philippine-type languages constituent order tended to become fixed, with the pivot either in preverbal or clause-final position, leaving the non-pivot argument regularly after the verb, a sequence which was then grammaticised.¹⁴ The only direct evidence we have for this, however, is from Ratahan (Himmelman II), where there is a patient-voice verb + actor phrase (but the pivot position is not fixed).

¹² Arka and Manning (1998), discussing Indonesian, suggest that an undergoer-voice clause with two core arguments should properly be called an ergative construction, not a passive. This is important in Indonesian and the other languages mentioned here, where there are two transitive voices (actor and undergoer) and a 'true' intransitive passive in which the actor is an oblique.

¹³ Note that there is apparently no verb phrase in closely related Sasak, although the patient must follow a verb of the form *N-√* in the Ngeno-Ngene dialect (Wouk, Austin 2001).

¹⁴ One might expect fixed order to develop in Philippine-type languages that had lost their phrase-markers. However, this does not seem to have happened, at least not as suggested here, in Bonggi, Sama Bangingi' or SS Sinama.

The generalisations presented here do not hold as widely for Indonesian-type languages as the corresponding generalisations do for Philippine-type languages. Indonesian-type languages are grammatically far more diverse than Philippine-type languages.

3.2 Verbal morphology in Indonesian-type languages

Although the voice and applicative systems of many Indonesian-type languages form a matrix with the cell structure similar to the one in Table 3, we find reflexes of quite a wide variety of morphemes in the cells of these matrices, suggesting that different sets of languages have made different innovations (usually simplifications) at different times. One means of sorting out the history and subgrouping of Indonesian-type languages is to examine these voice and applicative systems and to see what interstage protolanguages we are constrained to reconstruct. Mead has done this for Proto Bungku-Tolaki and Proto Kaili-Pamona, with interesting results.

3.2.1 Mead's reconstructions of Sulawesi verbal morphologies

Mead's basic argument is straightforward: van den Berg (1996) reconstructed a Proto Celebic, putative ancestor of all the Indonesian-type languages of Sulawesi except the South Sulawesi group.¹⁵ Van den Berg's Proto Celebic is very similar to the ancestor of one Sulawesi group, Proto Kaili-Pamona. However, the ancestor of another such group, Proto Bungku-Tolaki (Mead 1998), cannot be reconciled with van den Berg's Proto Celebic. Instead, Proto Celebic — or Proto Sulawesi — must be reconstructed such that at least Proto Kaili-Pamona and Proto Bungku-Tolaki — and hopefully other Indonesian-type Sulawesi groups — can be derived from it. If this cannot readily be done, one would have to infer that these groups do not share an immediate common ancestor but represent separate arrivals in Sulawesi.

The system reconstructed by Mead for Proto Kaili-Pamona is given in schematic form in Table 4. Cliticised pronouns reflecting the PMP genitives (GEN) are shown, as their presence is essential to the interpretation of the forms in the Kaili-Pamona (and other Indonesian-type) languages.¹⁶ It is evident from Martens' notes on Uma (1988) that Proto Kaili-Pamona must also have had a location applicative in *-i, so I have added this to Table 4.

Table 4: Proto Kaili-Pamona verb forms (Mead, Chart 1)

	Active	Passive
undergoer		
irrealis	maN-√	GEN=√
realis	naN-√	ni-√=GEN
locative undergoer		
irrealis	maN-√-i	GEN=√-i
realis	naN-√-i	ni-√-i=GEN

¹⁵ Other Sulawesi languages thereby excluded are those of the Sangiric, Minahasan and Gorontalo groups, which are Philippine-type languages.

¹⁶ I have omitted intransitive forms here for the sake of space and simplicity.

The system reconstructed by Mead for Proto Bungku-Tolaki is given in Table 5:

Table 5: Proto Bungku-Tolaki verb forms (Mead, Chart 14)¹⁷

	Active: indefinite U	Active: definite U	Passive
independent	moN -√	< um >√	< in >√
atemporal	GEN= poN -√	GEN=√	< in >√[=GEN]

If Proto Kaili-Pamona and Proto Bungku-Tolaki are indeed descended fairly immediately from a common protolanguage, then it must have had a matrix like that in Table 6 which differentiated irrealis, realis and atemporal, as Mead suggests. The Proto Kaili-Pamona and Proto Bungku-Tolaki systems would then represent different reductions of the Proto X system.

Table 6: Verb forms in Proto X

	Active non-specific U	Active specific U	Passive
undergoer			
irrealis	maN -√	< um >√	GEN=√
realis	naN -√	?	< in >√[=GEN]
atemporal	paN -√	√	—
locative undergoer			
irrealis	maN -√- i	?	GEN=√- i
realis	naN -√- i	?	< in >√- i [=GEN]
atemporal	paN -√- i	√- i	—

One untidiness in Table 6 is that there are two ‘actives’, one in **maN*-√, the other in **um*√. Zobel, §5.2, provides an explanation for this. In Philippine-type languages the actor voice is used in independent clauses where the patient is non-specific and in dependent clauses where the syntax (e.g. relativisation) requires it (Ross, §2.3.1). Under Zobel’s analysis of various Indonesian-type languages, these two functions became separated: **maN*-√ occurred in independent clauses with a non-specific patient or no patient (it was antipassive), **um*√ in dependent clauses where a specific patient was allowed. That this analysis also applies to Sulawesi languages is suggested by the presence of this distinction in Proto Bungku-Tolaki and the fact that in *Tukang Besi*, <*um*>√ only occurs in dependent clauses (Donohue, §9.1).

