East Nusantara:
typological and areal analyses

Pacific Linguistics 618

Pacific Linguistics is a publisher specialising in grammars and linguistic descriptions, dictionaries and other materials on languages of the Pacific, Taiwan, the Philippines, Indonesia, East Timor, 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 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.

FOUNDING EDITOR: Stephen A. Wurm

EDITORIAL BOARD: John Bowden and I Wayan Arka (Managing Editors), Mark Donohue, Nicholas Evans, David Nash, Andrew Pawley, Malcolm Ross, Paul Sidwell, Jane Simpson, and Darrell Tryon

EDITORIAL ADVISORY BOARD:

Karen Adams, *Arizona State University*
Alexander Adelaar, *University of Melbourne*
Peter Austin, *School of Oriental and African Studies*
Byron Bender, *University of Hawai'i*
Walter Bisang, *Johannes Gutenberg-Universität Mainz*
Robert Blust, *University of Hawai'i*
David Bradley, *La Trobe University*
Lyle Campbell, *University of Utah*
James Collins, *Universiti Kebangsaan Malaysia*
Bernard Comrie, *Max Planck Institute for Evolutionary Anthropology*
Soenjono Dardjowidjojo, *Universitas Atma Jaya*
Matthew Dryer, *State University of New York at Buffalo*
Jerold A. Edmondson, *University of Texas at Arlington*
Margaret Florey, *Monash University*
William Foley, *University of Sydney*
Karl Franklin, *SIL International*
Charles Grimes, *SIL International*
Nikolaus Himmelmann, *University of Münster*
Lillian Huang, *National Taiwan Normal University*
Marian Klamer, *Universiteit Leiden*
Harold Koch, *The Australian National University*
Frantisek Lichtenberk, *University of Auckland*
John Lynch, *University of the South Pacific*
Patrick McConvell, *Australian Institute of Aboriginal and Torres Strait Islander Studies*
William McGregor, *Aarhus Universitet*
Ulrike Mosel, *Christian-Albrechts-Universität zu Kiel*
Claire Moyse-Faurie, *Centre National de la Recherche Scientifique*
Bernd Nothofer, *Johann Wolfgang Goethe-Universität Frankfurt am Main*
Bambang Kaswanti Purwo, *Universitas Atma Jaya*
Ger Reesink, *Radboud University Nijmegen*
Lawrence Reid, *University of Hawai'i*
Jean-Claude Rivierre, *Centre National de la Recherche Scientifique*
Melenaite Taumoefolau, *University of Auckland*
Tasaku Tsunoda, *University of Tokyo*
John Wolff, *Cornell University*
Elizabeth Zeitoun, *Academica Sinica*

East Nusantara:
typological and areal analyses

edited by
Michael C. Ewing and Marian Klamer

Pacific Linguistics
School of Culture, History and Language
College of Asia and the Pacific
The Australian National University

Published by Pacific Linguistics
School of Culture, History and Language
College of Asia and the Pacific
The Australian National University
Canberra ACT 0200
Australia

Copyright in this edition is vested with Pacific Linguistics

First published 2010

National Library of Australia Cataloguing-in-Publication entry:

Title:	East Nusantara: typological and areal analyses / edited by Michael C. Ewing and Marian Klamer.
ISBN:	9780858836105 (pbk.)
Notes:	Includes bibliographical references.
Subjects:	Papuan languages. Austronesian languages.
Other Authors/ Contributors:	Ewing, Michael C. Klamer, Margaretha Anna Flora. Australian National University. Research School of Pacific Studies. Dept. of Linguistics.
Dewey Number:	499.22

Typeset by Julie Manley
Cover image: woven Ikat fabric from Mangili, East Sumba, depicting a cock fight.
Cover design by Julie Manley/Addcolour Digital Pty Ltd
Printed and bound by Addcolour Digital Pty Ltd, Fyshwick, Canberra

Table of contents

Glossing conventions vii

1 The languages of East Nusantara: an introduction 1
MARIAN KLAMER, LEIDEN UNIVERSITY AND
MICHAEL C. EWING, UNIVERSITY OF MELBOURNE

2 Towards a phonological overview of the vowel and consonant systems of East Nusantara 25
JOHN HAJEK, UNIVERSITY OF MELBOURNE

3 Notes on Waima'a intonational structure 47
NIKOLAUS HIMMELMANN, RUHR-UNIVERSITÄT BOCHUM

4 Prefixation of arguments in West Papuan languages 71
GER REESINK, RADBOUD UNIVERSITY NIJMEGEN

5 Person marking, verb classes, and the notion of alignment in Western Pantar (Lamma) 97
GARY HOLTON, UNIVERSITY OF ALASKA FAIRBANKS

6 Agentive alignment in Central Maluku languages 119
MICHAEL C. EWING, UNIVERSITY OF MELBOURNE

7 Reduced pronouns and arguments in Sou Amana Teru, Ambon 143
SIMON MUSGRAVE, MONASH UNIVERSITY

8 Voice and valency in Tetun Dili 165
CATHARINA WILLIAMS-VAN KLINKEN, DILI INSTITUTE OF TECHNOLOGY

9 Grammaticalisation of asymmetrical serial verb constructions in Klon 185
LOUISE BAIRD, LEIDEN UNIVERSITY

10 One item, many faces: 'come' in Teiwa and Kaera 203
MARIAN KLAMER, LEIDEN UNIVERSITY

11 Negation in Central Moluccan languages 227
MARGARET FLOREY, UNIVERSITY OF MELBOURNE

12 Hawu and Dhao in eastern Indonesia: revisiting their relationship 251
CHARLES E. GRIMES, AUSTRALIAN NATIONAL UNIVERSITY AND SIL INTERNATIONAL

13 Nouns and verbs in Magey Matbat 281
BERT REMIJSEN, UNIVERSITY OF EDINBURGH

Glossing conventions

1p	1st person plural
1pe	1st person plural exclusive
1pi	1st person plural inclusive
1s	1st person singular
2	2nd person (unmarked for singular or plural)
2p	2nd person plural
2s	2nd person singular
3d	3rd person dual
3nh	3rd person non-human
3p	3rd person plural
3ph	3rd person plural human
3pnh	3rd person plural non-human
3s	3rd person singular
3sh	3rd person singular human
3snh	3rd person singular non-human
4p	4th person plural (switch reference)
4s	4th person singular (switch reference)

A	agent
ABL	ablative
ACC	accusative
ACT	actor
ADMON	admonitive
AL	alienable
ALL	allative
ANT	anterior
APPL	applicative
ART	article
ASP	aspectual particle
AUX	auxiliary
BEN	benefactive
BNDRY	boundary
Ca	consonant + /a/ partial reduplication
CAUS	causative
CLF	classifier
CMP	complementizer
COLL	collective
COMP	comparative
COMPL	completive
CONJ	conjunction
CONT	continuative
COP	copula
DAT	dative
DECL	declarative
DEF	definite marker
DEI	deictic
DEM	demonstrative
DET	determiner
DIM	diminutive
DIR	directional
DIST	distal
DTOP	different topic (in discourse)
DUR	durative
EMPH	emphatic
ERG	ergative marker
EVID	evidential
EXCL	exclamation
EXIST	existential
f	free pronominal paradigm

F	feminine
FOC	narrow focus particle
FUT	future
GEN	genitive
GNR	generic
HORT	hortative
IMM	immediate future
IMP	imperative
INAL	inalienable
INAN	inanimate
INCL	inclusive
INS	instrument
INTR	intransitive
INTS	intensifier
IPFV	imperfective
IRR	irrealis
LANDW	landward
LM	lexical marker
LNK	linker
LOC	locative
M	masculine
MAL	Malay loan
MID	middle voice
MOD	modality
NEG	negative
NM	noun class marker
NMLZ	nominalizer
NPST	non-past
OBJ	object
OBL	oblique
OBV	obviative
ORD	ordinal
PAT	patient
PFV	perfective
PL	plural
PN	proper name
POL	polite
POSS	possessive
PRED	predicator / predicate marker
PRN	pronoun
PROH	prohibitive
PROX	proximal
RDP	reduplication
REC	recipient
REL	relative
RECP	reciprocal
RES	resumptive pronoun
RLS	realis
SBJ	subject
SEAW	seaward
SEQ	sequential marker
SG	singular
SIM	simultaneous marker
STAT	stative
STOP	same topic
TAG	question tag
TOP	topic
U	undergoer
VAL	valence increasing prefix
VIS	visible
WHQ	wh- question
YNQ	yes/no question

1 *The languages of East Nusantara: an introduction*

MARIAN KLAMER AND MICHAEL C. EWING

1 Introduction

'East Nusantara', the name used in the title of this book, refers to the islands of eastern Indonesia and East Timor. 'Nusantara' is a term that has come to refer to the Indo-Malaysian archipelago generally, without reference to national borders.[1] For the purpose of this volume, we define East Nusantara as a geographical area that extends from Sumbawa to the west, across the islands of East Nusa Tenggara, Maluku[2] including Halmahera, and to the Bird's Head of New Guinea in the east (see Map 1). In the northwest, the area is bounded by Sulawesi.[3]

In East Nusantara, some 400 languages are spoken (see Gordon 2005), most of which are endangered in terms of numbers of speakers, and the majority of which have not yet been described (Florey 2005). Linguistically, this geographic region displays great genetic diversity, being the meeting ground of languages belonging to the Austronesian and Papuan language families. Yet, similarities cut across many of these languages, giving rise to the notion of a linguistic area or *Sprachbund.* In this introduction chapter, we first present a brief history of the region and an overview of recent research that has had East Nusantara in its scope. This will serve as a general background for the chapters on individual languages that constitute the rest of this volume, summarised in the second section of this introduction.

---

1 An illustration of this use of the term in a recent publication is Jones (2007: x).

2 Nusa Tenggara (West and East) and Maluku are the comtemporary term for what have been called the Lesser Sundas and the Moluccas in earlier research. The Lesser Sundas comprise Bali, West Nusa Tenggara (Lombok and Sumbawa) and East Nusa Tenggara (Komodo, Flores, Solor islands, Alor-Pantar islands, Sumba, Timor).

3 In the literature, there is no consensus on the exact geographic delimitations of the East Nusantara region. While East Nusa Tenggara and Maluku (including Halmahera) are generally included, the precise topic of study determines whether (parts of) New Guinea are also considered to be part of it. Ross (2005: 15, footnote 2), in a study on the genetic subgrouping of Papuan languages, treats mainland New Guinea separate from East Nusantara. In contrast, Donohue (2007), in a study on word order in Austronesian, defines the area of 'east Nusantara', or the area where 'eastern 'Indonesian'' languages are spoken, as including New Guinea (2007: 350, 352).

Michael C. Ewing and Marian Klamer, editors.
Typological and areal analyses: contributions from East Nusantara, 1–24. Pacific Linguistics, 2010
©. This edition vested with Pacific Linguistics.

2 Introduction to the study of East Nusantara and its languages

2.1 Prehistory of East Nusantara

The earliest evidence of humans in what is now the Indonesian archipelago dates from about 40,000 years ago (BP). These people are most probably the ancestors of modern Melanesians, Australian Aborigines and the Negrito communities of the Malay Peninsula and the Philippines.[4] Due to climatic changes, human settlement became concentrated in the drier eastern part of the archipelago.

During the Pleistocene period, which lasted until approximately 11,000 BP, the landmasses of Australia and New Guinea were joined in a single continent that geologists and others refer to as Sahul. The islands of western Indonesia then formed a sub-continental peninsula, called Sunda or Sundaland. The islands between these two prehistoric land masses are referred to as Wallacea. Birdsell (1977) hypothesises that Sahul was populated by at least three groups of different people, at times when sea levels allowed relatively easy crossing between Sunda and Sahul. The scenarios that make up his proposal all include the possibility of a connection between populations in Wallacea (including East Nusantara) and New Guinea.

In the highlands of New Guinea, around 9,000 BP, the Melanesians made a major technological breakthough when they developed agriculture, which sustained much more densely settled communities than their previous hunting and gathering had done. This strengthened the Melanesian presence in the east of the archipelago, while the western and central regions, remained relatively sparsely populated (Cribb 2000:29-30).

The expansion of the Austronesians started around 5,000 BP. They moved southwards from Taiwan (Blust 1985, 1995, 1999) to the northern Philippines. In the late fifth and fourth millennia BP, a wave of migrants went west into Borneo and Sulawesi, and later toward Java, Sumatra, Peninsular Malaysia, and Vietnam. Other founders moved east and south into Maluku and East Nusa Tenggara. Austronesian speakers arrived in the Timor-Alor-Pantar area some 3,500 years BP. Additional movements occurred between 4,000-3,500 BP, via Halmahera to the east, skimming the north coast of New Guinea, to the Bismarck Archipelago, into Oceania, arriving in Melanesia around 3,500-3,300 BP (Bellwood 1997:105, Cribb 2000:30, Pawley 2005:95-96, among others).

Despite this Austronesian expansion, there is evidence that non-Austronesian peoples have remained in many of these areas, including the Timor-Alor-Pantar region, Maluku including Halmahera, and mainland New Guinea. Preliminary research shows a genetic connection between people living in these areas. For example, Reesink (2005:203) refers to a study by Capelli et. al. (2001) which included a population sample from the Bird's Head. Its results identified a haplogroup of the Y chromosome that is mainly restricted to Melanesia. Outside Melanesia it has a high frequency in Alor, and Capelli et al. (2001) relate this to the presence of Papuan languages in the region of Timor and the smaller islands of Alor and Pantar**.** Further, Kayser et al. (2003) found four haplogroups on the Y-chromosome that most likely arose in Melanesia, before the Austronesian expansion. They have a distribution of high frequencies in the Highlands of New Guinea, while three of them are also found in East Nusa Tenggara and Maluku, with higher frequencies in Papuan speaking populations than in Austronesian speaking groups. Therefore human genetic

[4] Although the date for initial occupation of New Guinea and Australia is still unresolved (Veth et al. 1998: 162), it is generally agreed that the first humans arrived no later than 40,000 BP, possibly going back to 50,000 BP.

studies support an old connection between the non-Austronesian languages spoken on the islands of East Nusantara and those of the Papuan mainland.

From a linguistic perspective, there is general consensus that the non-Austronesian populations found in Maluku and East Nusa Tenggara must have predated the arrival of the Austronesian speaking populations.[5] In fact, 'the first meeting of Austronesian and Papuan speakers was thus perhaps in the Timor area' (Ross 2005:18). Ross (2005) presents evidence that the homeland of the Trans New Guinea family is located somewhere in the eastern highlands of New Guinea. If the outlier languages in Pantar-Alor belong to the Trans New Guinea family, as he hypothesises, then they must have spread (as a result of language shift, or by means of peoples' migrations starting about 6,000 years BP, Ross 2005:41) from east to west, from the eastern part of the New Guinea mainland, all the way to Pantar, Alor, and Timor, reaching the area perhaps 4,500-4,000 BP (Bellwood 1997:123, Pawley 1998:684-5, Ross 2005:42). While the non-Austronesian languages that are spoken in Halmahera and Alor-Pantar today may point to an old (probably ancient) connection between East Nusantara and the New Guinea mainland, no Papuan languages are spoken in Central Maluku. However, these islands are located directly between the Papuan languages of North Halmahera in the north, those of the Timor region in the south, and those of the Bomberai peninsula in the east. Given the presence of Papuan languages around it, it is not unreasonable to posit an earlier Papuan presence in Maluku as well (see Donohue and Grimes 2008).

We should add that there is no reason to assume that the present-day Papuan languages in East Nusantara are the descendants of a single group of prehistoric populations or are the result of a single wave of migrations. Rather, it is far more plausible that they constitute a complex mix of prehistoric populations and various east-west migrations. Moreover, within historic times, there have also been numerous migrations between the various islands of East Nusantara.

As one illustration of this latter point, consider Makasai, Oirata, and Fataluku – the Papuan languages in the eastern part of East Timor. There is clear evidence that these languages post-date the arrival of the Austronesians. One type of evidence is archaeological, and comes from rock art motifs found in various archaeological sites in East Timor. Most of these sites are found on the eastern part of East Timor (see O'Connor 2003, Figure 1, p. 97), in areas that are currently populated by communities speaking a Papuan language. However, the rock art motifs found in these sites show significant stylistic affinities with painted art elsewhere in the Western Pacific. In the Pacific sites, the art co-occurs with Austronesian settlements that postdate the Austronesian expansion (O'Connor 2003:109). What this suggests is that the eastern part of East Timor was previously occupied by speakers of (an) Austronesian language(s), and that non-Austronesian speakers moved into that area in historic times (2,000 years ago, or later, O'Connor 2003:118). Anthropological evidence presented in McWilliam (2007) confirms this: Fataluku, a Papuan language on the eastern tip of East Timor, was adopted into an Austronesian speaking culture that already existed there before.

A much more recent migration of Papuan speakers is presented in Bouman (1943:484), who reports the oral tradition according to which the Tanglapui (Papuan) in east Alor descend from immigrants from Timor, and (in 1943) had come to Alor about 15 generations previously. Bouman also reports that the coastal populations in central and

---

[5] However Pawley is cautious enough to say that careful study of the internal diversity of the languages of Timor-Alor-Pantar is needed to settle the issue (Pawley 2005: 102).

east Alor are descendants of immigrants from Kisar, Timor, as well as Ende (Bouman 1943:485). Oral traditions like this may suggest that populations (and their languages) as they are found in certain locations today do not necessarily descend from ancient populations in that same location.

Clearly, the migratory and linguistic interactions between non-Austronesian and Austronesian populations in East Nusantara have been ongoing and complex. There is general consensus that the Papuan populations in East Nusantara are pre-historic and predate the arrival of the Austronesians. Nonetheless, an individual non-Austronesian language in a particular place may be the result of migration that took place in historic, or even recent times.

2.2 A brief history of linguistic studies in East Nusantara languages

For more than a century observers have noted that the languages spoken in East Nusa Tenggara Islands and Maluku are somehow different from those spoken to the west and the east. Brandes (1884) proposed the 'reversed' order of [possessor-possessum] or [Gen-N] as a criterion to separate the eastern and western languages as genealogical subgroups.[6] His basic order was thus N-Gen, as illustrated in (1), while the 'reversed' Gen-N order is illustrated in (2).

(1)	*bapakku,*	*bapak saya*		
	father-1sPoss	father 1sPoss	'my father'	Standard Indonesian

(2)	*au ami*		
	1s father	'my father'	Ambai (Silzer 1983)

A line drawn between Sulawesi and Maluku, and through Flores, became known as the 'Brandes line', and represented the division of Austronesian languages into west and east. The 'preposed genitive' was used by others (for example van der Veen 1915) as a diagnostic for non-Austronesian languages (for example van der Veen 1915).[7]

Following a practice that was common in the 19th and early 20th century, languages were classified into geographically-based linguistic groups, and the following four linguistic groups were postulated for Austronesian: from Indonesian in the west, via Melanesian on New Guinea, to Micronesian and Polynesian in Oceania.[8] Dempwolff (1934-38) placed the geographical line between 'Indonesian' versus 'Melanesian' languages along the western side of New Guinea. However, Capell (1944-45:19-20) observed that the characteristics of the languages of Timor would place them in the Melanesian group, and therefore agreed with Friederici (1913) that a more westerly line of demarcation, running between Sumba and Timor, across Flores, and between Sulawesi and

---

[6] In subsequent decades, word order as a basis for classifying geneological subgroups became controversial (Jonker 1914, Cowan 1952).

[7] Grimes (1991: 287, 495-506) suggests that the 'reversed Genitive' order in Austronesian languages is due to contact with non-Austronesian languages of the area.

[8] The link between linguistic grouping and geographical location was already argued by Otto Dempfwolff (1934) to be linguistically untenable. He demonstrated that the languages of Melanesia, Polynesia and Micronesia are members of a single subgroup, known today as the Oceanic subgroup (see Tryon 1995: 20).

Maluku, would be more appropriate. In later years, Dyen (1965), Haudricourt (1965) and Dahl (1976) seemed to group the languages of Maluku with the languages of the west rather than with those of Melanesia (see also Ross (1995a) and Grimes (2000) for additional details and references on the history of Austronesian studies).

Through the 1970s and 1980s, linguistic studies in East Nusantara were mainly concerned with subgrouping the languages of Maluku.[9] However, the overall linguistic situation of East Nusantara with its approximately 400 languages remained rather understudied until well into the 1990s. This led to Darrell Tryon's observation that '[Eastern Indonesia] remains perhaps the least known area in the Austronesian world today' (Tryon 1995:6).

Since the beginning of the 1990's, this situation has started to change with the publication of several grammars of East Nusantara languages, including Grimes (1991), van Minde (1997), Klamer (1998), van Klinken (1999), Dol (1999), van Staden (2000), Bowden (2001), Williams-van Klinken et. al. (2001), Baird (2002, 2008), De Vries (2007), van Engelenhoven (2005), van den Heuvel (2006), Kratochvíl (2007), Klamer (2010). An additional 15 languages are currently being described and/or documented by scholars in various research projects, including the following (moving roughly west to east): Rongga (Arka) and Palu'e (Donohue) in Flores, Helong (Bowden) and Bahasa Kupang (Jacob) in West Timor; Western Pantar (Holton) and Kaera (Klamer) in East Pantar; Sawila (Kratochvíl) in East Alor; Waima'a (Bowden, Hajek and Himmelmann), Makalero (Huber), Bunaq (Schapper) and Fataluku (Stoel, van Engelenhoven) in East Timor; and Allang (Ewing), Haruku (Florey) and Sou Amana Teru (Musgrave) in Central Maluku.

2.3 Genetic divisions in East Nusantara

2.3.1 Grouping the Austronesian languages of East Nusantara

The history of the Austronesian family (in particular the Oceanic subgroup) is now quite well understood, and an acceptable correlation of archaelogical and linguistic events has been achieved for much of its history (see Blust 1995b, Ross 1995, Bellwood 1997, Kirch 1997, Pawley 2002, Ross 2005, Reesink 2005 for overviews and references). In this section we focus on what is currently known about the genetic (or genealogical) divisions in East Nusantara.

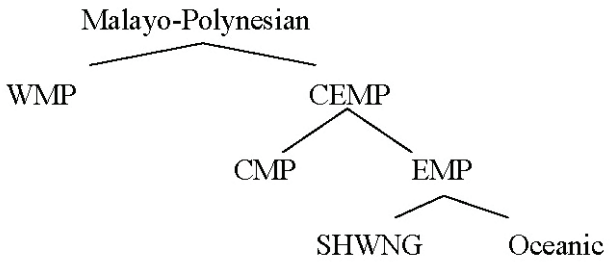
Blust (1993) was the first attempt to study the overall relationships of the Austronesian (AN) languages spoken in the Maluku and East Nusa Tenggara. This lead to the classification of the members of the subgroup of Central Malayo-Polynesian (CMP) languages as part of the genetic tree of the Austronesian languages spoken in Indonesia and East Timor, represented in (3).[10]

---

[9] This may have been due to the existence of a body of grammatical descriptions of a variety of Halmaheran and Central/South Moluccan languages written by (mainly) Dutch missionaries in the last quarter of the 19th C and the first quarter of the 20th C; see for example the literature overview in van Staden (2000), Holton (2003).

[10] WMP=Western Malayo-Polynesian, CEMP=Central-Eastern Malayo-Polynesian, CMP = Central Malayo-Polynesian, SHWNG= South Halmahera-West New Guinea.

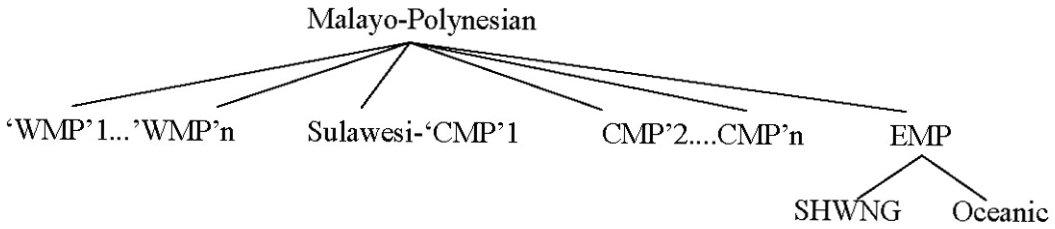
(3)



From its conception, the CMP subgroup has been considered problematic, because innovations that convincingly group the the CMP languages together to the exclusion of others, are lacking (Blust 1993, Ross 1995). Blust (1993) assumes that the patchy distribution of innovations is due to the fact that the CMP or CEMP languages are descendants of a chain of distinct dialects. For more than a decade, the existence of CMP or CEMP languages have been quietly accepted, although scholars working on individual languages of the area (for example Bowden 2001:12) have questioned its validity, pointing to the paucity of data available to Blust at the time.

In a recent paper, Donohue and Grimes (2008) take issue with the view that there has been a 'CMP' or 'CEMP' linkage. After a careful re-examination of the phonological and semantic features that Blust (1993) proposed as innovations defining CEMP and CMP, they conclude that these innovations are not exclusive to the languages in the CMP- or CEMP-area. While they agree with Blust that a linkage explains the patchy distributions of the innovations, they suggest that this linkage is much larger than the one proposed in Blust (1993) and also includes a large number of Austronesian languages in the WMP area (in particular some languages of Sulawesi), as well as languages from Formosan areas. In other words, they advocate a subgrouping as in (4) (Donohue and Grimes 2008:116).

(4)



Their conclusion is that the linguistic macro-history of eastern Indonesia, where Blust's WMP/CEMP border is said to be found, requires much more detailed investigation – a conclusion that most people working in the area will subscribe to. In fact, any further discussion of the status of CMP or CEMP may be impossible until more detailed bottom-up subgroupings have been proposed, using the detailed materials on the (putative) CMP- and EMP-languages that have become available during recent years. Such subgroupings would also have to take into account the complex role of diffusion through language contact between non-Austronesian and Austronesian speaking people in East Nusantara (see §2.1 above).

2.3.2 Grouping the Papuan languages of East Nusantara

The term 'Papuan' is generally used as a cover term for the perhaps 800 languages spoken in New Guinea and its vicinity that are not Austronesian (Ross 2005:15), and it is considered synonymous with 'non-Austronesian' (NAN). The label 'Papuan' says nothing about the genealogical ties between the languages.

Papuan languages are both lexically and morphosyntactically a highly heterogeneous group, and, due to lack of shared vocabularies, the familiar methods of lexical comparison are hard or impossible to apply in comparative studies of these languages (for discussion and references, see Foley 1986, 2000). This in itself is not a surprise, since most successful reconstructions in other language families go back only as far as approximately 6,000 to maximally 10,000 years (Nichols 1998:128), and have benefited from both archaeological and historical linguistic evidence. By contrast, the language(s) from which the present-day Papuan languages descend may have been present in East Nusantara for some 40,000 years. This is far too long ago to apply the comparative method.

The location, diversity and associated archaeology of the Papuan languages suggest that they have generally been in situ much longer than the Austronesian languages (see Bellwood 1997, Pawley 2005, Ross 2005 and the discussion in §2.1). Within the heterogeneous group of Papuan languages, various geneaological units have been suggested. Wurm (1982) proposed five major phyla of 'Papuan' languages, as well as six minor ones and a number of isolates. More conservative estimates (for example Foley 1986) suggest that there are at least 60 different families (some consisting of only a few members or even isolates) for which genealogical ties cannot be established yet. The largest family of Papuan languages for which there is general agreement is the Trans New Guinea (TNG) family, with about 300 languages (Ross 1995b, 2005). With two million speakers, this family comprises about half the Papuan speaking population (Foley 2000:363), but it represents only a tiny fraction of the genealogical variation found in Papua.

Within East Nusantara as defined for this volume, a conservative estimate gives five distinct families of Papuan languages, as follows.[11]

The Bird's Head has three families as well as three isolates. The three families are: East Bird's Head family (Voorhoeve 1975, Reesink 2002, 2005), West Bird's Head family (Voorhoeve 1987), Hatam and (extinct) Mansim (Reesink 2002, 2005). The three isolates are: Mpur (Odé 2002a,b), Maybrat (Dol 1999), Abun (Berry & Berry 1999)

North Maluku contains one family, the North Halmahera family, with four subgroups/languages (Voorhoeve 1987, 1989): Galela, Tobelo (Holton 2003), Pagu; Sahu; Tidore (van Staden 2000), Ternate; West Makian.

And finally, the Trans New Guinea family in East Nusantara includes the South Bird's Head, with Inanwatan (Voorhoeve 1975, Wurm 1982, Berry and Berry 1987, De Vries 2004); the West Trans New Guinea linkage with West Timor (=Bunaq)-Alor-Pantar; East Timor; West Bomberai; Wissel Lakes; and Dani (Ross 2005).

It is outside the scope of this introduction to present the motivations for all these language groupings (but see the references cited). The grouping that is perhaps the least clearly motivated is that of the Trans New Ginea family in Nusantara (Chapter 5) and we will therefore elaborate on its motivation here. The main evidence for affiliating the Trans New Guinea languages in New Guinea with those spoken in Timor-Alor-Pantar is the

---

[11] This list is from Klamer et. al. (2008), see also the discussion and references cited there.

pronominal evidence presented in Ross (2005:35-36), summarised below. The West TNG group is labelled a 'linkage' (in the sense of Ross 1988:9-11) because it is assumed to have resulted from the gradual diversification of (part of) an earlier dialect chain, and not from a discrete proto-language. A linkage is characterised by a pattern of overlapping innovations, but in the languages at hand there are extremely few such overlaps. The languages of the Dani, Wissel Lakes, West Bomberai and East Timor microgroups all reflect an innovation whereby **ani* 'I' has replaced pTNG **na*. But this innovation is not reflected in the West Timor-Alor-Pantar microgroup. On the other hand, the West Bomberai, East Timor and West Timor-Alor-Pantar microgroups all reflect an innovative form **bi* 'we' (Ross 2005:36). It is this overlapping pattern that might suggest the connection of a dialect chain.

A few cognates of the proto-TNG pronouns occur in Klon (West Alor, Baird 2008), Adang (West Alor, Haan 2001), Abui (Central-West Alor, Kratochvíl 2007), Teiwa (Klamer 2010) and Western Pantar (Holton 2007 and this volume), but the evidence is thin.[12] More detailed bottom-up reconstructions of Alor-Pantar language groups are needed before any higher level affiliation can be proposed with more certainty. In sum, if and how the non-Austronesian languages of East Nusantara are affiliated to those on the New Guinea mainland is still an unsettled issue.

2.4 Typological divisions in East Nusantara

2.4.1 The typology of Austronesian languages in East Nusantara

In the past, typological characterisations of Austronesian languages either concerned the characteristics of Western Austronesian versus Oceanic languages (for example, Clark 1990 and Tryon 1995), or the characteristics of the Austronesian languages spoken on New Guinea in contrast to the Papuan languages in their vicinity (for example, Voorhoeve 1994, Ross 1996, Foley 1998). The main reason for not considering the typological features of East Nusantara languages as such was lack of data.[13]

Grammars of East Nusantara languages published in the 1990s were the main source for the initial list of typological features proposed by Klamer (2002) to characterise the Austronesian languages of the Central/Eastern Indonesia.[14] Features proposed in that paper were scrutinised, and further debated in Donohue (2004) and Klamer (2004), or were shown to be inadequate characterisations in related, or subsequent research (for example Himmelmann 2005, Klamer et. al. 2008, Florey, this volume). Some features however survived, and are listed in an updated, cumulative list in (6) below.

Himmelmann (2005) is based on an impressive amount of data from a wide range of non-Oceanic Austronesian languages. His proposal is to divide these languages into two major typological groups: one group of 'symmetrical voice' languages, which include the

---

[12] The lexical evidence for assigning the West Timor-Alor-Pantar languages to the TNG family is also weak (see Pawley 1998: 683; 2001; Klamer et. al., 2008). Pawley (2001, 2005) contains about 200 reconstructed proto-TNG forms.

[13] Other references on typological differences between Papuan and Austronesian languages include Ross (2001), who discusses the contact between Papuan and the Oceanic languages in North West Melanesia, and Dunn et. al. (2005), who contrast Oceanic languages and Papuan languages spoken in the east of mainland New Guinea and the islands extending east to the middle of the Solomons.

[14] That paper dealt with the geographical area east of Lombok and west of Papua, up to and including south-eastern Sulawesi.

Phillipine-type and Indonesian-type languages, and which are predominantly found in western AN languages; and another group of 'preposed possessor' languages (referred to as Gen-N 'Genitive-Noun' below), including the Austronesian languages of Timor, Maluku and West Papua as well as the Pidgin-Derived Malay varieties (Himmelmann 2005:113). His two typological groups contrast on the following features:

(5) Two major typological groups in the non-Oceanic Austronesian languages (Himmelmann 2005:175).

Symmetrical voice languages	**Preposed possessor languages**
Symmetrical voice alternations	No or asymmetrical voice alternations
N-Gen in adnominal constructions	Gen-N in adnominal constructions[15]
No morphosyntactic distinction between alienably/inalienably possessed items	Morphosyntactic distinction between alienably/inalienably possessed items
Few/no differences between narrative and equational clauses	Clear-cut differences between narrative and equational clauses
Person marking only sporadically attested	Person marking prefixes or proclitics for S/A arguments
Numerals/quantifiers precede head	Numerals/quantifiers follow head
Negators in pre-predicate position	Clause-final negators
V-initial or SVX	V-second or -final

While Himmelmann (2005) contrasts the languages from the north-western part of the archipelago with those spoken in the south-east, Donohue (2007) distinguishes Northern, Southern, Western and Eastern groups in his typology of Austronesian word order characteristics. For present purposes we are only interested in his Eastern group (which includes the Austronesian languages as far east as eastern mainland Papua New Guinea). In this group we find the following constituent orders: Gen-N, N-Numeral, Verb-Object, and Subject-Verb. The first two of these mark the line dividing the Western from the Eastern group (Donohue 2007:381). In the Eastern group, the order of nominal modifiers/specifiers with respect to the noun shows mixed patterns: adjectives, demonstratives, relative clauses, numerals and adpositions may either precede or follow nouns. There is general agreement that this may reflect various degrees of substratal influence from Papuan languages (Grimes 1991, Voorhoeve 1994, Ross 2001, Klamer et. al. 2008).

As a summary, we present a list of features that are found in many of the Austronesian languages of East Nusantara in (6):

(6) Features characterizing Austronesian languages in East Nusantara (Klamer 2002, Himmelmann 2005, Donohue 2007, Klamer et. al. 2008).

[15] The Gen-N criterion refers to the most common or unmarked order found in possessive constructions. That is, it is not required that all possessive constructions in a preposed possessor language show the order Gen-N, and conversely, non-preposed possessor languages may optionally allow a Gen-N order.

Phonology	Prenasalised consonants
	Roots are generally CVCV
	- dispreference for homorganic consonant clusters
	- dispreference for closed syllables, creation of open syllables
	Metathesis
Morphology	No productive voice system on verbs
	Agent/subject indexed on verb as prefix/proclitic
	Morphological distinction between alienable/inalienable nouns
	Left-headed compounds
	Inclusive/exclusive distinction in pronouns
Syntax	Verb-Object order
	Prepositions
	Gen-Noun
	Noun-Numeral order
	Clause-final negators[16]
	Clause-initial indigenous complementisers
	Absence of a passive construction
	Formally marked adverbial/complement clauses
Other	Parallelisms without stylistic optionality

Note that some of these features also occur in the Papuan languages of East Nusantara, (compare (7) below), which (again) points to the complex role that diffusion – as result of contact between Autronesian and Papuan peoples and their languages in the past – has played in shaping the languages of East Nusantara. It also implies that the features in (6) should not be used to define genealogical distinctions between Austronesian and non-Austronesian languages (see Ross 2003, Klamer 2003).

2.4.2. The typology of Papuan languages

The Papuan languages share a number of characteristics, of which Foley (2000) and Aikhenvald and Stebbins (2007) give recent overviews. Without intending to be exhaustive, we list only a few of the more general points here.

The great majority of Papuan languages have only a single liquid phoneme (while in the Austronesian languages, by contrast, a phonemic distinction between /r/ and /l/ is virtually universal). Papuan languages exhibit sophisticated noun classification systems, and commonly mark gender (Foley 2000:371), but case marking is less common. Most Papuan languages have at least one bound pronominal for subjects, and this may be a prefix or a suffix, although it usually is a suffix (Foley 2000:377). Syntactically, Papuan languages are overwhelmingly head-final, with OV constituent order, final negations, final conjunctions, and postpositions. Also typical is clause chaining, often with some

[16] Florey (this volume) questions whether this is a characteristic typological feature of preposed possessor languages as per Himmelmann (2005). We still include the feature here because, while not all Austronesian languages of East Nusantara have clause final negation, it is a cross-linguistically uncommon, non-Austronesian feature and if found in an Austronesian language suggests the language is from East Nusantara. (Note that we do not imply to say that the feature is unique for East Nusantara: in Oceanic languages 'We also find a large number of languages where the grammaticalised negator is clause-final...' (Lynch, Ross & Crowley 2002: 88)).

concomitant switch reference system, and a morphological contrast between 'medial' and 'final' verbs (Pawley 2005:91). Many Papuan languages make extensive use of serial verb constructions (Foley 2000:385, Aikhenvald and Stebbins 2007:252-253, and the references cited there), clause chaining, switch reference systems, and/or a formal distinction between 'medial' and 'final' clauses. These Papuan features are summarised in (7):

(7) Typical features of Papuan languages
(Foley 1986, 2000, Pawley 2005, Aikenvald and Stebbins 2007)

Phonology	No distinction between *r* and *l*
Morphology	Marking of gender
	Subject marked as suffix on verb
	No inclusive/exclusive distinction in the pronominal paradigm
	Morphological distinction between alienable and inalienable nouns
Syntax	Object-Verb
	Subject-Verb
	Postpositions
	Gen-Noun
	Clause-final negators
	Clause-final conjunctions
	Clause-chaining, switch reference, medial vs. final verbs
	Serial verb constructions

To what extent do the Papuan languages of the islands in East Nusantara pattern like the languages of mainland Papua? The Papuan languages of Timor, Alor and Pantar and North Halmahera (among others, Tobelo, Pagu, Galela; see Holton 2003:2-3, and the references cited there) share the general head-final character of Papuan languages: they generally have OV as the unmarked word order, have post-predicate negations, and often their indigenous conjunctions are clause-final. Other Papuan features found in the languages of East Nusantara are the Gen-N order,[17] and the distinction between alienable and inalienable possession – the latter is absent in some North Halmaheran languages. Finally, the East Nusantara Papuan languages also have rich arrays of serial verbs.

However, these languages differ from the Papuan features in (7) in that they have a phonemic *r*/*l* contrast, little derivational morphology, and no adpositions – or just one or two that are cognate to verbs in serial verb constructions (see Baird this volume, Klamer this volume). The most westerly outliers, such as Klon, Abui, Adang, Teiwa and Kaera have no clause chaining or switch-reference system and no morphological contrast between medial and final verbs. Nouns do not inflect for number, gender or case. Thus in quite a number of features these outliers do not follow the Papuan characteristics listed in (7).

Sahu, Ternate, Tidore and West Makian in Halmahera have even more Austronesian features than the other Papuan languages of East Nusantara; for example, they are verb-medial, and have prepositions (see van Staden 2000:19, 22).

Most of the Papuan languages in East Nusantara have an inclusive-exclusive opposition in the first person plural. Such a pronominal distinction is a general feature of Austronesian languages, reconstructed even for Proto-Austronesian, and is not generally

[17] Western Pantar has both orders: Gen-N and N-Gen (Gary Holton, p.c.).

found in Papuan languages spoken in the interior of New Guinea. It seems that the inclusive/exclusive distinction in the pronominal paradigm of languages of East Nusantara is therefore the most noticeable Austronesian feature that diffused into these languages (Klamer et. al. 2008).

In general, we might say that the lexicon and morpho-syntax of Papuan languages in East Nusantara are rather unlike those of mainland New Guinea. Why this is so we do not know, but the following may be an explanation. Regarding time-depth, we know that many of the Papuan languages in East Nusantara have been separated from their mainland relatives for at least 6,000 years (see §1.2 and Pawley 2005). This is a period long enough to allow many autonomous developments to take place (indeed, to develop a new language family), so it is only to be expected that the separation resulted in different lexicons and typological profiles for the Papuan languages in East Nusantara and those of the mainland.

2.5. Is East Nusantara a linguistic area?

Given the long-standing contact between Austronesian and non-Austronesian speaking communities in East Nusantara explained in the sections above, can we say that it is a linguistic area? A linguistic area (or *Sprachbund*, Trubetzkoy 1928) is:

> '...a geographical region containing a group of three or more languages that share some structural features as a result of contact rather than as a result of accidence or inheritance from a common ancestor.' (Thomason 2001:99).

Put in a different way:

> 'The term linguistic area refers to a geographical area in which, due to borrowing and language contact, languages of a region come to share certain structural features.' (Campbell 1998:299-300)

It is undoubtedly the case that many (if not all) present-day languages spoken in East Nusantara have experienced influence from other languages in the area through various contact situations, both past and present, resulting in the diffusion of features. In Halmahera, for example, Tidore has a mainly non-Austronesian lexicon, and is therefore classified as such, but it has an 'impressive' number of Austronesian grammatical features as well (van Staden 2000:24). On the other hand, in Alor and Pantar the Austronesian language Bahasa Alor (or Alorese) has an Austronesian lexicon, but shares an interesting set of morpho-syntactic features with (substrate) Papuan languages that today are spoken in its surroundings (Klamer forthcoming).

The Bird's Head in Papua is another area where languages share grammatical structures but have wildly different vocabularies. This is why adequate cognate sets cannot be constructed, and applying the comparative method is virtually impossible when comparing these languages (see also Voorhoeve 1987, Foley 1998, Reesink 1996, 1998, 2002).

In East Nusantara, diffusion thus takes place across genetic boundaries, as well as within them. Some of the features that are typically found in today's Austronesian languages of East Nusantara (such as the 'preposed possessor' Gen-N construction and the distinct marking of alienable and inalienable possession), are supposed to have derived from earlier substrate Papuan languages (Reesink 2005, Klamer et. al. 2008), while there is also evidence that Austronesian languages have influenced the structure of Papuan languages, for example in marking a distinction between inclusive and exclusive pronouns. This is the type of evidence presented in Klamer et. al. (2008) to argue that East Nusantara

is indeed a 'linguistic area': an area that would include Halmahera and the Bird's Head as its core, and which radiates outwards to first include the Maluku and Alor-Pantar, and then Timor. This area was formed by several waves of diffusion, taking place at different points in time and going in various directions.

However, since the research for that paper took place (roughly between 2002-2004), general research in areal typology has made it more and more clear that there are severe conceptual and methodological problems with the notion of a 'linguistic area' in the definitions of Thomason and Campbell given above. We will mention a few of these problems here, and refer to Muysken (2008) for a more complete summary. One major problem appears to be the issue of how the features that are relevant in defining an area are selected – why is one feature considered to be definitive for the area, and not another? Another problem relates to the language sample: how are the languages that are being compared sampled from the total set of languages spoken in the area? It is often the case, as in East Nusantara, that random selection is not possible due to lack of sources, so that the sample is determined on other bases, such as the availability of printed records and the experience or knowledge of the researchers doing the research. This, of course, results in a sample that will always be biased in some way. Finally, how do we know which clusters of features are significant in establishing a linguistic area, and which clusters are not? In other words, how do we evaluate or weigh the relevance of certain clusters of features? Again, this often depends on the subjective choice of the researchers involved. And finally, one would like to have historical research confirming that there was contact in the linguistic area proposed. But in the case of East Nusantara, written records of the history of the islands (especially of the period before European colonisation) are virtually absent, and records of language stages preceding those of today are often missing completely.

Because the process of defining a linguistic area is generally problematic, and because written historical records of East Nusantara islands are virtually absent, the exact characteristics as well as the boundaries of this area will probably remain elusive. Nonetheless, we believe that there is evidence that suggests that the linguistic contact zone in East Nusantara is a linguistic area – with Halmahera and the Bird's Head as its core and radiating outwards to Maluku, Alor-Pantar, and Timor. The research presented in this volume then adds to the growing data base of linguistic knowledge about languages in this region and contributes to our developing understanding of typological trends in the area.

3. Summary of the chapters in this volume

The chapters in this volume cover a range of topics including phonology, alignment systems and argument encoding, serial verb constructions, and negation. Two additional chapters look at a broader range of linguistic features within individual languages, thus highlighting the systematic way that these features interact. At the same time, several overarching themes cut across these individual contributions, making each of them relevant to issues raised in the preceding discussion of East Nusantara as a (possible) linguistic area. Among these is the recurring issue of diffusion and influence between Austronesian and non-Austronesian languages in the area. As laid out above, there is little doubt as to the long and ongoing contact between speakers of these different languages groups, yet the extent to which this interaction influences language structures is not always self-evident and is often in need of closer scrutiny. Another recurring theme is the need for closer examination of the data used for making typological claims, both in terms of

quantity and quality. Many of the chapters in this volume contribute to the ongoing development of typological analyses by either tapping into data from an expanded number of languages or by taking a more detailed approach to analyses of particular phenomena, or both, in order to question or refine typological claims that have been made in the past.

Chapters in this volume take up this debate by providing, in some cases, detailed analyses based on individual languages, and in other cases by providing broader comparative studies. This volume is particularly rich in previously unpublished new data from recent fieldwork, and thus expands our understanding of the typological diversity, stability and spread of linguistic features in this geographically delimited area. Indeed one of the recurrent themes across many of the chapters presented in this volume is the need for a greatly expanded database, including more detailed documentation and analysis of a much wider range of languages, in order to develop more accurate typological conclusions.

In the area of phonology, **Hajek**'s contribution presents a typology of the vowel and consonant systems of East Nusantara, based on a sample of 70 languages from four distinct families, including both Austronesian and Non-Austronesian languages. It shows that many of the typological features of East Nusantara have not been observed in Maddieson's (2005a-g) overviews, which while including a large sample covering a very wide area, nonetheless included very few languages from East Nusantara. This demonstrates the importance of a detailed sampling for a specific region in conjunction with higher-level typological investigations. As well as identifying a number of typologically unusual phenomena found in certain East Nusantara languages, Hajek also highlights the importance of language contact with Malay/Indonesian and with Portuguese in East Timor. The effect of borrowing on the segment inventories of some languages is still limited, while for others, it is already potentially enormous.

Himmelmann focuses on a single language, presenting new data on intonation in Waima'a, an Austronesian language spoken on the Northeast coast of Timor island. A striking feature of Waima'a intonation, compared, for example, to the better known intonational structures in European languages, is the lack of an accentual or prominence-lending tone. The lack of accentual tone is a feature which has been reported for other languages in the larger region, including Javanese and certain varieties of Malay. Similar to these languages, Waima'a also appears to lack lexical accents. Languages that lack both accentual tone and lexical accent have for some time been treated marginally, but there is growing evidence that such languages are more common in the world than previously recognised. Tadmor (2000, 2001) in fact speculates that lack of lexical accent is a widespread feature of languages in western Indonesia (extending, roughly, from Sumatra to Bali and including Kalimantan), while languages in the East Nusantara often have regular penultimate lexical accents. The case of Waima'a suggests that the western pattern may occur at least as far east as East Timor.

The first chapters to examine morpho-syntax focus on alignment systems and argument encoding. These comprises five chapters: case studies of particular languages as well as chapters with a comparative perspective. Each chapter explores how argument-marking systems align with syntactic, semantic and/or discourse features of the language. These explorations are dynamic in that they look at changing systems and examine motivations for the patterns that are observed. **Reesink** conducts a survey of coding strategies for nominal arguments (subjects, objects and possessors) in a heterogeneous set of non-Austronesian languages of western Papua. He identifies three typologically defined subgroups, each containing both related and unrelated languages. The first group is typologically unusual in that these languages have object prefixes, while subjects can only

be expressed by independent forms. In these languages object prefixes also encode the single argument of involuntary predicates and are identical with possessor prefixes. Languages of the second subgroup have both subject and object prefixes and also display a correspondence between object marking and inalienable possessor affixes. Additionally, for languages of this second subgroup that have a split-S system, there is a correspondence between object prefixes and the marking of the single argument of uncontrolled intransitive verbs. Unlike the first two subgroups, which have SOV basic word order, the third subgroup involves SVO languages that have only subject affixation, with a tendency for these subject affixes to corresponding to affixes denoting inalienable possession. These findings suggest that the West Papuan region may not be merely geographical as first put forward, but may in fact imply the existence of a linguistic area, showing similarities with the preposed-possessor Austronesian languages in the region.

The remaining chapters on alignment and argument structure each present a detailed analysis from a particular language. Recurring issues correspond with those that arise in Reesink's comparative study, including grammatical alignment in the marking of core arguments within the clause and its implication for other aspects of the grammar, such as the presence or absence of voice alterations and the relationship with possessive paradigms. The complex pronominal affix systems of the non-Austronesian languages of Alor-Pantar can be particularly hard to characterise, and **Holton** looks specifically at the previously undescribed Western Pantar. This language has two distinct but overlapping systems of pronominals which instantiate grammatical alignment through the choice of independent pronoun and pronominal prefix on the verb. The system of full pronouns is largely semantically motived, producing an agentive system, while the pronominal prefixes may index any argument role, with different semantic word classes exhibiting a different alignment patterning. The choice of full and prefixed forms is complex and the mapping of grammatical roles onto pronominal prefixes yields seven verb classes, based on complex interaction of person, transitivity, grammatical role and obligatority of prefixes. Holton shows that a concerted attempt to apply traditional notions of alignment – such as nominative-accusative or ergative-absolutive – proves unhelpful in these languages, and he proposes a semantic analysis in which the distribution of argument forms aligns with different semantic classes of verbs.

Ewing presents a discussion of agentive alignment in Allang, an Austronesian language of Central Maluku. The high number of undergoer intransitives (compared to closely related languages) and valency-changing mechanisms that produce undergoer-intransitive clauses suggests that agentive alignment is well integrated into the Allang grammatical system. This discussion is followed by a survey of agentive alignment in other Central Maluku languages. The similarities in alignment systems in the languages examined do not correspond to the genetic relationships known to hold between these languages. While agentive systems have been observed in many languages of East Nusantara (and this has been suggested as an areal feature), the evidence from languages of Central Maluku supports recent suggestions in the literature that agentive alignment systems should not be used as diagnostic of genetic or areal relationships, as they seem to arise spontaneously and easily (Mithun 2008, Klamer 2006, 2008).

Musgrave investigates the use of reduced pronouns as argument markers in Sou Amana Teru, another Austronesian language of Central Maluku. These reduced forms are closely related to full pronouns, and both full and reduced forms can be used to indicate possessive and complements of prepositions, in addition to being argument markers, the function examined here. As argument markers, these reduced forms occur before the verb in the

role of A or S and as an enclitic on the verb in the role of O. The use of reduced bound forms for A and S together with full forms has the appearance of a morphological cross-referencing system. However, because these reduced forms can either be procliticised on verbs or encliticised on pre-verbal material in the verb phrase, and because their realisation is phonologically determined, the resulting constructions are syntactic phrases rather than morphological constituents. Musgrave hypothesises that Sou Amana Teru is moving towards becoming a more analytic system. This process appears to be, at least in part, due to the ongoing process of language shift in which younger speakers are reanalysing many aspects of the grammar under the influence of Indonesian/Malay, currently their dominant language.

The interplay between alignment, grammatical relations and the presence of diathesis is central to **Williams-van Klinken's** study of contemporary Tetun Dili; an especially interesting language because it is undergoing dramatic changes due to its rapidly expanding domains of use and due to influences of other languages. This chapter examines syntactic transitivity, which can be difficult to establish in Tetun due to a lack of morphology marking changes in valency, the fact that understood arguments can be left unexpressed and the flexibility of word order. Thus, when a clause contains a verb which implies two arguments (that is semantically transitive), and one of the arguments is not explicitly expressed, should this be considered syntactically transitive or intransitive? In her data Williams-van Klinken identifies a subset of verbs that are semantically transitive, but regularly occur with only one argument. She then uses a number of tests to see whether the question of syntactic transitivity is answerable. She identifies a subset of verbs for which an intransitive analysis is appropriate, while another subset of verbs are considered transitive with a fronted O argument. Williams-van Klinken also goes on to question whether O-fronted constructions might be analysed as a type of passive construction, most likely a result of influence form Portuguese and/or Indonesian.

Two chapters on serial verb constructions (SVCs) include case studies from three Papuan languages: Klon, spoken in western Alor, Teiwa, from north-western Pantar, and Kaera, from north-east Pantar. Both chapters describe grammatical processes of reanalysis by which verbs in serial constructions become grammatical morphemes. **Baird's** chapter presents new data on SVCs in Klon, using Aikhenvald's (2006) notions of symmetrical and asymmetrical SVCs. She identifies eight classes of serial verb constructions: three symmetrical, in which both verbs are from a non-restricted open class, and five asymmetrical ones, in which one of the verbs is from a restricted class and modifies the other, non-restricted verb. The symmetrical SVCs either convey sequences of events, describing the manner in which something is done, or are a kind of lexicalised parallelism. In the discussion of Klon asymmetrical SVCs, Baird pays special attention to the grammaticalisation processes involved. Some of these are aspectual in nature, while others – involving the verb *mi* 'to be at, place' – form locative or temporal constructions. In all cases it is the syntactic positioning and semantic structure of the verbs that have enabled a reanalysis into grammatical items.

In her chapter on Teiwa and Kaera serial verb constructions, **Klamer** discusses probable cognates of Klon *mi*: the oblique marker *mi* in Kaera, which is possibly historically derived from the transitive location verb *ming* in this language, and the transitive location verb *me'* in Teiwa. After presenting a sketch of the grammatical structures and the argument-encoding properties of Kaera and Teiwa, the chapter focuses on the analysis of the multi-functional deictic verb *ma* 'come (here)' in these two languages. In Teiwa and Kaera, *ma* is used as a deictic verb, a change-of-state verb, and to

mark intentions and imperatives. In Teiwa, but not in Kaera, it is also used as an oblique marker. Klamer argues that the different functions are all contextualised meanings rather than lexical meanings; that is, the interpretation of *ma* shifts with its grammatical context but retains a unitary semantic core. Comparing the functions of *ma* in Teiwa and Kaera, Klamer finds that in both languages *ma* functions to mark 'movement in time', but only in Teiwa does *ma* function as an oblique marker; Kaera marks obliques with *mi,* the cognate of Klon *mi* 'to be at, to place', discussed in Baird (this volume).

Negation is another linguistic domain that has been prominent in typological discussions of the languages of East Nusantara. **Florey** presents data from eight related Austronesian (CMP) languages of Maluku, which help to expand our understanding of the range of negation types in this region. Clause-final negation is an uncommon type among languages of the world, but it occurs frequently in languages of East Nusantara and Western Papua. Florey provides additional data from a cluster of languages in Maluku, examining word order and functional range of different negators. She finds that both cross-linguistically and language-internally, these Moluccan languages exhibit a number of different negative constructions, including pre-predicate negation, post-predicate and clause-final negation, and 'embracing' negation. This analysis suggests that an investigation of negation focused solely on primary negators is not sufficiently nuanced and that further evidence concerning clause-final negation may be found through a cross-linguistic examination of complex negatives. These data indicate that not all Moluccan languages have final negations (cf. Reesink 2002), and that it may not be a characteristic typological feature of preposed possessor languages (Himmelmann 2005) (but see our footnote 17). The interaction between negatives and other clausal operators, such as aspect and mood, is also explored. These languages discussed by Florey exhibit a peference for clause-final modifiers, similar to that noted in Magey Matbat in the contribution by Remijsen, and the interaction between these particles and negation may be one reason that pre-verbal negation has been attracted to clause-final position in these languages.

The volume closes with two chapters that present broader descriptions of various subsystems within the grammars of three different languages in the region. **Grimes** asks whether Austronesian Hawu and Dhoa (CMP) – spoken on three small islands in the Sabu Sea to the west of Timor Island – are dialects or separate languages, using new data and new analyses, thus picking up on several features not yet described by other researchers. Earlier observers looked at similarities in phonologies and lexicon, and while some concluded these were dialects of one language, others claimed they were separate languages. Grimes goes beyond by comparing entire subsystems of the two languages – including the sound systems, the systems of personal, spatial, temporal and referential deixis, the negation systems, verbal inflection patterns, the shape and structural properties of adverbials and prepositions, basic clausal syntax, phrase structure, interclausal relations, and question words – concluding that Hawu and Dhao represent typologically quite distinct languages. This exercise suggests that it can be necessary to compare functors and grammatical subsystems in their entirety to get a realistic picture of how great the differences are between two languages.

Remijsen presents a description of the morphological and syntactic behaviour of nouns and verbs in Magey Matbat, an Austronesian (SHWNG) language of Misol Island, West Irian Jaya Province. He pays particular attention to the typological features discussed by Himmelmann (2005), including possessive marking, numerical classifiers, tense aspect mood marking, verb serialisation and verb classes. As with many languages of East Nusantara, Magey Matbat has a robust system of alienable and inalienable marking in

possessive constructions. In all cases, the possessor precedes the possessum; inalienably possessed head nouns are additionally inflected with a suffix, while alienably possessed head nouns are not. In addition, various verbs with a meaning related to a physical or emotional state take subject inflections which are identical to inalienable possession markers. In closely related languages it remains controversial as to whether such constructions are best classed as verbs or nouns, but Remijsen makes the case that in Magey Matbat these forms are on a diachronic continuum towards becoming regular verbs, and seem to be further along this route than in other related languages. Similar undergoer intransitive constructions are mentioned in the contributions by Reesink, Musgrave and Ewing. All other verbs in Magey Matbat fall into four classes based on paradigms of onset consonants. As in the chapters by Baird and Klamer, Remijsen also discusses serial verb constructions in terms of function and possible grammaticalisation pathways. In Magey Matbat, SVCs take two different forms, co-lexicalised serialisations which produce a conceptual description of an event, and situations where an element in the serialisation has a more grammatical function. Some historical serialisations have fully grammaticalised as prepositions. At the clause level, it is noted that tense, aspect, modality, negation, and indeed most clause level modification – including politeness, questions, and commands – are marked with clause final particles. This preference for clause final marking is found in a number of languages in the area, and also forms part of the survey of negation in Central Maluku languages found in the chapter by Florey. Addressing the issue of genetic affiliation, Remijsen finds that the typological features of Magey Matbat have clear similarities with Austronesian languages in general, and with those of Central and Eastern Indonesia in particular. Yet similarities between Magey Matbat and the Papuan languages of the mainland of New Guinea also exist and Matbat is considered part of transitional area of shared Austronesian and Papuan characteristics.

While the overall organisation of this volume is driven by the themes outlined above, the chapters also cluster around issues of genetic and areal relationships. Chapters that deal explicitly with Non-Autronesian languages are those on Klon (Baird), Western Pantar (Holton), Teiwa and Kaera (Klamer), and additionally the survey of fifty West Papuan languages (Reesink). The remaining chapters focus on Austronesian languages, with the exception of Hajek, which is a typological survey covering both Austronesian and non-Austronesian languages. There is also an emphasis among these chapters on three regions of East Nusantara in particular. The chapters by Himmelmann and Williams-van Klinken focus on languages of Timor Island, including both Indonesian Timor and Timor-Leste. Languages of East Nusa Tenggara outside of Timor Island are represented by the chapters of Baird, Holton, Klamer and Grimes. Remijsen examines a single language of Misol Island in Papua, whereas Reesink's contribution looks at West Papuan languages more widely. A number of languages from Maluku, especially Central Maluku are discussed in the contributions by Ewing, Florey and Musgrave.

Finally, a concern for comparison is apparent through many of these chapters. This appears most strongly in the contributions that are framed in terms of typological comparisons. These include the chapter on phonology by Hajek, the chapter on alignment systems by Reesink, and the chapter on negation by Florey. Klamer and Grimes each look at pairs of closely related languages, asking questions which compare the development and the distribution of particular grammatical features: serial verb constructions in the case of Klamer and a range of grammatical subsystems in the case of Grimes. Ewing examines alignment in one particular language of Maluku, but places this within a broader survey, while the remaining chapters on alignment systems are all focused on single languages.

The strong focus on presenting new data from a range of previously under-documented languages in the region also provides valuable input for further comparative work. Taken together these chapters demonstrate the significance of East Nusantara as region of linguistic enquiry. At the same time, they highlight the importance of ongoing investigations, both empirical and theoretical, in this still under-documented linguistic region in order to help us continue to refine the notion of East Nusantara as linguistic area.

References

Aikhenvald, Alexandra Y., 2006. Serial Verb Constructions in a typological perspective. In R.M.W. Dixon, ed., *Serial Verb Constructions: A Cross-Linguistic Typology*, 1–87. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. and Tonya Stebbins, 2007. Languages of New Guinea. In O. Miyaoka, O. Sakiyama and M. Krauss, eds, *Vanishing Languages of the Pacific*, 239–266. Oxford: Oxford University Press.

Baird, Louise, 2002. A grammar of Kéo: An Austronesian language of East Nusantara, Canberra: The Australian National University. PhD dissertation,

— 2008. *A grammar of Klon: a Non-Austronesian language of Alor, Indonesia.* Canberra: Pacific Linguistics.

Bellwood, Peter, 1997. *The prehistory of the Indo-Pacific Archipelago* (2nd edition). Honolulu: University of Hawaii Press.

Berry, Kevin and Christine Berry, 1987. A Survey of the South Bird's Head Stock. *Workpapers in Indonesian languages and cultures* 4:81–117. Abepura: SIL/Department of Education and Culture, Percetakan Universitas Cenderawasih.

— 1999. *A Description of Abun, a West Papuan language of Irian Jaya*. Canberra: Pacific Linguistics.

Birdsell, Joseph B., 1977. The recalibration of a paradigm for the first peopling of greater Australia. In Jim Allen, Jack Golson and Rhys Jones, eds, *Sunda and Sahul: Prehistoric studies in Southeast Asia, Melanesia and Australia*, 113–167. New York: Academic Press.

Blust, Robert A, 1985. The Austronesian homeland: a linguistic perspective. *Asian Perspectives*, 26:45–67.

— 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics*, 32:241–293.

— 1995. The prehistory of the Austronesian speaking peoples: a view from language. *Journal of World Prehistory*, 9:453–510.

— 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In E. Zeitoun and P. Jen-Kuei Li, eds, *Selected papers from the Eight International Conference on Austronesian Linguistics*, 31–94. Tapei: Academia Sinica.

Bouman, M.A, 1943. De Aloreesche dansplaats. *Bijdragen Taal- Land- en Volkenkunde*, 102:481–500.

Bowden, John, 2001. *Taba, Description of a South Halmahera language*. Canberra: Pacific Linguistics.

Brandes, J.L.A, 1884. Bijdragen tot de vergelijkende klankleer der westerse afdeeling van de Maleisch-Polynesische taalfamilie. Utrecht: University of Utrecht. PhD dissertation.

Campbell, Lyle, 1998. *Historical linguistics: an introduction*. Edinburgh: Edinburgh University Press.

Capell, Arthur, 1944–45. Peoples and languages of Timor, Part Two of *Oceania*, 15:19–48.

Capelli, Christian, James Wilson, Michael Richards, Michael Stumpf, Fiona Gratrix, Stephen Oppenheimer, Peter Underhill, Vincenzo Pascali, Tsang-Ming Ko and David Goldstein, 2001. A predominantly indigenous paternal heritage for the Austronesian-Speaking peoples of insular Southeast Asia and Oceania. *American Journal of Human Genetics*, 68:432–443.

Clark, Ross, 1990. The Austronesian languages. In Bernard Comrie, ed., *The major languages of East and South East Asia*, 173–184. London: Routledge.

Cowan, Hendrik, 1952. Genitief constructie en Melanesische talen. *Indonesië*, 5:307–313.

Cribb, Robert, 2000. *Historical Atlas of Indonesia.* Honolulu: University of Hawaii Press.

Dahl, Otto C., 1976. *Proto–Austronesian*. Second revised edition. Oslo: Scandinavian Institute of Asian Studies Monograph Series 15.

Dempwolff, Otto, 1934–38. *Vergleichende Lautlehre des Austronesischen Wortschatzes* (3 volumes). *Beifhefte Zeitschrift für Eingeborenensprachen* 15, 17, 19. Berlin: Dietrich Reimer Verlag.

Dol, Philomena, 1999. A grammar of Maybrat, Leiden: Leiden University. PhD dissertation.

— 2007. Word order in Austronesian: from north to south and west to east. *Linguistic Typology* 11, 349–391.

Donohue, Mark and Søren Wichmann, eds, 2008. *The typology of semantic alignment*. Oxford: Oxford Universit Press.

Donohue, Mark, and Charles E. Grimes, 2008. Yet more on the position of the languages of eastern Indonesia. *Oceanic Linguistics* 47, 1:114–158.

Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley, and Stephen C. Levinson, 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science*, 309:2072–2075.

Dyen, Isidore, 1965. A lexicostatistical classification of the Austronesian languages. *International Journal of American Linguistics*, Memoire 19.

Engelenhoven, Aone van, 2005. *Leti, a language of Southwest Maluku*. Leiden: KITLV.

Foley, William .A, 1986. *The Papuan languages of New Guinea*. Cambridge: University of Cambridge Press.

— 1998. Towards understanding Papuan languages. In Jelle Miedema, C. Odé, and R.A.C. Dam, eds, *Perspectives on the Bird's Head of Irian Jaya, Indonesia*, 503–518. Amsterdam: Rodopi.

— 2000. The languages of New Guinea. *Annual Review of Anthropology*, 29:357–404.

Florey, Margaret, 2005. Language shift and endangerment. In K. Alexander Adelaar and Nikolaus P. Himmelmann, eds, *The Austronesian languages of Asia and Madagascar*, 43–64. London: Routledge.

Friederici, George, 1913. Wissenschaftliche Ergebnisse einer amtlichen Forschungsreise nach dem Bismarck–Archipel im Jahre 1908, III: Untersuchungen uber eine melanesische Wanderstrasse. *Mit aus den Deutschen Schutzgebieten. Ergänzungsheft no. 7.* Berlin: Mittler.

Gordon, Raymond G., Jr., ed., 2005. *Ethnologue: Languages of the World*, 15th edition. Dallas, Tex.: SIL International. Online version: http: //www.ethnologue.com.

Grimes, Charles E., 1991. The Buru language of Eastern Indonesia. Canberra: The Australian National University. PhD dissertation.

Grimes, Charles E., ed., 2000. *Spices from the East: Papers in languages of eastern Indonesia.* Canberra: Pacific Linguistics.

Haan, Johnson Welem, 2001. The grammar of Adang: a Papuan language spoken on the island of Alor – East Nusa Tenggara – Indonesia. Sydney: The University of Sydney. PhD dissertation.

Haspelmath, Martin, Matthew S. Dryer, David Gil and Bernard Comrie, eds, 2005. *The World Atlas of Language Structures.* Oxford: Oxford University Press.

Haudricourt, Andre, 1965. Problems of Austronesian comparative phonology. *Lingua*, 14:315–329.

Himmelmann, Nikolaus P., 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics, in K. Alexander Adelaar and Nikolaus P. Himmelmann, eds, *The Austronesian languages of Asia and Madagascar,* 110–181. London: Routledge/Curzon.

Holton, Gary, 2003. *Tobelo*. Munich: Lincom.

— 2007. Pronouns and pronominal prefixes in Alor-Pantar (Trans New Guinea). Paper presented at the Workshop on the Languages of Papua, Manokwari.

Jonker, Johann C.G., 1914. Kan men de talen van den Indischen archipel eene westelijke en eene oostelijke afdeeling onderscheiden? *Verslagen en Mededeelingen der Koninklijke Akademie van Wetenschappen* (4de reeks) 12:314–417.

Jones, Russell, ed., 2007. *Loan words in Indonesian and Malay*. Leiden: KITLV Press.

Kayser, Manfred, Silke Brauer, Gunter Weiss, Walt Schiefenhövel, Peter Underhill, Peidong Shen, Peter Oefner, Mila Tommaseo–Ponzetta, and Mark Stoneking, 2003. Reduced Y–chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *American Journal of Human Genetics*, 72: 281–302.

Kirch, Patrick V., 1997. *The Lapita peoples: ancestors of the Oceanic world*. Oxford: Blackwell.

Klamer, Marian, 1998. *A grammar of Kambera*. Berlin/New York: Mouton De Gruyter.

— 2002. Typical features of Austronesian languages in Central/Eastern Indonesia. *Oceanic Linguistics*, 41:363–383.

— 2003. Rejoinder to Malcolm Ross's squib. *Oceanic Linguistics*, 42: 511–513.

— 2004. East Nusantara: Genetic, areal, and typological approaches. *Oceanic Linguistics*, 43:240–244.

— 2006. Split–S in the Indonesian area: forms, semantics, geography. Paper presented at the Tenth International Conference on Austronesian Linguistics (ICAL 10), 17–20 January 2006. Puerto Princessa City, Palawan, Philippines. (http://www.sil.org/asia/Phillipines/ical/papers.html)

— 2008. The semantics of Semantic Alignment in eastern Indonesia. In Mark Donohue and Søren Wichmann, eds, *The typology of active–stative languages*, 221–251. Oxford: Oxford University Press.

— *A grammar of Teiwa.* Berlin/New York: Mouton de Gruyter, in press.

— 2010. A short grammar of Alorese. MS. Leiden University.

Klamer, Marian, Ger Reesink and Miriam van Staden, 2008. East Nusantara and the Bird's Head as a linguistic area. In Pieter Muysken, ed., *From Linguistic Areas to Areal Linguistics,* 95–149. Amsterdam: Benjamins.

Klinken, Catharina van, 1999. *A grammar of the Fehan dialect of Tetun, an Austronesian language of West Timor*. Canberra: Pacific Linguistics.

Kratochvíl, Frantisek, 2007. *A grammar of Abui, a Papuan language of Alor*. Utrecht: LOT dissertations.

Lynch, John, Malcolm Ross and Terry Crowley, eds, 2002. *The Oceanic Languages*. Richmond: Curzon Press.

Maddieson, Ian, 2005a. Consonant inventories. In Haspelmath *et al*, eds, 10–13.

— 2005b. Glottalised consonants. In Haspelmath *et al*, eds, 34–37.

— 2005c. Lateral consonants. In Haspelmath *et al*, eds, 38–41.

— 2005d. Uvular consonants. In Haspelmath *et alI*, eds, 30–33.

— 2005e. Voicing and gaps in plosive systems. In Haspelmath *et al*, eds, 26–29.

— 2005f. Voicing in plosives and fricatives. In Haspelmath *et al*, eds, 22–25.

— 2005g. Vowel quality inventories. In Haspelmath *et al*, eds, 14–17.

McWilliam, Andrew, 2007. Austronesian in linguistic disguise: Fataluku cultural fusion in East Timor. *Journal of Southeast Asian Studies*, 38:355–375.

Minde, Don van, 1997. Melayu Ambong. Leiden: Leiden University. PhD dissertation.

Miedema, Jelle, Cecilia Odé and Rein A.C. Dam, eds, 1998. *Perspectives on the Bird's Head of Irian Jaya, Indonesia*, 655–690. Amsterdam: Rodopi.

Milhum, Marianne, 2008. The Emergence of Agentive Patient Systems in Core Argument Marking. In Donohue and Wichman, eds, 297–334

Muysken, Pieter, 2008. Introduction: Conceptual and methodological issues in areal linguistics. In P. Muysken (ed), *From Linguistic Areas to Areal Linguistics*. Amsterdam: Benjamins.

Nichols, Johanna, 1998. The origin and dispersal of languages: Linguistic evidence. In N. G. Jablonski and L. C. Aiello, eds, *The Origin and Diversification of Language*, 127–170. San Francisco: Memoirs of the California Academy of Sciences, Number 24.

O'Connor, Susan, 2003. Nine new painted rock art sites from East Timor in the context of the Western Pacific region. *Asian Perspectives*, 42:96–128.

Odé, Cecilia, 2002a. A sketch of Mpur. In Ger P. Reesink, ed., *Languages of the Eastern Bird's Head*, 45–107. Canberra: Pacific Linguistics.

— 2002b. *Mpur prosody: An experimental–phonetic analysis with examples from two versions of the Fentora Myth*. Endangered Languages of the Pacific Rim [ELPR Kyoto: Publication series A1–003].

Pawley, Andrew K., 1998. In Jelle Miedema, *et al*, eds, 655–690

— 2002. The Austronesian dispersal: people, language, technology. In P. Bellwood and Colin Renfrew, eds, *Examining the language/farming dispersal hypothesis*, 251–273, Cambridge: McDonald Institute for Archaeological Research.

— 2005. The chequered career of the Trans New Guinea hypothesis. In Pawley *et al*, eds, 67–107.

Pawley, Andrew K., Robert Attenborough, Jack Golson and Robin Hide, eds, 2005. *Papuan pasts: Cultural, linguistic and biological histories of Papuan–speaking peoples*. Canberra: Pacific Linguistics.

Reesink, Ger P., 1996. Morpho-syntactic features of the Bird's Head languages. In Ger P. Reesink, ed., *Studies in Irian languages, NUSA*, 40:1–20.

— 1998. The Bird's Head as Sprachbund. In Jelle Miedema, *et al*, eds, 1998:603–642.

— 2002. Clause-final negation: structure and interpretation. *Functions of Language,* 9:239–268.

— 2005. West Papuan languages: Roots and development. In Pawley, *et al*, eds, 185–218.

Reesink, Ger P., ed., 2002a. *Languages of the Eastern Bird's Head*. Canberra: Pacific Linguistics.

Ross, Malcolm D., 1988. *Proto Oceanic and the Austronesian languages of western Melanesia.* Canberra: Pacific Linguistics.

— 1995a. The great Papuan pronoun hunt: recalibrating our sights. In Connie Baak, Mury Bakker, and Dick van der Meij, eds, *Tales from a concave world: liber amicorum Bert Voorhoeve,* 139–168. Leiden: Department of Languages and Cultures of South–East Asia and Oceania, Leiden University.

— 1995b. Some current issues in Austronesian linguistics. In Darrell T. Tryon, ed., *Comparative Austronesian Dictionary* 1, 45–120. Berlin: Mouton de Gruyter.

— 1996. Contact-induced change and the comparative method: cases from Papua New Guinea. In Mark Durie and Malcolm Ross, eds, *The comparative method reviewed: regularity and irregularity in language change*, 180–217. New York: Oxford University Press.

— 2001. Contact-induced Change in Oceanic Languages in North-West Melanesia. In Alexanda Y. Aikhenvald and R.M.W. Dixon, eds, *Areal diffusion and genetic inheritance*, 134–166. Oxford: University Press.

— 2003. Typology and language families: A comment on Klamer's 'Typical features of Austronesian Languages in C/Eastern Indonesia'. *Oceanic Linguistics*, 42:506–11.

— 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In Pawley, *et al*, eds, 15–65.

Silzer, Peter J., 1983. Ambai: An Austronesian language of Irian Jaya, Indonesia. Canberra: The Australian National University. PhD dissertation.

Tadmor, Uri , 2000. Rekonstruksi Aksen Kata Bahasa Melayu (The Reconstruction of Word Accent in Malay). In Bambang Kaswanti Purwo and Yassir Nasanius, eds,

Pertemuan Linguistik Pusat Kajian Bahasa dan Budaya Atma Jaya Ketiga Belas (PELBBA) 13, 153–167. Jakarta: Unika Atma Jaya.

— 2001. Phonological Convergence in the Malay Archipelago. Paper presented at the Annual Meeting of the Deutsche Gesellschaft für Sprachwissenschaft (DGfS), Leipzig.

Thomason, Sarah G., 2001. *Language Contact: an Introduction*. Edinburgh: Edinburgh University Press.

Trubetzkoy, Niholay S., 1928. Proposition 16. In *Actes de 1er Congrès international de linguistes*, 17–18. Leiden: A.W. Sijthoff's Uitgeversmaatschappij.

Tryon, Darrell T., 1995. The Austronesian languages. In Darrell T. Tryon, ed., *Comparative Austronesian Dictionary: An Introduction to Austronesian Studies. Part 1: Fascile*, 5–44. Berlin: Mouton de Gruyter.

van den Heuvel, Wilco, 2006. *Biak: a description of an Austronesian language of Papua*. Utrecht: LOT dissertations.

van den Veen, Hendrik, 1915. De Noord–Halmahera'se taalgroep tegenover de Austronesiese talen. Leiden: Leiden Uniersity. PhD dissertation.

van Staden, Miriam, 2000. Tidore, A linguistic Description of a Language of the North Moluccas. Leiden: Leiden University. PhD dissertation.

Veth, Peter, Matthew Spriggs, Ako Jatmiko, and Susan O'Connor, 1998. Bridging Sunda and Sahul: The archaeological significance of the Aru Islands, southern Moluccas. In G.–J. Bartstra, ed., *Bird's Head approaches*, 157–177. Rotterdam: Balkema. [Modern Quarternary Research in Southeast Asia 15.]

Voorhoeve, Clemens L., 1975. *Languages of Irian Jaya: checklist. Preliminary classification, language maps, wordlists*. Canberra: Pacific Linguistics.

— 1987. The non-Austronesian languages in the North Moluccas. In E.K.M. Masinambow, ed., *Halmahera dan Raja Empat sebagai Kesatuan majemuk*, Buletin LEKNAS, Vol. II, No. 2, 1983; published 1987: 13–39. Jakarta: Lembaga Ekonomi dan Kemasyarakatan Nasional, Lembaga Ilmu Pengetahuan Indonesia.

— 1989. The Masked Bird: Linguistic Relations in the Bird's Head Area. In Paul Haenen and Jan Pouwer, eds, *Peoples on the move*, 78–101. Nijmegen: Centre for Australian and Oceanic Studies.

— 1994. Contact-induced change in the non-Austronesian languages in the north Moluccas, Indonesia. In Tom Dutton and Darrell T. Tryon, eds, *Language Contact and Change in the Austronesian world*, 649–674. Berlin: Mouton de Gruyter.

Vries, Lourens de, 2004. *A short grammar of Inanwatan, an endangered language of the Bird's Head of Papua, Indonesia*. Canberra: Pacific Linguistics.

Williams-van Klinken, Catharina, John Hajek and Rachel Nordlinger, 2002. *Tetun Dili: A grammar of an East Timorese language.* Canberra: Pacific Linguistics.

Wurm, Stephen A., 1982. *The Papuan Languages of Oceania*. Tübingen: Narr.

2 *Towards a phonological overview of the vowel and consonant systems of East Nusantara*

JOHN HAJEK

1 Introduction

The East Nusantara region is an area of great linguistic diversity. While Austronesian languages dominate most of the area, it also includes many so-called non-Austronesian or Papuan languages. To date little attention has been paid to typological analysis, from a phonological perspective, of the sound systems of the region. In this study we provide the first results of an ongoing phonological-typological study of the area. East Nusantara is defined for our purposes in this study as extending from Sumbawa to the west across the Lesser Sunda Islands, Timor, Maluku including Halmahera, and as far east as the island of Biak in Cenderawasih Bay. Sulawesi, the western Sunda Islands and the Indonesian province of Papua (with the exception of the wider Bird's Head area including Biak) are excluded from consideration. Our sample includes seventy-one languages drawn from the following families (with relevant sub-groupings in brackets):

(1)	Austronesian	(Central Malayo-Polynesian, including Tetum, Leti, Alune)
		(Eastern Malayo-Polynesian, including Sawai, Biak, Ambai)
	Non-Austronesian:	
	Trans-New Guinea	(Mairasi-Tanahmerah, including Mairasi
		South Bird's Head-Timor-Alor-Pantar
		. South Bird's Head, including Arandai
		. Timor-Alor-Pantar, including Makasae, Teiwa)
	East Bird's Head	(including Sougb, Meah)
	West Papuan	(Halmahera, including Ternate, Sahu)
		(Bird's Head, including Hatam, Mpur)

All members of the Central Malayo-Polynesian (CMP) sub-branch of Austronesian (current estimate: 162 languages) are found within the geographical confines of East Nusantara. Not surprisingly, CMP dominates East Nusantara both in terms of number of languages and geographical spread. With respect to the Eastern Malayo-Polynesian (EMP) sub-branch spoken across New Guinea and the Pacific, only the relatively small South

Michael C. Ewing and Marian Klamer, editors.
Typological and areal analyses: contributions from East Nusantara, 25–46. Pacific Linguistics, 2010
©. This edition vested with Pacific Linguistics.

Halmahera-West New Guinea sub-grouping (SHWNG, 39 languages) is spoken in East Nusantara. The Trans-New Guinea (TNG) phylum which dominates New Guinea is represented in East Nusantara by the relatively small South Bird's Head-Timor-Alor-Pantar (SBH-TAP) sub-grouping (32 languages) and Mairasi-Tanahmerah. The West Papuan (WP, 26) and very small East Bird's Head (EBH, 3) families are spoken exclusively within East Nusantara.

Our sample is divided as follows:

(2)	Austronesian:			44
		CMP		37
		SHWNG		37
	Non-Austronesian:			27
	TNG			12
		Mairasi-Tanahmerah		1
		SBH-TAP		11
			SBH	2
			TAP	9
	EBH			3
	WP			12
		North Halmaheran		8
		Bird's Head		4

Previous observations about phonological systems and patterns in East Nusantara are restricted – with discussion usually brief and limited to Austronesian languages in the area, including Himmelmann (2005), Klamer (2002, 2004), and Donohue (2004). While there has been some debate about areal versus typological treatments of the region and the possible value each might have (see in particular Klamer 2002, 2004 and Donohue 2004), the purpose of this study is to make some initial, phonologically oriented typological observations that might prove useful for further study, without trying – at least at this stage - to define East Nusantara or any part of it as a specific linguistic area. Our observations involve both internal comparison amongst the languages of East Nusantara but also, and to a greater degree, external comparison with languages around the world (or occasionally, as we shall see, within the very large Austronesian family). The discussion here centres mainly on synchronic observations about consonant systems, albeit with relatively brief discussion of vowel systems in the EN region. There is only limited, and at this stage still very preliminary, reference to possible genetic patterning.

The specific phonological comparisons discussed below are those often discussed in the general phonological literature, including size and nature of vowel and consonant inventories, the nature and type of basic consonant categories, including stop consonants, fricatives and nasals, as well as more specific questions of typological interest, such as the presence of labial-velar and uvular stops. The selection of these items then allows us to assess, specifically for East Nusantara, some of the results of Maddieson's (2005a/b/c/d/e/f/g) typological survey of the world's vowel and consonant systems. We note his survey is not intended to focus on any specific geographical area as small as East Nusantara, and the number of languages from this region is relatively small. This latter fact increases the likelihood, however, that his sample will not allow for specific phenomena to be observed in East Nusantara, even if they are actually present there. Such an outcome is indeed confirmed by many of the results presented below.

2 Vowel Systems: Nature and Size

Vowel systems in East Nusantara are relatively small, and show little of the diversity that characterizes the consonant systems of the region. Most involve only five vowel systems of the basic kind, that is, /i e a o u/, and no language has fewer than five vowels. The quality of mid-vowels in these five vowel systems can vary across languages. Although these are commonly both high-mid or true mid across languages, they are low-mid /ɛ ɔ/ in some, including Inanwatan (TNG, BH), and mixed in height in others, including low-mid /ɛ/ and high-mid /o/ in Galela (WP, NH).

A much smaller number of languages have six vowels of different quality – usually involving an additional central vowel, including /i e ə a o u/ in Sika (AN, NT). Some have seven vowels, usually as a result of height distinction amongst mid vowels, including /i e ɛ a ɔ o u/ in Southern Mambae (AN, ET).

Larger vowel systems are very unusual in East Nusantara – and are accounted for in most cases by a vowel length contrast (eg. Kei [AN, SM], Adang [TNG, AP]). It is possible to reanalyse vowel length contrast in other ways, including long vowels treated as underlying sequences of identical short vowels. In so doing, the number of such large systems drops accordingly. Kedang (AN, NT) is highly unusual for the region, since its large twelve vowel system is the result of a contrast between plain and breathy voice vowels, including

(3) i i̤ u ṳ
ɛ ɛ̤ o o̤
æ æ̤ a a̤

Vowel nasalization, whether allophonic or contrastive, does not seem to be particularly characteristic of the region. Exceptionally, Kolana (TNG, AP) has a stable surface contrast between nasal and oral vowels, including [madã] ‘wear a garment on the upper body’ and [mada] ‘cook’. However, for reasons of morphophonemics, Donohue (1996) prefers to treat nasal vowels in this language as derivable from an underlying /Vn/ sequence. Elsewhere, in a relatively small number of languages, vowels may be nasalized by adjacent nasal consonants. In East Timorese languages, such as Tetum Dili (AN), Waima’a (AN) and Makasae (TNG), final /Vn/ tends to surface variably as [ṽⁿ] with a weakly articulated velar nasal prone to deletion.