3.2.2 A hypothetical extension of Mead’s work

Did a system like Proto X in Table 6 actually exist, as Mead supposes? There is evidence that it did. *Balantak*, a language belonging to another Sulawesi group, *Saluan*, reflects the three-way contrast between PMP neutral, perfective and atemporal forms quite clearly. The

¹⁷ U = undergoer; [] = occurred in some contexts.

relevant forms are shown in Table 7. There is a contrast among passive forms, reflecting PMP patient pivot forms, between $\sqrt{-on}$ (< PMP neutral $*\sqrt{-\partial n}$), $ni-\sqrt{\quad}$ (< PMP perfective $*\langle in \rangle\sqrt{\quad}$) and $GEN=\sqrt{\quad}$ ($\sqrt{\quad}$ < PMP atemporal $*\sqrt{-a}$).¹⁸ The contrast is obvious in Balantak because it retains contrasting reflexes of $*\sqrt{-\partial n}$ and $*\sqrt{-a}$. This enables us to recognise in turn that Proto Kaili-Pamona $GEN=\sqrt{\quad}$ 'passive irrealis' (Table 4) also reflects the atemporal (PMP $*\sqrt{-a}$). This inference is supported by the preposed actor clitic reflecting the PMP genitive. It indicates that the verb that follows it was once atemporal, because it was subordinate to an auxiliary. The genitive was formerly encliticised to the auxiliary, and later became procliticised to the following verb (Ross, §4.2).

Table 7: Balantak verb forms (Busenitz 1994)

	Active	Passive
undergoer		
irrealis	mVN-$\sqrt{\quad}$	$\sqrt{-on}$[=GEN]
realis	nVN-$\sqrt{\quad}$	ni-$\sqrt{\quad}$ (1S: GEN= $\sqrt{\quad}$)
locative undergoer		
irrealis	mVN-$\sqrt{-i}$	$\sqrt{-i-on}$[=GEN]
realis	nVN-$\sqrt{-i}$	ni-$\sqrt{-i}$ (1S: GEN= $\sqrt{-i}$)
benefactive undergoer		
irrealis	mVN-$\sqrt{-kon}$	$\sqrt{-kon-on}$[=GEN]
realis	nVN-$\sqrt{-kon}$	ni-$\sqrt{-kon}$ (1S: GEN= $\sqrt{-kon}$)

Table 8: A partial PMP paradigm

	Active		Passive
undergoer			
neutral	maN-$\sqrt{\quad}$	$\langle um \rangle\sqrt{\quad}$	$\sqrt{-\partial n}$
perfective	m$\langle in \rangle$aN-$\sqrt{\quad}$	$\langle umin \rangle\sqrt{\quad}$	$\langle in \rangle\sqrt{\quad}$
atemporal	paN-$\sqrt{\quad}$	$\sqrt{\quad}$	$\sqrt{-a}$
locative undergoer			
neutral	—	—	$\sqrt{-an}$
perfective	—	—	$\langle in \rangle\sqrt{-an}$
atemporal	—	—	$\sqrt{-i}$
circumstantial undergoer			
neutral	—	—	i-$\sqrt{\quad}$
perfective	—	—	i-$\langle in \rangle\sqrt{\quad}$
atemporal	—	—	$\sqrt{-\acute{a}n}$

¹⁸ The suffix $*-a$ 'patient voice, atemporal' seems to have been lost entirely in Indonesian-type languages.

Can we say then that Proto X is the ancestor of all the Indonesian-type groups of Sulawesi? The answer is, no, not quite. Balantak also reflects two applicatives **-i* 'locative' and **-akən* 'benefactive'. The blanks under 'Active' in the PMP matrix in Table 8 have been 'filled in' with non-PMP morpheme combinations. Perhaps these blanks had already been filled in Proto Sulawesi, but the evidence from a language of the more northerly Tomini-Tolitoli group, Lauje (Himmelman II), speaks against this. The Lauje paradigm in Table 9 contains an interesting asymmetry, as Himmelman notes: there is a full set of forms with the benefactive applicative suffix *-aʔe*, but the gaps have not been filled in with the locative applicative suffix *-i*. This suggests that Lauje is, so to speak, on the border between a Philippine-type language and an Indonesian-type language, and that Proto Sulawesi was on that border too. It also reminds us that the two applicative suffixes have different origins. Locative *-i* originates in PMP **-i* 'location pivot, atemporal'. Benefactive *-aʔe* is descended from a captured preposition **akən*, still a preposition in Indonesian *akan*. What the Lauje evidence suggests is that the filling of the gaps took place in two stages. First, **akən* was captured, giving the benefactive undergoer forms. Then by analogy the functions of **-i* were extended to complete the locative paradigm.¹⁹

This story goes against the received wisdom in two ways: (i) it is usually assumed that the gap-filling took place in one go and (ii) **-akən* is sometimes treated by analysts as if it were an allomorph of PMP **-án* 'circumstantial pivot, atemporal'. Lauje speaks against (i), and there is no evidence that a reflex of PMP **-án* was ever present in Proto Sulawesi. It looks instead as if the circumstantial voice had already vanished, and the capture of **-akən* was a separate and later event.²⁰

Table 9: Lauje verb forms (Himmelman II, Table 3)

	Active	Passive
undergoer		
irrealis	moN-√	no-√=GEN (1S: GEN=√)
realis	noN-√	<in>√=GEN
locative undergoer		
irrealis	—	no-√-i=GEN (1S: GEN=√-i)
realis	—	<in>√-aŋ=GEN
benefactive undergoer		
irrealis	moN-√-aʔe	no-√-aʔe=GEN (1S: GEN=√-aʔe)
realis	noN-√-aʔe	<in>√-aʔe=GEN

Lauje also attests another fact about Proto Sulawesi, namely that the PMP contrast between locative indicative **-an* and locative atemporal **-i* survived into Proto Sulawesi.²¹ However, this contrast was soon eliminated in favour of **-i* in closely related Pendau (Quick,

¹⁹ Matters are complicated by the fact that PMP also had a locative preposition **i*, and this may also have played a role in this development.

²⁰ Zobel, §6.1, makes the opposite inference for his Proto Nuclear Malayo-Polynesian: **-i* had already spread to the active in PNMP, but **-akən* was the outcome of later drift. See also below, §3.2.3.

²¹ The contrast is also reflected in Totoli, another Tomini-Tolitoli language.

Figure 2), where the gaps were also filled in as they were in Proto Kaili-Pamona and Balantak.

This leaves us with a reconstruction of Proto Sulawesi verbal morphology perhaps like the one in Table 10. However, my purpose here is not to propose a reconstruction but to show how even limited morphosyntactic data can quickly enlighten our understanding of historical developments. Indeed, a reconstruction of this kind needs to be complemented by the reconstruction of the forms of its member morphemes and a search for innovations shared by all languages descended from Proto Sulawesi, and Mead, §6, addresses these issues.

Table 10: Verb forms in Proto Sulawesi (?)

	Active non-specific U	Active specific U	Passive
undergoer			
irrealis	maN-√	<um>√	√-ən
realis	naN-√	?	<in>√[=GEN]
atemporal	paN-√	√	GEN=√
locative undergoer			
irrealis	—	?	√-i
realis	—	?	<in>√-an[=GEN]
atemporal	—	√-i	GEN=√-i
benefactive undergoer			
irrealis	maN-√-akən	<um>√	√-akən
realis	naN-√-akən	?	<in>√-akən[=GEN]
atemporal	paN-√-akən	√	GEN=√-akən

Mead notes a morphological innovation, **ki*= 'first person plural exclusive actor enclitic' that would include the South Sulawesi group in the larger Sulawesi grouping.²² Although there seems to be a (largely unspoken) consensus that South Sulawesi should be excluded, this shared innovation suggests its inclusion. An inspection of data in South Sulawesi languages in Campbell (1989) and Friberg (1991) suggests a basic Proto South Sulawesi system like that in Table 11, i.e. one in which only earlier irrealis forms have survived (it also had locative and benefactive applicative suffixes reflecting **-i* and **-akən*). A more difficult task will be to reconstruct the morphosyntactic history of *Tukang Besi* (Donohue) and the other members of the putative *Muna-Buton* group.

Table 11: Verb forms in Proto South Sulawesi (?)

Active intransitive	Active transitive	Passive
maN-√	<um>√	GEN=√

²² It would also include *Embaloh* of Borneo, apparently a South Sulawesi outlier (Adelaar 1994).

It has sometimes been assumed in the literature that Proto Sulawesi had no close relationship to the languages of the Philippine-type groups Sangiric, Minahasan and Gorontalo in the north of the island. The conservatism of Lauje, however, and its similarity in many ways to Sangiric Ratahan (Himmelman II) suggests that this assumption should be revisited — although it may yet prove to be valid. In any event, as Mead points out, the shared protolanguage of even just the Kaili-Pamona and Bungku-Tolaki languages must have been more Philippine-like than any of its descendants.

3.2.3 Zobel's Proto Nuclear Malayo-Polynesian

Zobel makes a proposal which goes much further than Table 10. He suggests that all Indonesian-type languages²³ are descended from a 'Proto Nuclear Malayo-Polynesian' (PNMP) with a possible homeland in Sulawesi (Zobel, §7). The innovations defining Nuclear Malayo-Polynesian (NMP) include what appear to be the same three features as Wolff (1996) uses to characterise Indonesian-type languages (Ross, §4.2). They are (Zobel, §5 and §6):

- (1) a paradigm of proclitics to the verb — GEN= in the tables above (Zobel's innovation 1);
- (2) filling the blanks in the matrix with new morpheme combinations, specifically **maN-√-i* and **um-√-i* (Zobel's innovation 3);
- (3) loss of the neutral/perfective — or irrealis/realis — distinction (part of Zobel's innovation 4; like Wolff, Zobel recognises that this has not affected all Indonesian-type/NMP languages).