Overall, preliminary statistical analysis shows little difference in the average size of vowel systems in East Nusantara: 6.2 with no difference between Austronesian and non-Austronesian languages. With respect to the former, the average is 6.1 for CMP languages and 6.7 for the smaller number of SHWNG languages. As for the latter, there is a range from 5 [EBH] to 6.9 [TNG]) but some care should be taken in interpreting this difference: the number of EBH languages is very small and they are closely related and in close physical proximity to each other. As a result, they show none of the internal variation in vowel systems that can be found within other groupings elsewhere in East Nusantara. Within the world-wide context, we note that EN average of 6.2 is lower than the global average of 8.5 vowels per language but intermediate between the 5.5 average reported for insular Southeast Asia (Malaysia, Philipines and Indonesia including East Nusantara) and 7.1 for New Guinea/Oceania (see Hajek 2007:209, and also Hajek 2004a).

3 Consonant Inventory Size

The average consonant inventory size (excluding loans) for the East Nusantara region is 16 – this is well below the worldwide average of 22.7. The EN average falls within what Maddieson (2005a) defines as the moderately small range (15-18). The consonant system of Fordata (16, AN, SE Maluku) is given here in (4) as an example of average size systems in the area.

(4) Fordata

	t̪		k	ʔ
b		d		
f		s		h
v				
m		n	ŋ	
		l		
		r		
w			j	

The smallest consonant systems involve only ten native phonemes, and are found, for example, in Leti (AN, S. Maluku) and in Sougb (EBH, BH), as seen in (5).

(5) Leti

p	t	k
	d	
	s	
β		
m	n	
	l	
	r	

Sougb

p	t	c	k	
ɸ	s			h
m	n			
r				

By far the largest consonant inventory is found in Waima'a (AN, ET) which has thirty-one phonemes.

(6) Waima'a

p	t		k	ʔ
p^h	t^h		k^h	
p'	t'		k'	
b	d		g	
			s	h
			ʔs	
m	n			
m̥	n̥			
ʔm̥	ʔn̥			
			l	
			l̥	
			ʔl̥	
			r	
			ʔr	
w		j		
w̥				
ʔw̥				

Waima'a is spoken at the border of an Austronesian-Papuan contact zone in East Timor, which leads one to immediately suspect system expansion through borrowing. However, this appears not to have been the case. Its immediate neighbours have relatively small and

very similar-looking systems. Galoli (AN) has thirteen consonant phonemes, while Makasae (TNG) has fourteen:

(7)

Galoli				Makasae			
	t	k	ʔ	p	t	k	ʔ
b	d	g		b	d	g	
	s		h		s		h
m	n			m	n		
	l				l		
	r				r		
w				w			

Neither Galoli nor Makasae shows any trace of the complexity that characterizes Waima'a. Its system involves a range of consonant types (including voiceless ejectives and other glottalized segments) that are unusual for the area and which appear to be the result of internal development.

Overall, there is no difference between Austronesian (16) and non-Austronesian (15.9), but there is some variation within each. Austronesian CMP languages have 16.4 consonants but SHWNG only 14.1. East Bird's Head (EBH) have only 12.35 consonants while West Papuan have 17, and TNG languages fall in the middle (15.3). Again, some care is needed in interpretation of these averages, given (a) the small number of EBH languages; and (b) the significant internal variation within Austronesian, including Leti (10) and Waima'a (31), both spoken in close proximity to each other. The EN average is well below the world-wide average of 22.4 consonants per language, but, again as for vowels, intermediate between the 18.5 average for insular Southeast Asia, mostly to the west, and the relatively low 14.1 average for New Guinea/Oceania directly to the east (see Hajek 2007: 209). These results suggest some kind of areal transition and progressive reduction in inventory sizes from west to east through East Nusantara.

4 Stops and stop systems

4.1 Stop contrasts

Most languages in the region have a two-way contrast between plain voiced and voiceless unaspirated stops, including Buru (AN, Maluku) /p b t d k g/. However, there is great variation across the area. Nuaulu (AN, Maluku) and Mai Brat (WP, BH) share the smallest system with only three voiceless stops /p t k/. A number of languages have three-way contrasts: in addition to plain voiced and voiceless stops, the additional set typically involves prenasalized stops (including Abun [WP]) or implosives (including Ngadha [AN], Sahu [WP], Sawu [AN]). But three-way systems without plain voiced stops are also found, for instance in Kambera /p ᵐb ɓ/.

The most complex systems are found in Waima'a (AN), Kéo (AN) and Manggarai (AN) which each have four-way systems, as in (8). While each has plain voiceless and voiced sets, overlap is only partial. Waima'a is the most aberrant, with aspirated and ejective stops – highly unusual for the region (see below for further discussion).

(8) a. Waima'a

p	t	k	ʔ
p^h	t^h	k^h	
p'	t'	k'	
b	d	g	

b. Kéo

p	t	k	ʔ
b	d	g	
ʔb	ʔd		
^{m}b	^{n}d	^{ŋ}g	

c. Manggarai

p	t	k	ʔ
^{m}p	^{n}t	^{ŋ}k	
b	d	g	
^{m}b	^{n}d	^{ŋ}g	

Many languages around the world which contrast stop voicing do not have the complete basic six-member set /p t k b d g/. It has long been observed that in some languages plain /p/ or /g/ is absent (and more rarely /p g/ simultaneously) and phonetic explanations can be found for these patterns (see Maddieson (2005b) for details). We certainly find examples of absent /p/ or /g/ in EN, as in (9a, b). As for the absence of both and only /p g/, Maddieson (2005b) reports it for only 3/566 languages (0.53%) around the world, and none for EN. Yet there are at least three examples in our East Nusantara sample (to which we could add Selaru if borrowed /p/ is excluded), as in (9c).

(9) a. missing /p/

	t	k	(including Galoli [AN], Kola [AN], Meah [EBH])
b	d	g	

b. missing /g/

p	t	k	(including Ambai [AN], Banda [AN], Mpur [EBH])
b	d		

c. missing /p g/

t	k	(including Fordata, Kei, West Tarangan [all AN])
b	d	

Other systems which involve the absence of plain stops other than /p g/ are also found in East Nusantara, including:

(10) other gaps

p	t		(including Tugun [AN], /b d/ only in loans)
		g	
p	t	k	(including Roma [AN])
	d		

With respect to the number of different places of articulation in stop systems, they normally fall within a range of three to five (see examples 4-10 above). Dhao (AN, NT) is somewhat exceptional in having stops, all native, at seven places, including retroflex /ɖ/ and palatal /ƒ/.

(11)	p	t	tʃ			k	ʔ
	b	d	dʒ	ɖ		g	
		ɗ			ʃ	ɠ	

4.2 The presence of prenasalized and implosive voiced consonants

Klamer (2002) and Himmelmann (2005) have pointed to the presence of prenasalized and implosive or implosive-like consonants in Austronesian languages in EN, specifically in the Lesser Sundas (Nusa Tenggara). Our survey confirms a concentration of both kinds of phoneme types in that sub-region. In general, languages present only one type or the other, including Manggarai (AN) which has only prenasalized segments but no implosives, while Bima has only voiced implosives. However, as previously noted, Kambera (AN) is very unusual in contrasting both types, without having plain voiced stops:

(12)	Kambera	ᵐb	ⁿd	ᵑg
		ɓ	ɗ	

As Klamer (2002) and Himmelmann (2005) also note, the presence of these types of segments is also characteristic of a wider area, in particular parts of Sulawesi and Borneo.

Prenasalized voiced stops are also found sporadically elsewhere in EN, including in Abun (WP, BH) as well as in Yamdena (AN, SE Maluku), which has /ⁿd/ alongside /ᵐp/. Elsewhere, they are occasionally reported to be allophones of plain voiced stops, as in Makasae (TNG, ET), Woisika (TNG, AP) and Arandai (TNG, BH).

Prenasalization of obstruents other than voiced stops is very rare in EN. Manggarai (AN, NT) is unusual for contrasting a wide range of voiced and voiceless prenasalized stops and affricates: /ᵐb ⁿd ᵑg ᵐp ⁿt ᵑk ⁿdʒ ⁿtʃ/. Each of these segments also contrasts with its plain equivalent.

Implosives in the Lesser Sundas are contrastive in Bima, Kambera, Ngadha and Sawu (all AN, NT). However, implosive phonemes are also found elsewhere: in Sahu (WP, Maluku), and Blagar (TNG, AP), while implosive allophones are reported in some varieties of Sika (AN, NT) and in Woisika (TNG, AP). Maddieson's (2005b) survey shows no trace of implosives in EN as defined here.

Of implosive consonants in EN, bilabial /ɓ/ is the most frequent and occurs in all (seven) languages reported to have them. This distribution is consistent with general typological observations about a preference for implosives to be bilabial (Ladefoged and Maddieson 1996: 82). Palatal and velar implosives are much rarer and are found only in Dhao (AN, NT) and Sahu (WP, N. Maluku) which each have four implosives: /ɓ ɗ ƒ ɠ/. Although some caution is necessary (see below), it does appear that prenasalization of voiced stops in EN is concentrated in the Lesser Sunda area, and is collaterally associated with the presence of implosives. That they should co-occur is not surprising – they are both strategies that favour the maintenance of voicing in stops – either through nasal venting or oral cavity expansion – before oral release. In prenasalization, nasal venting allows air to pass through the larynx facilitating vocal fold vibration required for voicing, while maintaining oral closure. In the case of implosives, lowering of the larynx associated with glottal constriction has a similar effect (see also Ladefoged and Maddieson 1996: 50-51).

Within this same sub-region we find other, phonetically related, types of voiced stops that should be included in any discussion of prenasalization/implosiveness in this same

area, such as preglottalized voiced stops in Keo /ʔb ʔd/, alongside prenasalized /ᵐb ⁿd ᵑg/ and breathy voiced stops in Lio /bɦ dɦ gɦ/. In Palu'e (AN, NT) preglottalization and implosion appear combined in /ʔɓ/. Baird (2002b) also makes the point that preglottalized stops in Kéo correspond to implosives in shared cognates in Ngadha.

4.3 Aspirated and ejective stops, and voiceless and glottalized sonorants

While prenasalized and implosive stops are not infrequent in parts of East Nusantara, contrastive aspiration is very rare. Only Hatam (WP, BH) and Waima'a (AN, ET) in our sample are known to contrast aspirated stops with other stops. Hatam has no plain voiceless stops, while contrastive aspiration is only one mechanism employed contrastively in Waima'a, as seen previously an (6) and (8). This language is striking also for the presence of voiceless ejective stops (Hajek & Bowden 2002), as in (13):

(13)	*p'ari*	'big'	*t'aku*	'to chew'	*k'ama*	'to claw'
	p'ehe	'husked rice'	*t'odo*	'to pull'	*k'odo*	'mange'

These are formed by tightly closing and compressing the space between the stop closure in the oral cavity and at the glottis. On the release of the stop closure, there is often what appears to be a popping sound. Ejective stops are widespread in different parts of the world, including Caucasus, the Americas and Africa, but are vanishingly rare across the Asia-Pacific region (see map provided by Maddieson 2005b:36). Of the more than 1200 members of the Austronesian language family, only Waima'a and Yapese (spoken in Micronesia) are known to have developed ejective stops. The fact that the same two languages have also developed glottalized sonorants and fricatives (see (6) above) is consistent with a well-established correlation between the presence of ejectives and glottalized consonants in individual languages (see Maddieson, 2005) and Clements (2003) for details). Waima'a is truly exceptional in our sample: sonorants are subject to an unusual three-way contrast: (a) plain voiced; (b) plain voiceless; and (c) glottalized, as seen in the following (given in orthographic form):

(14)	**plain**		**voiceless**		**glottalized**	
	maa	'tongue'	*mhare*	'dry'	*m'ala*	'flea'
	naa	'body'	*nhada*	'fall'	*n'ani*	'weave'
	laa	'sail'	*lhae*	'slice'	*l'ada*	'ginger'
	raa	'blood'			*r'aa*	'snatch'

To Waima'a we can add only the marginal exception of Sika (AN, NT), which has the glottalized lateral sonorant /ʔl/.

4.4 Labial-velar and uvular stops

While the labial-velar glide /w/ occurs frequently in East Nusantara, labialized stops are very rare. /k^{w}/ contrasts with /k/ in Alune (AN, Maluku), and in Dobel (AN, S. Maluku) /k^{w}/ contrasts with /p b t d/, but there are no plain velar stops. Labial-velar stops /kp gb/ are commonplace in large parts of neighbouring New Guinea (Foley 1986:61). In East Nusantara only Adang (TNG, AP) is known to have /gb/ alongside /b g/, while [gb], in optional variation with [k^{w}], is a word-final allophone of /k/ in Sougb. In Yamdena,

Mettler & Mettler (1990) treat [gb] as an optional variant for /b/ in word-initial position, although there is some evidence it may in fact be fully contrastive, as in [bobole] 'tomorrow' v. [gboti] 'carrying basket'. Similarly, /t/ appears to have a highly unusual alveolar-velar variant also restricted to word-initial position, as in [ktimpe] 'kind of bamboo'. Although more information is required, an alternative analysis, which seems just as feasible, would be to accord [gb], and possibly also [kt], with independent phonemic status in this language.

Uvular place of articulation is not normally associated with languages of East Nusantara. Indeed, Maddieson (2005d) shows no trace of uvulars throughout any part of Nusantara – west or east. However, we do find examples in our EN sample. In Ma'ya (AN, Raja Empat Islands spoken off BH) voiceless /q/ and voiced /G/ take the place of absent /k/ and /g/ respectively. Elsewhere, in Tehit (WP, BH) and Teiwa (TNG, AP) /k q/ are contrastive, but there is no voiced uvular.

5 Other consonant types

5.1 Fricatives and affricates

In general, fricative and affricate systems are relatively simple in East Nusantara and show little of the complexities we find for stops. However, typologically unusual patterns can still be found.

Almost all languages in the sample have at least one fricative – typically voiceless /s/. Few languages in our sample lack native /s/. In Kambera and Sawu (both AN, NT) historical /s/ has lenited to /h/. Arandai (TNG, BH) unusually has no voiceless fricatives, but it does have voiced /β ð ɣ/ in its system. In Roma (AN, S. Maluku) the only sibilant is retroflex /ʂ/. Palatal-alveolar /ʃ/ is also rare and never appears without /s/: it is native only in Abun (WP, BH), but occurs as a loan in Kéo (AN, NT) and Tetun Dili (AN, ET). Velar /x/ is also rare. The only voiceless fricatives, other than /s/, to appear with any frequency are glottal /h/ and one of two labials, labio-dental /f/ or bilabial /ɸ/. Indeed almost all EN languages contrast /s/ with a glottal and/or labial fricative, such that systems of the types /s h/, /s f/ and /s f h/ are widespread.

While a voicing contrast in stops is commonplace, the same cannot be said for fricatives. Voiced fricatives are relatively unusual in East Nusantara: only sixteen languages have some kind of native voiced labial (/v/ or /β/), while /z/ and /ɣ/ are each native to only three.

Affricates are barely present in East Nusantara. We find only one example of /ts/ (Fataluku [TNG, ET]) and only ten languages have /tʃ/ as part of their inherited system (but see below). /dʒ/ (and rare prenasalized /^{n}dʒ/) is, with the exceptions of /bβ d̪ð/ in Dhao, the only voiced affricate found in the entire sample.

Pharyngeal fricatives /ħ ʕ/ are rare in the world's languages, and are reported to occur primarily amongst Caucasian and Semitic languages (Ladefoged and Maddieson 1996:167–169). It is, therefore, surprising to find pharyngeal fricatives also in East Nusantara: voiceless /ħ/ contrasts with glottal /h/ in Teiwa (TNG, AP) while voiced /ʕ/ has limited word-initial distribution in Dhao (AN, NT). Its presence in this language is undoubtedly related to a wider areal phenomenon in the Lesser Sundas and Timor noted by Hajek (forthcoming), in which some kind of word-onset glottal constriction (usually [h] or [ʔ]) has developed in vowel-initial position.

A contrast involving labio-dental and bilabial fricatives, either /f ɸ/ or /v β/, is rare amongst the world's languages, but has been reported in a small number of African

languages, including Ewe (Ladefoged and Maddieson 1996: 139–141). Unusually, we find two examples of contrast in our sample: /f ɸ/ in Kola (AN, S. Maluku) and /v β/ in Sika. Matters are further complicated in Sika (AN, NT) by an additional contrast with a labiodental flap /vb/:

(15) /βoter/ 'I buy' /vo:ter/ 'we buy' /vbo:ter/ 'I stand pole in ground'
{ø-voter} {m-voter} {ʔ-voter}

Sika is here a particularly marked typological oddity: it is the only known language with a labial flap of any kind outside of Africa (see Olson and Hajek 2003 for a detailed survey). Although /vb/ also occurs within morphemes, as in /vbəvbe/ 'to hunt', its primary source in Sika is a morphophonemic process that involves the fusion of prefixed person marker <ʔ> (1st sg) and the following root-initial consonants. In the case of underlying (<ʔ + v>) and (<ʔ + r>), there is unity of results with a flapped articulation in each case: /vb/ and /ɽ/ (Donohue, forthcoming).

5.2 Nasals

All East Nusantara languages have bilabial /m/ and coronal /n/. If a third nasal is contrastive, it is usually /ŋ/, found in more than three quarters of the sample. At least ten languages, including Adang (TNG, AP), have a four-member system /m n ɲ ŋ/. Tetun Dili (AN, ET) is the only exception to this pattern: it has borrowed /ɲ/ from Portuguese, but [ŋ] has no phonemic status. It appears only as an allophone of /n/ before velar stops and in word-final position.

5.3 Liquids

Lateral and rhotic systems are even simpler throughout most of EN. The overwhelming majority of languages in the region contrast one lateral and one rhotic, typically /l/ v. /r/. A small number of languages, geographically dispersed (including Leti [AN], Alune [AN], Tehit [WP]) have only /l/, while the presence only of a rhotic but not a lateral is not uncommon in and around the Bird's Head area (AN and NAN), as in Ambai (AN), Meah (EBH), Mor (AN), and Mai Brat [WP]. In Hatam (WP) /r/ has lateral and rhotic allophones, while in Biak (AN) /r/ is clearly the only native liquid phoneme. The lateral /l/ has been introduced in loans, as in /kapal/ [kapal] ~ older [kapar] < Malay *kapal* 'boar', but it also appears as a pragmatically marked allophone of /r/. The complete absence of any kind of phonemic liquid occurs rarely, but is noted in Inanwatan (TNG) and Mpur (EBH), both in the Bird's Head area.

In most cases the rhotic is a tap or a trill. However, Abun has only an alveolar glide /ɹ/ and Liana (AN, Maluku) has only the retroflex flap /ɽ/. The presence of more than one rhotic or lateral is very unusual. Sika (AN, NT) has /r ɽ/, with the latter, as noted above, derivable from an earlier preglottalization of /r/. Tetun Dili (AN, ET) has an unstable contrast between tap and trill /ɾ r/ that tends to merge in the speech of most speakers. It is also the only language in the entire sample to have palatal /ʎ/ alongside /l/. Both trill and palatal lateral are borrowed from Portuguese, as in /kareta/ 'car' < Port. *carreta,* and /ʒuʎu/ 'July' < Port. *julho*. As previously noted, Waima'a (AN, ET) is truly exceptional: it contrasts plain /l r/ with glottalized /ʔl ʔr/ and voiceless /l̥/ (but no voiceless rhotic).

6 The Impact of Loans and Loan Phonology

In any survey of phonological systems in East Nusantara, one issue to be borne in mind is that of phoneme loans. It is clear that language contact, with Portuguese in East Timor but especially with Malay/Indonesian throughout East Nusantara, has led and is leading to the frequent borrowing of sounds, especially consonants, from these languages. Languages differ greatly in the extent to which they are open to loans and the extent to which these might be assimilated or integrated. Moreover, many sources do not include loans in the basic consonant inventories of the languages they describe. Klamer (1998) for instance, excludes loans from the basic phoneme set for Kambera, but notes the unstable appearance of Malay/Indonesian /s/ (often assimilated to native /h/) in a language whose native /s/ was lost in the late 19th century.

In general, for the purpose of this study we have tended to follow the preference of sources, and have made only occasional reference to loans. However, it needs to be noted that the process of borrowing is ongoing, is intensifying in its effect, and will become more apparent over time. Speakers of local EN languages have increasing exposure, through media, education, and officialdom, to Indonesian and Portuguese which have official status in Indonesia and East Timor respectively. Local varieties of lingua franca Malay, related to Indonesian, are also widely spoken in parts of East Nusantara.

For some languages there is still no evidence of inventory expansion through borrowing, such as for example Adang (TNG, AP), Bunak (TNG, ET), Kola (AN, SE Maluku), but in others the impact is already potentially enormous. The native consonant system of Leti (AN, SW Maluku) is very small with only ten members, but an additional five phonemes appear in loans (listed in brackets at (16)):

(16)	p		t	(c)	k
	(b)		d		(g)
	β				
		s			(h)
	m		n		(ŋ)
		l			
		r			

Indonesian/Malay is the source for frequent borrowing of affricates /dʒ/ and /tʃ/ (eight and five languages respectively). Alternatively they may also appear in EN languages in related form as palatal stops /ɟ/ and /c/ (five and six languages respectively). The other consonants most prone to borrowing are labiodential /f/ and palatal /c/ (eight and seven languages respectively). The source for each is also usually Indonesian/Malay, although Portuguese influence is clearly discernible in East Timor.

The most extreme case of borrowing in East Nusantara is undoubtedly Tetun Dili (AN, ET). Extensive contact with Portuguese (and to a lesser extent Malay) has led to a doubling of inventory size – with an increase from eleven to twenty-two members as at (17). /p g w j/ can be found in Portuguese and Malay loans, while the remaining borrowed items are taken from Portuguese. The presence of European-like fricatives, palatals and tap-trill contrast gives the language a very different feel to that of other languages in East Nusantara.

(17)

(p)	t			k	
b	d			(g)	
f	s	(ʃ)			h
(v)	(z)	(ʒ)			
m	n		(ɲ)		
	l		(ʎ)		
	(r)				
	ɾ				
(w)			(j)		

The extent of lexical borrowing from Portuguese (and Malay) is so great that borrowed phonemes must be considered part of the Tetun Dili consonant inventory. Traditionally, the East Timorese elite resident in Dili are bilingual Tetun Dili-Portuguese speakers. As a result, borrowing into Tetun Dili has extended beyond individual phonemes to include the integration of phonological processes as well, including optional palatalization of /s/ and /z/ before consonants and phrase-finally, and optional /s/-voicing at word-boundaries between vowels:

(18) a. *s* and *z-palatalszation*
loans: /festa/ [festa] ~ [feʃta] ‘party’
/ezmola/ [ezmola] ~ [eʒmola] ‘begging’
native: /has/ [ha:s##] ~ [ha:ʃ##] ‘mango’

b. *s-voicing*
loans: /estadus unidus/ [estaduz unidus] ‘United States’
native: /lis asu/ [lis asu] ~ [liz asu] ‘garlic’

Tetun Dili, which is the preferred lingua franca in East Timor, and which has now achieved co-official status in the country, is also mediating the transfer of Portuguese phonemes into the phonologies of other East Timorese languages. This process occurs through the increasing diffusion of Tetun Dili (ultimately Portuguese) lexical loans. Younger speakers of Waima’a are, for instance, more comfortable with maintaining /v z ʒ ʃ ʌ ɲ/ in loans, although the historical pattern has been for assimilation to native /b s d s l n/ respectively. Over the longer term, as knowledge of Tetun Dili and Portuguese continues to improve throughout the country, we can foresee full integration of all of these Portuguese phonemes into Waima’a (and other East Timorese languages) – and, as a result, an even larger consonant inventory for Waima’a, already the largest and most complex in the region.

7 Conclusions and Future Directions

Although our investigation of phonological systems in East Nusantara is to this point somewhat preliminary and largely restricted to consonant systems, it does provide a useful overview of a range of segmental phenomena and types – both in specific languages and over the wider EN region. In some cases, the general pattern seems to be fairly similar throughout, including the distribution of nasals, laterals and rhotics. In other cases, the situation is much more complex, including stop systems and the distribution of implosive and prenasalized segments – with a concentration (albeit not exclusive) of the last two in

the Lesser Sundas. We have also been able to identify the presence of a number of unusual consonant types – with respect to EN itself but also to the wider geographical area. These include the presence of ejective stops, glottalized sonorants and fricatives in Waima'a, as well as pharyngeals in Teiwa; uvulars in Ma'ya, Teit and Teiwa; labial-velar stops in Adang and Yamdena and the labial-dental flap in Sika.

Our results also highlight the value of detailed sampling for a specific region: it allows for phenomena to be observed that higher-level typological investigation and sampling is not well equipped to capture. Future investigation of East Nusantara will focus on expanding the current sample in order to help shift the focus from general phonological typology to a more locally oriented analysis aimed at teasing out more clearly and reliably the areal and genetic aspects of phonological patterning. Further expansion into the areas surrounding East Nusantara of the type carried out for this study would also be useful in helping us understand the extent to which observations made here are useful in defining all or part of EN as a linguistic area that can be delimited from neighbouring regions within and around East Nusantara. At this stage, however, we can report with respect to vowel and consonant inventory sizes that EN appears to be somewhat transitional between smaller vowel systems and large consonant systems of insular Southeast Asia and more complex vowel systems but much more reduced consonant inventories to the east in the wider New Guinea/Oceania region.

Appendix A: East Nusantara language sample per family, with subgrouping and location

(For the list of abbreviations see below.)

Language	Family	Subgrouping	Location
Alune	AN	CMP	W Seram, C Maluku
Ambai	AN	SHWNG	Ambai, BH
Atoni	AN	CPM	W Timor
Baikenu	AN	CMP	E Timor
Banda	AN	CMP	Kei, S Maluku
Biak	AN	SHWNG	Biak, BH
Bima	AN	CMP	Bima, NT
Buru	AN	CMP	Buru, C Maluku
Dhao	AN	CMP	Sabu, W Timor
Dobel	AN	CMP	Aru, SE Maluku
Fordata	AN	CMP	N Tanimbar, SE Maluku
Galoli	AN	CMP	E Timor
Irarutu	AN	SHWNG	Bomberai, BH
Kaiwai	AN	CMP	Bomberai, BH
Kambera	AN	CMP	Sumba, NT
Kedang	AN	CMP	Lembata, NT
Kei	AN	CMP	Kei, SE Maluku
Kéo	AN	CMP	Flores, NT
Kisar	AN	CMP	Kisar, S Maluku
Kola	AN	CMP	Aru, S Maluku
Larike	AN	CMP	Ambon, C Maluku

Language	Family	Subgrouping	Location
Leti	AN	CMP	Leti, S Maluku
Liana	AN	CMP	Seram, C Maluku
Lio	AN	CMP	Flores, NT
Manggarai	AN	CMP	Flores, NT
Ma'ya	AN	SHWNG	Raja Ampat, BH
Mor	AN	SHWNG	Mor, Cendrawasih Bay, BH
Ngadha	AN	CMP	Flores, SC
Nuaulu	AN	CMP	Seram, C Maluku
Palu'e	AN	CMP	Palu'e, NT
Roma	AN	CMP	Roma, S Maluku
Roti	AN	CMP	Rote, W Timor
S Mambae	AN	CMP	E Timor
Sawai	AN	SHWNG	Halmahera, N Maluku
Sawu	AN	CMP	Sabu, W Timor
Selaru	AN	CMP	Tanimbar, S Maluku
Sika	AN	CMP	Flores, NT
Taba	AN	SHWNG	Makian, N Maluku
Tetun Dili	AN	CMP	E Timor
Tetun Fehan	AN	CMP	E Timor
Tugun	AN	CMP	Wetar, SW Maluku
W Tarangan	AN	CMP	Aru, S Maluku
Waima'a	AN	CMP	E Timor
Yamdena	AN	CMP	Yamdena, SE Maluku
Meah	EBH		E BH
Mpur	EBH		E BH
Sougb	EBH		E BH
Adang	TNG	Timor-Alor-Pantar	Alor Pantar
Arandai	TNG	South Bird's Head	BH
Blagar	TNG	Timor-Alor-Pantar	Alor Pantar
Bunak	TNG	Timor-Alor-Pantar	E and W Timor
Fataluku	TNG	Timor-Alor-Pantar	E Timor
Inanwatan	TNG	South Bird's Head	S BH
Kabola	TNG	Timor-Alor-Pantar	Alor Pantar
Kolana	TNG	Timor-Alor-Pantar	BH
Mairasi	TNG	Mairasi-Tanahmerah	S BH
Makasae	TNG	Timor-Alor-Pantar	E Timor
Teiwa	TNG	Timor-Alor-Pantar	Alor Pantar
Woisika	TNG	Timor-Alor-Pantar	Alor Pantar
Abun	WP	Family level isolate	BH
Galela	WP	North Halmaheran	Halmahera, N Maluku
Hatam	WP	Hattam	E BH
Mai Brat	WP	Central Bird's Head	C BH
Makian, W	WP	North Halmaheran	Halmahera, N Maluku
Pagu	WP	North Halmaheran	Halmahera, N Maluku
Sahu	WP	North Halmaheran	Halmahera, N Maluku
Tabaru	WP	North Halmaheran	Halmahera, N Maluku
Tehit	WP	West Bird's Head	SW BH

Language	Family	Subgrouping	Location
Ternate	WP	North Halmaheran	Ternate, N Maluku
Tidore	WP	North Halmaheran	Tidore, N Maluku
Tobelo	WP	North Halmaheran	Halmahera, N Maluku

Appendix B: East Nusantara Language Sample in alphabetical order with sources

Abun	Berry (1995)
Adang	Haan (2001); Hajek (2004b), field notes
Alune	Taguchi and Taguchi (1990)
Ambai	Silzer (1983)
Arandai	Voorhoeve (1985)
Atoni	Steinhauer (1993)
Baikenu	Hull (2003)
Banda	Kaikerissa, Kasihuw and Tamaela (1996)
Biak	van den Heuvel (2006)
Bima	Tama, *et al* (1996)
Blagar	Steinhauer (1991, 1977)
Brat (also: Mai Brat, Maybrat)	Dol (1999)
Bunak	Friedberg (1978)
Buru	Grimes (1992)
Dhao	Grimes (2009)
Dobel	Hughes and Hughes (1989)
Fataluku	Campagnolo (1973)
Fordata	Marshall (2000)
Galela	Shelden (1989)
Galoli	Hull (2003b)
Hatam	Reesink (1999)
Inanwatan	de Vries (2004)
Irarutu	Matsumura (1991)
Kabola	Stokhof (1987)
Kambera	Klamer (1998)
Kedang	Samely (1991)
Kei	Tetelepta, *et al* (1985)
Keo	Baird, L. (2002a), (2002b)
Kisar	Christensen and Christensen (1992)
Kowiai	Walker and Walker (1990)
Kola	Takata and Takata (1992)
Kolana	Donohue (1996)
Larike	Laidig (1992), Laidig and Laidig (1995)
Leti	van Engelenhoven (1995)
Liana	McCollum (1992)
Lio	Levi (1978)
Mai Brat/ Maybrat	Dol (1999)

Mairasi	Peckham (1991)
Makasae	Hull (2005)
Makian, West	Voorhoeve (1982)
Mambae, South	Hull (2003)
Manggarai	Verheijen and Grimes (1995)
Ma'ya	van der Leeden (1993)
Meah	Gravelle and Gravelle (1991)
Mor	Laycock (1978)
Mpur	Odé (2002)
Ngada	Moore (1980), Djawanai (1977)
Nuaulu	Bolton (1989)
Pagu	Wimbish (1992)
Palu'e	Donohue (n.d.)
Roma	Steven (1991)
Roti	Fox and Grimes (1995)
Sahu	Severn (1995)
Sawai	Whisler (1992)
Sawu	Walker (1982)
Selaru	Coward (1989), Coward and Coward (2000)
Sika (also Sikka)	Donohue (in press), Rosen (1986)
Sougb	Reesink (2002)
Taba (Makian Dalam)	Bowden (1998)
Tabaru	Kotynski. (1988a, (1988b)
Tarangan, w	Nivens (1992
Tehit	Flassy and Stokhof (1979,) Flassy (1981)
Teiwa	Klamer (2005)
Ternate	Watuseke (1991)
Tetum	van Klinken, Hajek, and Nordlinger (2002)
Tetun Fehan	van Klinken (1999)
Tidore	van Staden (2000); Pikkert and Pikkert (1995)
Tobelo	Rugebregt, *et al* (1986; Holton (2003)
Tugun	Hinton (1991)
Waimaha	Hajek, *et al* (2006)
Woisika	Stokhof (1979)
Yamdena	Mettler and Mettler (1990)

List of Abbreviations

AN	Austronesian
BH	Bird's Head
C	Central
CMP	Centro-Malayo-Polynesian
E	East
EBH	East Bird's Head
EMP	Eastern Malayo-Polynesian

EN	East Nusantara
ET	East Timor
N	North
NAN	Non-Austronesian
NM	North Maluku
NT	Nusa Tenggara
SBH	South Bird's Head
SBH-TAP	South Bird's Head-Timor-Alor-Pantar
SHWNG	South Halmahera-West New Guinea
SM	South Maluku
TAP	Timpor-Alor-Pantar
TNG	Trans-New Guinea
W	West
WP	West Papuan

References

Baird, Louise, 2002a. Kéo. *Journal of the International Phonetic Association,* 32:93–97.

— 2002b. A grammar of Kéo: An Austronesian language of East Nusa Tenggara. Canberra: The Australian National University. PhD dissertation.

Berry, Keith, 1995. A description of the Abun language phonology and basic morpho-syntax. Melbourne: La Trobe University. MA dissertation.

Blevins, Juliet, 2004. The mystery of Austronesian final consonant loss. *Oceanic Linguistics*, 43:208–213.

Bolton, Rosemary A., 1989. Nuaulu Phonology. In Wyn D. Laidig, ed., *Workpapers in Indonesian Languages and Cultures*, 7:89–119. Ambon: Summer Institute of Linguistics.

Bowden, John, 1998. *Taba: Description of a South Halmahera language*. Canberra: Pacific Linguistics.

Campagnolo, Henri, 1973. La langue des Fataluku de Lórehe (Timor Portugais). Paris: Université René Descartes. PhD dissertation.

Christensen, John and Sylvia Christensen, 1992. Kisar phonology. In Donald A. Burquest and Wyn D. Laidig, eds, *Phonological studies in four languages of Maluku,* 33–65. Dallas: Summer Institute of Linguistics.

Clements, George N., 2003. Feature economy in sound systems, *Phonology*, 20:287–333.

Coward, David and Naomi Coward, 2000. A phonological sketch of the Selaru language. In Charles E. Grimes, ed., *Spices from the East: Papers in eastern Indonesian languages*, 9–54. Canberra: Pacific Linguistics.

Coward, Naomi E., 1989. A phonological sketch of the Selaru language. In Wyn D. Laidig, ed., *Workpapers in Indonesian languages and culture*s, 7:1–42. Ambon: Summer Institute of Linguistics.

Djawanai, Stephanus, 1977. A description of the basic phonology of Nga'da and the treatment of borrowings. *NUSA: Linguistic studies of Indonesian and other languages of Indonesia*, 5:10–18.

Dol, Philomena H., 1999. A grammar of Maybrat. A language of the Bird's head, Irian Jaya, Indonesia. University of Leiden. PhD dissertation.

Donohue, Mark, 1996. Kolana: Syllables, (small) words and CV. Unpublished. The Australian National University.

— 2004. Typology and linguistic areas. *Oceanic Linguistics* 43:221–239.

— forthcoming. Sikka notes. In Hein Steinhauer, ed., special number of *NUSA: Linguistic Studies of Indonesian and Other Languages of Indonesia*.

— n.d. The prosodic nature of Palu'e. Unpublished ms. National University of Singapore.

Engelenhoven, Aone van, 1995. A description of the Leti language (as spoken in Tutukei). Leiden University. PhD dissertation.

Flassy, Don A.L. and W.A.L. Stokhof, 1979. A note on Tehit (Bird's Head – Irian Jaya). *NUSA: Linguistic studies of Indonesian and Other Languages of Indonesia*. 7:35–83.

Flassy, Don A.L., 1981. *Struktur bahasa Tehid*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Foley, William A., 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.

Fox, James J. and Charles E. Grimes, 1995. Roti. In Daryl T. Tryon, ed., *Comparative Austronesian dictionary: An introduction to Austronesian studies*, part 1: fascicle 1, 611–622. Berlin and New York: Mouton de Gruyter.

Friedberg, Claudine, 1978. *Comment fut franchée la liane céleste et autres textes de literature orale bunaq (Timor, Indonésie)*. Paris: Centre national de la recherche scientifique.

Gravelle, Gilles and Gloria Gravelle, 1991. Meah phonology. In Margaret Hartzler and LaLani Wood, eds, *Workpapers in Indonesian Languages and Cultures*, No. 10, 1–36. Jayapura: Summer Institute of Linguistics.

Greenberg, Joseph, 1978. Some generalisations concerning initial and final consonant clusters. In Joseph Greenberg, ed., *Universals of Human Language*, 243–279. Stanford: Stanford University Press.

Grimes, Charles E, 1992. The Buru language of eastern Indonesia. The Australian National University, Canberra. PhD dissertation.

— 2010. Hawu and Dhao in eastern Indonesia: Revisiting their relationship. This volume.

Haan, John W., 2001. The grammar of Adang: A Papuan language spoken on the island of Alor, East Nusa Tenggara – Indonesia. The University of Sydney. PhD dissertation.

Hajek, John, 2004a. Consonant inventories as an areal feature of the New Guinea-Pacific Region: Testing Trudgill's hypotheses. *Linguistic Typology* 8:343–350.

— 2004b. Adang field notes.

— 2007. Sound systems of the Asia-Pacific: some basic typological observations. In Peter K. Austin and Andrew Simpson, eds, *Endangered Languages, Linguistische Berichte Sonderheft* 14, 207–222, Hamburg: Helmut Buske.

— forthcoming. Breathy voice as an areal feature of Nusa Tenggara and East Timor. In Hein Steinhauer, ed., special number of *NUSA: Linguistic Studies of Indonesian and Other Languages of Indonesia*.

Hajek, John and Nikolaus Himmelmann, 2006. Notes on Waima'a Phonology. Unpublished manuscript. University of Melbourne.

Hajek, John and John Bowden, 1999. Taba and Roma: Clusters and geminates in two Austronesian languages. In *Proceedings of the 13th International Congress of Phonetic Sciences*. San Francisco, 143–146.

— 2002. A phonological oddity in the Austronesian area: ejectives in Waimoa, *Oceanic Linguistics* 41:222–224.

Haspelmath, Martin, Mathew S. Dryer, David Gil and Bernard Comrie, eds, 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.

Heuvel, Wilco van den, 2006. *Biak, description of an Austronesian language of Papua*. Utrecht: Netherlands Graduate School of Linguistics (LOT).

Klinken, Catharina L. van, 1999. *A grammar of the Fehan dialect of Tetun. An Austronesian language of West Timor*. Canberra: Pacific Linguistics.

Himmelmann, Nikolaus, 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In Nikolaus Himmelmann and Karl A. Adelaar, eds, *The Austronesian Languages of Asia and Madagascar*, 110–181. London: Routledge.

Hinton, Bryan D., 1991. Aspects of Tugun phonology and syntax. Arlington: The University of Texas. PhD dissertation.

Holton, Gary, 2003. *Tobelo*. Munich: Lincom Europa.

Hughes, Jock and K. Hughes, 1989. A phonology of Dobel. In Wyn D. Laidig, ed., *Workpapers in Indonesian Languages and Cultures*, 7:43–76. Ambon: Summer Institute of Linguistics.

Hull, Geoffrey, 2003a . *Baikenu*. Dili: Instituto Nacional de Linguística.

— 2003b. *Galoli*. Dili: Instituto Nacional de Linguística.

— 2003c. *Southern Mambai*. Dili: Instituto Nacional de Linguística.

— 2005. *Makasai*. Dili: Instituto Nacional de Linguística.

Kaikerissa, O., J. Kasihuw and J. Tamaela, 1996. *Struktur bahasa Banda*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Klamer, Marian, 1998. *A grammar of Kambera*. Berlin/New York: Mouton de Gruyter.

— 2002. Typical features of Austronesian languages in central/eastern languages in Central/Eastern Indonesia. *Oceanic Linguistics* 41:363–383.

— 2004. East Nusantara: Genetic, areal and typological Approaches. *Oceanic Linguistics* 43:240–244.

— 2005. A grammar of Teiwa. Unpublished manuscript. Leiden University.

Kotynski, Edward A., 1988a. Tabaru phonology and morphology. In D. C. Derbyshire, ed., *Work papers of the Summer Institute of Linguistics*. University of North Dakota session, 32:143–216.

— 1988b. Tabaru phonology and morphology. Grand Forks: University of North Dakota. MA thesis.

Ladefoged, Peter and Ian Maddieson, 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Laidig, Carol J., 1992. Segments, syllables and stress in Larike. In Donald A. Burquest and Wyn D. Laidig, eds, *Phonological studies in four languages of Maluku*, 67–126. Dallas: Summer Institute of Linguistics.

Laidig, Wyn D. and Laidig, Carol J., 1995. A synopsis of Larike phonology and syntax. In Wyn D. Laidig, ed., *Descriptive studies in languages of Maluku, Part II*, 19–42. Badan penyelengaraan seri NUSA, Universitas Katolik Indonesia Atma Jaya, Jakarta.

Laycock, Donald A., 1978. A little Mor. In Stephen A. Wurm and Lois Carrington, eds, *Second international conference of Austronesian linguistics: Proceedings. Fascicle 1. Western Austronesian*, 285–316. Canberra: Pacific Linguistics.