To these Zobel adds:

- (4) functional differentiation between **maN-√* as antipassive in independent clauses with a non-specific patient or no patient and **um√* in dependent clauses with a specific patient (Zobel's innovation 2);
- (5) conversion of **in√* from a perfective to a passive marker (part of Zobel's innovation 4).

Innovations 1, 4 and 5 are reflected in Table 10.

Zobel's is the first attempt I am aware of to tackle the history of Indonesian-type languages as a block by an analysis of their morphosyntax. Given the difficulty — due to lexical borrowing — of using phonology and lexicon to do this job, I am convinced that morphosyntax must be used, and Zobel presents us with an impressively wide-angle view of the morphosyntax of Indonesian-type languages, informed by data from a large, well-distributed corpus of languages.²⁴

²³ He draws a line in Borneo between Malayic and Tamanic (both Indonesian-type) and the rest (presumably Philippine-type).

²⁴ The difficulty of subgrouping the more westerly Indonesian-type languages is clear if one compares the numerous attempts which have been made, none of which agrees with another. They include Dyen (1965), Nothofer (1975, 1985, 1988), Blust (1988), and, with regard to Malayic, Blust (1988) and Adelaar (1991, 1992a).

Innovation 1 appears at first sight to be the same as Wolff's formulation of the innovation of proclitics, but it isn't: Wolff's formulation is turned on its head. Wolff, like Himmelmann (1996), believes that GEN= $\sqrt{\quad}$ (procliticisation of a former genitive actor pronominal to the former patient voice atemporal) was innovated bit by bit, starting with the first person singular (as in Lauje and Karo Batak), then spreading to the second person singular (as in Pendau), and only rarely spreading further. Zobel, §5.1, takes the contrary view, believing with van den Berg (1996) that the whole paradigm was inherited into PNMP from a Philippine-type ancestor, and that the PNMP innovation applied not only to 'patient focus' atemporals but also also to 'actor focus' and intransitive atemporals.

Table 12 represents an attempt for comparison's sake to reproduce Zobel's PNMP paradigm in the same format as the tables above.²⁵

Table 12: A partial PNMP paradigm (Zobel, Table 13)

	Actor focus	Actor participle	Patient focus	Gerundive
undergoer				
non-past	maN- $\sqrt{\quad}$	<um> $\sqrt{\quad}$	$\sqrt{\quad}$ (?)	$\sqrt{\quad}$ -ən
past	m<in>aN- $\sqrt{\quad}$	<umin> $\sqrt{\quad}$	<in> $\sqrt{\quad}$	
atemporal	GEN=paN- $\sqrt{\quad}$	—	GEN= $\sqrt{\quad}$ (?)	
locative undergoer				
non-past	maN- $\sqrt{\quad}$ -i	<um> $\sqrt{\quad}$ -i	$\sqrt{\quad}$ -i	$\sqrt{\quad}$ -an
past	m<in>aN- $\sqrt{\quad}$ -i	<umin> $\sqrt{\quad}$ -i	<in> $\sqrt{\quad}$ -an	
atemporal	GEN=paN- $\sqrt{\quad}$ -i	—	GEN= $\sqrt{\quad}$ -i	
circumstantial undergoer				
non-past	maN- $\sqrt{\quad}$ -an	<um> $\sqrt{\quad}$ -an	$\sqrt{\quad}$ -an	i- $\sqrt{\quad}$
past	m<in>aN- $\sqrt{\quad}$ -an	<umin> $\sqrt{\quad}$ -an	(i-)in- $\sqrt{\quad}$	
atemporal	GEN=paN- $\sqrt{\quad}$ -an	—	GEN= $\sqrt{\quad}$ -an	

Innovation 1 raises three issues. First, were proclitics were innovated once, in PNMP, as Zobel suggests, or is this an areal feature that has been spread by contact or that results from parallel inheritances from various Philippine-type languages? For Zobel's case to be supported, a reconstruction of the PNMP proclitic set is needed (or an account of why the proclitics of some languages do not reflect the set), as well as an explanation of why languages that have proclitics in only one or two singular persons should have them in just those persons.

Second, was GEN= $\sqrt{\quad}$ innovated bit by bit? The answer to this is associated with the answer to the first question. If this is an areal feature resulting from contact, then we might expect to find fragmentary paradigms. But if it reflects a single (PNMP) inheritance or multiple inheritances, then we would expect to find whole paradigms, as Philippine-type languages generally have whole paradigms. We would attribute partial paradigms to attrition, assuming, for example, that the Pendau irrealis passive marker *ra-* (Quick, §2) reflects the earlier third

²⁵ Forms with question marks are not included in Zobel's Table 13. I have added the GEN= proclitics in a way that I hope matches Zobel's intentions.

person plural member of the proclitic paradigm. Again, the answer to the question will entail morphological reconstruction.

The third issue is whether the paradigm, partial or otherwise, has been generalised from the passive to the active and intransitive in individual languages, or whether this innovation had already occurred in one fell swoop in PNMP. If the latter, then languages which have proclitics only in the passive are the result of independent parallel limitations on proclitic use. Here, it seems to me, the burden of proof lies with Zobel: by what mechanism did this generalisation of proclitic function occur in PNMP? By Occam's Razor, the inheritance of genitive actor markers in the passive only (albeit, perhaps, the inheritance of the whole paradigm), and their later generalisation to the active and the intransitive in particular languages, is a preferable reconstruction.

Zobel's proposal needs to be tested in detail, in particular by the reconstruction of the forms of its member morphemes, including both the proclitic set and derivational morphemes. For example, Zobel adopts the assumption that PMP **-án* 'circumstantial voice atemporal' was generalised to the new 'actor focus' forms in PNMP as **-an*. The evidence for this is limited, and above I suggested that it had been lost in Proto Sulawesi, and that subsequently **akən* had been captured as an enclitic. As far as I know, all benefactive undergoer applicatives in Sulawesi reflect **-akən*, and so do Karo Batak *-kən* (Norwood), Nias *-ʔo* (Sundermann 1883), Lampung *-ko* (Schröter 1937; Walker 1976), Sundanese *-kin* (Hanafi 2001), Old Javanese *-aken* and modern Javanese *-ake* (Poedjosoedarmo), Madurese *-aghi* (Stevens 1968) and standard Indonesian *-kan*. Languages that may reflect PNMP **-an* are certain Malayic languages, namely Minangkabau *-an* and Kendayan Dayak *-atn* (Adelaar 1984), and Balinese and Sasak, which have beneficiary applicative *-in* rather than *-i*, so their morphological history is in need of further investigation.²⁶

The fact that reflexes of **-akən* are so pervasive in Indonesian-type languages leaves one wondering whether they are to be explained by drift, as Zobel assumes, or by contact, or whether they reflect an innovation in an interstage language like PNMP. There are two objections to the last suggestion. First, Adelaar (1992a) finds no convincing grounds for reconstructing Proto Malayic **(a)kAn* because its reflexes are phonologically irregular, some Malayic lects apparently reflect **-an*, and Banjar Hulu *-akan* is suffixed to verbs that already have *-i*. Malayic would be a subgroup within NMP. The second objection is that a cognate Proto Oceanic **-akin* is reconstructable, and Indonesian-type and Oceanic languages do not belong to the same branch of the Austronesian genealogical tree, unless one accepts Zobel's revision of it (his Figure 2). For the moment, the reflexes of **-akən* remain a matter for further research.

I suggested above that **maN-√-i* etc. should not be reconstructed for Proto Sulawesi. This would also mean that it should not be reconstructed for the higher-order interstage PNMP. However, the fact that there seems to be no language with a reflex of **-akən* that lacks a reflex of **maN-√-i* other than Lauje implies that the gaps in the matrix were filled with **-i* forms first, resulting in the reinterpretation of **-i* as an applicative and laying the

²⁶ One thing is clear: Balinese and Sasak share enough common innovations to guarantee that they form a subgroup, and it is reasonably clear that Sumbawa belongs here, too. It is *not* clear that they subgroup with Javanese, as is often assumed. Sasak and Sumbawa also reflect the actor proclitic paradigm (Wouk), implying that Balinese once had it too, although there is no sign of it today.

ground for the capture of **-akən* as a second applicative. The absence of forms reflecting **maN-√-i* etc. in Lauje, which I had taken to reflect their absence in Proto Sulawesi, could alternatively be explained by loss as a result of contact between Lauje speakers and speakers of Philippine-type languages. This would then imply the addition of **maN-√*, **naN-√* and **paN-√* to the reconstruction of Proto Sulawesi in Table 10.

3.2.4 Indonesian-type languages outside Sulawesi

A number of languages outside Sulawesi have voice matrices which could be derived fairly readily from this modified Proto Sulawesi matrix. Again, the point is not that Proto Sulawesi is their common ancestor, but rather that something like Proto Sulawesi or Zobel's PNMP could eventually prove to be a common reconstructable ancestor for these languages.

The Karo Batak optional passive marker *i-* (Table 13) evidently reflects **in*, as it has an archaic alternant *ni-*. Karo Batak has neutralised irrealis and realis sets, largely in favour of the realis (but with irrealis **-i*).

Table 13: Karo Batak verb forms (Norwood)

	Active	Passive
undergoer	N-√	[i-]√=GEN (1S/P: GEN=√)
locative undergoer	N-√-i	[i-]√-i=GEN (1S/P: GEN=√-i)
benefactive undergoer	N-√-kən	[i-]√-kən=GEN (1S/P: GEN=√-kən)

The Old Javanese matrix in Table 14 resembles the Proto Sulawesi matrix, maintaining irrealis active forms and realis passives.