Leeden, A. C. van der, 1993. *Ma'ya: A language study. Jakarta*: Lembaga Ilmu Pengetahuan Indonesia.

Levi, Ferdinandus, 1978. A preliminary study of Lionese. Yogyakarta: Sanata Dharma Teacher's Training Institute. B.A. thesis.

Maddieson, Ian, 1983. *Patterns of sounds*. Cambridge: Cambridge University Press.

— 2005b. Glottalized consonants. In Haspelmath *et al*, eds, 34–37.

— 2005c. Lateral consonants. In Haspelmath *et al*, eds, 38–41.

— 2005d. Uvular consonants. In Haspelmath *et al*, eds, 30–33.

— 2005e. Voicing and gaps in plosive systems. In Haspelmath *et al*, eds, 26–29.

— 2005f. Voicing in plosives and fricatives. In Haspelmath *et a*l, eds, 22–25.

— 2005g. Vowel quality inventories. In Haspelmath *et al*, eds, 14–17.

Marshall, Craig H., 2000. A phonological description of Fordata. In Charles E. Grimes, ed., *Spices from the East: Papers in languages of Eastern Indonesia*, 181–235. Canberra: Pacific Linguistics.

Matsumura, Takashi, 1991. Irarutu phonology. In Margaret Hartzler and LaLani Wood, eds, *Workpapers in Indonesian Languages and Cultures*, 10:37–74. Jayapura: Summer Institute of Linguistics.

McCollum, Frank, 1992. Liana phonology sketch. Ambon: Summer Institute of Linguistics. Unpublished manuscript.

Mettler, Toni and Heidi Mettler, 1990. Yamdena phonology. In Wyn D. Laidig, ed., *Workpapers in Indonesian Languages and Cultures*, 8:29–79. Ambon: Summer Institute of Linguistics.

Moore, Patrick, 1980. Nga'da phonology. In Paz B. Naylor, ed., *Papers from the Second Eastern Conference on Austronesian Languages. Austronesian Studies*, 203–212. Ann Arbor: Ann Arbor Centre for South and South East Asian Studies, University of Michigan.

Nivens, Richard, 1992. A lexical phonology of West Tarangan. In Donald A. Burquest and Wyn D. Laidig, eds, *Phonological studies in four languages of Maluku*, 127–227. Dallas: Summer Institute of Linguistics.

Odé, Cecilia, 2002. A sketch of Mpur. In Ger P. Reesink, ed., *Languages of the Eastern Bird's Head*, 45–107. Canberra: Pacific Linguistics.

Olson, Kenneth and John Hajek, 2003. Cross-linguistic insights on the labial flap, *Linguistic Typology*, 7:1–30.

Peckham, Lloyd, 1991. Mairasi phonology. *Workpapers in Indonesian languages and cultures*, 10:111–145. Jayapura: Summer Institute of Linguistics.

Pikkert, Joost J. J. and Cheryl M. Pikkert, 1995. A first look at Tidore phonology. In Wyn D. Laidig, ed., *Descriptive studies in languages of Maluku. Part II*, 43–70. NUSA: Linguistic studies of Indonesian and other languages of Indonesia, 38. Jakarta: Universitas Katolik Indonesia Atma Jaya.

Reesink, Ger P., 1999. *A grammar of Hatam*. Canberra: Pacific Linguistics.

— 2002. A grammar sketch of Sougb. In Ger P. Reesink, ed., *Languages of the Eastern Bird's Head*, 181–275. Canberra: Pacific Linguistics.

Rosen Joan M., 1986. Phonemes, verb classes and personal endings in Maumere. *NUSA: Linguistic studies of Indonesian and other languages of Indonesia*, 25:39–69.

Rugebregt, E. J. F., O. Kakerissa, J. Pentury, C. C. Seumahu and J. J. Anakotta, 1986. *Struktur bahasa Tobelo*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Samely, Ursula, 1991. *Kedang (Eastern Indonesia). Some aspects of its grammar*. Hamburg: Helmut Buske Verlag.

Severn, John A., 1995. The phonemic syllable in Sahu: a computer-aided phonological analysis. In Wyn D. Laidig, ed., *Descriptive studies in languages of Maluku, Part II* , 71–87. NUSA: Linguistic studies of Indonesian and other languages of Indonesia, 38. Jakarta: Universitas Katolik Indonesia Atma Jaya.

Shelden, Howard, 1989. Galela phonemes and stress. *Workpapers in Indonesian Languages and Cultures*, 7:77–88. Ambon: Summer Institute of Linguistics.

Silzer. Peter J., 1983. Ambai: An Austronesian language of Irian Jaya, Indonesia. Canberra: The Australian National University. PhD dissertation.

Staden, Miriam van, 2000. Tidore: A linguistic description of a language of the North Moluccas. Leiden: University of Leiden. PhD dissertation.

Steinhauer, Hein, 1977. 'Going' and 'coming' in the Blagar of Dolap (Pura-Alor-Indonesia). *NUSA: Linguistic studies of Indonesian and other languages of Indonesia*, 4:49–59.

— 1991. Demonstratives in the Blagar language of Dolap (Pura, Alor, Indonesia). In T. Dutton, ed., *Papers in Papuan linguistics. No. 1*, 177–221. Canberra: Pacific Linguistics.

— 1993. Notes on verbs in Dawanese (Timor). In Ger P. Reesink, ed., *Topics in descriptive Austronesian linguistics*, 131–58. *Semaian* 11. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit te Leiden.

— 1996. Synchronic metathesis and apocope in three Austronesian languages of the Timor area. In Suwilai Premsrirat, ed., *Pan-Asiatic Linguistics. Proceedings of the Fourth International Symposium on Languages and Linguistics*, January 8–10, 1996, 471–492. Salaya: Institute of Language and Culture for Rural Development. Mahidol University.

Steven, Lee A., 1991. The phonology of Roma, an Austronesian language of Eastern Indonesia. Arlington: The University of Texas. MA dissertation.

Stokhof, W. A. L., 1979. *Woisika II: Phonemics*. Canberra: Pacific Linguistics.

— 1987. A short Kabola text (Alor, East Indonesia). In Donald C. Laycock and W. Winter (eds) A *World of language: Papers presented to Professor S. A. Wurm on his 65th birthday,* 631–648. Canberra: Pacific Linguistics.

Taguchi, Yushin and Takako Taguchi, 1990. A phonology of Alune. In Wyn D. Laidig, ed., *Workpapers in Indonesian languages and cultures*, 8:95–127. Ambon: Summer Institute of Linguistics.

Takata, Masahiro and Yuko Takata, 1992. Kola phonology. In Donald A. Burquest and Wyn D. Laidig, eds, *Descriptive studies in languages of Maluku*, NUSA: Linguistic studies of Indonesian and Other Languages of Indonesia, 34:31–46.

Tama, I. W., Sukayana, I. N., Partami, N. L. and Z. M. Hamidsyukrie, 1996. *Fonologi bahasa Bima.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Tetelepta, J., E. A. Apituley, J. Kasihuw, E. Wattimury and I. Talakua, 1985. *Struktur bahasa Kei*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Verheijen, J. A. J. and Charles E. Grimes, 1995. Manggarai: Introduction and wordlist. In Daryl T. Tryon, ed., *Comparative Austronesian dictionary: An introduction to Austronesian studies, part 1: fascicle 1*, 585–92. Berlin/New York: Mouton de Gruyter.

Voorhoeve, Clemens L., 1982. The West Makian language, North Moluccas, Indonesia: A fieldwork report. In Clemens L. Voorhoeve, ed., *The Makian languages and their neighbours*, 1–74. Canberra: Pacific Linguistics.

— 1985. *Some notes on the Arandai language, South Bird's Head, Irian Jaya*. Irian 13:3–40.

Vries, Lourens de, 2004 . *A short grammar of Inanwatan, an endangered language of the Bird's Head of Papua, Indonesia*. Canberra: Pacific Linguistics.

Walker, Alan T., 1982. *A grammar of Sawu*. NUSA: Linguistic studies of Indonesian and other languages of Indonesia, 13.

Walker, Roland and J. Walker, 1990. Kowiai phonology. Jayapura: Universitas Cendrawasih and Summer Institute of Linguistics.

Watuseke, F. S., 1991. The Ternate language. In T. Dutton, ed., *Papers in Papuan linguistics, No. 1*, 223–244. Canberra: Pacific Linguistics.

Whisler, Ronald, 1992. Phonology of Sawai. In Donald A. Burquest and Wyn D. Laidig, eds, *Phonological studies in four languages of Maluku*, 7–32. Dallas: Summer Institute of Linguistics.

Williams-van Klinken, Catharina, Hajek, John and Rachel Nordlinger, 2002. *Tetun Dili: A grammar of an East Timorese Language*. Canberra: Pacific Linguistics.

Wimbish, Sandra G., 1992. Pagu phonology. In Donald A. Burquest and Wyn D. Laidig, eds, *Descriptive studies in languages of Maluku*. NUSA: Linguistic studies of Indonesian and other languages of Indonesia, 34:69–90.

3 *Notes on Waima'a intonational structure*

NIKOLAUS HIMMELMANN

1 Introduction

This chapter presents a first attempt at analysing the intonational structure of Waima'a, an endangered Austronesian language spoken in East Timor. More precisely, it deals with the variety of Waima'a spoken in the village of Caisido and parts of the neighboring town of Baucau.[1]

The analysis makes use of the autosegmental-metrical framework where intonational contours are represented as strings of high (H) or low (L) tonal targets (Ladd 1996). In accordance with their function within an intonation unit (IU), three different types of tonal targets are distinguished within this framework (*T* here stands for *tonal target*, the cover term for H and L targets):

1. T*: accentual (or prominence-lending) tone, which occurs on lexically accented syllables and usually marks focal information (these tones are also known as (*intonational*) *pitch accents*).
2. T-: phrase accent (or edge tone), that is a major prominence which occurs at a fixed distance from the edge of an intonation unit (e.g., the first syllable or the penultimate syllable of an IU).

[1] For further information on the language and its setting, including the dialect situation, see the archive of the Waima'a documentation project at http://www.mpi.nl/DOBES/.

Many thanks to Maurício C.A. Belo, John Bowden, John Hajek and Alex V. Tilman, my main collaborators in this project. Very special thanks to Maurício, the native speaker on the documentation team, who recorded most of the elicited data forming the empirical basis of this chapter. Further information and full acknowledgements for the Waima'a project can be found on the website just mentioned.

I am very grateful to Ruben Stoel for sharing his innovative work on Javanese and Manado Malay and especially for his very helpful and productive comments on the first draft of the analysis presented here. Similarly, Bob Ladd was of great help in developing the current analysis and provided pertinent criticism on the second draft. Many thanks also to John Hajek, Michael Ewing and René Schiering for very useful comments on earlier versions of this chapter.

This work was made possible by a research professorship funded by the Volkswagen foundation and I am most grateful for this very generous support.

Michael C. Ewing and Marian Klamer, editors.
Typological and areal analyses: contributions from East Nusantara, 47–69. Pacific Linguistics, 2010
© This edition vested with Pacific Linguistics.

3. T%: boundary tone, a pitch excursion which occurs on the first or last syllable(s) of an IU, regardless of whether or not the syllables are lexically accented.

Perhaps the most major difference between Waima'a intonation and the much better known intonational structures in European languages[2] pertains to the fact that there appear to be no accentual tones (no T*) in Waima'a. That is, the basic structure of the Waima'a intonation unit looks as depicted in Figure 1.

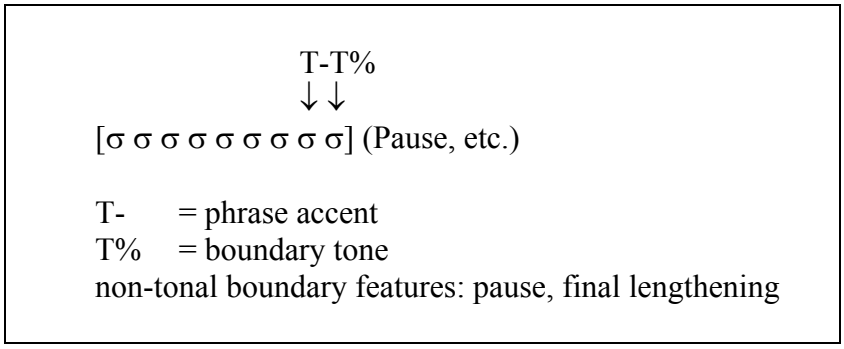


Figure 1: Basic structure of Waima'a intonation unit

A Waima'a IU thus obligatorily consists of a phrase accent on the penultimate syllable and a final boundary tone. The main part of this chapter will describe the tonal patterns that may occur in these positions (section 3) and the optional occurrence of a lower-level intonational constituent at the beginning of an IU (section 4). Section 5 briefly shows that IUs may include syntactic structures of various sizes.

The absence of accentual tones is not a unique feature of Waima'a. Instead, this is a feature which has also been reported for other languages in the area, including western varieties of Malay and Javanese (cp. Tadmor 2000, van Zanten et al. 2003, Stoel 2006, 2007). As in these other languages, the absence of accentual tones raises the question of whether Waima'a lexical items carry lexical accents (are lexically stressed), and the evidence presented in section 6 points in the direction that Waima'a in fact lacks lexical accents. Waima'a would thus appear to belong to the type of languages which have neither lexical accent nor lexical tone. To date, this type of languages has played a somewhat marginal role in the literature on prosodic typology, but there is growing evidence from languages around the world that this type is in fact not marginal at all, as briefly discussed in the final section.

Before turning to the prosodic analysis proper, however, section 2 will briefly present the rather complex Waima'a segment inventory. This is needed because various segment types cause considerable perturbations with regard to fundamental frequency, and thus are relevant to the interpretation of the data presented here.

In concluding these introductory remarks, a note on the database for this study and how the data was collected will be in order. The analysis is based primarily on two sets of elicited mini-discourses targeting sentence mood and information structure. Examples include the Waima'a equivalents of exchanges such as:

---

[2] See Ladd (1996), the contributions on major European languages in Hirst and Di Cristo (1998) and Jun (2005), and table 16.2 in Jun (2005a) for exemplification of typical European systems.

(1) (In the market): What are you looking for? (I am looking for) vegetables.
(2) Have you ever eaten a snake? No, I am afraid of snakes.

One set consisted of 15 items and was recorded with 14 speakers, 9 female and 5 male. The second set consisted of 16 items. To date, data for 5 speakers (that is 3 female and 2 male) have been processed for this set. This basic set was complemented by recordings of short word lists with two speakers (one male, one female), each consisting of 3-7 items presented in different orders and spellings, and approximately 150 intonation units taken from the corpus of spontaneous naturalistic speech compiled by the Waima'a documentation team.

The elicitation of the data was complicated by the fact that there is no well-established writing tradition for Waima'a. Consequently, the prompts for the mini discourses could not be presented in writing but rather had to be rehearsed in advance with the speakers who then had to *enact* the scene. All examples were first rehearsed and recorded with Waima'a team member Maurício Belo as the main speaker, with the author operating the video camera. All further rehearsing and recording was done by Maurício, who thus appears in all recordings. The author was usually not present at these further recordings, which were mostly done open air in the village. For the short word lists, reading Waima'a was rehearsed with two speakers who are fully literate in Tetum and Malay (one of them was again Maurício Belo) and all recordings were done indoors in an office setting by the author. See Himmelmann (2006) and Himmelmann and Ladd (2008) for further details on data gathering procedures.

Obviously, this basic set-up is prone to lots of interferences and noise of various kinds, and it will thus not come as a surprise that less than 50% of all recorded data were actually usable for the analysis. The corpus of utterances which were plotted (F0, waveform) and segmented into syllables for this study consists of approximately 1150 items. The program used for creating and managing this corpus was EMU (http://www.shlrc.mq.edu.au/emu), the figures included in this chapter were created with the statistical computing freeware program *R* (http://www.R-project.org.).[3]

2 Waima'a segments causing F0 perturbations

While the Waima'a vowel inventory is unremarkable and straightforward (5 vowels /i/, /e/, /a/, /o/, /u/),[4] the consonant inventory is rather large and unusual for an Austronesian language. As shown in Table 1, it includes glottalised and aspirated series for most manners of articulation in addition to the simple voiced and/or voiceless series commonly found in many Austronesian languages. Segments are given here in the practical orthography used throughout this chapter (IPA symbols are added in [ ] where they differ from the practical orthography). Segments in parentheses only occur in loans or are marginal to the system. For a fuller exposition of the phoneme inventory, including detailed commentary regarding phonetic realization, see Hajek and Himmelmann (2006). For more detail regarding glottalised segments, see also Hajek and Bowden (2002), Hajek and Stevens (2005) and Stevens and Hajek (2004).

---

[3] Special thanks to Emina Kurtić and Jan Strunk for creating and managing this resource. Jan Strunk prepared all the plots shown in the following figures.

[4] Phonetically long vowels arise when two like vowels occur in adjacent syllables (that is (C)V.V) as in *mee* 'red' or *khaa* 'eat'.

Table 1: Waima'a consonant inventory

		Labial	**(Post-) Alveolar**	**Velar**	**Glottal**
Stops	voiceless unaspirated	(p)	t	k	' [ʔ]
	voiceless aspirated	(ph) [p^h]	th [t^h]	kh [k^h]	
	voiceless ejective	p'	t'	k'	
	voiced	b	d	g	
Fricatives	plain	(f)	s		h
	glottalised		s' [$s^ʔ$]		
Nasals	plain	m	n		
	voiceless	mh [m̥]	nh [n̥]		
	glottalised	m' [$m^ʔ$]	n' [$n^ʔ$]		
Laterals	plain		l		
	voiceless		lh [l̥]		
	glottalised		l' [$l^ʔ$]		
Rhotic	plain		r [ɾ]		
	glottalised		r' [^{ʔ}r]		
Approximants	plain	w [w/ɥ]	j		
	voiceless	wh [ẘ]			
	glottalised	w' [$w^ʔ$/$ɥ^ʔ$]			

Of major import for current concerns is the fact that glottalised and aspirated segments may cause considerable perturbations in fundamental frequency which may go well beyond the usual microperturbations observable for voiced and voiceless stops and fricatives (that is that voiced stops tend to lower F0, while voiceless stops tend to raise F0, etc., which are standard effects also observable in Waima'a; see Laver (1994:431–546) for detailed discussion). Thus, when inspecting the F0 extractions provided here to illustrate claims about intonational structure, it will be important to keep in mind that some of the major changes in F0 observable in the plots are not related to intonational targets, and thus are in fact hardly detectable auditorily. Instead, they are caused by a preceding glottalised or aspirated segment. Figure 2 illustrates this for the example given in (3), rendered by a female speaker.

(3) *Da. Aku thaku khaa sawa.*
NEG 1s afraid eat snake
'No! I am afraid of eating snakes'

In Figure 2 there is a clearly observable jump up in F0 after both the alveolar and the velar voiceless aspirated stops (*th* and *kh*), more than 60 Hz in the first instance, close to 80 in the second. Both these jumps are caused exclusively by the stops, and are hardly perceptible auditorily. Auditorily, pitch is low and flat after the initial rise-fall on the negative *da* (which constitutes an IU of its own). It is low and flat right up to the vowel *a* in the first syllable of the final word *sawa* where a clearly perceptible jump up to a higher

target occurs, followed by a rise in the final syllable of the unit. That is, F0 for the penultimate syllable *sa* is also somewhat misleading in that there is no perceptible fall in pitch.

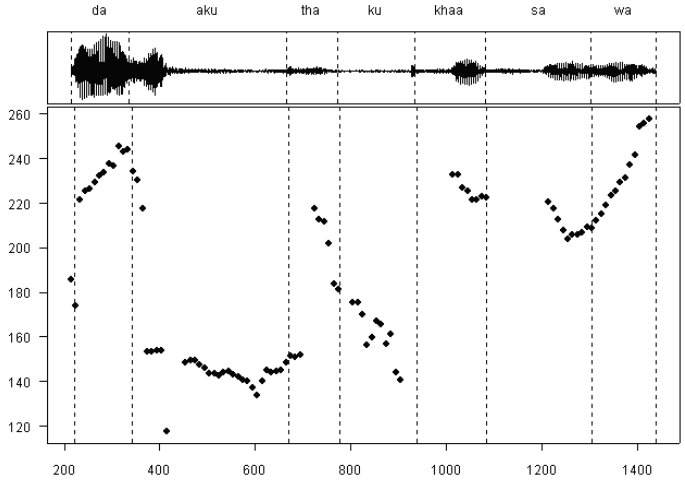


Figure 2: Waveform and F0 for example (3), female speaker ASB

Such major F0 excursions after glottalised and aspirated segments are of course well known from the literature on tonogenesis (cp. Beckman 1986:31f, Yip 2002:35-38). In this respect, it will be useful to note that in Waima'a there is considerable intra- and interspeaker variation with regard to the articulation of these segments. Figure shows another rendering of example 3 by the same speaker. Here, the initial *th* in *thaku* is rendered almost like a voiced stop and hence no upward F0 excursion occurs. Also, the jump after *kh* is much less prominent (less than 40 Hz).

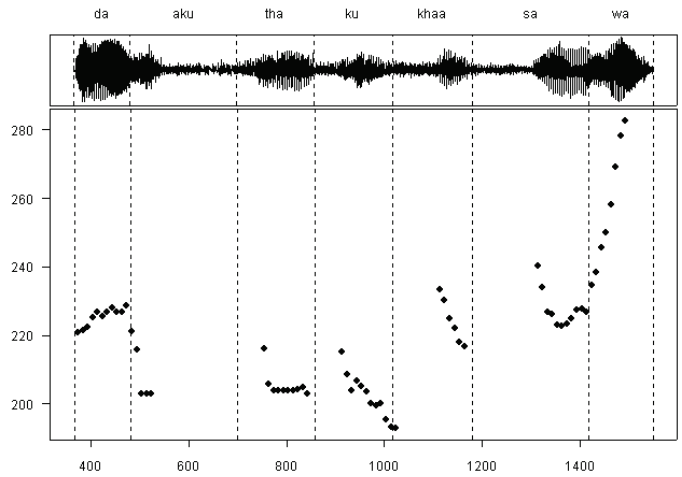


Figure 3: Waveform and F0 for example (3), alternative rendering by female speaker ASB

Figure 4 shows the same example rendered by a male speaker. Here the jumps in F0 after the aspirated initial consonants are rather moderate and much more like the minor jumps found after plain voiceless stops. (This speaker also uses a different intonation

pattern for this utterance, ending on a final fall rather than a rise, but this is not of import for the matters under consideration here.)

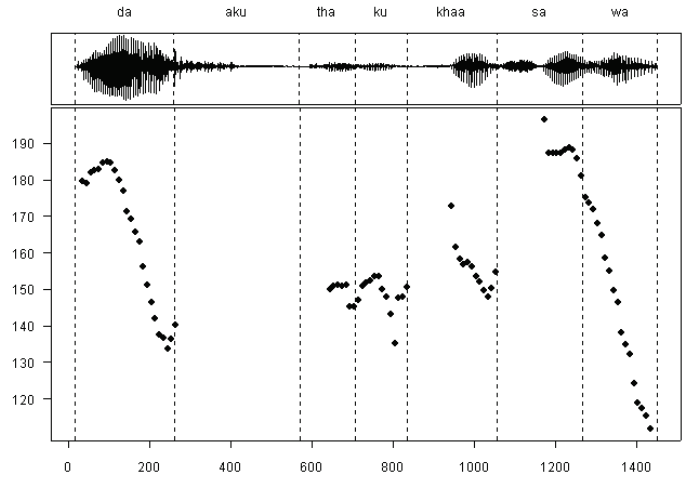


Figure 4: Waveform and F0 for example (3), male speaker MB

3 Tonal patterns at the right edge of an IU

3.1 The major, unmarked pattern: H-L%

The most common intonation pattern in the database begins mid-level in the speaker's current register, continues level along the middle range, rises on the penultimate syllable and ends on a final fall to the bottom of the current range. In terms of the autosegmental framework used here, this pattern can be analyzed as consisting of an unmarked or default initial boundary tone, followed by a H(igh) phrase accent on the penultimate syllable and a final L(ow) boundary tone, in short: H-L%. Figure 5 illustrates this contour on the basis of example (4). Note in particular that F0 is almost flat up to the penultimate syllable, where a clear rise occurs, reaching its peak towards the end of the syllable. The following fall is articulated with very strongly and quite abruptly decreased loudness, a widespread, but by no means universal feature in the database.

(4) *ne de kara haru lumu*
3s NEG like shirt reen
's/he doesn't like the blue shirt'

Before continuing the discussion of intonation contours, it will be useful briefly to take note of another major characteristic of Waima'a connected speech seen in Figure 5, that is a pervasive tendency to reduce syllables. In the current example (and also in the next one), the final vowel in *haru* 'shirt' is omitted. Syllable reduction, however, is not restricted to final syllables but may, in principle, occur in any syllable and may pertain to complete syllables rather than just to the nucleus, as further illustrated in section 6 below.

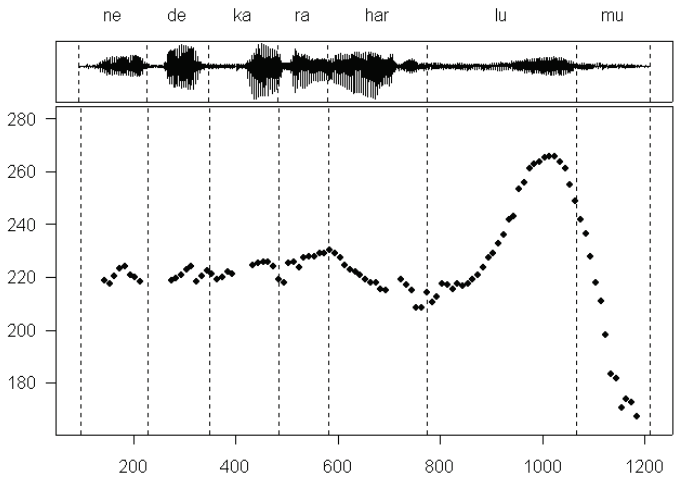


Figure 5: Waveform and F0 for example (4), female speaker SB

Returning to the discussion of tonal patterns at the right edge of an Waima'a IU, if the final word in an intonation unit ends on two identical vowels which are phonetically realised as a single long vowel (that is (C)V.V -> [(C)V:]), then phrase accent and boundary tone both occur on this phonetically long vowel, as illustrated by example (5).[5]

(5) *ne kara hile haru mee*
3s like again shirt red
's/he likes the red shirt'

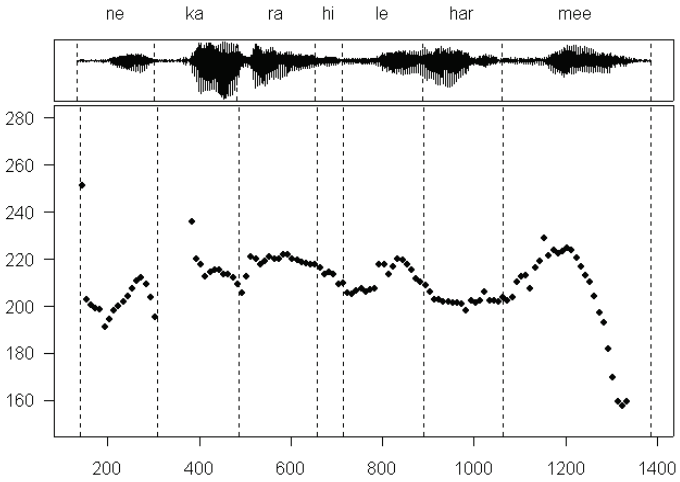


Figure 6: Waveform and F0 for example (5)5), female speaker SB

As seen in the preceding two examples (Figure 5 and Figure 6), the rise characterizing the phrase accent usually starts at the beginning of the penultimate syllable, if its onset is

[5] No syllable boundary is indicated in waveform and F0 in words involving a phonetically long vowel. Alternatively, one could analyze the contrast involved here in terms of an opposition between monomoraic and bimoraic syllables, making the mora rather then the syllable the tone-bearing unit in Waima'a. At this point it is unclear whether the analysis of phonetically long vowels as bimoraic or disyllabic provides for an overall simpler statement of Waima'a phonology (see Hajek and Himmelmann 2006 for further discussion).

voiced. If the onset of the penultimate syllable is voiceless, then there is no rise, but rather a jump to the high target, followed by the fall of the final boundary tone, which may already begin in the penultimate syllable. Compare example (6) and also example (3) and Figure 4 above.

(6) *karita soke mata lo kii la basara*
car crash dead ASP person LOC market
'(Oh, you haven't heard yet:) A car killed someone in the market!'

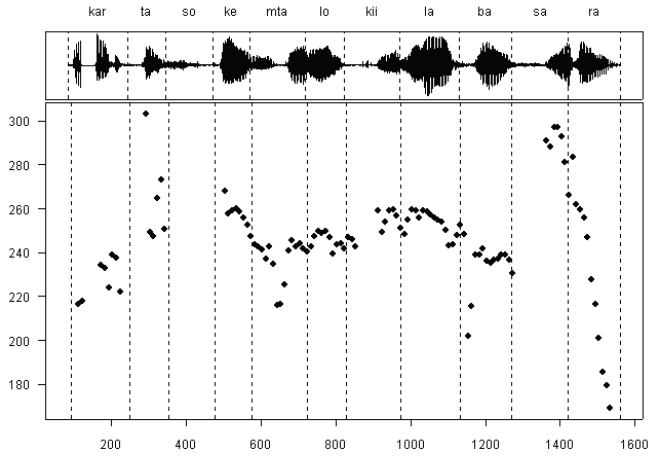


Figure 7: Waveform and F0 for example (6), female speaker ASB

Phrase-final function words appear to allow for two different realizations. Either the function word is realised as a short monosyllable, as in the short command *to'e di!* (sit ALLATIVE) 'sit down!' shown in Figure 8. Here the phrase accent occurs on the final syllable of the verb *to'e*, the allative marking *di* carrying the falling boundary tone.[6]

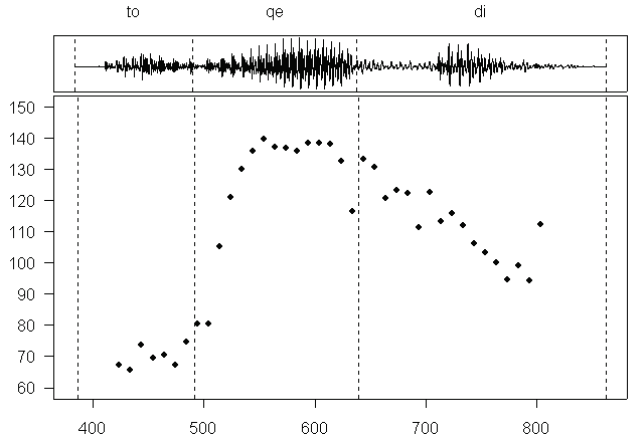


Figure 8: Waveform and F0 for *to'e di* 'sit down!', male speaker ACS[7]

[6] Allative *di* regularly occurs in some types of commands in Waima'a. It is currently still unclear whether this is best analyzed as extended multifunctionality of a high frequency function word or rather represents a case of homonymy which would then warrant an independent gloss as imperative marker.

[7] Note that glottal stop, which is represented orthographically as <'>, appears as <q> in this and some of the following figures due to limitations of the database (EMU) and plotting program (R) used.

Alternatively, the function word is realised with a (phonologically) long vowel, thus counting as disyallbic, in which case both phrase accent and boundary tone occur on the function word. This is illustrated by the second rendering of *to'e di!* in Figure 9, where both tonal targets occur on the considerably lengthened function word *di*.

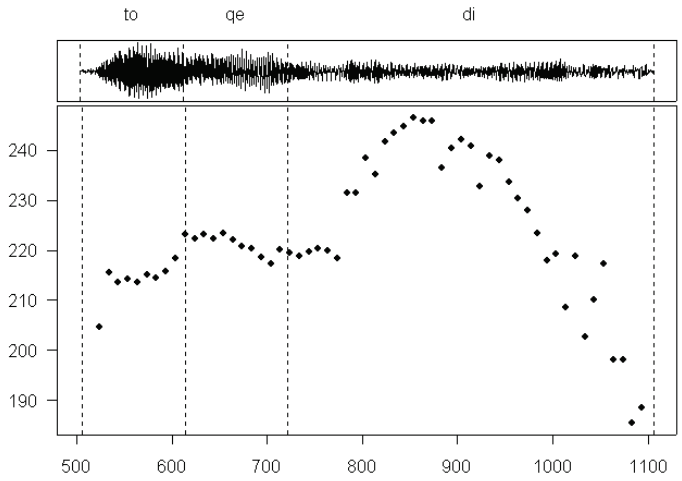


Figure 9: Waveform and F0 for *to'e di* 'sit down!', female speaker SB

At this point, it is not clear which factors determine the choice between these two options of rendering function words. Note that content words of the shape CV.V such as *khaa* 'eat', *kii* 'person' or *mee* 'red' also vary considerably in the length of the vowel, sometimes approaching the length of a short monosyllable. To date, however, no examples for an alternation similar to the one just described for function words has been observed for such content words when they occur in phrase-final position. Instead, when they occur phrase-finally, both phrase accent and boundary tone occur on them, as illustrated for *mee* in Figure 6 above.

As seen in the preceding examples, the basic pattern H-L% is used in simple declarative main clauses and in commands. In commands, the final fall is often cut somewhat short as seen in Figure 8.

3.2 Marked patterns at the right edge: H-H% and L-H%

In addition to the basic pattern discussed in the preceding section, two other patterns are well attested in the corpus. One of these is a minor variant of the basic pattern in that the H phrase accent is followed by a high boundary tone (H%) rather than a low one. This pattern occurs in wh-questions, if the wh-word occurs in final position as in example (7). (Note that here the intervocalic glottal is not realised, a widespread phenomenon in East Timorese languages.)

(7) *ka loo se'i?*
2s make what
'what are you doing?'

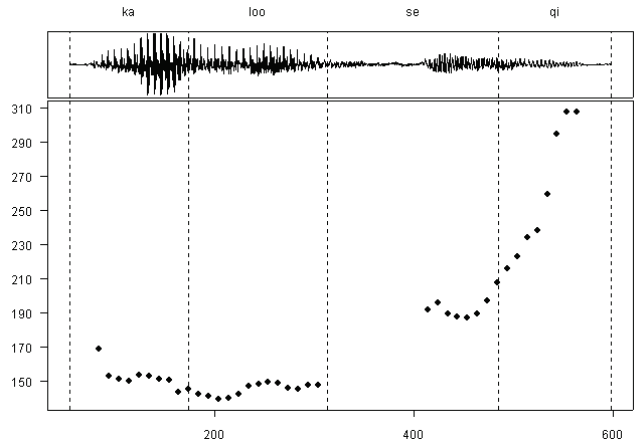


Figure 10: Waveform and F0 for example (7), male speaker MB

The most common use of H-H%, however, is as a 'continuer', signalling that the current utterance is part of a sequence which will be followed by more. Thus, it is the usual pattern found for preposed adverbial clauses (8), non-final members of lists (9) and the like.

(8) *antaun*[8] *ale'e* *anu-ata* *wuo-telu* *ke* *lheo* *la* *umo*
then child female CLF-three DEM arrive LOC house
'Then, when the three girls arrived at home, …'

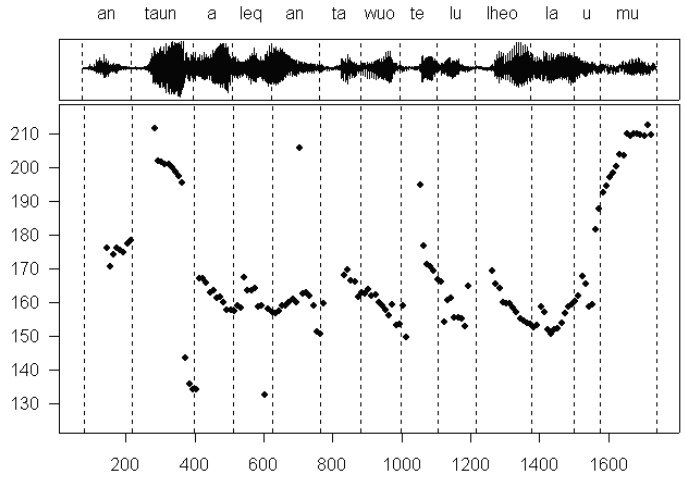


Figure 11: Waveform and F0 for example (8), male speaker JCB

(9) *do'e* *kareta::*
EXIST car
'he had a car, (a motorbike, an airplane, everything)'

[8] The initial temporal conjunction *antaun* here forms an IU by itself, as further discussed in section 5 below.

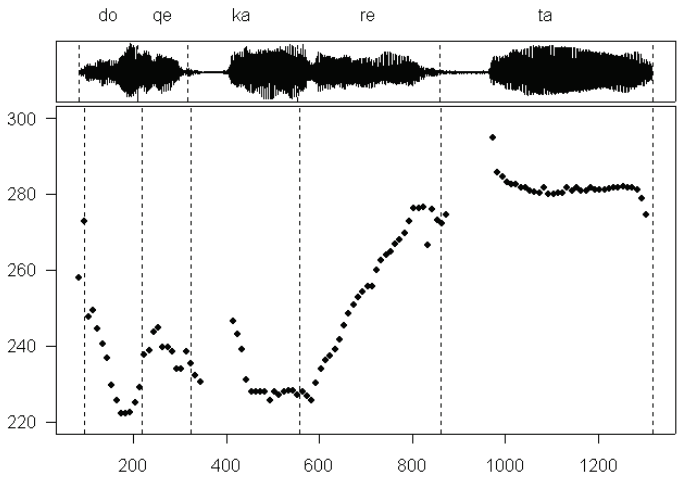


Figure 12: Waveform and F0 for example (9), female speaker OCB

While there is a general tendency for the final high boundary tone to level out flatly once the high target has been reached early in the final syllable, it should be noted that the very long flat stretch at the end of example (9) (on the final syllable of *kareta:*:; the colons here indicating conversational lengthening) is due to hesitation and not part of the H-H% pattern (the speaker is searching for further items to list in order to illustrate the wealth of the king the story is about). Lengthening the unit-final syllable with a flat pitch is the major hesitation or pause-filling strategy in western Austronesian languages (cp. Streeck 1996).

In contrast to the two patterns discussed so far, the third pattern to be introduced now involves a low phrase accent. That is, the pitch stays flat up to the final syllable on which a rather steep rise occurs. This pattern is analyzed here as a L- phrase accent followed by a H% boundary tone. It is the usual pattern for polar questions as in (10), among other functions.

(10) *Kii* *dai* *mai* *lo?*
person foreign come ASP
'has the foreigner arrived yet?'

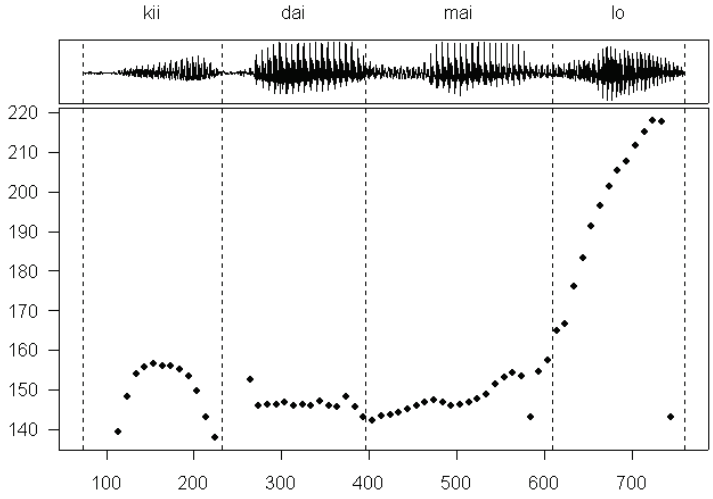


Figure 13: Waveform and F0 for example (10), male speaker JAJ

If the final two syllables consist of a vowel cluster ((C)V.V), pitch stays low in the first part of the cluster and only rises towards the end, as seen in (11). Compare Figure 14 with Figure 6 where there is a clear pitch movement also on the first part of a vowel cluster.

(11) *lonau aisa'i de mai?*
why yesterday NEG come
'why didn't you come yesterday?'

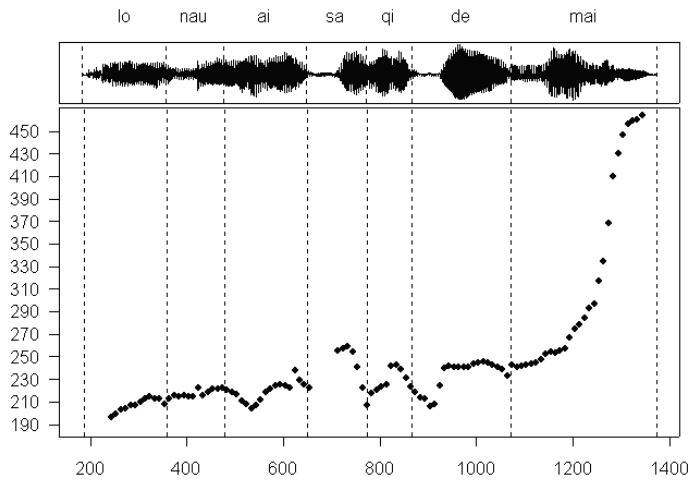


Figure 14: Waveform and F0 for example (11), female speaker JMB

As the preceding example shows, the L-H% pattern is not restricted to polar questions but may also occur with some types of information question, in particular those introduced with *lonau* 'why'. Information questions asking for the identity of participants with questions words for 'who' or 'what' usually involve another contour illustrated with (14) below.

The L-H% pattern also occurs in contexts which are widely known as 'deaccenting' contexts in the literature (e.g., Lambrecht 1994:248f, Ladd 1996:175-179), though this is not yet well understood. Thus, for example, apart from polar questions, the pattern is also attested in the corpus for utterances which contain answers to a preceding question and where, importantly, the answer largely consists of the same lexical material as the question, as in the exchange in (12).[9]

(12) *ga tabaku tasa lo?*
2s.POSS tobacco cooked ASP
'Is your tobacco ripe already?'

Da! aku de kore tabaku.
NEG 1s NEG plant tobacco
'No! I don't plant tobacco.'