Table 14: Old Javanese verb forms (Poedjosoedarmo, Tables 5 and 6)

	Active	Passive
undergoer	<um>√ [m]aN-√	<in>√=GEN
locative undergoer	<um>√-i [m]aN-√-i	<in>√-an=GEN
theme undergoer	<um>√-aken [m]aN-√-aken	<in>√-aken=GEN
beneficiary undergoer	p<um>a-√-aken	p<in>a√-aken=GEN

As Poedjosoedarmo remarks, the relationship between Old and modern Javanese is not well understood, and comparing Tables 14 and 15 does not elucidate a great deal. Modern Javanese *N-√* may reflect either Old Javanese *[m]aN-√* or a conflation of the latter with *um>√* (as Beratha 1992 proposes for Balinese). Modern Javanese *√-en* appears to reflect PMP **-ən*, where there is no evidence of it in Old Javanese, and the modern Javanese passives appear to have been remodelled on the basis of Classical Malay.

Table 15: Modern Javanese verb forms (Poedjosoedarmo, Tables 2 and 3)²⁷

	Active	Passive
undergoer		
irrealis	N-√	di-√=GEN (1S, 2S: GEN=√)
imper/subjunc	N-√-a	√-en
desiderative	—	1S: GEN=√-e
locative undergoer		
irrealis	N-√-i	di-√-i=GEN(1S, 2S: GEN=√-i)
imper/subjunc	(N-√-an-a)	√-an-a
desiderative	—	1S: GEN=√-an-e
benefactive undergoer		
irrealis	N-√-ake	di-√-ake=GEN (1S, 2S: GEN=√-ake)
imper/subjunc	(N-√-(k)n-a)	√-(k)n-a
desiderative	—	1S: GEN=√-(k)n-e

The modern Javanese imperative/subjunctive marker *-a* may just possibly reflect the PMP actor voice projective, but this seems unlikely, as it is generalised to all voice forms of the imperative/subjunctive.

Finally, the Ngeno-Ngene Sasak matrix in Table 16 reflects a generalisation of actor clitics to the nasal forms. Wouk notes that these may be procliticised to the following verb, or encliticised to a preceding auxiliary, much as Wolff (1996) reconstructs for archaic Philippine-type languages.

Table 16: Ngeno-Ngene Sasak verb forms (Wouk)

	Nasal	Oral
undergoer	[GEN]=N-√	GEN=√
locative undergoer	[GEN]=N-√-in	GEN=√-in
benefactive undergoer	[GEN]=N-√-ang	GEN=√-ang

The Sasak applicative suffixes are not obviously cognate with those in most other Indonesian-type languages, but are identical to those in Balinese. The Balinese matrix is identical to Sasak, except that it lacks the clitics and retains the identification of nasal and oral forms with active and passive respectively.

3.2.5 Chamorro, Palauan and Oceanic

The usefulness of an approach like Zobel's for subgrouping is demonstrated by his arguments about the place of the two outliers Chamorro and Palauan in the Malayo-Polynesian genealogical tree. Their place in the tree has long been a puzzle, and Zobel shows that they are both Indonesian-type languages and must have their origins among languages of

²⁷ Forms in parentheses are from Dudas (1976) and Robson (1992).

this type rather than of the Philippine type. I have rearranged Zobel's analysis of Chamorro as Table 17, roughly in the same format as earlier tables. Although there have been a number of reassignments of functions,²⁸ the relationship of this matrix to those in Table 10 or Table 12 is clear enough. Features which speak for this being an Indonesian-type language are (i) the locative applicative *-i*, (ii) the presence of GEN= (extended in its scope to **in*√), and (iii) the division between the antipassive function of *maN*-√ and the participle function of *um*√.

Table 17: Chamorro verb forms (Zobel, Tables 8 and 9)

	Antipassive	Participle	Active	Passive
undergoer				
realis	maN -√	⟨ um ⟩√	GEN=√	GEN=⟨ in ⟩√
irrealis	faN -√	—	GEN=√	—
locative undergoer				
realis	maN -√- i	—	GEN=√- i	GEN=⟨ in ⟩√- i
irrealis	faN -√- i	—	GEN=√- i	—

Reid (In press) believes that Chamorro represents a single-member first-order subgroup of Malayo-Polynesian and presents evidence of its resemblances to Philippine languages (and of its borrowings from Oceanic Micronesian languages). He is writing in response to an earlier version of Zobel's contribution to this volume, but he does not deal in detail with the three features I have just mentioned, and these seem to me more compelling than the Philippine resemblances (which may be attributed to shared inheritance). Blust (2000) examines the phonological history of Chamorro, and, although he believes that Chamorro originated in the Philippines, he is unable to show that it subgroups with Philippine languages. This leaves Zobel's hypothesis as one which deserves further investigation.

The case for Palauan also being an Indonesian-type language is not as strong as the Chamorro case, because the functional reassignments that are required are more complex and less obviously motivated, as comparison of Table 18 with Tables 10 and 12 reveals (see Zobel, Table 12). However, GEN= occurs here too, albeit with a much altered distribution, and, significantly, there is again a division between the antipassive function of *maN*-√ and transitive function of *um*√. Zobel also offers *-akl*, a fossilised reflex of **akən*, as evidence of Palauan's Indonesian origins.

Table 18: Palauan verb forms (Zobel, Tables 11 and 12)²⁹

	Antipassive	Active	Gerundive
indicative non-past	meN -√	⟨ m ⟩√	√- el
indicative past	m ⟨ il ⟩ eN -√	⟨ il ₂ ⟩√	—
subjunctive non-past	oN -√	√	—
subjunctive past	uleN -√	⟨ (i) l ₂ ⟩√	—

²⁸ The origins of the Chamorro functions are: antipassive < active w non-specific patient, participle < active with specific patient, active (transitive) < passive (transitive), passive < GEN= + passive, realis < irrealis, irrealis < atemporal.

²⁹ Actor proclitics (GEN=) are not shown in the table, as their use is syntactically determined.

Zobel includes the whole of Central–Eastern Malayo-Polynesian, including Oceanic, in his NMP grouping. This remains open to question, as work on the subject prefixes of Oceanic languages suggests rather strongly that there were two sets of prefixes at a stage immediately prior to Proto Oceanic, one reflecting the PMP pivot pronouns, the other the PMP genitives. This has been taken to mean that the immediate forerunner of Proto Oceanic was a Philippine-type language (Lynch, Ross & Crowley 2001:Ch.4; Kikusawa 2000). Palauan also has sets of prefixes from each of these sources: this might either suggest that Palauan has a Philippine-type origin after all, or that Proto Oceanic has an Indonesian-type origin.

3.2.6 *Other verbal morphemes in Indonesian-type languages*

The tendency of Philippine languages to mark intransitive verbs according to the semantic role of their single argument (§ 2.1) often carries over into Indonesian-type languages as a contrast between actor argument (= unergative) verbs, reflecting **maN-√* or **N-√*, and unprefixated undergoer argument (= unaccusative) verbs. This is exemplified by Karo Batak, Balinese, Ngeno-Ngene Sasak (Austin 2001). Some languages, like Lampung (Walker 1976), Pendau, the Bungku-Tolaki and Kaili-Pamona languages, and Tukang Besi, retain a reflex of the PMP stative intransitive prefix **ma-*.

There are several verbal affixes with quite widespread reflexes in Indonesian-type languages which appear to play a different role from the one they have in Philippine-type languages. They include **ka-* ‘intransitive passive’ and **taR-* ‘involuntary intransitive passive’. These need further investigation with regard both to their functions and their distribution.

4 Post-Indonesian languages

Above I defined ‘Post-Indonesian languages’ as those in which the morphology of voice still occurs, but the voice system has broken down and the former voice-marking affixes no longer mark alternate pivot choices.

In Riau Indonesian (Gil) the old Malay voice-marking morphology still occurs in the shape of *N-√* and *di-√*, but its function is no longer to mark actor and patient voices, i.e. to designate the pivot. Instead, as Gil shows, *N-√* simply indicates that the clause *has* an actor, and *di-√* that it has a patient.

The Ngeno-Ngene dialect of Sasak (Wouk) seems to be undergoing a parallel development, although it has not progressed as far along this path as Riau Indonesian. There is no longer a clear voice distinction between *N-√* and *√*, i.e. there is no apparent reason to identify a pivot with either form, except (i) in a passive construction where the verb takes the enclitic *=ne* and the actor is an oblique with the preposition *siq*,³⁰ and (ii) in relative clauses, where one may argue that it is the actor pivot of *N-√* and the undergoer pivot of *√* that is relativised. The actor clitic is spreading from the *√* form to the *N-√* form (Table 16). The *N-√* form is still likely to have a non-specific patient, but this is a tendency, not a rule.

This is sharp contrast with Balinese, where the voice system remains fully operative. The Meno-Mene and Meriaq-Meriku dialects of Sasak also seem to retain more of the voice

³⁰ There is a similar construction in Balinese, the function of which is to mark the actor as definite.