[9] As seen very clearly in this example, the negation marker in Waima'a Caisido occurs in two variants *da* and *de*. *da* is used for emphatic negation and whenever the negation constitutes a speech act by itself (as when answering simply with *no!*). In other varieties of Waima'a the negative marker appears to be *da* in all environments.

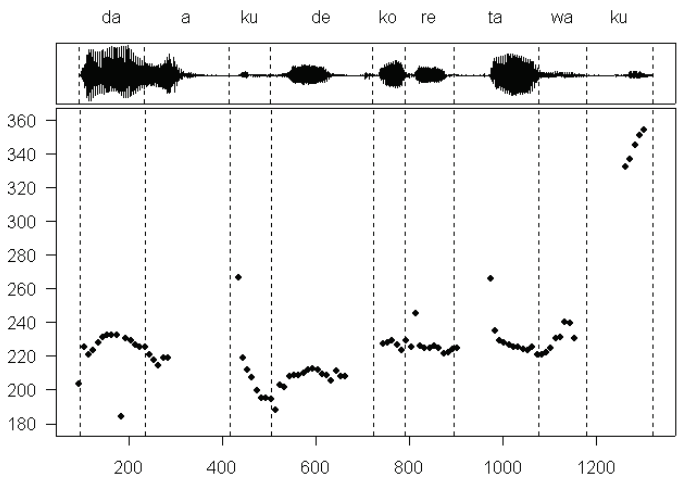


Figure 15: Waveform and F0 for answer part of example (12), female speaker ASB

Since in the preceding examples the pitch is mostly low and flat throughout the IU, it may not be immediately obvious why the pattern is analyzed as involving an initial low target (*L*-H%), rather than just as a simple high boundary tone (H%). This analysis is motivated by the fact that without this initial low target one would have to allow for the option that this pattern is realised by a continuous pitch rise throughout the IU (at least as one of a number of realizational alternatives) because the final H% would provide the only pitch target in the whole utterance. However, no pitch tracks with continuously rising pitch are attested in the database. Hence, the initial L in L-H% is required to ensure that the final rise only ever occurs on the very final syllable.

Further evidence for this analysis is provided by examples where a final L-H% follows an initial rise, as in Figure 17 below (= example (14)). In this example, pitch rises to a high target early in the IU, then falls continuously throughout, and sharply rises again only in the very last syllable. Without assuming an initial low target as part of the final pattern, this pitch track could not be explained, because then pitch would be predicted to stay high throughout the entire unit.

A more difficult point to argue is the question of whether this pattern indeed involves a phrase accent (L-) on the penultimate syllable. The alternative would be to claim that it simply involves a single final boundary tone LH% on the final syllable. Clear phonetic evidence to decide this issue is not available at this point.[10] The current analysis is based on the purely phonological consideration that it allows to preserve the generalization that all IU-final patterns in Waima'a involve the combination of a phrase accent and a boundary term. Further research may show this to be an overgeneralization.

[10] Unlike the H- phrase accents in the H-L% and H-H% patterns which clearly render the penultimate syllables prominent to the ears of a Germanic speaker, there is no such clearly perceptible prominence on the two final syllables of the L-H% pattern. This pattern rather sounds like a strong rise on a final unaccented syllable in a Germanic language. But this would appear to be exactly the kind of example where a non-native speaker analyst tends to become a victim of her or his native prosodic system. Consequently, one important piece of evidence in this regard would come from perceptual testing with native speakers.

3.3 Summary: tonal targets at the right edge

So far, the following three patterns for tonal targets at the right edge of a Waima'a IU have been described:

H-L%	declarative main clause with unmarked information structure
H-H%	wh-question with wh-word in final position; 'continuing', that is non-final unit in a sequence (as in lists and clause chains)
L-H%	polar question, exclamation, 'deaccenting' (response which repeats major lexical items from preceding utterance)

These three patterns account for a substantial number of the IUs in the corpus, but not for all of them. They all have a H target at the end of an IU, either on the penultimate or ultimate syllable or both. However, the corpus also contains a number of examples where no such H target is discernible, as in Figure 16.

(13) *aku soru kai*
1s plane wood
'I am planing wood' (answer to question 'what are you doing?')

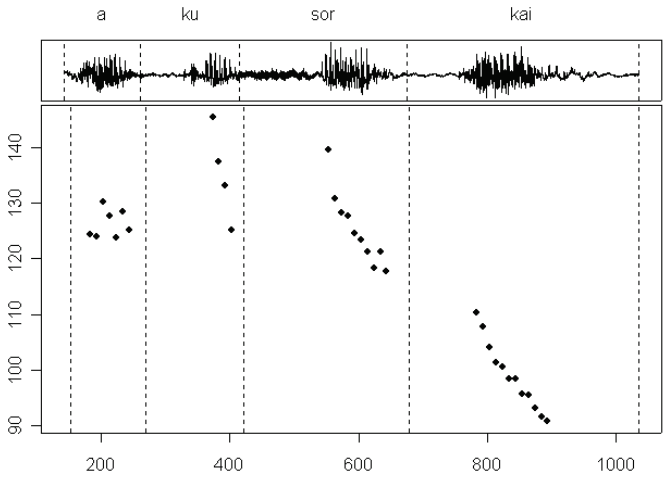


Figure 16: Waveform and F0 for example (13), male speaker DB

Here pitch appears to fall continuously throughout the IU after an initial H on the first word (this initial H is discussed in the next section). This fall could be analyzed as involving a low phrase accent (L-) followed by a low boundary tone (L%), thus resulting in a nicely symmetrical system, involving two phrase accents and two final boundary tones, which can be freely combined with each other. But apart from the fact that the rather few examples for a possible L-L% pattern found so far involve a large amount of segmentally caused perturbations, which considerably complicates the intonational analysis, it is also unclear what the functional distribution of this pattern would be.

4 Intermediate level phrases?

Most IUs in the database start somewhere in the middle range of the current register. This is analyzed here as the default option for which no special tonal target needs to be

identified. All the preceding examples, with the exception of examples (6) and (13),[11] illustrate this default option. The exceptions involve a rise on the initial word also seen in the following example:

(14) *sie* *ne* *lau* *aku?*
who FOC look.for 1s
'who is looking for me?'

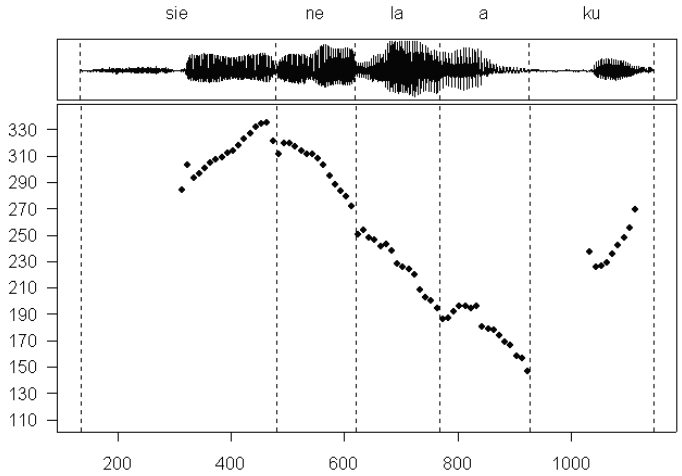


Figure 17: Waveform and F0 for example (14), female speaker S

In this example, the pitch rises on the first word and then continuously falls to the L(ow) target of the final L-H% pattern (as already discussed at the end of section 3.2). There are two possible analyses for this initial rise. One option is an initial boundary tone (%H). Alternatively, one could consider this to involve a lower-level prosodic constituent such as an accentual or intermediate-level phrase, marked by a final boundary tone of its own (H-). In his analyses of Manado Malay and Banyumas Javanese, Stoel (2005, 2006, 2007) makes considerable use of such smaller constituent units, and it appears likely that this is a more widespread phenomenon in western Austronesian languages.

Once again, however, the data set used for this preliminary report lacks the crucial data needed to decide this issue. All clear examples of major pitch excursions not occurring at the end of an IU involve the kind of initial rise seen in example (14) above, and the rise always occurs on the first word only, which always constitutes a complete syntactic constituent (for example, an NP). Hence it remains to be seen, whether the rise is strictly aligned to the initial boundary (and hence a phrase-initial boundary tone) or to the end of a phrase-initial constituent (and hence marking the end of an intermediate-level phrase).

Initial rises regularly occur when wh-words are phrase-initial (as in example (14)). Similarly, the negative imperative marker *deme'e* 'don't!' is prosodically made prominent in this way. An initial rise is also found when the first word contains new information (as in example (6)). However, it is not confined to instances where the unit-initial word arguably constitutes, or at least belongs to, the focus domain of the clause. Example (13) is an example where the initial word (*aku* 'I') is not focal, and there are quite a few examples of this type in the database (another one is in Figure 18 below). Thus, while the

[11] Note that the negation marker *da* occurring initially in some of the other examples is always realised in an IU of its own, usually with the H-L% pattern.

first-mentioned uses would support the intermediate-level phrase analysis, the overall functional distribution of the initial rising pattern also does not provide unequivocal evidence for either analysis. In the following discussion of prosodic chunking (in section 5), we will employ the intermediate-level phrase analysis primarily for reasons of ease of exposition. There is no evidence in the database for major non-final pitch excursions other than this initial rise.

5 Prosodic chunking

As in most other languages, Waima'a IUs are typically co-extensive with grammatically definable units such as clauses, noun phrases or prepositional phrases. The mapping between prosodic and grammatical units has not yet been investigated in detail, but it will be useful to note that speakers have to make choices as to how much segmental material they include in a single IU. Thus, the same speaker rendered the following example in two quite different ways:

(15) *tamba* *aisa'i* *aku* *bira*
because yesterday 1s sick
'because I was sick yesterday'

Figure 18 shows the rendering of example (15) in a single IU, with an initial intermediate level phrase ending on H- and a final H-L% (that is [[tamba H-] aisa'e aku bira H-L%]). In Figure 19, on the other hand, the example is chunked into three IUs, the first two of which end on a high boundary tone, the last one on L%:

(16) [tamba H-H%] [aisa'e H-H%] [aku bira H-L%]

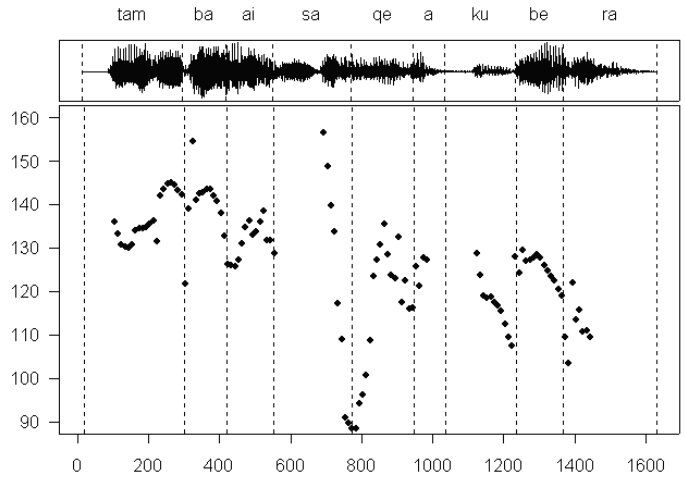


Figure 18: Waveform and F0 for example (15), single IU, male speaker ACS

In some instances, especially with regard to clause initial conjunctions, it is not always straightforward to decide whether the unit begins with an initial intermediate level phrase or whether the conjunction is presented in an IU of its own. Thus example (8) and Figure 11 above could be analyzed in two ways:

(17) a) [antaun H-H%] [ale'e anu-ata wuo-telu ke lheo la umo H-H%]
b) [[antaun H-] ale'e anu-ata wuo-telu ke lheo la umo H-H%]

In this example, the main reason for choosing analysis (a) (that is two IUs) pertains to the fact that there are quite clear indications for a new onset after *antaun*. Among other things, there is a considerable jump down in pitch after *antaun*, quite similar to the jumps seen after *tamba* and *aisa'i* in Figure 19. That is, the lack of onset phenomena, including the impossibility of a (short) pause at the boundary, distinguishes intermediate-level phrase boundaries from the kind of IU boundaries illustrated here with Figure 19.

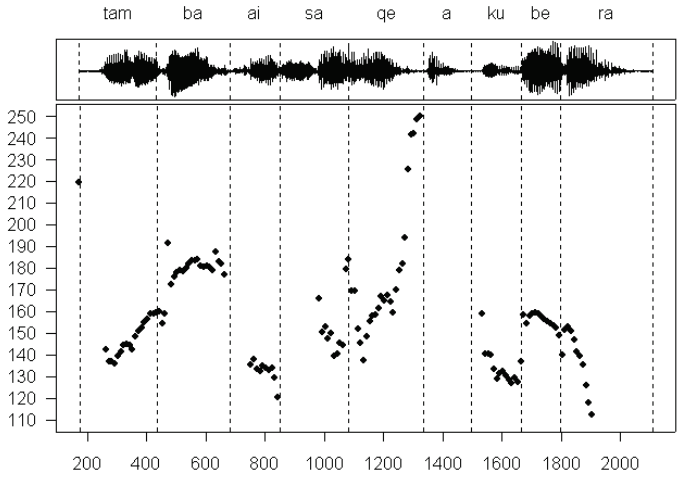


Figure 19: Waveform and F0 for example (15), 3 IUs, male speaker ACS

6 What about lexical accent?

Viewed from the perspective of Standard Average European, the most remarkable feature of the Waima'a intonational system as sketched in the preceding paragraphs is the apparent absence of accentual tones (or pitch accents), that is intonational pitch changes which are linked to a lexically accented syllable, usually represented by T* in the autosegmental-metrical analysis. This could mean that lexical accents are 'ignored' for intonational purposes, an option further discussed in the following section. Alternatively, Waima'a could lack lexical accent altogether and in this regard would be similar to languages such as French, Javanese or many varieties of Malay.[12] While it is not yet possible to provide a definitive solution to this issue here, there is some evidence which supports the latter alternative. Thus, when looking at words recorded in isolation, on first appearance it would seem that they all are regularly accented on the penultimate syllable, as seen in Figure 20 for the word *ria-kuko* 'nighttime'.

[12] See further below section 7. The status of a genuinely lexical accent in French continues to be a matter of debate (as opposed to the uncontroversial regular phrase accent). See Hayes (1995:24), Di Cristo (1998) and Gussenhoven (2004:Chapter 13) for pertinent comments and references.

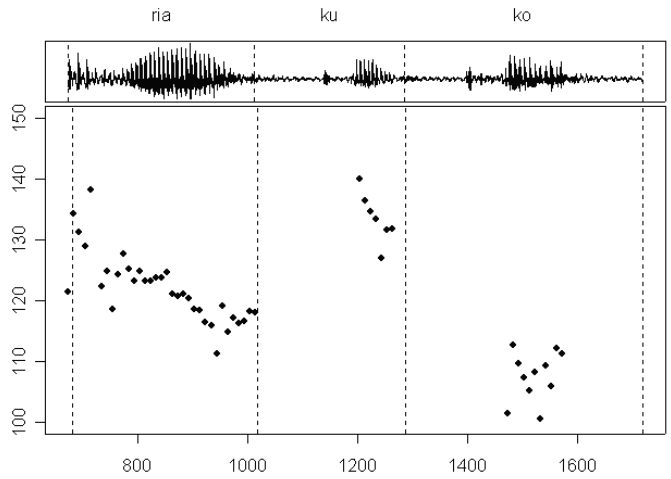


Figure 20: Waveform and F0 for *ria-kuko* 'nighttime', male speaker MB

Here the penultimate syllable *ku* is clearly made prominent acoustically as well as auditorily because of the major pitch rise which occurs on it. Note, however, that this syllable is auditorily not prominent in terms of duration.

What is more, this syllable can easily be omitted, as seen in the alternative rendering of *ria-kuko* shown in Figure 21. Cross-linguistically, various types of reduction, including complete omissability, are generally a characteristic of unaccented syllables, while accented syllables, if at all omissible, can be reduced or omitted only in very specific circumstances.[13] In Waima'a, in principle any syllable in a given item can be reduced or omitted, and while the factors governing the reduction processes are still unclear, it is clear that no syllable is exempt from these processes (that is reduction processes do not provide direct evidence for lexical accent).

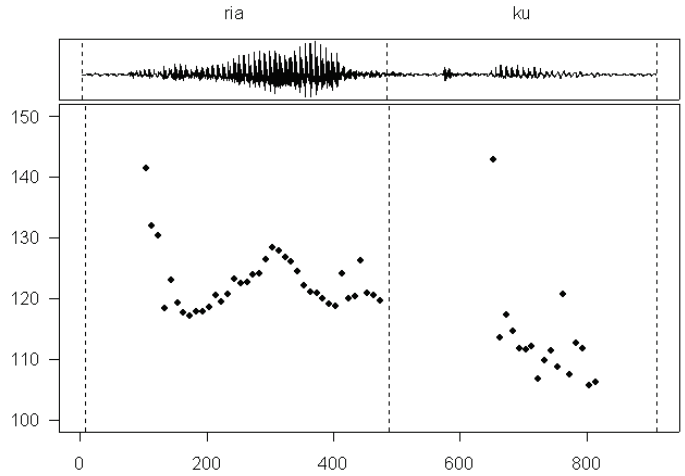


Figure 21: Waveform and F0 for *ria-kuko* 'nighttime', male speaker MB

[13] An instance of a rather narrowly delimited possibility for omitting a stressed syllable nucleus occurs in some forms of gemination found in Oceanic languages, as discussed by Blust (2001). Note that in Waima'a the complete syllable can be omitted, not just its nucleus. Furthermore, gemination is not of relevance here.

Taking these observations together, it becomes clear that the pitch rise which lends prominence to the penultimate syllable in Figure 20 is due to utterance-level intonation since words in isolation of course also have to be produced with some kind of utterance-level intonation. Hence, what Figure 20 and Figure 21 both actually show is the unmarked declarative pattern H-L% extensively exemplified in section 3.1 above.

That this is the correct interpretation is supported by a comparison of different renderings of the same word in isolation and within an IU in non-final position. Figure 22 shows the word *basara* 'market' spoken in isolation, with a clear high pitch target in the penultimate syllable *sa*. In Figure 23 the same word occurs as part of the utterance *laka to basara di* (go already market ALL) 'go to the market!'. Here, all syllables of *basara* are of almost equal pitch height, *sa* being in no way more prominent than either the preceding or the following syllable.[14] The unit-final rise-fall (H-L%) here occurs on the allative particle *di* (cp. also the discussion for Figure 8 and Figure 9 above).

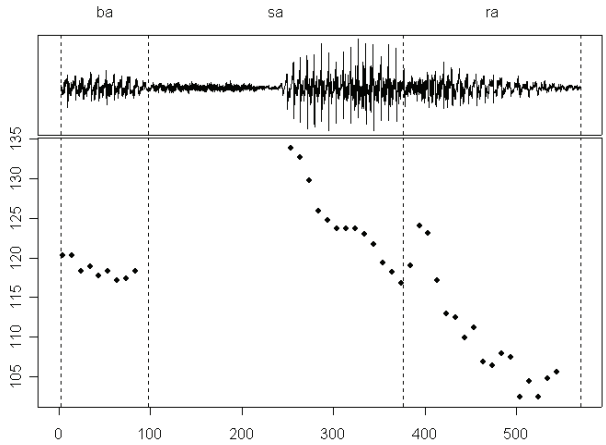


Figure 22: Waveform and F0 for *basara* 'market', male speaker MB

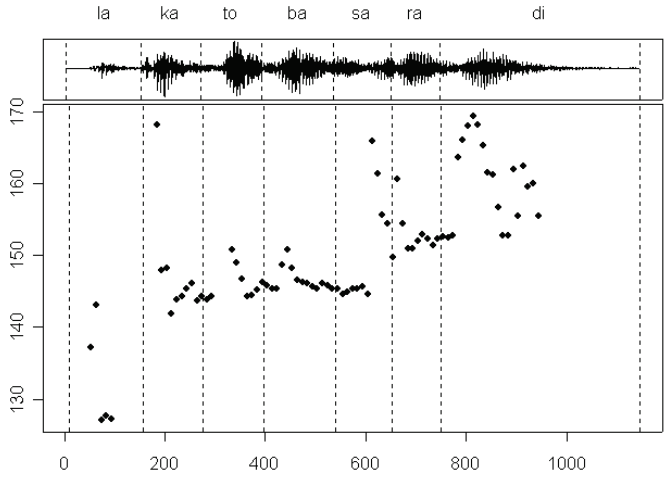


Figure 23: Waveform and F0 for *laka to basara di* 'go to the market!', male speaker MB

[14] The very minor pitch differences visible in Figure 23 are due to segmentally induced perturbations and are not perceivable auditorily.

An alternative explanation for the lack of prominence for *basara* observed in Figure 23 would be to claim that Waima'a words regularly loose their lexical accent when occurring in phrase-medial position. Such a claim, however, would only make sense if independent evidence for lexical accent in Waima'a could be found. To date, no prosodic phenomenon has come to light which would be strongly suggestive of lexical accent, but this is an issue which needs further research. Among other things, it will be necessary to have a closer look at how Waima'a accommodates the different accentual patterns occurring on Portuguese loanwords, which make up a substantial part of the Waima'a lexicon.

7 Conclusion

While still preliminary and incomplete, the analysis sketched in the preceding sections suggests that the Waima'a prosodic system is characterised by the following three core characteristics:

1. It has no lexical tone.
2. It has no lexical accent and hence also no intonational pitch accent.[15]
3. Intonation units are primarily characterised by a phrase accent on the penultimate syllable and a phrase-final boundary tone .

While the lack of lexical tone contrasts is not an unusual feature in crosslinguistic perspective, so far the claim that a language lacks both lexical tone and lexical accent has not been made very often. However, as already noted in the preceding section, it has been made a number of times for languages spoken in the western parts of Indonesia, including various varieties of Malay (van Zanten et al. 2003, Tadmor 2000, 2001, Stoel 2007) and Banyumas Javanese (Stoel 2006). Tadmor (2000, 2001) in fact speculates that the lack of lexical accent is a widespread feature of languages in western Indonesia (extending, roughly, from Sumatra to Bali and including Kalimantan), while eastern languages often have regular penultimate lexical accents. The case of Waima'a suggests that the western pattern may actually occur as far east as East Timor.

In her survey of prosodic systems, Jun (2005a:444) lists French, Bengali and Korean as languages lacking lexical accents as well as lexical tone, which shows that this phenomenon is not confined to Indonesia and East Timor. However, at this point at least, it is not clear that these three languages have much else in common with regard to their prosodic systems. Similarly, there are many obvious differences between these systems and the Waima'a system. The domain for phrase accents in French, for example, is much smaller than in Waima'a, often only containing a single word or even just a part of it. Consequently, phrase accents in French are of much higher frequency in spontaneous speech than the Waima'a ones. Korean, on the other hand, has a much more complex system of both initial and final boundary tones delimiting accentual phrases which also are units of considerably smaller size than the Waima'a IU. While it is much too early to tell, this raises the suspicion that lack of lexical accent and tone are possibly not very important typological parameters, but instead define a rather heterogeneous class of languages which happen to lack these features.

---

[15] The implication of course only holds if (intonational) pitch accent is defined as a change in pitch aligned with a lexically stressed syllable. Gussenhoven (2004:Chapter 13) analyses French as a language without lexical accent ('an intonation-only language', 2004:253), but with variable 'pitch accents'.

Instead, it may be more interesting and promising to compare the Waima'a system with other languages where apart from a limited number of boundary tones, at most one major pitch change occurs in units of a similar size as IUs in Waima'a. The Papuan language Kuot (Lindström 2002, Lindström and Remijsen 2005) and the West African language Wolof (Rialland and Robert 2001) have been described in this way. For both languages, it is claimed that at least some lexical items have lexical accents, the primary phonetic exponent of which is duration, but that this accent is 'ignored' by intonation, as Lindström and Remijsen put it in the title of their joint paper on Kuot. That is, unlike in Standard Average European, this lexical accent does not interact with intonation. Most importantly, it does not attract the pitch changes characterizing different intonational contours. Instead and as in Waima'a, Kuot and Wolof intonation patterns only involve pitch changes which are definable with regard to the boundaries of an IU, the major events generally occurring on the unit-final syllable or syllables. Consequently, at least at first glance and without any deeper probing, the structure of intonation units is quite similar across the three languages. Most importantly, the contours of the most frequent and unmarked type in each language are essentially flat, with major pitch changes being confined to the last one or two syllables.

Similarly, and returning to the Indonesian archipelago, it would seem obvious that Manado Malay (as analysed in Stoel 2005 and 2007) on the one hand, and Banyumas Javanese and Waima'a on the other have much more in common with each other than any of them with prototypical European lexical-accent languages such as English, German or Dutch, despite the fact that Manado Malay is said to have lexical accents. The common feature again pertains to the fact that accentual pitch excursions in Manado Malay may occur at most once per IU so that – oversimplifying a bit and ignoring marked intonational patterns – one could hypothesise that a single major pitch excursion per intonation unit constitutes one of the core features of the intonation systems of a larger number of languages in the area. It remains to be seen whether the further segmentation of intonation units into lower level accentual or intermediate-level phrases, as proposed by Stoel for Manado Malay as well as Banyumas Javanese, constitutes another characteristic areal feature (for Waima'a, see the preliminary remarks in section 4 above). If so, the typical size and frequency with which IUs are segmented into such smaller units would constitute the major parameter for variation across the area.

References

Beckman, Mary E., 1986. *Stress and non-stress accent*. Dordrecht: Foris.

Blust, Robert A., 2001. Some remarks on stress, syncope, and gemination in Mussau. *Oceanic Linguistics*, 40:143–150.

Di Cristo, Albert, 1998. Intonation in French. In Hirst and Di Cristo, eds, 195–218.

Gussenhoven, Carlos, 2004. *The phonology of tone*. Cambridge: Cambridge University Press.

Hajek, John and John Bowden, 2002. A phonological oddity in the Austronesian area: ejectives in Waimoa. *Oceanic Linguistics*, 41:222–224.

Hajek, John and Nikolaus P. Himmelmann, 2006. Notes on Waima'a phonology and orthography. Waima'a DoBeS archive, <http://www.mpi.nl/DOBES/>.

Hajek, John and Mary Stevens, 2005. On the acoustic characterization of ejective stops in Waimaa, *INTERSPEECH-2005*, 2889–2892.

Hayes, Bruce, 1995. *Metrical stress theory*. Chicago: The University of Chicago Press.

Himmelmann, Nikolaus P., 2006. Prosody in language documentation. In Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel, eds, *Essentials of language documentation*, 163–181. Berlin: Mouton de Gruyter.

Himmelmann, Nikolaus P. and Robert D. Ladd, 2008. Prosodic description: an introduction for fieldworkers. *Language Documentation and Conservation*, 2:244–274.

Hirst, Daniel and Albert Di Cristo, eds, 1998. *Intonation Systems*. Cambridge: Cambridge University Press.

Jun, Sun-Ah, 2005a. Prosodic typology. In Sun-Ah Jun, ed., 430–458.

Jun, Sun-Ah, ed., 2005. *Prosodic Typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Ladd, D. Robert, 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

Lambrecht, Knud, 1994. *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Laver, John, 1994. *Principles of phonetics*. Cambridge: Cambridge University Press.

Lindström, Eva, 2002. Topics in the grammar of Kuot, a non-Austronesian language of New Ireland, Papua New Guinea. Stockholm: Stockholm University (available online at http://www.ling.su.se/STAFF/evali/thesis/Kuot-PhD.html). PhD dissertation.

Lindström, Eva and Bert Remijsen, 2005. Aspects of the prosody of Kuot, a language where intonation ignores stress. *Linguistics*, 43:839–870.

Rialland, Annie and Stéphane Robert, 2001. The intonational system of Wolof. *Linguistics*, 39:893–939.

Stevens, Mary and John Hajek, 2004. A preliminary investigation of some acoustic characteristics of ejectives in Waima'a: VOT and closure duration, *Proceedings of the Tenth Australian International Conference on Speech Science and Technology,* Sydney, 277–282. Available online at http://www.assta.org.

Stoel, Ruben B., 2005. *Focus in Manado Malay*. Leiden: CNWS Publications.

— 2006. The intonation of Banyumas Javanese, *Proceedings Speech Prosody 2006*, 827–830. Dresden: TUD Press.

— 2007. The intonation of Manado Malay. In Vincent J. van Heuven and Ellen van Zanten, eds, *Prosody in Indonesian Languages*, 117–150. Utrecht: LOT.

Streeck, Jürgen, 1996. A little Ilokano grammar as it appears in interaction. *Journal of Pragmatics*, 26:189–213.

Tadmor, Uri, 2000. Rekonstruksi Aksen Kata Bahasa Melayu (The Reconstruction of Word Accent in Malay). In Bambang Kaswanti Purwo and Yassir Nasanius, eds, *PELBBA 13 (Pertemuan Linguistik (Pusat Kajian) Bahasa dan Budaya Atma Jaya Ketiga Belas)*, 153–167. Jakarta: Unika Atma Jaya.

— 2001. Phonological Convergence in the Malay Archipelago. Paper presented at the Annual Meeting of the Deutsche Gesellschaft für Sprachwissenschaft (DGfS) in Leipzig.

Yip, Moira, 2002. *Tone*. Cambridge: Cambridge University Press.

Zanten, Ellen van, Robert W.N. Goedemans and Jos J. Pacilly, 2003. The status of word stress in Indonesian. In Jeroen M. van de Weijer, Vincent J. van Heuven and Harry G.

van der Hulst, eds, *The Phonological Spectrum II*: *Suprasegmental Structure*, 151–175. Amsterdam: Benjamins.

4 *Prefixation of arguments in West Papuan languages*

GER REESINK[1]

1 Introduction

The term 'West Papuan', as used in this chapter, is meant to be not more than a geographic label.[2] It refers to the heterogeneous fifty Non-Austronesian languages spoken in the western region of the wider Papuan area (Reesink 2005 and 2006). The Papuan expanse stretches from the islands Timor, Alor and Pantar of Nusantara in the west, to the Solomon Islands in the east (see for example Foley 1986, 2000, Pawley 2007 and Ross, 2005). The term 'Papuan' is widely used for Non-Austronesian languages of the greater New Guinea area, but implies no genealogical relationship (Foley 1986, 2000, Ross 2005). Nevertheless, it can be reasonably assumed that all Papuan languages (of whatever stock) originally had SOV word order.[3] At present there are only three areas where Papuan languages with constituent order other than SOV are found: in the western region, some languages of North-Halmahera and most of the Bird's Head, most languages of the Torricelli family north of the Sepik River, and in the eastern Papuan region, the languages of the Bismarck Archipelago and Bilua of the Solomon Islands.

Although the West Papuan languages are divided into a number of unrelated families and isolates (Reesink 2005:186–187; see Appendix for map and affiliations of languages

[1] The research of this chapter has been carried out as part of the Spinoza research program Lexicon and Syntax, under the direction of Pieter Muysken (Radboud University Nijmegen) at Leiden University, and more recently, the NWO programme 'Breaking the time barrier: Structural traces of the Sahul past', under direction of Pieter Muysken and Steve Levinson (Max Planck Institute for Psycholinguistcs, Nijmegen). I am very grateful to Louise Baird, František Kratochvíl and Marian Klamer for their permission to use their unpublished data. I thank Louise Baird, Michael Dunn, Gary Holton, Marian Klamer, and Carl Rubino for comments on an earlier version. Although they may not agree with the final product, their comments have helped my thinking.

[2] The term 'West Papuan' may call to mind a putative genealogical grouping, as proposed by H. K. J. Cowan (1957) and A. Capell (1975), but it is used here in analogy to East Papuan, likewise a term originally used by Wurm for a putative 'phylum'. Just as the term 'Papuan' does not suggest a genealogical unity, the adjectives west and east are only meant as geographic labels. Whether some distant relationship, either genealogically or typologically, exists between these heterogeneous groups remains a matter of further research, see further the conclusion.

[3] See Mark Donohue (2005) for an alternative view.

Michael C. Ewing and Marian Klamer, editors.
Typological and areal analyses: contributions from East Nusantara, 71–95. Pacific Linguistics, 2010
© This edition vested with Pacific Linguistics.

discussed in this chapter), all of them are somewhat typologically unusual in their verbal affixation. World-wide, languages have a preference for suffixing (Dryer 2005), but all of these West Papuan languages have prefixes to indicate subject and/or object on the verb and possessor on the possessed noun. Except for Inanwatan of the South Bird's Head (de Vries 2004), none of these Papuan languages has verbal marking for tense. A number of Papuan languages of the Bird's Head and some languages of the North Halmahera family (for example, Sahu, West Makian), as well as some of the Timor-Alor-Pantar family (for example, Adang and Klon) also lack verbal morphology for aspect and mood. Yet all of them mark at least one core argument by verbal prefix, with the exception of Abun (Berry and Berry 1999), which only employs full noun phrases or free pronouns for subject and object. This means that a sizeable number of West Papuan languages do not follow the universal tendency (Enrique-Arias 2002:11) that 'presence of inflectional subject agreement in a language implies the presence of either aspect, tense, or mood'.

Yet another tendency, namely, that no language has object agreement that does not also have subject agreement (Siewierska and Bakker 1994), is contradicted by genealogically unrelated SOV languages in the West Papuan area. As we will see, the languages of the South Bird's Head family and most of the North Halmahera languages do indeed have both object and subject agreement. But languages of different genealogical groups, Alor-Pantar and Yawa, only have an object prefix on the verb. These languages (for example, Yawa and some of Alor-Pantar) together with some North Halmahera languages, have a split-S system: roughly speaking, active events that are under control have S marked as A (the agent-like argument of a transitive verb), while stative verbs or dynamic verbs with clear affect on the sole argument have S marked as UNDERGOER, that is, by means of an object prefix.

The position of verbal and possessive affixation is clearly congruent with the existing constituent order in the clause. The morphological shape of the prefixes is closely related to the free pronouns found in these languages. In addition, in OV languages of this area the possessive prefixes are formally the same or very similar to object prefixes; while in VO languages, there is a remarkable similarity between possessive prefixes on the noun and subject prefixes on the verb. These facts call to mind Givón's (1971:413) well known claim 'Today's morphology is yesterday's syntax'.

In this chapter I will present verbal and possessive marking in the various West Papuan families, firstly to show how some general tendencies are only partly followed in this geographic area, and secondly to illustrate the formal similarity between possessive prefixes with either object or subject prefixes on verbs; thirdly the exposition of object prefixes will include a cursory discussion of so called Split-S systems in different groups of Papuan languages.

Section 2 will discuss SOV languages that have only object prefixes, which are identical to possessor prefixes. Section 3 will present SOV languages that have both subject and object prefixes, in this order, on verbs. Section 4 is concerned with Papuan SVO languages that have only subject prefixes. The conclusion will summarize some general typological tendencies and comment on the origin of morphology.

2 SOV languages and o-V verbal affixation

In this section I will review languages that only mark object (UNDERGOER) as a prefix on the verb. Actor subjects are realized as noun phrase or free pronoun. Undergoer prefixes may also be used on uncontrolled or involuntary events and states. Furthermore,

obligatorily (or: inalienably) possessed nouns employ prefixes that are identical to (one of) the undergoer prefixes, see also (Klamer and Kratochvíl 2006). The languages that are characterized by the combination of features given in the title of this section concern the Alor-Pantar (AP) languages[4] and Yawa, spoken on Yapen Island in the Cenderawasih Bay.

The Papuan languages of Timor, Alor and Pantar are clearly genealogically related, though perhaps divided into a few subfamilies. On the basis of their pronominal forms, as well as the use of ablaut for number (*a* for singular; *i* for plural), they appear to belong to the widespread Trans New Guinea family (Ross 2005: 35).

For at least some AP languages the notion of object is rather complex, for example, in Abui (Kratochvíl 2007) and Klon (Baird 2008) three (Abui) or four (Klon) different UNDERGOERS have to be distinguished on the basis of the prefix vowel. While the choice of the prefixes on transitive verbs is largely lexically determined in Klon, some verbs allow alternative markings, with concomitant semantic changes. In Abui some transitive verbs can take any of the three prefixes. It is impossible to do justice to the complexity of undergoer marking in these languages here, but I will provide some examples to illustrate:

(i) the free pronoun used to express the actor subject,

(ii) the undergoer as prefix on a transitive verb,

(iii) the undergoer prefix as the only argument of an uncontrolled predicate and

(iv) how some undergoer prefixes are identical to the possessive markers for inalienable and alienable possession.

In Abui the vowel *a* is characteristic of singular free pronouns, marking the subject of a transitive verb, or the only argument of a controlled intransitive, as in (1) and (2), and of prefixes indexing the UNDERGOER, as most affected participant, as in (3).[5]

(1) *Na bataa tukong.*
1s(A) wood cut
'I cut wood.'
Abui (Kratochvíl 2007:79)

(2) *Na furai.*
1s(A) run.CONT
'I am (was) running.'
Abui (Kratochvíl 2007:202)

(3) *Naana di na-wel .*
older.sibling 3s(A) 1s(U)-wash
'My older sibling bathes me.'
Abui (Kratochvíl 2007:204)

The same prefix is also used to mark the only argument of some unaccusative verbs, as in (4), and the possessor of obligatorily marked possessed nouns, as in (5).

---

[4] On the advice of Marian Klamer I restrict my comments to the Papuan languages spoken on these two islands, rather than refering to the whole Timor-Alor-Pantar group, since it is yet unclear how the Papuan languages of Timor relate to those in Alor and Pantar. Secondly, Gary Holton (personal communication) informs me that some TAP languages, notably Lamma on Alor, permit subjects to be referenced as prefixes as well.

[5] Throughout this chapter, where necessary, I indicate subject prefixes as (A) and object as (U).

(4) Na-lal-e Na-rik.
1s(U)-laugh-IPFV 1s(U)-ill
'I am laughing.' 'I am ill.'
Abui (Kratochvíl 2007:204)

(5) Na-ne na-min.
1s(U)-name 1s(U)-nose
'My name.' 'My nose.'
Abui (Kratochvíl 2007:141,143)

A second UNDERGOER requires a prefix with *e*. This references a less affected participant, such as theme, location and benefactive, as in (6). It is also used to reference the possessor in alienable constructions, as in (7).

(6) *Di tifi do he-buka he-faaling.*
3A television PROX 3s.LOC-switch.on 3s.LOC-listen
'He switched on the television to watch it.'
Abui (Kratochvíl 2007:196)

(7) *Simon he-fala.* *Ne-lui*
Simon 3s.LOC-house 1s.LOC-knife
'Simon's house.' 'My knife.'
Abui (Kratochvíl 2007:145)

The alienable possessive marking applies also to two body part nouns, *toku* 'leg' and *pikai* 'head' and all kinship terms, as illustrated in (8).

(8) *Ne-wil ne-toku he-fahat.*
1s.LOC-child 1s.LOC-leg 3s.LOC-embrace.COMPL
'My child embraced my leg.'
Abui (Kratochvíl 2007:197)

Stative predicates may have their single argument corss-referenced by an *e*-prefix, as in (9).

(9) *He-beka.* *He-lunga.*
3s.LOC-be.bad 3s.LOC-long
'It is bad.' 'It is long.'
Abui (Kratochvíl 2007:201)

The third UNDERGOER relation concerns an animate or inanimate goal, requiring the vowel *o*. Inanimate goals occur rarely, but animate goals are often recipients. This form is also used to mark the single argument of verbs denoting physiological experiences, as in (10). It is not used to mark a possessive relation.

(10) *Simon no-dik* *No-lila*
Simon 1s.REC-prick 1s.REC-hot
'Simon is tickling me.' 'I feel hot.'
Abui (Kratochvíl 2007:201)

In Klon (Baird 2008) we find similar, although not the same, alignments. Free pronouns are used as actor subjects of transitive verbs (11) and the largest class of intransitive predicates.

(11) *Ga ool méd* .
3s(A) woman take
'He took a wife.'
Klon (Baird 2008:67)

There are three sets of undergoer prefixes in Klon. The singular forms are given in Table 1, which includes a clitic-like pronoun as class III, which will be ignored in my account, since it allows material between it and the verb and thus functions at a different level.

Table 1: Klon singular pronominal forms (Baird 2008:67)

Class	**Free**	**I**	**II**	**III**	**IV**
1s	*na(n)*	*n-*	*no-*	*nin=*	*ne-*
2s	*a(n)*	V-/Ø	*o-*	*in=*	*e-*
3s	*ga(n)*	*g-*	*go-*	*gin=*	*ge-*

Class II undergoer prefixes with the vowel *o* is the most frequent, occurring with more than 50 percent of the transitive predicates, class I occurs with about 30 percent and IV only with about four percent transitives. Each of these undergoer prefixes is also used to reference the single argument of verbs that denote involuntary events or states.

Here are examples of class I, as undergoer of a transitive predicate (12), as single argument of an intransitive (13), and as prefix marking inalienable possession (14).

(12) *Koh ini awa g-eh nang.*
finish 3ns(A) again 3(I)-feed NEG
'Then they didn't feed her anymore.'
Klon (Baird 2008:31)

(13) *Uruut béq ma n-edan na ete hil agai [....].*
deer pig come 1s(I)-scared 1s(A) tree climb go
'Deer and pig came, I was scared, I climbed a tree.[...]'
Klon (Baird 2008:31)

(14) *N-ooi.*
1s(I)-mother
'My mother.'
Klon (Baird 2008:91)

Class IV prefixes are the least commonly used on transitive verbs, an example is given in (15).

(15) *Ni* *e-qad* *a* *agai* *de* *nge-moi.*
1e(A) 2s(IV)-come 2s(A) go CONJ 1e(IV)-help
'We come to you and you go to help us.'
Klon (Baird 2008:74)

They can also be used to reference single arguments of involuntary predicates, as in (16), and are used as free possessive pronouns for alienable possession, as in (17).

(16) *Ge-wet.* *Ge-eneh*
3s(IV)-urinate 3s(IV)-calm
'S/he urinates.' 'S/he is calm.'
Klon (Baird 2008:74)

(17) *Anus* *ge* *kuur* *ele* *ihih* *mteh [...].*
Anus 3POSS dog 3d get.up stand
'Anus and his dog got up, stood [and....].'
Klon (Baird 2008:54)

As we saw in Abui, Klon allows some alternative marking for a number of verbs. As Baird puts it, 'Class IV prefixes are the 'default' prefix for intransitive verbs that typically take an Actor argument, but under the right semantic conditions the single argument is treated as an Undergoer' (2008:73). This would suggest a distinction as in (18), with Class IV focusing on the affectedness of the sole participant. This function of Class IV is further illustrated by the contrasting examples, (19) and (20), involving *hrak* 'hot', but now alternating with Class II, the most commonly used prefixes to reference undergoers in transitive predicates.

(18) *Ge-wet.* *Ga* *wet*
3s(IV)-urinate 3s(A) urinate
'S/he urinates.' 'S/he urinates (intentionally).'
Klon (Baird 2008:52)

(19) *Adaq* *ne-hrak*
fire 1s(IV)-hot
'The fire makes me (unbearably) hot.'
Klon (Baird 2008:76)

(20) *Mdi* *no-hrak*
sun 1s(II)-hot
'The sun heats me up.'
Klon (Baird 2008:76)

To illustrate a third AP language, here are examples from Adang (Haan 2001) to show the formal similarity between free pronouns, used as actor subject of transitives (21), undergoer prefix of transitives (22), and as experiencer undergoer on a transitive verb that has an inanimate subject, (23), similar to the Klon examples (19)-(20), and finally, the marker of inalienable possession, which elides its vowel preceding a vowel-initial stem, cf.(24).

(21) *Na bel beh.*
1s(A) dog hit
'I hit the dog.'
Adang (Haan 2001:151)

(22) *John na-puŋ.*
John 1s(U)-hold
'John caught me.'
Adang (Haan 2001:234)

(23) *Fil na-pa=am.*
urine 1s(U)-feel=PFV[6]
'I have started to feel like urinating.'
Adang (Haan 2001:235)

(24) *Na-taang.* *N-at*
1s.POSS-hand 1s.POSS-mouth
'My hand.' 'My mouth.'
Adang (Haan 2001:135)

Inalienably possessed nouns inflecting as in (24) include fourteen other nouns, such as *nafel* 'my ears', *napot* 'my thighs'. A number of other body part nouns, including 'head' employ a thematic vowel [ɛ], rather than the common [a]: *nɛ-lafung* 'my head'. This form contrasts with what Haan (2001:163) calls 'Genitive pronouns', used for alienable possession. These genitive pronouns come in two series, one with the vowel [ɔ] just marks alienable possession, while the series with the vowel [*e*] expresses emphasis or contrast. The latter may be used with inalienably possessed nouns to stress a contrastive possessor, as in:

(25) *N-e n-ɛʔai ʔɛ maru dai.*
1s-GEN 1s-eye NEG blind EVID
'These eyes of mine (e.g. as in contrast with those of someone else) are not blind yet.'
Adang (Haan 2001:164)

The pronominal paradigm of AP languages shows its likely genealogical connection with the Trans New Guinea family (Ross 2005:29)[7] as illustrated by the forms in Table 2, comparing the proposed proto-forms and the reflexes found in Teiwa (Klamer forthcoming) and Nggem of the Dani family of the Papua Highlands (Etherington 2002).[8]

---

[6] Following Haan's (2001:79) analysis of *am* 'PFV' and *eh* 'PROG' as clause-final clitics, I mark PFV as clitic, rather than as suffix.

[7] Ross (2005:36) notes: 'The languages of the Dani, Wissel Lakes, West Bomberai and East Timor microgroups all reflect an innovation whereby **ani* 'I' has replaced Proto TNG **na*, but this innovation is not reflected in the West-Timor-Alor-Pantar microgroup.'

[8] I only give possessive prefixes for Nggem. As Foley (2000:377) notes, many TNG languages, including Dani (a close relative of Nggem), do not permit the object prefix to be directly attached to verb stems, but to a secondary auxiliary-like verb, compounded to the verb root, with a resulting shift in position of the object affix on the verb. Etherington (2002: 113) states that there are four primary verbs that indicate the object by direct affixation to the root, three of which do this by transparently prefixing the forms given in

Obviously there are some differences between the paradigms of Teiwa and Nggem, but the similarities indicate that the languages may well belong to the same family. The single consonants in Teiwa are prefixed to vowel-initial stems. In Nggem the stem consonant or its following vowel determines the prefix vowel, while certain nouns add *-g-* between the prefix and the noun. There is no clear indication of an alienable-inalienable distinction in Nggem. Teiwa and other AP languages not only distinguish inclusive and exclusive in first person, but also have extra categories for third person reference, not shown here.

Table 2: Pronouns in Teiwa and Nggem, reflecting the Trans New Guinea paradigm

	Proto-TNG[a]	**Teiwa**			**Nggem**	
		subject short form	object free	object/possessor prefix	free	possessor prefix
1s	**na*	*na*	*na'an*	*n*(*a*)-	*an*	*n*V-, *nag*
2s	**ŋga*	*ha*	*ha'an*	*h*(*a*)-	*hat*	*h*V-, *hag-*
3s	**[y]a/ua*	*a*	*ga'an*	*g*(*a*)-, *gə-*	*at*	V-, *ag-*
1e	**ni/nu*	*ni*	*ni'in*	*n*(*i*)-	*nit*	*nen*V-, *nenag-*
1i		*pi*	*pi'in*	*p*(*i*)-	*nit*	*nen*V-, *nenag-*
2p	**ŋgi*	*yi*	*yi'in*	*y*(*i*)-	*hit*	*hen*V-, *henag-*
3p	**i*	*i, a*	*iman*	*g*(*i*)-, *ga-*	*it*	*en*V-, *enag-*

a Ross also reconstructs a few additional forms: Plural non-singular suffix * *n[V]-* and Inclusive *-*m-*, and for Dual *-*li/*-*t* and *-*p-* respectively. The last form may have a reflex in AP languages, such as *pi* in Teiwa.

The facts in Yawa are remarkably similar to the ones in the Alor-Pantar languages shown above. Yawa has one set of pronouns for subjects of transitives (A) and active intransitive (S_A) verbs (Jones 1986:42; also Reesink 2005:211), as illustrated with the form *sy(o)* 'I' as free pronoun in (26) and as proclitic before a vowel-initial stem in (27).

(26) *Syo Mariana r-aena.*
1s(A) Mariana 3s.F(U)-see
'I am/was looking at Mariana.'
Yawa (Jones 1986:41)

(27) *Sy=oyowa.*
1s(A)-cry
'I cry/was crying.'
Yawa (Jones 1986:41)

A different set of pronominal forms is used to reference animate objects by verbal prefixes, which are identical to possessive prefixes on inalienable nouns, as witnessed by the 1s(U) forms *in-* in (28) and (29).

Table 2 as possessive prefix: *wat* 'strike, kill', *ha* 'see'and *ombat* 'think, believe'. The Dani languages, like many other TNG languages, also have subject suffixation on the verb. Thus, the comparison here focusses only on the morphological similarities between the Alor-Pantar and Dani families.

(28) *Mariana mo in-aena.*
Mariana 3s.F.ERG 1s(U)-see
'Mariana is/was looking at me.'
Yawa (Jones 1986:41)

(29) *In-aneme.*
1s(U/POSS)-hand
'my hand.'
Yawa (Jones 1986:45).

And it is these object/possessor prefixes that are found on a small class of uncontrolled verbs. Jones (1986:43) lists ten verbs denoting mainly mental and some bodily experiences, including items such as *anumambe* 'be sad', *anode* 'be happy', *anatantonawe* 'think' and *anikakowa* 'forget', and *anaidudu* 'be exhausted' and *awabea* 'yawn' as illustrated in (30).

(30) *In-awabea.*
1s(U)-yawn
'I yawn/was yawning.'
Yawa (Jones 1986:41).

Alienable possession in Yawa is expressed by what appears to be a verbal element *-a*, prefixed by the subject marker for active intransitive verbs *sy(o)-* '1s(A)', related to the free pronoun, as in (31).

(31) *Weti sy-a ana-syora yamo, syopi no naije.*
so 1s(A)-POSS NMLZ-speech TOP arrive LOC there
So as for my speech, it's finished.
Yawa (Linda Jones, unpublished texts)

In summary, this section has provided evidence of SOV languages, including a number of Alor-Pantar languages, possibly members of the large Trans New Guinea family, and unrelated Yawa of the Cenderawasih Bay, that have an object prefix as the only argument affix on a transitive verb. The actor subject (A) can only be realized by a noun (phrase) or free pronoun. Thus, they do not follow the cross-linguistic tendency that object affixation entails subject affixation.

Secondly, the object (UNDERGOER) prefix is also used to cross-reference the sole argument of uncontrolled (involuntary) predicates. The class(es) of such verbs is/are much larger in the AP languages than in Yawa.

Thirdly, in all these languages object prefixes and possessor prefixes are identical or very similar. It is tempting therefore to analyse inalienable possession in such languages as verbal constructions for uncontrolled states. It is not quite clear how these constructions relate to the different source schemas Heine (1997) has proposed for possessive constructions. Both Location and Goal schemas are possible sources for these attributive possessive constructions (Heine 1997:103), in which the possessor is a locative or object complement: (It) locates Y at/to X. In AP languages, alienable and inalienable possession are distinguished by different undergoer prefixes, apparently signaling less affectedness for alienable possession, as in Abui. In Yawa, alienable possession is expressed by forms similar to forms for S_A , indicating a connection with volition or control.

3 SOV and s-o-V verbal affixation

This section will look at the argument affixation on verbs in SOV languages that do adhere to the general tendency claiming that object affixation entails subject affixation. Again, I will show the formal similarity between the object affix and the possessor reference in possessive constructions. A third issue concerns the expression of stative and patientive arguments.

There are two areas with languages whose verbal affixation corresponds to the clausal word order SOV: most of the North-Halmahera family and the South Bird's Head family, of which Inanwatan is a good representative. de Vries (1998:651) suggests a possible genealogical connection between Inanwatan and the Trans New Guinea language Marind on the basis of their gender morphologies and the verbal affixation pattern, referring to Foley (1986:138), who at that time stated that 'Marind is the only Papuan language I know which consistently exhibits A-U-V [= Actor-Undergoer-Verb]'.

As this section shows, there are more Papuan languages that mirror the canonical clausal constituent order SOV in their verbal morphology. This observation does not directly illustrate Givón's maxim, quoted in the introduction. Obviously, the present order of the prefixes for subject and object cannot be directly related to the present or historical nominal constituent order, since the subject of the clause would always be separated by the object (and other peripheral constituents) from the predicate. Presumably, the attachment of a pronominal subject is the result of a resumptive pronoun which was grammaticalized, and in some cases morphologically fused with the object prefix. I will return to this issue in the conclusion.

Although the Papuan languages of North Halmahera form a well-established genealogical family (Voorhoeve 1988; Holton 2008), the languages show considerable variation in various respects. The most obvious one is that a number of southern languages (West Makian, Tidore, and Sahu) have shifted to a SVO order, and two of these, West-Makian and Tidore, have no object prefix on the verb (anymore).

As a representative of the North Halmahera family, let us consider some examples from Galela. Galela has different sets of prefixes for subject (A and S_A) and for object (U) for all person-number, except first person exclusive (both are *mi*) and second person plural (both are *ni*). Howard Shelden (1991: 162) claims it is a split-S language (see also Foley 2005:409). It has the same subject marker as verbal prefix for A of transitive verbs, preceding the object prefix (32), as for S_A of controlled intransitive verbs, as in (33).

(32) *Una wo-mi-sepa.*
3S.M 3S.M(A)-1e(U)-kick
'He kicked us.'
Galela (H. Shelden 1991:168)

(33) *O Robi wo-tagi.*
ART Robby 3S.M.(A)-walk
'Robby is walking.
Galela (D. Shelden 1998:6)

For the class of uncontrolled stative verbs (H. Shelden 1991:166), denoting physical states, experiences and attributes (adjectival notions), object prefixes are used to reference S; compare the object prefix *wi* in (34) and (35). In other words, languages of the NH family exhibit a semantic alignment of core arguments. This means that there are two

patterns for marking the sole argument of intransitive verbs: active or agentive arguments of intransitive verbs are marked differently from stative or patientive (undergoer) arguments.

(34) *Una mi-wi-ngapo.*
3s.M 1e(A)-3s.M(U)-hit
'We hit him.'
Galela (H. Shelden 1991:168)

(35) *Wi-maḍe*
3s(U)-embarrassed
'He is embarrassed.'
Galela (H. Shelden 1991:167)

Thus, the uncontrolled verbs in Galela agree with the unrelated AP languages and Yawa in marking the experiencing undergoer by a/the object prefix.

It appears that this construction has relatively recently developed from a fully transitive construction with an impersonal 3s actor subject marked as well, given the explicit description provided by van Baarda (1908:81) a century ago, when he compares Galela constructions such as (36) with Dutch and German 'Dative-Subject' constructions.

(36) *I-ni-sapi*
3s(A)-2s(U)-hungry
You are hungry. (lit. 'it hungers you')
Galela (van Baarda 1908:81)

The form *i* is recognized by Howard Shelden (1991:163) as a subject prefix referring to non-humans, not distinguishing number, and a homophonous form as 1s.OBJECT, see also Deirdre Shelden (1986:234), compare (37) and (38).

(37) *O kaso i-goli.*
ART dog 3s.NEUTRAL-bite
'A dog bites.'
Galela (D. Shelden 1986:236)

(38) *I-sapi*
1s(U)-hungry
'I am hungry.'
Galela (H. Shelden 1991:167)

With 1s.OBJECT also being *i*, the 3s NON-HUMAN SUBJECT *i* is apparently absorbed in (38), which van Baarda indicates by a diacritic on the same form, *ï sàpi*, suggesting a long or rearticulated vowel. But this impersonal subject marker has disappeared altogether in the course of a century.[9]

---

[9] Thus, Galela does indeed resemble the unergative (= controlled) versus unaccusative (= uncontrolled) split found in Austronesian languages of Maluku (Foley 2005:409). While Foley states that the split in Maluku languages is essentially based on aspectual properties of the predicate, states are unaccusatives versus performed events that are unergative, Klamer (forthcoming) rightly modifies this claim by showing that in some languages, for example Larike and Dobel, the opposition between volition (control)

There are other North Halmahera languages, in which this 3s ACTOR is still fully productive. For example, Holton (2008) shows that 'Tobelo requires the *i-* prefix with all semantically intransitive stative verbs'.[10] He shows that erosion of prefixes, as witnessed in the examples from Galela, is widespread in various North Halmahera languages. Even subject actor prefixes need no formal expression if the discourse context allows easy recoverability (Wimbish 1991:146; Holton 2008). Holton's conclusion is that, although in some present-day languages of the North Halmahera family there is no longer morphological evidence for a semantic alignment, there is still evidence reflecting a semantic alignment in the protolanguage.

Wimbish (1991) for Pagu and Visser and Voorhoeve (1987) for Sahu make no mention of a Split-S phenomenon, and examples throughout their descriptions suggest that the subject prefixes on transitive verbs are also used on uncontrolled statives. Consider the following examples, (39) and (40) for Pagu, and (41)–(42) for Sahu.

(39) *O ngo Sarah o gula mo-ki-kula ma ngoa-ngoak-ika.*
ART F Sarah ART sugar 3s.F(A)-3p(U)-give ART RDP-child-PFV
'Sarah gave sugar to the children.'
Pagu (Wimbish 1991:32)

(40) *Muna mo-lamok.*
3s.F 3s.F(A)-big
'She is big.'
Pagu (Wimbish 1991:32)

In Sahu, as in Galela, Pagu and other North Halmahera languages, there are two sets of prefixes whose combinations on transitive verbs result in a number of phonological assimilations (Visser and Voorhoeve 1987:28): Subject 1s is *to*, while object is *ri*. In combination with 2s object *ni*, the vowel of the subject prefix assimilates: *ti-ni*, as in (41).

(41) *Ngoi ti-ni-molo'ara.*
I 1s(A)-2s(U)-marry
'I'll marry you.'
Sahu (Visser and Voorhoeve 1987:30)

Uncontrolled statives, like *sau'u* 'hot', *malata* 'cold' and *sisi'di* 'ill', use the subject prefixes, illustrated in (42).

---

and lack thereof can be reflected by A encoding of S of states (Larike) or P (UNDERGOER) encoding of S of events (Dobel). In Galela it is essentially uncontrolled states that have S marked as P; Shelden (1991:166) specifies (i) physical states, such as *kiolo* 'sleep', (ii) experiences, such as *tiiki* 'cough [unintentionally]', and (iii) attributes, such as *tubuso* 'heavy'. But the widespread Split-S systems in Austronesian, whether due to inherent verbal aspect or to (lack of) control on the part of the single argument, show that the argumentation given in Reesink (2005:191-194) that 'experiential verbs' in Austronesian languages of Maluku and Cenderawasih Bay are the result of contact with Papuan languages was not very felicitous, to put it mildly.

[10] Holton (2008:27-28) quotes work from Kern (1891, 1892) showing that for some verbs, such as *toosa* 'angry', the *i-* prefix is present throughout the paradigm, while others, such as *pereki* 'old', are defective, evidencing erosion of the actor prefix.

(42) *Ngoi la to-malata.*
I DEGREE 1s(A)-cold
'I am very cold.'
Sahu (Visser and Voorhoeve 1987:57)

But, in fact, in both languages, there are traces of the original North Halmahera alignment.[11] For Pagu, Wimbish (1991:43) remarks 'the object agreement markers may be used alone (without a preceding subject marker) when the subject is an experiencer or patient, rather than an agent', providing (43) as an example.

(43) *O panyakit wi-daen.*
ART disease 3s.M(U)-experience
'He has a disease.'[12]
Pagu (Wimbish 1991:43)

In other words, just as Galela, Pagu has lost the actor prefix *i-*, which is still available in Tobelo. Whereas Holton (20008) analyses the *i*-prefix on stative verbs in Tobelo as a 'non-referring pleonastic experiencer subject', I think this example exhibits the same construction as the century-old Galela and present-day Tobelo in having an inanimate actor-subject acting on a human object, in spite of the gloss 'experience' in (43). I will return to this issue in the conclusion.

It is not clear from the various descriptions of North-Halmahera languages whether they make a distinction between alienable and inalienable possession. Possessive pronouns in Pagu (Wimbish 1991:18–19) and Sahu (Visser and Voorhoeve 1987:39) are morphologically more related to the object than to the subject prefixes, but they are employed for all kinds of possessive relations:

(44) *Ngoi a-ri wala.*
I POSS-1s(U) house
'My house.'
Sahu (Visser and Voorhoeve 1987:57)

There is some indication of inalienable possession in NH languages, but this is not marked by either object or subject prefixes found on verbs. In various languages, such as Pagu (Wimbish 1991:20) and Sahu (Visser and Voorhoeve 1987:37), there is a form *ma* which is used to link a part to its whole. This appears to be a general feature of the North Halmahera family. Holton (2003:31) states for Tobelo, 'Nouns which are inherently possessed or stand in an inherent relationship to another entity are marked with the relational noun marker *ma-* [rather than the default noun marker o, that accompanies any noun in citation form]'. Such nouns include kinship terms and nouns referring to body parts, regardless of whether they are attached to the body (Holton 2003:20–21). These nouns may also receive the general possessive prefix indicating number and person of the

---

[11] I will only provide some comments on Pagu; see Holton (2008) for the argument that Sahu too has traces of the original semantic alignment system.

[12] Wimbish identified the vowel *i* of object prefixes as a separate morpheme, indicating 'humanness'; her paradigm of subject-object agreement markers (1991:43) shows that *i* is homophonous for 1s(U) and 3s.NONHUMAN(A), as in other NH languages; I therefore take her object prefixes as cognate with those in Galela, Tobelo, etc.

possessor, and even be optionally preceded by an independent pronoun prefixed with *to-*, compare (45) and (46).

(45) *Ma-hiranga.*
NM-opposite.sex.sibling
'Man's sister; woman's brother.'
Tobelo (Holton 2003:21)

(46) *(To-ngohi) ngo-ahi-hiranga.*
POSS-I F-1s.POSS-opposite.sex.sibling
'My[male ego] sister.'
Tobelo (Holton 2003:32)

However, this does not establish *ma-* exclusively as an inalienable possession marker, because the rules for using *ma-* encompass many other discourse considerations, such as references to earlier established participants. Another condition concerns the case of non-human possessors, as in (47).

(47) *Ma-ode (to-enanga) ma-inomo*
NM-pig POSS-3.non-human NM-food
'The pig's food.'
Tobelo (Holton 2003:32)

We can conclude, then, that languages of the North Halmahera family exhibit a close formal similarity between the object prefix[13] on transitive verbs, the single argument on uncontrolled verbs and the person-number index of the general possessive pronouns, which are presented as free forms in some descriptions, while in others they are reported as prefixes.

Although the South Bird's Head family is unrelated to North-Halmahera languages, Inanwatan (de Vries 1996, 2002, 2004) also has SOV as its constituent order and both subject and object prefixes on transitive verbs, as illustrated by *me-rá-wi-de* '3SUBJ-1s.OBJ-give-PAST' in (48), showing the common pattern in Papuan languages, which is that object prefixes are the ones cross-referencing the recipient of 'give'.

(48) *Áwoge méqaro naguáre áwoge* ***dínasi****-da-wo* *nágo-wo*
again house one again service-POSS-be.3s.F one-CONN

me-rá-wi-de ***buat*** ***nanti*** ***páakai****-béqewu* *mó* *ré-re-sa.*
3SUBJ-1s(P)-give-PAST for later use-in.order.to(P) there 1s-sleep-FUT
'And they gave me also a house of the (health) service to live there in the future.'
(de Vries 2004:36; **bold** items are loans from Malay)

There does not seem to be a special construction for uncontrolled predicates in Inanwatan. The language does make a distinction between alienable and inalienable possession, the latter involving possessor prefixes that formally resemble the object

[13] This claim holds only for those languages that still mark both subject and object by prefixes to the verb, of course. As stated earlier, a number of the southern NH languages no longer use an object prefix, as they also have shifted from SOV to SVO order.

prefixes, as illustrated in Table 3, based on the forms in de Vries (2004:27, 29, 36). Object prefixes for first and second person singular contain the vowel *a*, while the bound subject forms have the vowel *e*. I deliberately include also the free pronouns used as subject or object in Table 3, since they appear to consist of the bound object forms as prefixes to other elements, quite similar to the free possessive pronouns.

Table 3: Pronominal forms in Inanwatan

	SUBJ free	**Bound**	**OBJ free**	**Bound**	**Possessive free**$^{+}$	**Possessive prefix**
1s	*ná-iti*	*ne-*	*ná-we*	*na-*	*ná-ridó-wo*	*na-*
2s	*á-iti*	*e-*	*á-we*	*a-*	*a-ridó-wo*	*a-*
3s.M	*í-tigi*	*me-*	*	-	*tigidá-wo*	-
3s.F	*í-tigo*	*me-*	*	-	*tigáe-so*	-
1e	*ní-iti*	*nige-*	*ní-we*	*ni-*	*ní-rido*(*-wo*)	*ni*(*da*)-
1i	*dá-iti*	*ge-*	*í-we*	*i-*	*dá-ro*(*-wo*)	*da-*
2p	*í-iti*	*ge-*	*í-we*	*i-*	*í-rido-wo*	*i*(*da*)-
3p	*í-ti-ga*	*me-*	*	-	(*tig*)*a-wo/áo*	-

* Enclitic demonstratives function as personal pronouns for third person objects, while the bound forms are absent or zero; the morpheme boundaries in this Table are not all supplied in the source text.

$^{+}$ The free possessive pronouns mark gender for all person-number categories; I have listed just the feminine forms. The third person singular forms express double gender: male or female possessor plus gender of the possessed noun; I give 3s.M possessor with FEMININE possessed, and 3s.F with MASCULINE possessed. Material in parentheses is optional.

Possessive prefixes are only used for first and second persons and are attached to nouns denoting body-parts, kinship terms and a few other nouns, such as *néro* 'name of female'and *néri* 'name of male', with the gender of nouns determined by the final vowel.

Here are examples that contrast inalienable possession *ná-wir-i* '1s(U)-belly-M' (49) and alienable possession, expressed by a pre-nominal possessive pronoun *na-ridó-wo* '1s-POSS.F-F.POSSESSEE' (50).

(49) *Ná-wir-i* *me-tutú-rita-bi.*
1s(U/POSS)-belly-M 3s(A)-hurt-DUR-M
'I [male] have pain in my belly (lit. my [male] belly it hurts).'
Inanwatan (de Vries 2004:30)

(50) *Owó-i* *ná-rido-wo* *méqaro-wo.*
that.F-S 1s(U)-F.POSS-F house-F
'That is my (female) house.'
Inanwatan (de Vries 1996:104-106)

In this section I have shown the verbal and possessive affixation patterns in SOV languages. Two unrelated groups of Papuan languages, those of North Halmahera and those of the South Bird's Head have both subject and object prefixes on transitive verbs. In both groups there is a formal similarity between the object prefix and the person-number marker indexing the possessor of possessive pronouns. The NH languages have a construction for uncontrolled predicates in which the experiencer is marked by the object prefix, which in some languages is preceded by an impersonal 3s subject, while this has

been eroded in other languages. Such constructions appear to be absent in Inanwatan of the South Bird's Head.

While the bound morphology is normally quite stable through genealogical descent, the word order in the clause is more easily affected by contact with languages from other lineages. While Inanwatan does not adhere to a strict V-final order in that oblique constituents do not always precede the final verb, a few North Halmahera languages have completely shifted from V-final to V-medial clausal order. This takes us to the next section.

4 SVO and s-V verbal affixation

In this section I will review West Papuan (see footnote 2) languages with SVO constituent order. Most of these have only subject prefixes on the verb, while a few languages allow encliticized pronominal objects as variant of free pronouns. None of them have tense morphologically marked on the verb, while only in the East Bird's Head family, constituted by Meyah (Gravelle 2002, 2004), Moskona (Gravelle-Karn 2010) and Sougb (Reesink 2002), there is some aspect and mood prefixation. In all these languages inalienable possession is marked by prefixes indexing person- number of the possessor. With some minor exceptions, these prefixes are identical to the verbal subject prefixes in Hatam and Sougb, see below.

The North Halmahera family is clearly related to the West Bird's Head family on the basis of pronominal forms and a number of lexical cognates. Possibly other Bird's Head languages belong to the same genealogical grouping (Reesink 2005, Ross 2005). As Voorhoeve (1987) argued, the SOV word order must have been original. But presumably under pressure of neighboring Austronesian languages, Sahu, Tidore, and West Makian of the North Halmahera family have shifted to a SVO order. While Sahu still maintains the combination of subject and object prefixes on the verb, Tidore (van Staden 2000) and West-Makian (Voorhoeve 1982) both have lost the cross-reference of the object, as illustrated in (51).

(51) *Una* *wo-falang* *mina* *nyao.*
3s.M 3s.M(A)-give 3s.F fish
'He gave her fish.'
Tidore (van Staden 2000:212)

The loss of object prefixes in these languages apparently was concomitant with the shift in constituent order and it implied the disappearance of the morphological construction for uncontrolled predicates still productive in Tobelo and Galela, as we saw in the previous section. Quoting van Staden's observation (2000:79) that the disappearance of actor prefixes is more advanced on stative (adjectival) verbs than on other verbs, Holton (2008) argues that the notional distinction of semantic alignment is still traceable in Tidore.

In Tidore, not only have the object prefixes disappeared altogether, actor and possessive prefixes are also easily omitted (van Staden 2000:79), although the possessive prefixes, clearly cognate with object prefixes in other NH languages (van Staden 2000:253), are less easily omitted than the actor-subject prefixes. As van Staden (2000:125-126, 253) shows, inalienably possessed nouns, including nouns referring to body parts, parts of plants and trees, kinship, and also words for 'name' and 'house', require a possessive prefix or an invariant marker *ma-*.

(52) *Una i-fayaa.*
3s.M 3s.M.POSS-wife
'His wife'or 'he has a wife'
Tidore (van Staden 2000:126)

(53) *Una ma-fayaa.*
3s.M INAL-wife
'His wife'or 'he has a wife'
Tidore (van Staden 2000:126)

This means that with inalienable nouns there is the choice between person/number marking and invariant *ma*, as in (52) and (53), while with alienable nouns, agreement between the possessor, which can be a free pronoun or noun (phrase), and a possessive pronoun is obligatory, as between *mi-ngofa* 'her children' and *na-* 'their' in (54).

(54) *Ica mi-ngofa na-guru.*
Ica 3s.F.POSS-child 3p.POSS-teacher
'The teacher of Ica's children.'
Tidore (van Staden 2000:250)

The configuration of SVO word order and concomitant subject prefixing is the one found in all Papuan languages of the Bird's Head, except, as shown in the previous section, in the South Bird's Head family (Reesink 1998, 2005). All North Halmahera and the Bird's Head languages with SVO order lack verbal marking of tense. Some aspect or modal categories are marked by affixation or cliticization in most NH languages. Of the SVO Bird's Head languages, only the East Bird's Head family (Meyah, Moskona and Sougb) have verbal prefixes indicating mood (irrealis) or aspect (durative in Meyah and Moskona), see Reesink (2002). The verbal subject prefix in these languages is identical to the possessive prefix on inalienable nouns, as well as on the possessive pronoun, thus most likely of verbal origin, used for alienable possession, as illustrated by Mpur examples.

(55) *An-det bar-(d)et pa=e?*
2s-eat thing-eat already=QUESTION
'Have you already eaten?'
Mpur (Odé 2002:54)

(56) a. *An-muk*
2s-name
'Your name.'
Mpur (Odé 2002:62)

b. *an-tar jan*
2s-POSS house
'Your house.'

This holds for most Bird's Head languages, although there is some variation. Hatam has zero for 3s subject marking, while 3s possessive is *ni-* for inalienables and *ni-de* for alienable possession (Reesink 1999:40). Meyah (Gravelle 2002; 2004) and Sougb (Reesink 2002), likewise have zero for 3s verbal subject and other forms for possessors. But the possessive linkers are, just as *tar* in Mpur, very likely of verbal origin, given the similarity between subject marker in (57) and the possessive marker on the possessive linker *de* for alienable nouns in Hatam, as in (58).

(57) *A-pim tut mindei=i?*
2s-cry along what=QUESTION
'Why are you crying?'
Hatam (Reesink 1999:69)

(58) A-de singau.
2s-POSS knife
'Your knife.' or 'You have a knife.'
Hatam (Reesink 1999:49)

Maybrat (Dol 1999:92–98) employs identical prefixes for verbal subject marking and possessor on inalienable nouns, while alienable possession is expressed by a post-nominal preposition phrase, as in (59).

(59) a. Fnia m-ao
woman 3s.F-foot
'The/a woman's foot.'

b. Amah ro t-atia
house POSS 1s-father
'My father's house.'
Maybrat (Dol 1999:94, 97)

In Moi and Tehit, members of the West Bird's Head family, subject marking is ubiquitous, as these languages express all prepositional notions as verbs that need obligatory subject prefixes, as witnessed by (60). The same prefixes are used on possessed nouns, obligatorily on nouns referring to body parts and kinship terms, while alienable nouns may have a post-nominal construction, as in (61). These languages have no further verbal morphology. More details on Moi can be found in Menick (1996, 2000) and Reesink (2008). For Tehit see Hesse (1995; 2000). Consider (60) illustrating Moi and (61) from Tehit.

(60) *Wu$_i$-suu* [*m-e-kem aali*]$_j$ *n$_j$-uu* *kii-suwe* *meele*
3s.M$_i$-give [3s.F-POSS-meat two]$_j$ 3p.NH$_j$-lie.down.p rope-k.o.aerial.root one

masaagu, *w$_i$-ewiyek* *Ø$_j$* *n$_j$-uu* *ma* *se* *a*
then 3s.M$_i$-hang (them)$_j$ 3p.NH$_j$-lie.down.p TOP PFV LINKER

[*poolos* *aana-Ø-e-n*]$_k$ *n$_k$-iini.*
[paddle VIS-near.addressee-p-3p.NH]$_k$ 3p.NH$_k$-go.p
'He put her two pieces of meat on one *suwe*-rope, then, given that he had already hung them, it was (on) those paddles over there.'
Moi (Menick 2000:18)

(61) *Om* *m-ase* *m-ak* *biele* *om* *m-a.*
3s.F 3s.F-sleep.s 3s.F-be.at.s garden 3s.F 3s.F-POSS.s
'She lives at her garden.'
Tehit (Hesse 2000:26)

The only languages in the Bird's Head that have a special construction for uncontrolled predicates are the East Bird's Head family, Meyah (and its close relative Moskona) and Sougb. In these languages a small closed class of predicates expresses physiological states, in which the experiencer is marked by an object enclitic/suffix. It is possible to

express the experiencer by a noun phrase or free pronoun in the clause-initial position, as in (62).

(62) *Dan ar-eb-ed.*
1s something-do-1s
'I am/was sick.' (lit. 'I, something do (affect) me.')
Sougb (Reesink 2002:208)

Whether such a construction warrants a split-S system, however, is doubtful, as it is clear that the free pronoun referring to the experiencer is not a subject, but rather a topic, with the generic noun *ar(a)* 'something' functioning as subject. Although the constituent as well as the affix order differ from the ones in Alor-Pantar and North Halmahera languages, it is clear that the experiencer is marked as object, while the verb involves an inanimate instigating actor, presumably without volition.

Summarizing, Papuan SVO languages of the Bird's Head are typologically rather unusual in that a number of them have subject prefixation as their only verbal morphology, lacking any tense, aspect, mood affixation. The same prefixes are used directly on inalienable nouns, and on possessive linkers (free possessive pronouns), used for alienable nouns. In these languages the distinction alienable-inalienable possession is determined by different noun classes. Special constructions for uncontrolled predicates are rare, limited to the East Bird's Head family.

5 Conclusion

This brief review of verbal and nominal marking of arguments in various West Papuan families has presented some typologically unusual features:

(i) a preponderance of prefixing,

(ii) marking of arguments in the absence of TAM morphology,

(iii) marking of object only, and

(iv) formal similarity of object prefix on transitive verbs, marking of undergoer on uncontrolled intransitives and possessor on nouns.

Perhaps these findings suggest that more than just a geographic label for the West Papuan languages is justified. While it is true that some of these features can also be found in Papuan languages of other regions, the West Papuan languages may form a typological areal group, rather closely geographically aligned along with the typologically defined 'preposed possessor' Austronesian languages, as Nikolaus Himmelmann (2005) identified them, see also (Klamer, *et al* 2008). On the basis of typological features found in WALS (Haspelmath *et al* 2005), Cysouw and Comrie (forthcoming) found 'a basic two-way distinction between Austronesian and West Papuan on the one side and the remaining 'Papuan' languages on the other side'. However, these general statements need some qualifications:

(i) Although argument affixation is overwhelmingly found as prefixes to verbs, some of the SVO languages allow object clitics as truncated variants of free pronouns.

(ii) A number of OV languages seem to have some suffixing to indicate aspect and mood. This issue is not easy to summarize, as clearly cognate forms are described in some grammars as suffixes, while in others they are analyzed as free particles or enclitics. As mentioned in the introduction, at least some AP and NH languages and the majority of SVO languages of the Bird's Head lack any further affixation. For Adang (Haan 2001:79) and Klon (Baird 2008) it is explicitly stated that aspect and mood are expressed by adverbs as clitics or particles. Similarly, the descriptions of Sahu (Visser and Voorhoeve 1987:25–27) and West Makian (Voorhoeve 1982:13) do not indicate that aspect or mood are affixed to the verb.

(iii) Although some AP languages may allow subjects to be indexed by verbal prefixes (Gary Holton, personal communication), it is clear that the majority of them, as well as unrelated Yawa, have only verbal prefixes for object but not for subject, on transitive verbs.

(iv) In all OV languages that mark S as single arguments of uncontrolled predicates, the pronominal prefixes are identical to those indexing objects of transitives, and they are transparently related to the possessor prefixes. Languages that have restructured completely to SVO order and subject prefixation, as found in some North Halmaheran and most Bird's Head languages, lack such morphological marking of uncontrolled intransitives, and simply have the subject prefix marking the possessor, as prefix directly on inalienable nouns and as prefix on a possessive linker, preceding the possessee.

These points are reminiscent of Givón's dictum quoted in the introduction that 'today's morphology is obviously related to the syntactic organization of the past and present'. However this relationship between morphology and syntax is not one of an immediate extrapolation to the original constituent order. As Comrie (1980:93) warns,

> 'The most obvious repercussion [...] for syntactic reconstruction is that such reconstruction must proceed with extreme care, since one cannot accept as a general principle that morpheme order reflects earlier (basic, or even possible) word order'.

Bearing in mind Comrie's warning, I would like to conclude with an attempt to explain the formal similarity between transitive object markers, the undergoer subject marking of uncontrolled predicates and prefixes indexing the possessor on a possessed noun, found in many of the West Papuan languages.

As the erosion of North Halmaheran 3s actor-subject *i* in Galela shows, the so-called Split-S system (semantic alignment) in these languages finds its origin in transitive constructions in which the (mostly) animate participant is the affected patient, marked by an object prefix. When the instigating element is no longer present, this participant remains as the only argument of the verb. This can then be re-analyzed as the undergoer subject, which in fact is marked as a regular object.

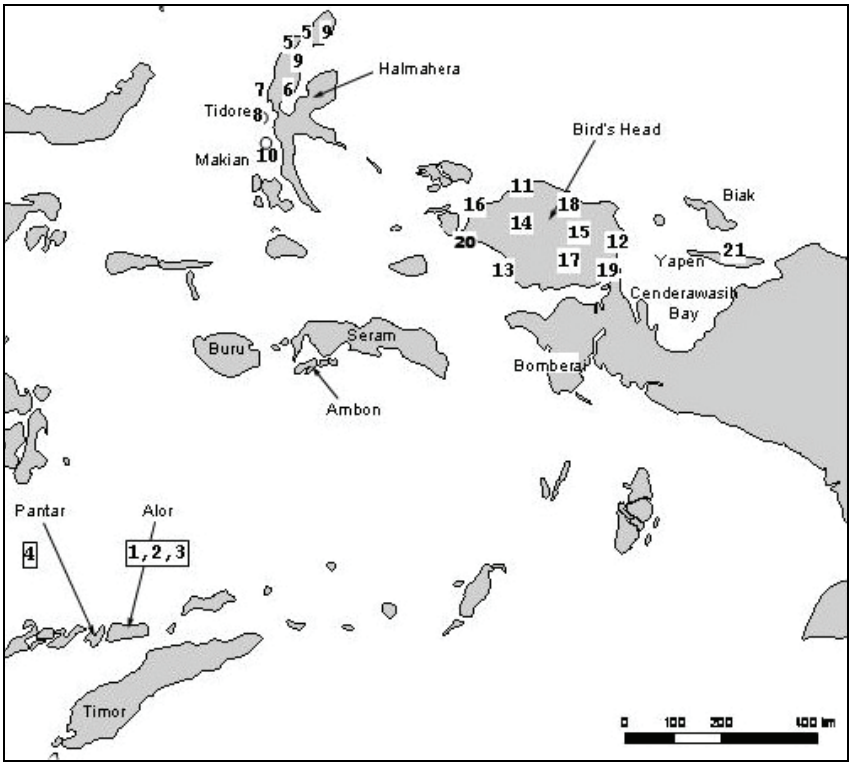
My suggestion is that this is the process that led to the Su markings in Yawa and AP languages, which in some cases, see Klon examples (19) and (20) and Adang (23) in section 2, still allows an inanimate instigating subject. As shown in section 4, Sougb of the East Bird's Head family, has a transparent transitive construction expressing uncontrolled affectedness but reflecting the clausal constituent order of SVO. With respect to the

undergoer prefix on possessed items, I hypothesised in the conclusion to section 2 that the possessor is the affected patient in erstwhile verbal constructions, based on a Location/Goal schema (Heine 1997). This hypothesis echoes the insightful observation van Staden made as an aside (2000:253, fn. 175) when she compared the Tidore possessive prefixes with their cognate forms in more conservative NH languages:

> '[...] one might even speculate that the possessive constructions find their origin in experiential constructions, also used in the N[orth-east] H[almahera] languages, for emotions or numbers. Analogous to 'it fears me' and 'it two-s me', we might find '(it) books me' [...]'.

The correspondence between subject prefixes and possessor-markings in SVO BH languages can also be related to (yesterday's) syntax. Possessors on inalienable nouns show great similarity with subject markers on verbs, the same prefixes are attached to verbal linkers forming possessive pronouns for alienable nouns, as shown in section 4. These languages appear to have formed their possessive constructions on what Heine calls the Action schema: X takes Y. If this hypothesis is correct, these languages contradict Heine's (1997:92) assessment that 'the Action Schema may give rise to have-constructions [...] or to belong-constructions, but never to patterns of attributive possession', as clearly such constructions can function both attributively and predicatively in languages like Meyah and Hatam.

Appendix



Map 1: West Papuan languages

Table 1: Sample of West Papuan languages considered; numbers refer to map

	Language	Affiliation	Source
1	Abui	Timor-Alor-Pantar	Kratochvíl 2007
2	Adang	Timor-Alor-Pantar	Haan 2001
3	Klon	Timor-Alor-Pantar	Baird 2008
4	Teiwa	Timor-Alor-Pantar	Klamer forthcoming
5	Galela	North Halmahera	van Baarda 1908
			D. Shelden 1986
			H. Shelden 1991; 1998
6	Pagu	North Halmahera	Wimbish 1991
7	Sahu	North Halmahera	Visser and Voorhoeve 1987
8	Tidore	North Halmahera	van Staden 2000
9	Tobelo	North Halmahera	Holton 2003
10	West Makian	North Halmahera	Voorhoeve 1982
11	Abun	West Bird's Head	Berry and Berry 1999
12	Hatam	Hatam-Mansim, Bird's Head	Reesink 1999
13	Inanwatan	South Bird's Head	de Vries 1996; 1998; 2002; 2004
14	Maybrat	isolate, Bird's Head	Dol 1999
15	Meyah	East Bird's Head	Gravelle 2002; 2004
16	Moi	West Bird's Head	Menick 1996; 2000; Reesink 2008
17	Moskona	East Bird's Head	Gravelle-Karn 2010
18	Mpur	isolate, Bird's Head	Odé 2002
19	Sougb	East Bird's Head	Reesink 2002
20	Tehit	West Bird's Head	Hesse 1995; 2000
21	Yawa	Cenderawasih Bay	Jones 1986; 1991; Reesink 2005

References

Baarda, M.J. van, 1908. *Leiddraad bij het bestudeeren van 't Galela'sch dialekt, op het eiland Halmaheira*. The Hague: Martinus Nijhoff.