- 1991, Ergativity, focus and verb morphology in several South Sulawesi languages. In Ray Harlow, ed. *VICAL 2: Western Austronesian and contact languages*, 103-130. Auckland: Linguistic Society of New Zealand.
- 1996, Konjo's peripatetic person markers. In H. Steinhauer, ed. 1996:137-171.
- Hanafi, Nurachman, 2001, A description of basic clause structure in Sundanese. In Austin, Blake and Florey, eds, 2001:121-142.
- Hanson, Chris, 2001, A description of basic clause structure in Bugis. In Austin, Blake and Florey, eds, 2001:143-160.
- Himmelman, Nikolaus P., 1996, Person marking and grammatical relations in Sulawesi. Steinhauer, ed., 1996:115-136.
- Huang, Lillian M., 1994a, Ergativity in Atayal. *Oceanic Linguistics*, 33:129-143.
- 1994b, *A study of Atayal syntax*. Taipei: The Crane.
- 1995, *A study of Mayrinax syntax*. Taipei: The Crane.
- 2000, Verb classification in Mayrinax Atayal. *Oceanic Linguistics*, 39:364-390.
- Kikusawa, Ritsuko, 2000, Reconstructing the actancy systems of Proto-Central Pacific and its daughter languages: Ergativity, accusativity and their diachronic development. PhD dissertation, University of Hawai'i.
- Lynch, John, Malcolm D. Ross and Terry Crowley, 2001, *The Oceanic languages*. Richmond: Curzon Press.
- Martens, Michael, 1988, Notes on Uma verbs. In Steinhauer, ed., 1988:239-256.
- Mead, David, 1998, *Proto-Bungku-Tolaki: reconstruction of its phonology and aspects of its syntax*. PhD, Rice University.
- Nothofer, Bernd, 1975, *The reconstruction of Proto-Malayo-Javanic*. VKI 73. The Hague: Martinus Nijhoff.
- 1985, The sub-grouping of the languages in the Javo-Sumatra Hesion: a reconsideration. *Bijdragen tot de Taal-, Land- en Volkenkunde*, 141:288-302.
- 1988, A discussion of two Austronesian subgroups: Proto-Malay and Proto-Malayic. In Ahmad and Zain, eds, 1988:34-58.
- Pastika, I Wayan, 1999, Voice selection in Balinese narrative discourse. PhD thesis, The Australian National University.
- Pecoraro, Ferdinando, 1979, *Eléments de grammaire taroko, précédés del la présentation de la vie et de la culture des taroko*. Paris: Association Archipel.
- Post, Ursula, 1992, *Binukid dictionary*. Studies in Philippine Linguistics 9/2. Manila: Linguistic Society of the Philippines & Summer Institute of Linguistics.
- Reid, Lawrence A., In press, Morphosyntactic evidence for the position of Chamorro in the Austronesian language family. In Robert Bauer, ed. *Collected papers in South-East Asian linguistics*. Canberra: Pacific Linguistics.
- Robson, Stuart, 1992, *Javanese grammar for students*. Monash Papers on Southeast Asia 26. Clayton, Vic.: Centre for Southeast Asian Studies, Monash University.
- Ross, Malcolm D., forthcoming, Notes on the prehistory and internal subgrouping of Malayic. In John Bowden and Nikolaus P. Himmelman, eds *Papers in Austronesian Linguistics*. Canberra: Pacific Linguistics.
- Rubino, Carl R.G., 2000, *Ilocano grammar and dictionary*. Honolulu: University of Hawai'i Press.

- Schachter, Paul, 1984, Sematic-role-based syntax in Toba Batak. In Schachter, ed., 1984: 122-149.
- Schachter, Paul, ed., 1984, *Studies in the structure of Toba Batak*. UCLA Occasional Papers in Linguistics 5.
- Schachter, Paul and Fe T. Otones, 1972, *Tagalog reference grammar*. Berkeley: University of California Press.
- Schröter, Richard, 1937, Versuch einer grammatischen Auslegung des Kernerling-Dialektes der Lampung-Sprache. Dissertation zur Erlangung der Doktorwürde [PhD dissertation], Hansische Universität Hamburg.
- Sneddon, James N., ed., 1989, *Studies in Sulawesi linguistics Part I, NUSA* 31.
1991, *Studies in Sulawesi linguistics Part II, NUSA* 33.
- Steinhauer, Hein, ed., 1988, *Papers in western Austronesian linguistics No.4*. Canberra: Pacific Linguistics.
1996, *Papers in Austronesian linguistics No.3*. Canberra: Pacific Linguistics.
- Stevens, Alan M., 1968, *Madurese phonology and morphology*. American Oriental Series 52. New Haven, Conn.: American Oriental Society.
- Sundermann, H., 1883, Kurze Formenlehre der Niassischen Sprache nebst einem syntaktischen Anhang. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde*, 28: 93-157.
- Tuuk, H.N. van der, 1971, *A grammar of Toba Batak*. The Hague: Martinus Nijhoff.
- Walker, Dale F., 1976, *A grammar of the Lampung language: the Pesisir dialect of Way Lima*. NUSA 2.
- Wechsler, Stephen and I Wayan Arka, 1998, Syntactic ergativity in Balinese: an argument-structure based theory. *Natural Language and Linguistic Theory*, 16:387-441.
- Wolff, John U., 1996, The development of the passive verb with pronominal prefix in western Austronesian languages. In Bernd Nothofer, ed. *Reconstruction, classification, description: festschrift in honor of Isidore Dyen*, 15-40. Hamburg: Abera.
- Woollams, G., 1996, *A grammar of Karo Batak*. Canberra: Pacific Linguistics.
- Wouk, Fay, 1986, Transitivity in Batale and Tagalog. *Studies in Language*, 10:391-424.
1999, Sasak is different: a discourse perspective on voice. *Oceanic Linguistics*, 38: 91-114.
- Zeitoun, Elizabeth and Lillian M. Huang, 2000, Concerning *ka-*, an overlooked marker of verbal derivation in Formosan languages. *Oceanic Linguistics*, 39:391-414.
- Zorc, R. David, 1977, *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. Canberra: Pacific Linguistics.