Baird, Louise, 2008. *A Grammar of Klon: a non-Austronesian language of Alor, Indonesia.* Canberra: Pacific Linguistics.

Berry, Keith and Christine Berry, 1999. *A Description of Abun, a West Papuan language of Irian Jaya*. Canberra: Pacific Linguistics.

Capell, A., 1975. The 'West Papuan Phylum': general, and Timor and areas further west. In S.A. Wurm, ed., *Papuan languages and the New Guinea linguistic scene*, 667–716. Canberra: Pacific Linguistics.

Comrie, Bernard, 1980. Morphology and word order reconstruction: problems and prospects. In J.Fisiak, ed., *Historical morphology*, 83–96. The Hague: Mouton.

Cowan, H. K. J., 1957. 'A large Papuan language phylum in West New Guinea, *Oceania*, 28:159–166.

Cysouw, Michael and Comrie, Bernard, forthcoming. New Guinea through the eyes of WALS. *Language and Linguistics in Melanesia*.

Dol, Philomena, 1999. A grammar of Maybrat. Leiden: Leiden University. PhD dissertation.

Donohue, Mark, 2005. Word order in New Guinea: dispelling a myth, *Oceanic Linguistics* 44:527–536.

Dryer, Matthew, 2005. Prefixing versus suffixing in inflectional morphology. In Martin Haspelmath, Matthew Dryer, David Gil and Bernard Comrie, eds, *The World Atlas of Language Structures*, 110–113. Oxford: Oxford University Press.

Enrique-Arias, Andrés, 2002. Accounting for the position of verbal agreement morphology with psycholinguistic and diachronic explanatory factors, *Studies in Language*, 26:1–31.

Etherington, Paul A., 2002. Nggem morphology and syntax. Darwin: Northern Territory University, Australia. MA thesis.

Foley, William A., 1986. *The Papuan languages of New Guinea*. Cambridge: University Press.

— 1998. Towards understanding Papuan languages. In Jelle Miedema, Cecilia Odé and Rien A.C. Dam, eds, *Perspectives on the Bird's Head of Irian Jaya, Indonesia*, 503–518. Amsterdam: Editions Rodopi.

— 2000. The languages of New Guinea, *Annual Review of Anthropology*, 29:357–404.

— 2005. Semantic parameters and the unaccusative split in the Austronesian language family, *Studies in Language*, 29:385–430.

Givón, T., 1971. Historical syntax and synchronic morphology: An archaeologists field trip, *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society*, 394–415.

Gravelle, Gilles, 2002. Morphosyntactic properties of Meyah word classes. In Ger Reesink, ed. *Languages of the eastern Bird's Head*, 109–180. Canberra: Pacific Linguistics.

— 2004. Meyah, an east Bird's Head language of Papua, Indonesia. Vrije Universiteit, Amsterdam. PhD dissertation.

Gravelle-Karn, Gloria, 2010. A grammar of Moskona: An East Bird's Head Language of Papua, Indonesia. Vrije Universiteit Amsterdam. PhD dissertation.

Haan, Johnson W., 2001. The grammar of Adang: a Papuan language spoken on the island of Alor - East Nusa Tenggara - Indonesia. Sydney: The University of Sydney. PhD dissertation.

Haspelmath, Martin, Dryer, Matthew, Gil, David and Comrie, Bernard, 2005. *The world atlas of language structures*. Oxford: Oxford University Press.

Hesse, Ronald, 1995. Syllable structure in Imyan Tehit, *Language and Linguistics in Melanesia*, 26:101–171.

— 2000. Tehit. In Ger Reesink, ed., *Studies in Irian Languages, Part II. NUSA Volume 47*, 25–33. Jakarta: Lembaga Bahasa, Universitas Katolik Atma Jaya.

Heine, Bernd, 1997. *Cognitive foundations of grammar*. Oxford: Oxford University Press.

Himmelmann, N. P., 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In K.A. Adelaar and N.P. Himmelmann, eds, *The*

Austronesian languages of Asia and Madagascar, 110–181. London: Routledge/Curzon.

Holton, Gary, 2003. *Tobelo*. München: Lincom Europa.

— 2008. The rise and fall of semantic alignment in North Halmahera, Indonesia. In Mark Donohue and Søren Wichman, eds, *The typology of semantic alignment,* 252–276. Oxford: Oxford University Press.

Jones, Linda K., 1986. The question of ergativity in Yawa, a Papuan language, *Australian Journal of Linguistics*, 6:37–55.

— 1991. Question words in Yawa. In Tom Dutton, ed., *Papers in Papuan linguistics No.1*, 97–105. Canberra: Pacific Linguistics.

Kern, Johan H. C., 1891. Opmerkingen over't Galelareesch naar aanleiding der Beknopte Spraakkunst van M.J. van Baarda. *Bijdragen Taal Land en Volkenkunde*, 40:493–530.

— 1892. Opmerkingen over't Galelareesch naar aanleiding der Beknopte Spraakkunst van M.J. van Baarda. *Bijdragen Taal Land en Volkenkunde*, 41: 15–138.

Klamer, Marian, 2008. The semantics of Semantic Alignment in eastern Indonesia. In Mark Donohue and Søren Wichman, eds, *The typology of semantic alignment*, 221–251. Oxford: Oxford University Press.

— forthcoming. *A grammar of Teiwa.* Berlin/New York: Mouton de Gruyter.

Klamer, Marian and Kratochvíl, František, 2006. The role of animacy in Teiwa and Abui (Papuan). *Proceedings of BLS 32*. Berkeley: Berkeley Linguistic Society.

Klamer, Marian, Ger Reesink and Miriam van Staden, 2008. East Nusantara as a linguistic area. In Pieter C. Muysken, ed., *From linguistic areas to areal linguistics*, 95–150. Amsterdam: John Benjamins.

Kratochvíl, František, 2007. A grammar of Abui. Leiden: Leiden University. Utrecht: LOT. PhD dissertation.

Menick, Raymond H., 1996. Verb sequences in Moi. In Ger P. Reesink, ed., *Studies in Irian languages, Part I. NUSA Volume 40*, 41–60. Jakarta: Lembaga Bahasa, Universitas Katolik Atma Jaya.

— 2000. Moi. In Ger P.Reesink, ed., *Studies in Irian languages, Part II. NUSA Volume 47*, 7–24. Jakarta: Lembaga Bahasa, Universitas Katolik Atma Jaya.

Odé, Cecilia, 2002. A sketch of Mpur. In Ger P. Reesink, ed. *Languages of the eastern Bird's Head*, 45–107. Canberra: Pacific Linguistics.

Pawley, Andrew, 2007. Recent research on the historical relationships of the Papuan languages, or, what does linguistics say about the prehistory of Melanesia. In Jonathan S. Friedlaender, ed., *Genes, language, and culture history in the Southwest Pacific*, 36–58. Oxford: Oxford University Press.

Reesink, Ger P., 1998. The Bird's Head as Sprachbund. In Jelle Miedema, Cecilia Odé and Rien A.C. Dam, eds, *Perspectives on the Bird's Head of Irian Jaya, Indonesia*, 603–642. Amsterdam: Rodopi.

— 1999. *A grammar of Hatam*. Canberra: Pacific Linguistics.

— 2005. West Papuan languages: roots and development. In Andrew Pawley, Robbert Attenborough, Jack Golson and Robin Hide, eds, *Papuan Pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*, 185–218. Canberra: Pacific Linguistics.

— 2006. West Papuan languages. In Keith Brown, ed., *Encyclopedia of Language and Linguistics 2nd Edition*, Vol. 13:555–558. Oxford: Elsevier Limited.

— 2008. Lexicon and syntax from an emic viewpoint, *Studies in Language*, 32:4, 866–893.

Reesink, Ger, ed., 2002. *Languages of the eastern Bird's Head*. Canberra: Pacific Linguistics.

Ross, Malcolm, 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In Andrew Pawley, Robbert Attenborough, Jack Golson and Robin Hide, eds, *Papuan Pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*, 15–65. Canberra: Pacific Linguistics.

Siewierska, Anna, and Bakker, Dick, 1994. The distribution of subject and object agreement and word order type. In Anna Siewierska, ed., *Eurotype Working papers* 6, 83–126.

Shelden, Deirdre, 1986. Topical and non-topical participants in Galela narrative discourse, *Papers in New Guinea linguistics*, 25:233–248. Canberra: Pacific Linguistics.

Shelden, Howard, 1991. Galela pronominal verb prefixes. In Tom Dutton, ed., *Papers in Papuan linguistics No.1*, 161–175. Canberra: Pacific Linguistics.

— 1998. Transitivity and Galela pronominal reference. SIL Electronic Working Papers. http://www.sil.org/silewp/1998/005/SILEWP1998-005.html. (accessed 15 November 2007)

Staden, Miriam van, 2000. Tidore, a linguistic description of a language of the North Moluccas. Leiden: Leiden University. PhD dissertation.

Visser, L. E. and C. L. Voorhoeve, 1987. *Sahu-Indonesian-English dictionary and Sahu grammar sketch.* Dordrecht: Foris.

Voorhoeve, C. L., 1982. *The Makian languages and their neighbours*. Canberra: Pacific Linguistics.

— 1988. The languages of the northern Halmahheran stock. *Papers in New Guinea Linguistics No. 26*, 181–209. Canberra: Pacirfic Linguistics.

Vries, Lourens de, 1996. Notes on the morphology of the Inanwatan language. In Ger P. Reesink, ed. *Studies in Irian languages Part I. NUSA Volume 40*, 97–127. Jakarta: Lembaga Bahasa, Universitas Katolik Atma Jaya.

— 1998. Some remarks on the linguistic position of the Inanwatan language. In Jelle Miedema, Cecilia Odé and Rien A.C. Dam, eds, *Perspectives on the Bird's Head of Irian Jaya, Indonesia*, 643–653. Amsterdam: Editions Rodopi.

— 2002. An Introduction to the Inanwatan language of Irian Jaya. In K. Alexander Adelaar and Robert Blust, eds, *Between worlds: linguistic papers in memory of David John Prentice*, 77–93. Canberra: Pacific Linguistics

— 2004. *A short grammar of Inanwatan, an endangered language of the Bird's Head of Papua, Indonesia.* Canberra: Pacific Linguistics.

Wimbish, Sandra Gay, 1991. An Introduction to Pagu through the analysis of narrative discourse. Arlington: The University of Texas at Arlington. MA thesis.

5 *Person marking, verb classes, and the notion of alignment in Western Pantar (Lamma)*

GARY HOLTON

1 Introduction

Much recent work on the languages of eastern Indonesia (for example Klamer 2008, Ewing, this volume) has focused on the identification of systems of argument alignment, particularly so-called active/agentive systems (Mithun 1991), which code single arguments of some intransitive clauses like agents of transitives, and others like patients of transitives. As is now well-understood, alignment patterns do not in general characterize languages but rather particular subsystems of a language. Thus, languages often exhibit splits in which different subsystems follow different alignment patterns. For example, in the Australian language Dyirbal, Dixon (1979) famously demonstrated a split based on person: first and second person pronouns follow a nominative-accusative pattern, while third person pronouns follow an ergative-absolutive pattern. Furthermore, just as alignment patterns are not always constant across a particular language, alignment patterns may also exhibit significant variation across genetically related languages. Ewing (this volume) discusses the varying degrees of instantiation of agentive systems in genetically related languages of Central Maluku; Klamer (2008) presents case studies of semantic alignment in nine languages of East Nusantara; and Holton (2008) describes the evolution of agentive systems within the closely related languages of North Halmahera.

Yet, argument alignment is not always so easily described in terms of systems, such as active/agentive, accusative or ergative. Indeed, in some languages even the identification of an alignment system can present significant challenges. The non-Austronesian languages of the Alor-Pantar (AP) group present just this challenge. Spoken on the islands of Alor and Pantar in the eastern Indonesian province of Nusa Tenggara Timur, the AP languages form a well-defined subgroup of some 15 to 20 languages exhibiting striking similarities in pronominal morphology.[1] Both nominative and agentive systems are found,

[1] While much of the literature includes certain non-Austronesian languages of Timor within a larger Timor-Alor-Pantar group (Stokhof 1975), the precise relationship of the Timor languages and the degree to which they may represent a substrate influence remains the subject of some controversy (Donohue and Schapper 2007). The (T)AP languages are in turn assumed to belong to Ross' (2005) West Trans-New Guinea linkage, though this relationship has also yet to be rigorously demonstrated.

Michael C. Ewing and Marian Klamer, editors.
Typological and areal analyses: contributions from East Nusantara, 97–117 . Pacific Linguistics, 2010
© This edition vested with Pacific Linguistics.

and splits in argument alignment are abundant. The alignment of independent pronouns in AP languages is relatively straightforward. However, for many of the AP languages the alignment of pronominal prefixes is difficult to characterize within traditional categories. In many of the western and central Alor languages the object of a transitive verb may be coded with a different pronominal prefix according to lexical verb class, leading to a split in the transitive object (cf. Kratochvíl 2007). Elsewhere, a single object prefix is the norm, but even so lexical verb classes play a role in determining the licensing of object prefixes. That is, not all verbs admit object pronominal prefixes. This is true for example of the Pantar languages Teiwa and Kaera (Klamer, this volume). In the western Alor language Klon, more than one pronominal prefix may be present (Baird 2008), while in the eastern Alor language Kula (Tanglapui) an inverse system has developed (Donohue 1996). The sheer variety of alignment systems in the Alor-Pantar region begs a closer look.

The goal of this chapter is to describe the argument alignment patterns in Western Pantar (Lamma), a previously undescribed AP language. Western Pantar (WP) exhibits a number of features which have not been found in other AP languages. Most notably, in WP certain intransitive verbs may index their single argument via the same pronominal prefix used to index the object of a transitive verb, reflecting what at first sight appears to be an ergative alignment system. However, by examining the distribution of pronominal prefixes across a range of lexical verb classes the putative strength of this particular alignment system can be seen to be less robust. Pronominal prefixes may index any argument role, and the alignment of prefixes across argument roles can only be explained by appeal to lexical verb classes. Hence, evidence from WP provides additional support for the notion of semantic verb classes implicit in extant descriptions of AP languages. While a detailed characterization of these classes remains beyond the scope of the current chapter, the WP data suggest that the distribution of pronominal prefixes in AP languages is better characterized in terms of such verb classes than in terms of traditional notions of argument alignment.

2 Western Pantar

Western Pantar (WP) is spoken by approximately ten thousand people on the island of Pantar, at the far western end of the AP language area. There are three principal dialects: Tubbe, Mauta, and Lamma. However, dialect variation is primarily lexical and does not extend to the pronominal systems described in this chapter. WP remains one of the most isolated of the AP languages. As of early 2007 the region lacked regular electricity, cellular phone service, and (broadcast or satellite) television. Although a road now connects portions of the WP-speaking region with the Austronesian-speaking region of the district capital on the north coast, no roads connect the WP region with other AP language regions (though a road does serve Maliang, an area of mixed WP (Mauta dialect) and Teiwa speakers). Perhaps because of this isolation, language use within the WP region remains vigorous. WP is the language of daily communication outside schools, churches, and government offices; and most children continue to be raised bilingual in WP and Alor Malay. Most of the data in this chapter derive from the Tubbe dialect and were collected during a total of 10 months of field work between June 2004 and July 2007, based on elicitation and on transcriptions of recordings of spontaneous narrative and conversation.

3 Person-marking

WP has two distinct paradigms of independent pronouns, which I refer to here as the actor (ACT) and undergoer (U) pronouns (see Table 1 below). In addition, WP has a single paradigm of bound pronominal prefixes. In order to describe alignment patterns I will use the heuristic primitives A (more agent-like argument of a transitive clause), S (single argument of an intransitive clause), and P (more patient-like argument of transitive clause). The constraints on the distribution of pronouns across A, S, and P arguments are fairly loose. Both undergoer independent pronouns and pronominal prefixes may index any of A, S, or P. However, there are no instances of an actor independent pronoun indexing a P argument.

Before proceeding to describe person-marking in detail it should be pointed out that with most verbs person and number reference are not obligatory. Both independent pronouns and bound pronominal prefixes may be omitted with zero reference. Thus, clauses with no explicit nominal referent are common, as in (1).[2]

(1) ang me golang ga
market LOC go.home PFV
'[S/he] came home from the market'

The referent of a clause such as (1) is generally inferred to be third person singular, however, other inferences are possible. For example, given the appropriate context (1) could also mean 'I came home from the market'. This could plausibly be uttered by a speaker entering a house upon return from the market. This meaning could of course be unambiguously expressed through the use of an independent pronoun.

(2) *nang ang me golang ga*
1S.ACT market LOC go.home PFV
'I came home from the market'

Moreover, a reference to first person could also be achieved via a pronominal prefix (though in this particular case use of a pronominal prefix is incompatible with the perfective aspect), as in (3).

(3) *ang me na-golang*
market LOC 1S-go.home
'I am coming home from the market'

While independent pronouns may occur optionally with many verbs, the use of pronominal prefixes is required with some verbs and is not possible with some other verbs. For example, the transitive verb *-sorang* 'sharpen' requires a pronominal prefix indexing the P argument, as in (4). A pronoun or pronominal prefix indexing the A argument (that

---

[2] Glossing follows Leipzig conventions with the following exceptions: the abbreviations ACT and U are used to refer to the actor and undergoer pronouns, respectively. Thus, the first person singular actor pronoun *nang* is glossed 1S.ACT. It is important to bear in mind that these labels refer to grammatical, not semantic categories. In addition, INCP is used to refer to the inceptive.

is, the more agent-like argument of the transitive clause) is not required and indeed is not present in this example.[3]

(4) *nang kalang diris wang ga-sorang*
1s.ACT spade whetstone on 3s-sharpen
'I am sharpening the spade on the whetstone'

In contrast, the transitive verb *lu'ung* 'cut' does not admit a pronominal prefix, as shown in example (5) below, as well as in (21) below.

(5) a. *nang yettu lu'ung*
1s.ACT tree cut
'I cut the tree'

b. **yettu na-lu'ung*
tree 1s-cut

c. **nang ga-lu'ung*
1s.ACT 3s-cut

The distinction between those verbs which admit pronominal prefixes and those which do not is lexically determined. In particular, certain transitive verbs require pronominal prefixes indexing the P argument, while other transitive verbs do not admit such prefixes, and yet others optionally allow pronominal prefixes.

The complete set of WP pronouns is shown in Table 1 below. As is the case across the AP group, these pronouns reflect the Proto-AP pattern of **n*-, **h*-, **g*- for first, second, and third persons, respectively, and the singular-plural distinction is marked via vowel quality. The undergoer pronouns are derived from the actor pronouns via the addition of a high front vowel, and the pronominal prefixes are reduced forms of the independent pronouns. The pronouns glossed 4s in Table 1 refer to fourth person, and are used in switch-reference to distinguish a distinct third person.[4] Two additional sets of pronouns, not shown in in Table 1, are used in adnominal possession.

Table 1: WP pronouns and pronominal prefixes

	actor	**undergoer**	**prefix**
1s	*nang*	*naing*	*na-*
2s	*hang*	*haing*	*ha-*
3s	*gang*	*gaing*	*ga-*
4s	*ang*	*aing*	*a-*
1pi	*ping*	*pi'ing*	*pi-*

[3] In contrast to the Pantar languages Teiwa and Kaera (Klamer, this volume), animacy does not play a role in the distribution of WP pronominal prefixes. Both animate and inanimate arguments may be indexed via a pronominal prefix.

[4] The WP fourth person pronouns are cognate with Adang obviative pronouns (Haan 2001). The choice of fourth person terminology derives from that used in Athabaskan and Eskimo linguistics (for example, Payne 1979) and is not intended to represent a theoretical claim regarding distinctiveness. Indeed, the function of the WP fourth person pronouns remains the subject of ongoing investigation.

	actor	**undergoer**	**prefix**
1pe	*ning*	*ni'ing*	*ni-*
1p PAUCAL	*taing*	*taing*	*ta-*
2p	*hing*	*hi'ing*	*hi-*
3p	*ging*	*gi'ing*	*gi-*

The actor and undergoer pronouns may occur alone, as in (6),

(6) a. *gang* *mising*
3s.ACT sit
'he is sitting'

b. *gaing* *massa*
3s.U tired
'he is tired'

or in an appositional construction together with a co-referential noun phrase, as in the examples in (7).

(7) a. *Hen* *gang* *mising*
Hen 3s.ACT sit
'Hen is sitting'

b. *aname* *ara* *sing* *gang* *mising*
person large this 3s.ACT sit
'this big (important) man is sitting'

Likewise, pronominal prefixes may also co-occur with a co-referential noun phrase or an independent pronoun. Additional examples are given in (7). In (8) the noun phrase *halia sing* 'that water' in the first clause is co-referential with the prefix *ga-,* and the argument *aname* 'person' in the second clause is co-referential with the second person plural prefix *hi-*. (Notice also that the first prefix indexes a P argument while the second indexes an S argument.)

(8) *halia* *sing* *i-ga-niaka* *allang* *yalu* *aname* *hi-yama* *kaising*
water that PROG-3s-see thus usual person 2p-come.toward scoop
'he saw that water which you people usually came to draw'

Pronominal prefixes are relatively rare in extended discourse. For example, the narrative text excerpt in (9) consists of eleven clauses, yet only two verbs occur with pronominal prefixes.

(9) *gabo* *ang* *a-raung* *kau* *ga'ung* *mising*
then 4s 4s-climb pandanus above sit
'afterward he climbed a pandanus tree and sat on top'

mising *prame* *kan* *kauwa* *hallang* *kuba* *ye* *ma*
sit long right NEG almost old.woman one come
'he didn't sit for very long before an old woman came along'

kuba *ginnu* *Kuba Tuma Kau*
old.woman this Kuba Tuma Kau
'this old woman was [named] Kuba Tuma Kau'

ai *kabbarung* *diggang* *gar* *ma* *halia* *sing* *a-kaising* *ga'ai*
louse bee carry.on.head as come water this 4s-scoop PFV
'she carried a beehive on her head as she came to scoop water'

ma *mising* *halia* *i-kaising* *hallang* *kau*
then sit water PROG-scoop almost pandanus

hissa *ye* *atang* *tasing*
seed one on.top fall
'while she was sitting scooping water a pandanus nut fell on her [head]'

In this case both pronominal prefixes are fourth person *a-*, here glossed 4s, used to indicate switch reference. The fourth person prefix is used to refer to a second third person, distinct from another third person referent.

3.1 Independent pronouns

Core participants of WP verbs may be indexed via independent pronouns. The choice of independent pronoun is motivated, though perhaps not completely determined, by participant semantics. Sufficiently controlling agent arguments are indexed with actor pronouns, while non-controlling arguments are indexed with undergoer pronouns. This results in an agent-patient system of alignment, in the sense of Mithun (1991). Actor pronouns code the grammatical agent, and undergoer pronouns code the grammatical patient. In particular, controlling arguments of intransitive verbs are coded with actor pronouns, as in (10).

(10) *nang* *mising*
1s.ACT sit
'I am sitting'

In contrast, non-controlling arguments of intransitive verbs are coded with undergoer pronouns.

(11) *naing* *massa*
1s.U tired
'I am tired'

Undergoer pronouns may also occur with active intransitive verbs when the referent is not sufficiently controlling.

Three broad classes of intransitive verbs can be distinguished, based on whether the verb in question occurs with actor pronouns but not undergoer; with undergoer pronouns but not actor; or with either actor or undergoer pronouns. Examples of the first type include intransitive verbs which select for highly controlling agents, as in (12). These verbs cannot occur with undergoer pronouns.

(12) *nang (*naing) birang* 'I speak'
*nang (*naing) biring* 'I run'
*nang (*naing) kalalang* 'I understand'

Examples of the second type include intransitive verbs with non-controlling participants, as in (13). These verbs generally cannot occur with actor pronouns.

(13) *naing (*nang) massa* 'I'm lazy'
*naing (*nang) ubah* 'I'm wet'
*naing (*nang) sisa* 'I'm dry'

Not every intransitive verb with a non-controlling participant refuses to admit actor pronouns. For some verbs it may be possible to force a reading with actor pronoun, as with *muddi* 'strong' in (20) below. Ultimately, the distinction between whether a verb takes actor or undergoer pronoun is lexically, not semantically determined. Indeed, verbs with seemingly identical semantics may take different pronoun sets. For example, the directional verbs *pia* 'descend' and *mia* 'ascend' occur with actor pronouns, as in (14) and (15).

(14) *nang (*naing)* *na-pia*
1S.ACT (*1S.U) 1s-descend
'I descended'

(15) *nang (*naing)* *na-mia*
1S.ACT (*1S.U) 1s-ascend
'I ascended'

However, the directional verbs *middang* 'come from below' and *yang* 'come from above' require undergoer pronouns, in spite of their semantic similarity to *pia* and *mia.*

(16) *naing (*nang)* *middang* *ga*
1S.U (*1S.ACT) come.from.below PFV
'I came up from below'

(17) *naing (*nang)* *yang* *ga*
1S.U (*1S.ACT) come.from.above PFV
'I came down from above'

The verbs *pia*, *mia*, *middang*, and *yang* function within a larger system of directional verbs, and there is no evidence that the last two of these verbs imply less control than the first two. Rather, the last two verbs have simply lexicalized with undergoer pronouns, perhaps reflecting an historical semantic shift from a less controlling, patientive form. While agent-patient systems may be semantically motivated, choice of pronoun is not directly determined by verbal or participant semantics (cf. Mithun 1991).

The third class of intransitive verbs may occur with either actor or undergoer pronoun, though for a given verb in this class one pronoun set is less marked. The following pair of examples, which differ only in the choice of actor versus undergoer pronoun, is illustrative. The verb in both cases is the active intransitive motion verb *pia* 'descend'. In (18) the use of a actor pronoun implies a controlling argument, thus yielding an unmarked reading with

this active verb. The use of a undergoer pronoun in (19) implies less control, indicating in this case that the referent is accompanying another person who may be in more control of the action, as captured by the use of the comitative in the English translation.

(18) *nang hoang me na-pia*
1s.ACT beach LOC 1s-descend
'I am going to the beach'

(19) *naing kan hoang me na-pia*
1s.U also beach LOC 1s-descend
'I too am going to the beach [with you]'

Crucially, the choice of actor versus undergoer pronoun is governed not by the verb itself but rather by the participant semantics. However, that is not to say that the choice between actor and undergoer pronouns is completely free. The verb *pia* 'descend' occurs in isolated, elicited forms with actor pronouns, as in (18). The alternate form with undergoer is marked, and this markedness gives rise to the implicature which yields the comitative interpretation. In this sense the WP independent pronoun system differs significantly from semantically aligned systems such as found in the West Papuan language Tobelo, where many intransitive verb roots occur freely with either actor or undergoer pronominal prefixes, with neither form being more marked (cf. Holton 2003:58).

In many cases the choice between actor and undergoer pronoun with an intransitive verb can be directly correlated to the degree of participant control. Where varying degrees of control can be attributed to the participant, an alternation is possible. For example, the forms with actor pronoun in (20) imply a degree of control on the part of the referent, while those with undergoer pronoun imply less control.

(20)

Actor Pronoun	Undergoer Pronoun
gang pali 'is he afraid?'	*gaing pali* 'he is afraid'
nang muddi 'I should be strong'	*naing muddi* 'I am strong'

Thus, the forms with undergoer pronouns are simple declarative statements, whereas the forms with actor pronouns are non-declarative. The question *gang pali?* allows the possibility that the referent may choose whether or not to be afraid, whereas the declarative statement *gaing pali* assumes that decision has been made, and is now beyond the referent's control. Similarly, the modal reading of *nang muddi* assumes that I still have a choice to be strong, for example, to carry on through difficult times. The non-modal *naing muddi* simply indicates that I am in a state of being strong, a state over which I have no control.[5]

Alternation between actor and undergoer pronouns is also possible with transitive verbs. Typically the more agentive, controlling argument of a transitive verb is coded with an actor pronoun, while the less agentive argument is coded with an undergoer pronoun, as in (21).

---

[5] An alternate interpretation of the use of actor pronouns with the irrealis mood forms, suggested by one of the reviewers of this chapter, is that actor pronouns are simply required in the irrealis mood, neutralizing the semantic contrast between control and non-control. This hypothesis certainly warrants further investigation.

(21) *nang* *gaing* *lu'ung*
1s.ACT 3s.U cut
'I cut him'

However, if neither argument of a transitive verb is sufficiently controlling, both may be coded with undergoer pronouns, as in (22).

(22) *naing* *gaing* *oswang* *aggi*
1s.U 3s.U outside take
'I took (coaxed) him outside'

In this example the use of a undergoer pronoun for the first person argument does not imply complete absence of control but rather a situation of less control than would be the case with an actor pronoun. Thus, the suggested gloss 'coaxed' rather than 'took'. As Mithun (1991) points out, such double undergoer constructions are common among agentive systems, reflecting the increased role of semantics in determining alignment patterns in such systems.

Finally, while most of the examples cited up to this point reflect an APV word order, alternate word orders are extremely frequent in natural discourse. For example, (23) reflects AVP word order, while (24) reflects PAV word order.

(23) *ning* *i-ga-niaka* *allang* *duang*
1pe.ACT prog-3s-see thus snake
'thus we saw the snake'

(24) *duang* *ning* *lu'ung* *gaterannang* *kanna* *ga* *ya*
snake 1pe.ACT cut completely already PFV and
'and we cut up the snake completely'

Thus, word order does provide cues as to the alignment of independent pronouns.

3.2 Pronominal prefixes

As is evident from Table 1, the WP pronominal prefixes are phonologically reduced forms of the independent pronouns.[6] Pronominal prefixes may occur with both intransitive or transitive verbs, but not all verbs admit pronominal prefixes. Where they do occur, pronominal prefixes may co-occur with a co-referential noun phrase or independent pronoun, as in the following examples.

(25) *nang* *duang* *ga-niaka*
1s.ACT snake 3s-see
'I saw a snake'

(26) *nang* *na-lama* *ta*
1s.ACT 1s-go IPFV
'I'm going'

---

[6] I use the term pronominal prefix here in a pre-theoretical sense, leaving open the question of whether these should be interpreted as agreement markers or pronominal arguments.

The choice between reference with independent pronoun, pronominal prefix or both is quite complicated, and depends in part on the verbal semantics. However, one characterization which is possible is that referents indexed via pronominal prefixes are less affected. Some evidence for this hypothesis can be gleaned from transitive verbs which do not permit P arguments to be indexed by both prefixes and pronouns. For such verbs an alternation is possible between pronoun and prefix, and there is a preference for independent pronoun over pronominal prefix with more highly affected referents. That is, the undergoer pronoun signals greater affectedness of P than does the pronominal prefix. Thus, the use of a pronoun with the verb *diti* 'stab' in (27a) indicates a more severe stabbing than does the form with a pronominal prefix in (27b).

(27) a. *nang* *gaing* *diti*
1S.ACT 3S.U stab
'I stabbed him' (severely)

b *nang* *ga-diti*
1S.ACT 3s-stab
'I stabbed him' (superficially)

The affectedness hypothesis extends to the use of pronominal prefixes with intransitive verbs as well. Many intransitive verbs may occur with a pronominal prefix indexing the single S argument, in addition to a co-referential independent pronoun or noun phrase. In such cases there is a tendency for pronominal prefixes to occur in irrealis contexts, often to express a desire or intention, as in (28a). In contrast, forms without the pronominal prefix, as in (28b), are more typically associated with realis contexts.

(28) a. *nang* *na-golang* *ta*
1S.ACT 1s-return IPFV
'I'm going to go home (but haven't yet)'

b. *nang* *golang* *ga*
1S.ACT return PFV
'I went home (already)'

This close association between the use of pronominal prefixes in imperfective contexts but not in perfective contexts is pervasive. However, modality not aspect is the crucial factor. Pronominal prefixes are found also with intransitive verbs in perfective, irrealis contexts, as in the negative form in (29).

(29) *na-baulung* *hallang* *kauwa*
1s-fall almost NEG
'I amost fell'

This close association of pronominal prefixes with irrealis contexts yields highly productive alternations between actor independent pronouns and pronominal prefixes, with prefixes used to express desire or intention.

(30) a. *nang tallang* 'I am swimming' *na-tallang* 'I want to, am about to swim'
b. *nang kale'e* 'I am vomiting (now)' *na-kale'e* 'I am going to vomit'

In middle constructions the pronominal prefix may be used together with a co-referential undergoer pronoun, as in (31).

(31) *naing* *na-ulang*
1S.U 1s-bathe
'I am bathing'

3.3 Summary of WP person-marking

In summary, person-marking in WP is instantiated via two distinct but overlapping systems of pronominals. Independent pronouns pattern according to an agentive or semantically aligned system. One set of pronouns is used to reference more actor-like arguments, while another distinct set is used to reference undergoer arguments. Pronominal prefixes can also be used to reference arguments, either with or without co-referential independent pronouns. However, the distribution of pronominal prefixes is much more complex, to the extent that a single pronominal prefix may reference arguments with any of the A, S, or P roles. In the following section, the distribution of these pronominal prefixes is discussed in some detail, and distinct verb classes are delineated according to varying alignment patterns of the prefixes.

4 Distribution of pronominal prefixes

Constraints on the occurrence of pronominal prefixes in WP are lexically determined and can thus be used to delineate verb classes. Not all verbs admit pronominal prefixes, and among those that do the argument referenced by the prefix may be limited to a certain subset of A, S, and P. Based on this distribution I recognize seven verb classes. It should be acknowledged from the start that this classification is somewhat ad-hoc. Others might recognize a smaller or larger number of verb classes. Indeed, in a previous chapter (Holton 2005) I recognized just five classes, lumping the Class III, IV, and V verbs (those for which the prefix is optional) together within a single class. The finer-grained classification presented here recognizes further restrictions on permissible macro-roles which can be indexed by the pronominal prefix. Other divisions are certainly possible. For example, the classification presented here essentially ignores transitivity, lumping intransitive and transitive verbs together within single classes. Thus, Class I consists of all verbs which do not admit pronominal prefixes, regardless of transitivity. The purpose of the classification here is to provide insights into alignment systems; hence the focus on constraints on the indexing of A, S, and P.

It should also be emphasized that while I have attempted in some cases to provide an ad-hoc semantic characterization of each verb class, the classes themselves are delineated strictly on morpho-syntactic grounds based on the distribution of pronominal prefixes. That said, semantics clearly has a role to play, in that argument structure and alignment patterns in WP are clearly based on semantic principles. Thus, Class II verbs with mandatory prefix indexing the P argument can be characterized as denoting highly transitive events with less affected P arguments, but this characterization derives ultimately from a characterization of the pronominal prefix as indexing low-affected arguments. In the following subsections I describe each verb class in turn.

4.1 Class I Verbs: no pronominal prefixes

Verbs in Class I are distinguished by their inability to occur with pronominal prefixes. This is a relatively large class of verbs which includes both intransitive-only verbs (including most statives) and transitive-only verbs, as in (32).

(32)

Class I intransitive verbs		Class I transitive verbs	
batta	'stupid'	*lu'ung*	'cut'
kutta	'fat'	*bugi*	'hit'
maba	'cold', 'well'	*komi*	'knead'
irpatang	'understand'	*aggi*	'take'
		tiggung	'add to'

This class includes in particular most stative intransitive verbs. That is, most stative intransitive verbs cannot occur with pronominal prefixes but must instead be indexed via undergoer pronouns.

(33) *naing maba ga (*na-maba)*
1S.U cold PFV
'I am recovered' (lit: 'I am already cold' [not feverish])

The semantics of transitive-only Class I verbs is more difficult to characterize, as there exist transitive verbs with similar meanings to Class I verbs which occur in other verb classes. For example, *kasi* 'split' belongs to Class VI (see §4.6 below), while *lu'ung* 'cut' belongs to Class I, and thus does not admit a pronominal prefix. One possible characterization of Class I transitive verbs is that they indicate a focus on the action itself rather than on its effect on the grammatical patient, as in (34), where the focus is on the activity of kneading rather than on its effect on the bread.

(34) *nang roti komi (*ga-komi)*
1S.ACT bread knead
'I am kneading bread'

However, the robustness of this tentative characterization has yet to be fully tested.

4.2 Class II Verbs: mandatory prefix indexes P

Class II verbs are transitive-only, and require a pronominal prefix obligatorily referencing the P argument. This is a large class of verbs of high transitivity; the P argument is very clearly affected and individuated (cf. Hopper and Thompson 1980). As such there is a slight tendency for the P argument to be third person inanimate; however, animate P arguments are also possible. The examples in (35) are representative.

(35) Class II (transitive-only) verbs

-lu	'chase'	*-sising*	'stare at'
-ammu	'suck'	*-bili*	'drag'
-ssi	'bite'	*-sidding*	'haul'
-pa'ang	'be different'	*-riang*	'care for'
-walaing	'hit'	*-boya*	'swing'
-walang	'wait for'	*-allas*	'persuade'

Class II verbs are bound forms, and thus cannot occur without a pronominal prefix. In citation form these verbs are given with third person prefix *ga-,* for example *gassi* 'bite it'. Class II verbs may also occur with independent pronoun co-referential with the pronominal prefix, but in such cases the pronoun has an emphatic reading.

4.3 Class III Verbs: optional prefix indexes 1/2 person S

Verbs in Class III are intransitive-only verbs which optionally index first and second person referents via pronominal prefixes. Verbs in this class generally describe dynamic activities, but not always. Some examples are given in (36).

(36) Class III (intransitive only) verbs

golang	'return'	*lama*	'walk'
ti'ang	'sleep, go to sleep'	*mudali*	'play'
biring	'run'	*mising*	'sit down'

Class III verbs typically occur without a pronominal prefix, as in (37). However, first and second person referents may optionally be indexed via a pronominal prefix, as in (38).

(37) nang bloko
1s.ACT jump
'I jumped'

(38) na-bloko
1s-jump
'I jumped'

The distinguishing criterial feature of Class III verbs is their failure to admit third person pronominal prefixes. Pronouns or full noun phrases must be used instead, as in (39).

(39) *gang bloko (*ga-bloko)*
3s.ACT jump
'he jumped'

4.4 Class IV Verbs: optional prefix indexes 1/2 person absolutive

Class IV verbs are ambi-valent. The intransitive form refers to an involuntary action or state. The corresponding transitive form has a causative sense, meaning that the A argument causes the P argument to carry out the action or come into the state denoted by the verb.

(40) Class IV (ambi-valent) verbs
olang 'fall over'
aroga 'angry'

Like Class III, this class is distinguished by the fact that its members do not admit third person prefixes, neither with intransitive (41) nor with transitive (42) forms.

(41) *gang olang ta (*gang ga-olang ta)*
3s.ACT fall.over IPFV
'he's about to fall over'

(42) *nang gi'ing olang (*nang gi-olang)*
1s.ACT 3p.U fall.over
'I laid them down (to sleep)'

With first and second person S arguments, the presence of a pronominal prefix is optional, as in (43). The use of the pronominal prefix is commonly associated with irrealis contexts, as in (44).

(43) *nang olang*
1s.ACT fall.over
'I fell over'

(44) *nang na-olang ta*
1s.ACT 1s-fall.over IPFV
'I'm about to fall over'

First and second person P arguments of Class IV verbs may be indexed via either independent pronouns or pronominal prefixes, but not via both, though there is a preference for the latter. For example, the form (46) with pronominal prefix referencing the P argument is more acceptable than that in (45).

(45) *nang haing (*h-)olang*
1s.ACT 2s.U (2s-)fall.over
'I laid you down (to sleep)'

(46) *nang (*haing) h-olang*
1s.ACT (2s.U) 2s-fall.over
'I laid you down (to sleep)'

First and second person A arguments of Class IV verbs are never expressed with pronominal prefixes; for these verbs pronominal prefixes are restricted to absolutive (S and P) arguments.

4.5 Class V Verbs: optional prefix indexes 1/2 person S and 3 person P

Class V verbs are also ambi-valent, and their behaviour is in many ways similar to those in Class IV. However, unlike Class IV verbs, verbs in Class V can occur with pronominal prefixes indexing third person arguments, though only in the P macro-role. As with Class IV verbs, the corresponding transitive form has a causative sense. Verbs in this class exhibit a high degree of semantic unity in that they generally denote involuntary states.

(47) Class V (ambi-valent) verbs
pali 'to be afraid'
natter 'to be standing'

With first and second person referents the intransitive form of these verbs may be expressed with either an independent pronoun (48) or a pronominal prefix (49). As with

Class IV intransitive verbs, the use of a pronominal prefix is associated with irrealis contexts.

(48) *nang pessing*
1s.ACT sneeze
'I sneezed'

(49) *na-pessing*
1s-sneeze
'I'm about to sneeze'

When a third person pronominal prefix occurs with a Class V verb it must be interpreted as indexing the P argument. The precise interpretation of such forms reflects a fair amout of lexicalization, as indicated by the gloss in (50). In order to express a third person S argument an actor independent pronoun is required, as in (51).

(50) *ga-pessing*
3s-sneeze
'sneeze him!' (i.e., 'wipe his nose')

(51) *gang pessing (*ga-pessing)*
3s.ACT sneeze
'he sneezed'

Curiously, Class V verbs do not admit first or second person pronominal prefixes to index P arguments. Thus, to express clauses such as 'he sneezed you' the undergoer independent pronoun *haing* must be used to refer to the second person.

4.6 Class VI Verbs: optional prefix indexes 1/2 person A or 3 person P

Verbs in this class are transitive only. An optional pronominal prefix may index either first and second person A arguments or third person P arguments. Some examples are given in (52).

(52) Class VI (transitive-only) verbs
-na 'eat'
-uddang 'buy'
-kasi 'split'

With Class VI verbs first and second person pronominal prefixes must index the A argument, as in (53). First and second person P arguments must be indexed with independent pronouns, as in (54).

(53) *attua hissa na-kasi*
coconut fruit 1s-split
'I split the coconut'

(54) *gang haing na*
3s.ACT 2s.U eat
'he ate you'

The distribution of prefixes on Class VI may reflect a frequency effect.[7] Class VI verbs are highly transitive, typically having A arguments which are high in agency and P arguments which are highly affected. Thus, the A argument of these verbs tends to be human and hence more likely to be first or second person, while the P arguments of these verbs tend to be third person non-human. In other words, the situation described in (53) is much more typical than that in (54).

---

[7] Thanks to Michael Ewing for calling this to my attention.

4.7 Class VII Verbs: optional second prefix indexes A

Class VII consists of verbs which may occur with two pronominal prefixes indexing the A and P arguments. There are very few verbs in this class, but all are high-frequency lexical items. These verbs are mostly transitive, requiring at a minimum one pronominal prefixes (indexing P), as in (55). However, some of the verbs in this class have free roots and may occur with zero, one, or two pronominal prefixes, as in (56).

(55) Class VII bound verbs
-ussar 'catch'
-niaka 'see'
-nia 'give'

(56) Class VII free verbs
tiaring 'close, guard'
asang 'say, speak, tell'

The ordering of pronominal prefixes within Class VII verbs is rigid. Where only one pronominal prefix occurs, it obligatorily indexes the P argument, as in (57).

(57) *gang* *na-niaka*
3s.ACT 1s-see
'he sees me'

With transitive-only verbs, when two prefixes occur, the order is also fixed. The first prefix indexes the A argument, while the second indexes the P argument, as in (58). However, the prefixes are not distinguished in form. These verbs consist of bound roots, which require a prefix indexing the P argument.

(58) *ke'e* *pi-ga-ussar*
fish 1p-3s-catch
'we are catching fish'

These verbs consist of bound roots, which require a prefix indexing the P argument. So the rigid prefix order can be explained by a layered morphology in which the prefix indexing the A argument is affixed to an already complex form containing a P prefix. In other words, with a verb like *ga-nia* the third person prefix *ga-* must refer to the P, not the A.

The situation is drastically different with the free verb roots in this class. As a free root *asang* may occur without a prefix, as in (59). And as with the other Class VII verbs, when *asang* occurs with a single prefix, that prefix indexes the P argument, as in (60).

(59) *o* *talle* *hang* *hala* *asang*
EXCL friend 2s.ACT FOC say
'Oh, friend, it's you that spoke'

(60) *ging* *biring* *wa* *wenang* *ga-asang*
3p.ACT run go old.man 3s-say
'they ran to tell the old man'

However, with two prefixes only the order P-A is possible, as in (61) and (62).

(61) *wee* *ga-i-tiaring*
fishpond 3s-4p-close
'they are guarding the fishpond'

(62) *nang* *ana* *ha-asang* *kauwa* *ing gob,* *ana*
1s.ACT now 2s-say NEG if day

dinni *ta* *ha-na-asang*
how.many before 2s-1s-say
'if I don't tell you now, how long will it be before I can tell you?'

The more 'normal' A-P prefix order is not possible with the above verbs, though there is some indication that younger speakers are beginning to use forms such as *naga'asang* 'I told him'. However, most speakers reject such forms in favour of either an independent pronoun *nang ga'asang* or the P-A prefix order.

5 Pronominal prefixes and alignment patterns

The distribution of pronominal prefixes described in the previous section can be difficult to interpret in terms of argument alignment patterns. If alignment is understood in terms of morphological coding of the A, S, and P arguments, then each verb class exhibits different alignment patterns. In some cases these individual patterns can be understood in a straightforward way. For example, Class II verbs index P arguments via pronominal prefixes; A and S arguments are not indexed. This reflects a nominative-accusative system which aligns the A and S roles. In contrast, Class IV verbs index S and P non-third person arguments via pronominal prefixes. These verbs are ambi-valent: the pronominal prefix indexes the P argument of transitive verbs and the S argument of intransitive verbs, reflecting an ergative-absolutive pattern.

Yet even within a single verb class the alignment pattern is not constant. For example, Class IV verbs employ pronominal prefixes only for non-third persons, thus the ergative system is reflected only in first and second person forms. This results in a type of person-based split which is rather odd cross-linguistically, since in split systems the ergative-absolutive pattern has been argued to be more motivated for non-speech act participants (cf. Mithun 1991; Mithun & Chafe 1999). When examined across all seven WP verb classes, attempts to discern a coherent pattern of alignment yield highly variable results. As summarized in Table 2, the alignment pattern of pronominal prefixes varies both by person and by verb class, and each class reflects a unique distribution of permissible A, S, and P roles across the various persons.

Table 2 succinctly summarizes the alignment of A, S, and P across each verb class and person category. It is necessary to break this information down by verb class and person category because each verb class (by definition) exhibits a different pattern of alignment and, moreover, for at least some classes, the patterns are different for different persons. To say that Class II verbs exhibit a nominative-accusative alignment pattern is equivalent to saying that pronominal prefixes index P arguments (but not A or S arguments) for all persons. Thus the 'P' entry for each person in Class II in Table 2 (for example) indicates, for each person in each verb class, which arguments can be indexed by a pronominal prefix.

Table 2: Arguments indexed by pronominal prefixes by verb class and person

	person		
Verb Class	1	2	3
I	-	-	-
II	P	P	P
III	S	S	-
IV	SP	SP	-
V	S	S	P
VI	A	A	P
VII	AP	AP	AP

Arguably, Table 2 would be easier to interpret if standard descriptive labels such as 'nominative-accusative' or 'ergative-absolutive' were assigned to each verb class. But such an approach is fraught with difficulty. Firstly, many of the patterns in the table cannot be readily described in terms of standard labels. For example, Class V verbs exhibit an alignment split based on person. Third persons follow a nominative-accusative pattern (P is indexed via a prefix; S and A are not). But the pattern for first and second persons groups S as opposed to A and P (S is indexed via a prefix; A and P are not). And we do not have a standard label to describe systems which group A and P as opposed to S.

More problematic is the fact that any attempt to force an interpretation of the distribution of pronominal prefixes across these verb classes in terms of the alignment of A, S and P yields results which are inherently unstable, since small perturbations in morphological structure yield huge variations in alignment patterns. For example, as discussed in the preceding paragraph, Class V verbs exhibit a split in alignment patterns based on person. First and second person S arguments are indexed via pronominal prefixes, while first and second person A and P arguments are not indexed on the verb. On the other hand, third person P arguments are indexed via pronominal prefixes, while third person S and A arguments are not indexed. Class V verbs thus group first and second person S as opposed to A and P, and third person P as opposed to S and A. The pattern with third person can be readily labeled nominative-accusative, while the pattern for first and second person is rather odd and more difficult to label. In any case, this sort of description belies the underlying sensitivity to semantics governing the choice of pronominal prefix.

For example, Class V verbs are ambi-valent, so that each verb in this class has both an intransitive and a transitive sense. Thus, the verb *pessing* can mean 'to sneeze' or 'to sneeze someone', that is 'to wipe someone's nose'. The transitive reading requires a highly controlling agent and a highly affected patient. Attributing this level of affectedness to a non-speech act participant is permissible, but attributing this level of affectedness to a speech act participant (first or second person) requires a undergoer independent pronoun, which cannot co-occur with a pronominal prefix. Thus, it is possible to express the statement 'he sneezed me,' provided first person reference is achieved with a pronoun rather than a prefix, as in (63).

(63) *gang naing (*na-)pessing*
3S.ACT 1S.U (1s-)sneeze
'he sneezed me'

With non-speech act participants either undergoer pronouns or pronominal prefixes are possible: *gaing pessing* or *ga-pessing* 'sneeze him!'. As this example demonstrates, the distribution of pronominal prefixes is at least in part motivated by the semantic parameter of affectedness.

In this regard it is interesting to compare WP with the Pantar language Nedebang. In Nedebang pronominal prefixes are generally restricted to indexing the P argument; however, there are some defective paradigms which show evidence of a system similar to that in WP, in which pronominal prefixes may index an S argument. For example, the paradigm for Nedebang *ola* 'go home' in (64) shows pronominal prefixes with first person S arguments.[8]

(64)	*nang n-ola*	'I go home'	ping *p-ola*	'we (inc) go home'
	hang ola	'you go home'	*ning n-ola*	'we (exc) go home'
	gang ola	'he or she goes home'	*ing ola*	'you (pl) go home'
			ging ola	'they go home'

Nedebang *ola* is cognate with WP *golang*, a Class III verb which optionally indexes first and second person S arguments via pronominal prefixes. These verbs share a restriction against the use of pronominal prefixes to index third person arguments, but WP has relaxed that restriction slightly to allow pronominal prefixes with second as well as first persons. If the pronominal prefix is viewed as a marker of low affected arguments, then Nedebang marks the first person arguments in (64) as being low affected, reflecting a kind of middle construction, as in returning to my own home (cf. Kemmer 1993). WP simply extends this interpretation to all speech act participants, including both first and second person. In this respect the semantic difference between Nedebang and WP is quite small, yet the consequences for argument alignment are quite large.

It should be acknowledged that the lexicon still plays an important role, in that verb classes reflect lexical not semantic groupings. Crucially, verb classes are not identical across the Alor-Pantar languages. There may even be significant variation across closely related languages. As Donohue (1997) laments regarding the Alor languages Kui and Kolana, 'annoyingly, although (some of) the pronominal forms are identical for the two languages, the verbs with which they occur aren't.' Furthermore, to say that the choice of pronominal prefix is semantically motivated is not to say that we can always provide a synchronic semantic explanation. Verb meanings may change over time, while the distribution of pronominal prefixes remains constant. Comparative information may eventually provide insight into the evolution of alignment systems in Alor-Pantar languages, but at the present time our understanding of the internal historical relationships remains in its infancy.

Indeed, to the extent that the present study is based on preliminary fieldwork, the conclusions presented here should be approached with some caution. What is clear is that traditional notions of alignment as operationalized via the alignment of A, S, and P arguments do little to elucidate the distribution of pronominal prefixes in WP. Rather, as I have argued for non-Austronesian languages elsewhere in East Nusantara, the multiplicity of formal alignment systems belies an underlying semantic unity. For example, in the closely-related North Halmaheran languages Tobelo and Galela, the choice of actor versus undergoer pronoun with intransitive verbs is governed by identical semantic principles

---

[8] Nedebang data are from the author's fieldwork in 2004.

(Holton 2008). However, due to differing morphological constraints Tobelo reflects a formally nominative-accusative system, while Galela reflects a semantically-aligned stative-active system. As more data become available, differences in formal alignment patterns between Western Pantar and other Alor-Pantar languages may similarly turn out to reflect structural constraints rather than actual differences in semantic categorization.

References

Baird, Louise, 2005. Doing the split-S in Klon. In J. Doetjes. and J. van der Weijer, eds, *Linguistics in the Netherlands*, 1–12. Amsterdam: John Benjamins.

— 2008. *A Grammar of Klon: A Non-Austronesian Language of Alor, Indonesia*. Canberra: Pacific Linguistics.

Dixon, R.M.W., 1979. Ergativity, *Language*, 55:59–138.

Donohue, Mark, 1996. Inverse in Tanglapui, *Language and Linguistics in Melanesia*, 27.101–18.

— 1997. Developments of affectedness marking in the languages of Alor (split, splat). Paper presented at the University of California Berkeley, 25 September.

Donohue, Mark and Antoinette Schapper, 2007. Towards a morphological history of the languages of Timor, Alor and Pantar. Paper presented at the Fifth East Nusantara Conference, Kupang, Indonesia, 2 August.

Haan, Johnson Welem, 2001. A grammar of Adang: a Papuan language spoken on the Island of Alor, East Nusa Tenggara, Indonesia. Sydney: The University of Sydney. PhD dissertation.

Holton, Gary, 2003. *Tobelo*. (Languages of the World/Materials 328). Munich: LINCOM Europa.

— 2005. Grammatical relations in Western Pantar. Paper presented at the Linguistic Society of America, Oakland, California, 6–9 January.

— 2008. The rise and fall of semantic alignment in North Halmahera, Indonesia. In Mark Donohue and Søren Wichmann, eds, *The Typology of Semantic Alignment*, 252–256. Oxford: Oxford University Press.

Hopper, Paul J. and Sandra A. Thompson, 1980. Transitivity in grammar and discourse, *Language*, 56:251–99.

Kemmer, Suzanne, 1993. *The Middle Voice*. (Typological Studies in Language 23). Amsterdam: John Benjamins.

Klamer, Marian, 2008. The semantics of semantic alignment in eastern Indonesia. In Mark Donohue and Søren Wichmann, eds, *The Typology of Semantic Alignment*, 221–251. Oxford: Oxford University Press.

Kratochvíl, František, 2007. A Grammar of Abui. Leiden, Leiden University. PhD dissertation.

Mithun, Marianne, 1991. Active/agentive case marking and its motivations, *Language*, 67:510–546.

Mithun, Marianne and Wallace Chafe, 1999. What are S, A, and O?, *Studies in Language*, 23:569–606.

Payne, Thomas E., 1979. Who's kissing who: The fourth person in Yupik Eskimo. In UCLA, ed., *Studies of Switch-Reference*, 65–88. Los Angeles: Dept. of Linguistics University of California.

Ross, Malcom, 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In A. Pawley, R. Attenborough, J. Golson and R. Hide, eds, *Papuan Pasts: Cultural, linguistic and biological histories of Papuan-speaking peoples*, 15–66. Canberra: Pacific Linguistics.

Stokhof, W. A. L., 1975. *Preliminary notes on the Alor and Pantar languages (East Indonesia)*. Canberra: Pacific Linguistics.

6 *Agentive alignment in Central Maluku languages*

MICHAEL C. EWING

1 Introduction

In agentive systems of argument alignment (Mithun 1991), also called split intransitive or split S systems in the literature (Dixon 1979, Merlan 1985), the single argument of an intransitive clause may be coded as either an actor or as an undergoer. Such systems of coding argument structure are found in a range of diverse languages of the world and continue to prove theoretically interesting to linguists. Recent papers by Klamer (2006) and Mithun (2008) suggest that the feature of agentive alignment, even when apparently common through a group of languages, cannot be assumed to be evidence of either areal or genetic relationships. In Indonesia, the existence of agentive systems has been suggested to be an areal feature of eastern Indonesian languages (Donohue 2004), but Klamer (2006) shows that in fact this feature is proportionately no more common in eastern than in western Indonesia, and in both regions numerous non-agentive languages also occur, thus suggesting split intransitivity has no special status in the grouping or comparing of languages in the region. This parallels Mithun's (2008) conclusion that among North American languages, although agentive systems are common, this should not be used as diagnostic of genetic relationships because the extent of their occurrence varies widely and such systems seem to develop spontaneously and easily. Mithun (2008) goes on to point out that this is contrary Nichols' (1992) generalization that the argument coding system of a language is a stable property of a language over time.

The aim of the present study is twofold. Firstly, it is to provide a fairly detailed presentation of agentive alignment in one Central Maluku language, Allang, in order to contribute further data to our knowledge of agentive languages. Because the Allang system is relatively robust, it provides a useful starting point for the second aim of the study, to look at argument coding more broadly across Central Maluku languages by surveying the alignment systems of other related languages in the Central Maluku group. Several but not all of these languages are found to have similar agentive systems. Yet the languages differ widely in the extent to which this system of argument alignment is integrated into their grammars, and the historical depth of these alignment systems cannot yet be determined for many of these languages. This adds further support to Klamer's (2006) and Mithun's (2008) suggestions that such systems tend to be sporadic and may not be reliable diagnostic markers of areal or genetic relationships.

Michael C. Ewing and Marian Klamer, editors.
Typological and areal analyses: contributions from East Nusantara, 119–141. Pacific Linguistics, 2010
© This edition vested with Pacific Linguistics.

2 Allang

The language of Allang is a variant of the Allang-Wakasihu-Larike language group, spoken on the western tip of Ambon Island in Maluku (See Laidig and Laidig 1991 and 1995 for descriptions of the Larike variety), and is a member of the West Piru Bay subgroup of East Central Maluku languages (Collins 1983). In the second half of the 19th century, Allang and neighbouring Liliboi were reported as among the few Christian villages on Ambon which still had a vital local language (van Hoëvell 1875:95). Beginning in the early 20th century, there has been a steady shift to Malay in Allang. The regional language is extinct in Liliboi and is is now moribund in Allang. The only remaining fluent speakers of Allang, out of a village population of over 4,000 people, are all over 60 years of age, and among those in this age category only about half (around 60 people) have good comprehension and production skills. The database used for this analysis consists of spontaneous conversational language and stories produced by speakers originally in Allang, augmented by data elicited through translation of stories and individual sentences from Malay.

3 Allang argument marking

3.1 Pronouns in Allang

Allang is a robust agentive language. It has paradigms of free pronouns, actor proclitics and undergoer enclitics as well as a paradigm of free possessive pronouns. These are differentiated for person, number and animacy as laid out in Table 1. In this study I will use the terms 'actor' and 'undergoer' to refer to the syntactic function of the clitic forms listed in Table 1 or their referents. I will use the terms A (more agent-like argument of a transitive clause), P (more patient-like argument of a transitive clause) and S (single argument of an intransitive clause) to refer to arguments of clauses and the ordering of these arguments. Thus in Allang, the A argument of a transitive clause is an actor and the P argument of a transitive clause is an undergoer, while the S of an intransitive clause may be either an actor or an undergoer.

Table 1: Allang pronoun paradigm

Person / Number	Free pronoun	Actor proclitic	Undergoer enclitic	Possessive pronouns
1s	*au, ai*	*au=*	*=au*	*aku*
2s	*ane*	*a=*	*=wa*	*amu*
3sh	*mane*	*mae=, me=*	*=ma*	*mana*
3nh		*i=*	*=a / =ya*	*ina*
1pi	*ite*	*ite=*	*=ite*	*iter*
1pe	*ami*	*ami=*	*=ami*	*amir*
2p	*imi*	*imi=*	*=imi*	*imir*
3p	*mati*	*mati=*	*=mati*	*matir*

Note that as well as being differentiated for person and number, among third person pronouns, a distinction is also made between human (h) and non-human (nh). Additionally, Allang makes no distinction between alienable and inalienable possession, and, unlike most other Central Maluku languages, it has a full set of independent possessive pronouns but no possessive clitic forms. The present-day systems of possessive

pronouns, partitive constructions and classifiers provide evidence that Allang did once have an alienable-inalienable system similar to that in other Central Maluku languages (Ewing 2005a). A close connection between argument coding and possessive forms has been noted for a number of agentive languages (see for example chapters by Reesink and Remijsen in this volume). As it happens, there appears to be no synchronic relationship between these two subsystems of the grammar in Allang. We will see, however, that in some of the other languages surveyed here a close resemblance between argument coding and possessive constructions is evident. Possessive paradigms are thus included for Allang as well as the other languages discussed below in order to provide information on the extent to which this relationship appears across these languages.

3.2 Transitive clauses

The following examples illustrate different configurations of transitive clauses in Allang. Because the occurrence of various free and bound forms are interrelated, examples employing lexical noun phrases, free pronouns and bound pronouns to code arguments are included.

(1) *Mansia ele-ti raha=wa.*
person other-PL stab=2s.U
'Some other people stabbed you.'

(2) *Mane ripi arnasi-nu na usa-na.*
3sh peel pineapple-LM[1] POSS skin-LM
'She peeled the pineapple skin.'

(3) *Urpae-nu ma mane na pasa ia-nu*
woman-LM DIST 3sh want grill fish-LM
'That woman is going to grill some fish.'

(4) *Lai Hina Mutua, mae=runu=a.*
arrive village Mutua 3sh.ACT=shoot=3nh.U
'Arriving at Mutua Village, he shot it (the arrow).'

(5) *Naci me=sali ia-nu.*
Naci 3sh.ACT-buy fish-LM
'Naci bought some fish.'

(6) *Mane me=patina=wa?*
3sh 3sh.ACT=know=2s.U
'He knows you?'

In these examples we see that A can be expressed by a lexical noun phrase, a free pronoun or both (1-3), by an actor proclitic (4), or by a combination of lexical noun phrase or free pronoun plus actor proclitic (5-6). P, when post-verbal, can be expressed by a

[1] LM 'lexical marker' refers to a semantically empty suffix, usually in the form -CV, occurring on many nouns and occasionally on lexemes from other word classes in Allang. Similar suffixes are found in many Central Maluku languages (see Musgrave, Ewing and Florey in press).

lexical noun phrase (2, 3, 5) or an undergoer enclitic (1, 4, 6) but never by both. All the clauses in the above examples have an A-V-P ordering of elements, whether understood with regards to the distribution of noun phrases or the use of clitics on the verb.

It is also possible in Allang for P to be expressed clause-initially as a free form, in which case a co-referential undergoer enclitic will also always occur on the verb. This is illustrated in (7) and (8).

(7) *Walita ria huho mati=hete-n=a wana.*
rope for fish.trap 3p.ACT=call-LM=3nh.U wana
'Rope for a fish trap, they call it *wana*.'

(8) *Uta-pu'u-nu racun ma ta bisa kanu=a.*
vegetable-mushroom-LM poison.MAL DIST NEG can.MAL eat=3nh.U
'Poisonous mushrooms cannot be eaten.'

If A is expressed in a P-fronted clause, it always occurs pre-verbally, that is, in the same position as A in a clause with a post-verbal P. The A, however, need not be explicitly expressed at all, as seen in (8). In these undergoer-fronted constructions, the V-P order is maintained at the level of clitics, and if the A is also explicit, a fully expressed A-V-P order is maintained. There is no grammatical passive or undergoer voice in Allang, and P-fronted clauses such as these can serve the functions of marking undergoer continuity through discourse or decreasing the prominence of an actor (see Musgrave this volume for a similar construction in Sou Amana Teru).

3.3 Intransitive clauses

Allang has an agentive alignment system in which the single argument of an intransitive clause is sometimes marked as an actor and sometimes as an undergoer. For the vast majority of intransitive verbs, the choice is lexically determined. Examples (9) through (12) illustrate a variety of actor intransitive clauses. The distributional possibilities of noun phrases, free pronouns and proclitics available to express an actor S are the same as those for the A of transitive clauses. Approximately two thirds of the intransitive verbs in the lexicon take actors.

(9) *La ami=sanapau nanala lai se lawa-i.*
and 1pe.ACT=walk continue arrive at garden-PL
'We walked continuously, arriving at the gardens.'

(10) *Mane lo'i intola pake sasina-nu*
3sh.ACT dance good use.MAL handkerchief-LM
'She dances beautifully with a handkerchief.'

(11) *Mama me=kene he asa-lu*
mother 3sh.ACT=go to market-LM
'Mother went to the market.'

(12) *I=rihu lou.*
3sn.ACT- fly far
'It (the balloon) flew far.'

Undergoer intransitive clauses are illustrated in (13) through (15). The distributional possibilities of noun phrases and enclitics available to express an undergoer S are the same as those for P. This includes the possibility of expressing an undergoer pre-verbally together with a coreferential enclitic. Approximately one third of the intransitive verbs in the lexicon take undergoers.

(13) a. *Paliri=wa?* b. *Ya, paliri=au.*
cold=2s.U yes cold=1s.U
'Are you cold?' 'Yes, I'm cold.'

(14) *Paliri mawo.*
cold cat
'The cat is cold.'

(15) *Mawo paliri=a.*
cat cold=3sn.U
'The cat, it's cold.'

3.4 Existentials

Allang has three existential forms, *se*, *he* and *ria*, and one negative existential, *tahi*. The forms *se* and *he* have extensively grammaticized, serving not only to predicate existence but also as general locative prepositions and as particles indicating realis mode. The presumed pathway of grammaticization would be from existential to locative and realis marker. In their predicating function, both *he* and *se* can be analysed structurally as actor intransitive verbs as they always occur with SV word order.

(16) *Mana napase se sala.*
3sh.POSS breath EXIST still
'He's still alive.' (Lit: 'His breath still exists.')

(17) *Aku ina la aku bapa he sala.*
1s.POSS mother and 1s.POSS father.MAL EXIST still
'My mother and father are still alive' (Lit: 'My mother and father still exist.')

The third existential, *ria*, can be analysed as an undergoer intransitive verb (18). It can occur in the full range of undergoer intransitive constructions, including with a post-verbal noun phrase, an enclitic or with a pre-verbal noun phrase together with enclitic.

(18) *Ria ani-n hute hama.*
EXIST wind-LM gentle there
'There's a gentle wind there.'

The word *ria* actually has a variety of meanings. These range from semantically richer uses such as 'to prepare' and 'to strike' to more generalized functions such as a preposition meaning 'for' and a relative clause marker. It can thus be seen to have participated in a wide ranging process of grammaticization in the development of Allang. The full story of

this grammaticization process is beyond the scope of the current study; however, one can hypothesize a pathway of grammaticization in which a verb like 'to prepare', which takes as its undergoer 'that which is prepared' (or 'brought into existence'), would be consistent with the undergoer structure of the more grammaticized from meaning 'EXIST'.

Allang has a negative existential, *tahi*, which is also an undergoer intransitive:

(19) *Tahi=au* *Onco.*
NEG.EXIST=1s.U Onco
'I wasn't there, Onco.'

4 Semantics and morphosyntax of the Allang semantic alignment system

4.1 The semantics of underived intransitive verbs

As seen above, Allang has a split S system that is characterized by a semantically based alignment of core arguments. Characteristics that are commonly expressed through agentive systems include animacy, aspect, agentivity and affectedness (Mithun 1991). Foley (2005) characterizes the semantic split found in the languages of Maluku as being based on aspect, specifically event vs. state or in some cases event and change into a state vs. being in a state. However, as I will show below, the split in Allang is broadly based on agentivity and affectedness rather than on aspect. Moreover, the semantic distinctions that have shaped the alignment have in fact produced a grammaticized system of lexicalized verb classes, which synchronically do not always respond to these semantic distinctions.

In Allang, the actor-undergoer distinction does not appear to have developed from notions of animacy or aspect. Actors of intransitive clauses can be either animate (20) or inanimate (21), as can undergoers (22) and (23).

(20) Animate Actor 'work'
Me=peseu *he* *lawa* *he* *lesta* *muli-nu.*
3sh.ACT-work at garden at house back-LM
'He worked in the garden behind the house.'

(21) Inanimate Actor 'go'
Motor *ma* *i=rnapau* *karena* *mati* *se* *una*
motorboat DIST 3nh.ACT=go because.MAL 3p.ACT RLS make

iya *ina* *wayer.*
good 3sn.POSS propeller
'The motorboat runs because they fixed the propeller.'

(22) Animate Undergoer 'surprised'
Ana *koile* *hitilia=ma.*
child small surprised=3h.U
'That small child was surprised.'

(23) Inanimate Undergoer 'wet'
Lopo *aku* *na* *pake-u.*
wet 1s.POSS POSS clothing-LM
'My clothes were wet.'

Similarly, actors can be the argument of both stative and eventive intransitive clauses (24) and (25), as can undergoers (26).

(24) State with Actor 'be tame'
Manua ma i=mahe.
chicken that 3s.ACT=tame
'That chicken is tame.'

(25) Event with Actor 'cross'
Ite musti lehi ndo Nusa Buono.
1i must.MAL cross seaward island Buono
'We must cross to Buono Island.'

(26) State with Undergoer 'be sleepy' - Event with Undergoer 'nod'
Ka-ta-tilo=ma nala ha-soi-soi mana ulo.
MID-Ca-sleep=3s.U until MID-nod-RDP 3.POSS head'
'He was sleepy to the point where his head was nodding.'

In Allang, present evidence suggests that the actor-undergoer distinction in intransitive clauses developed around the notions of (non-)agentivity and affectedness. An S argument that is both non-agentive and adversely affected is usually marked as an undergoer, as seen in the following examples.

(27) *Koho-ri ma na tapi isi=a.*
child-PL DIST POSS sarong lost=3nh.U
'The children's sarongs were lost.'

(28) *Rika=ya.*
broken=3nh.U
'It's broken'

(29) *Luarerene=ami.*
hungry=1pe
'We're hungry.'

It should be noted in passing that the semantically similar concept 'to be thirsty' is expressed using a possessive construction, which will be discussed below in Section 5.1.

Complementing these undergoer intransitive constructions in Allang, an S argument that is either agentive or not adversely affected is usually marked as an actor. The following examples of actor intransitives show agentivity through volitionality and control of action.

(30) *Hilu hina Alane ai=sanapau hale hina Waesihu.*
from village Allang 1s.ACT=walk transverse village Wakasihu
'I walked from Allang to Wakasihu'.

(31) *Mati alime-nu sala.*
3p play-LM still
'They are still playing.'

In the following examples, the S arguments of actor intransitive verbs are non-agentive but are not negatively affected.

(32) *Weil ma, i=rala kakoli.*
water DIST 3nh.ACT=flow twisting
'The stream flows in a twisting way.'

(33) *Ure ma i=ape.*
banana DIST 3nh.ACT=ripe
'The banana is ripe.'

In following example, the agent is negatively affected, but can be viewed as agentive, in that being drunk is usually brought on by one's own actions.

(34) *Undana ma me=tole karena me=kasili.*
man DIST 3s.ACT=noisy because.MAL 3s.ACT=drunk
'That man was being noisy because he was drunk.'

Although a semantic basis can be seen for the split in intransitive argument coding in Allang, the system is based on two firmly grammaticized verb classes, and not all cases can be understood synchronically on agentive grounds. Specifically, the class of actor intransitives also includes some verbs whose single participant is both non-volitional and adversely affected. For example, the notion 'to die' is often expressed by an undergoer intransitive in agentive languages (Mithun 1991), but in Allang *mata* 'to die, to be dead' is an actor intransitive, as seen in the second clause in (35).

(35) *Mae=roka=ya nala i=mata.*
3sh.ACT=slash=3nh.U until 3nh.ACT=dead
'He it slashed it (the pig) until it died.'

The notion 'to die' is similarly coded as an actor intransitive in other languages of Maluku, and Foley (2005) proposes that this is because of an aspectual focus on the change of state rather than the resulting state. While such an aspectual motivation may have been present earlier in the history of Allang, and thus account for the otherwise incongruous membership of *mata* in the class of actor intransitives, aspect as a motivating factor is not prominently visible in the present-day alignment system.

Semantic alignment systems often include a fluid S category where the same verb may take either an actor or an undergoer participant, with an appropriate semantic shift. Acehnese is a well-know example of a language which has both a lexically determined and a fluid split among intransitives (Durie 1985). Some semantic alignment systems may exhibit extensive fluid S, while others may exhibit none at all (see for example the survey in Klamer 2008). To date, I have identified only one intransitive verb in Allang which allows a fluid choice of case marking. The verb *piku* means 'hot' with an actor argument, but 'burnt' with an undergoer.

(36) *Ala ma i=piku*
rice that 3nh.ACT=hot
'The rice is hot.'

(37) *Luma-nu piku=a.*
house-LM hot=3nh.U
'The house burnt.'

A possible explanation for this alternation is that something that is hot is is by extension something that can burn (someone or something else). Nonetheless, the actor intransitive form does not synchronically imply any adverse situation: bath water that is comfortably hot (but does not burn) is described as *piku*. In contrast, the undergoer intransitive form always implies an adverse or destructive effect.

In general, the semantic alignment system of Allang is highly lexicalized. In cases where similar events might have S arguments with different degrees of agentivity, this may be marked lexically with different verbs in Allang. This is illustrated with two words meaning 'fall': *mulu* and *kate*. *Kate* is non-volitional, and connotes an adverse outcome. It takes an undergoer argument.

(38) *Au pese ta rua, kate=au.*
1s hold NEG strong, fall=1s.U
'I didn't hold on tightly and I fell.'

In contrast, *mulu* can be non-volitional but there is no adversity implied. Consequently its single argument is an actor.

(39) *Durenu ma i=mulu.*
durian DIST 3nh.ACT=fall.
'The durian fell.'

It is thus *mulu* that is the appropriate basis for a causative form, where causing to drop or fall is an intended action:

(40) *Hatu ma ria i=pa-mulu huho nehe meita waka-nu.*
rock that for 3nh.ACT=CAUS-fall fish.trap down sea bottom-LM
'The rock is for sinking the fish trap to the bottom of sea.'

The Allang semantic alignment system can be characterized as one in which the undergoer intransitive class is the marked category. This is because its contains fewer members than the actor class, and its membership is semantically more restricted. The actor intransitive category is the unmarked or default category, with a larger number of semantically more diverse members. (See Merlan 1985 for discussion of markedness in split intransitive systems).

4.2 The agentivity-altering prefixes *ha-* and *ka-*

Allang has two prefixes that both mark verbs with undergoer-intransitive argument structure: *ha-* and *ka-*. Both of these prefixes can function to derive undergoer-intransitive verbs from other types of verbs, and are therefore glossed as middle-deriving prefixes (MID). In addition, both prefixes also occur in lexemes that appear to be historically derived forms, but for which there is no longer any synchronic evidence for an independent meaning or function of the putative base form. In the lexicon currently available for

Allang, verbs with *ha-* include many more synchronically derivable forms (65% synchronically derived), while verbs with *ka-* are more likely to be historically derived forms that represent the only occurrence of their base forms (only 38% synchronically derived). Forms with the *ha-* prefix are much more frequent in the lexicon than forms with *ka-*, also suggesting the relatively greater productivity of *ha-*.

Verbs with *ha-* may be derived from transitive bases or actor-intransitive bases. They may also contrast with forms that have the causative suffix *pa-*. The transitive verb *lala* means 'untie'. As an undergoer-intransitive verb with the middle prefix *ha-*, *halala* means 'untied, unravelled, unkempt'.

(41) *Urpae-nu ma me=lala mana ulo-huto.*
woman-LM DIST 3sh.ACT=untie 2sh.POSS head-bun
'That woman untied her bun.'

(42) *Ana ma, ha-lala mana ulo wala-ri,*
child DIST MID-undo 3sh.POSS head hair-PL

karena mae=ta-rira=ya
because.MAL 3sh.ACT=NEG-comb=3sn.U
'That child's hair is a mess, because he doesn't comb it.'

The verb *hali* can be used as an actor-intransitive form meaning 'return' or transitively meaning 'turn over'. The undergoer-intransitive *hahali* means 'be upside down'.

(43) *Pea maka-kawe-ri, ite hali paheluke.*
finished person-marriage-PL 1pi return again
'After the wedding, we went back (home) again.

(44) *Mati hali=a la pa-tola=ya peluke.*
3p.ACT turn.over=3sn.U and CAUS-good=3sn.U again
'They turned it over and repaired it.'

(45) *Ha-hali=a.*
MID-turn.over-3sn.U
'It's upside-down.'

The base *tali'i* 'lean' occurs with both the causative prefix *pa-* and the middle prefix *ha-*. The form *patali'i* is an actor-intransitive whose single argument can be volitional.

(46) *Kalau au ntilo pa-tali'i, au nake-u i-ntola*
if.MAL 1s lie.down CAUS-lean 1s sleep-LM 3nh.ACT-pleasant
'If I lie down on an incline, I sleep well.'

With the middle prefix *ha-*, an S is unexpectedly or inappropriately at an angle.

(47) *Lile ma ha-tali'i=a.*
pillar DIST MID-lean=3nh.U
'The pillar is leaning'

Most undergoer-intransitive verbs with *ka-* appear to be historically derived forms. For example, there is currently no other form with *loto* other than *kalota* 'dry' nor any other form with *pusa* other than *kapusa* 'flat, deflated'.

(48) *Mati ta-bisa rue hama lou sebabe weil ka-loto=a.*
3p NEG-can.MAL stay there long because.MAL river MID-dry=3nh.U
'They couldn't stay there long because the river was dry.'

(49) *Balon kalau ka-pusa=ya, i=ta-bisa rihu.*
balloon if. MAL MID-deflated=3nh.U 3nh.ACT=NEG-can fly
'A balloon, if deflated, can't fly.'

A few verbs with *ka-* are clearly derived forms. Interestingly, all of these also include the use of *Ca-* partial reduplication in addition to *ka-*. Partial reduplication on its own derives nouns. However, for these examples with *ka-*, there are no independent forms with *Ca-*, although all the bases of these examples do occur independently or in other derived forms. Examples include *ntilo* 'lie down, sleep' with *katatilo* 'to fall asleep accidentally' and *uta-pu'u* 'mushroom' with *kapapu'u* 'appear suddenly'.

(50) *Ulanu puna ka-ta-tilo=au*
rain make MID-Ca-sleep=1s.U
'Rainy weather makes me tired'

(51) *Niko ka-pa-pu'u=ma se henei. Runi=ma se.*
earlier MID-Ca-pop.up=3h.U RLS here vanish=3h.U RLS

Ke rewa=ma.
go search.for=3h.U
'Earlier she appeared here. Now she's vanished. Go look for her.'

Like mono-morphemic undergoer intransitives, undergoer intransitive verbs with *ha-* and *ka-* also have a diverse range of semantics. This diverse semantics together with the relatively small number of examples in the lexicon means that it is difficult to identify a principle that differentiates the two suffixes semantically or functionally. There is the structural difference that *ka-* forms tend to be more historically frozen, and this itself in fact contributes to the difficulty in determining a semantic or functional difference between the two forms. So far, only one pair of forms with *ha-* and *ka-* on the same base have been identified. These exhibit the structural difference that the *ka-* form has *Ca-* partial reduplication. Additionally, the *ha-* form has a fully reduplicated base. Yet the two words have essentially the same meaning, so it is still not possible to determine a difference between the two prefixes.

(52) *Ha-silu-silu=mati.*
MID-sob-RDP=3p.U
'They were sobbing.'

(53) *mae=rani nala ka-sa-silu=ma*
3s.ACT=cry until MID-Ca-sob=3h.U
'He cried to the point he was sobbing.'

4.3 The Allang system of Argument expression

Allang argument coding is based on an actor-undergoer distinction, which is consistent across transitive and intransitive clauses. This distinction is consistent in its formal properties across transitive and intransitive clauses, in that there is more flexibility and variation in the realization of actors, whether transitive or intransitive, which may be expressed pre-verbally by free forms, proclitics, both or neither. This is in contrast to undergoers, which, in both transitive and intransitive clauses, generally must be expressed post-verbally by a free form or an enclitic but not by both. Semantically, the difference between actor and undergoer intransitives appears to have formed along lines of control and affectedness, but ultimately the distinction is now lexical in nature. There is virtually no flexibility in choice of expression. Undergoers are more restricted than actors, both structurally and semantically. While Allang does not have extensive verbal morphology, all the verbal prefixes in the language are geared around changing degrees of agentivity, resulting in the shifting of actor and undergoer roles. Allang verbal morphology – and clausal syntax more generally – does not appear to be very sensitive to either the issue of grammatical relations, as for example English, or to the information flow status of arguments, as for example Javanese (see Ewing 2005b). Instead, the actor-undergoer distinction appears to be one of the primary motivating factors in the overall grammar of argument-structure in Allang.

5 A look at other Central Maluku languages

This section examines case marking and argument coding in several related Central Maluku languages. The focus is on whether and to what extent these languages have an agentive alignment system. This survey of a set of closely related languages shows wide variation in agentive coding, and is consistent with the hypothesis that, although agent-patient case systems can be very deeply entrenched in individual languages, their presence or absence cannot be taken as diagnostic of relationships between languages.

5.1 Alune

Like Allang, Alune is an East Central Maluku language, but of the Three Rivers, rather than Priu Bay, branch (Collins 1983). The following table of Alune pronouns is based on Florey (2001, 2005).

Table 2: Alune pronoun paradigm

Person Number	Free pronoun	Actor proclitic	Undergoer enclitic	Alienable possessive proclitic	Inalienable possessive enclitic
1s	*au*	*'u=*	*='u*	*'u=*	*='u*
2s	*ale*	*a=*	*=mu*	*m(u)=*	*=mu*
3sh	*ile*	*(e)i=*	*=(n)i*	*((e)n)i=*	*=(n)i*
3sn	*ele*	*e=*	*=(l)e*	*((e)n)i=*	*=(l)e*
1pi	*ite*	*'i=/ma=*	*=ma*	*'i=/ma=*	*=ma*
1pe	*ami*	*'i=/ma=*	*=ma*	*'i=/ma=*	*=ma*
2p	*imi*	*mi=*	*=mi*	*mi=*	*=mi*
3ph	*sie*	*esi=*	*=si*	*si=*	*=si*
3pnh	*elu*	*u=*	*=(l)u*	*u=*	*=(l)u*

In Alune, alienable possession is marked with a proclitic on the head noun (54). Inalienable possession is marked with an enclitic on the head noun (55). Among conservative speakers of Alune, the morphological distinction between alienable possession and inalienable possession is still robust, although the system is changing in language shift contexts (Florey 1997, 2005).

(54) *Ami alena-'e atu-e ami 'i=ebe-ru.*
1pe narrate-APPL BEN-APPL 1pe 1p.POSS.AL=friend-PL
'We told (the story) to our friends.'

(55) *Au 'eu hole lo'o beta='u 'ai*
1s go visit ALL opp.sex.sibling=1s.POSS.INAL and

'wali='u mpai Rumah Tiga.
same.sex.sibling=1s.POSS.INAL DIR Rumah Tiga
'I went to visit my brother and sister at Rumah Tiga.'

Actors and undergoers can also be expressed with proclitic and enclitic forms respectively. These forms are almost identical with the proclitic and enclitic forms used to indicate alienable and inalienable possession. Differences do occur between actor and alienable proclitics, in second person singular and third person. The paradigms of undergoer and inalienable enclitics are identical. Use of actor and undergoer marking is illustrated in (56).

(56) *Mo'ai-a-ru esi=tneu au pene.*
male-PL-PL 3p.ACT=ask 1s PFV
'Men proposed to me.' (Lit: Men asked me.)

The vast majority of intransitive clauses take actor proclitics, suggesting that Alune has a fairly stable accusative system in which A and S are marked in the same way. Alune does, however, have a very limited number of predicates referring to certain experiential states which take a single argument marked by an enclitic.

(57) *Musu 'ai 'mala'e='u.*
hot and thirsty=1s
'It's hot and I'm thirsty.'

(58) *Au dila='u bei asu.*
1s fear=1s from dog
'I'm afraid of dogs.'

The construction exemplified above appears to be structurally ambiguus. Because inalienable possessive and undergoer enclitics are morphologically very similar in Alune, these clauses could be interpreted as either possessive constructions or undergoer intransitive constructions. There is a second, unambiguous construction in Alune to express certain experiential states. In (59), the concept 'be hungry' is expressed as a possessive construction meaning 'my stomach hurts'.

(59) *Au tia='u sene 'wate ele'i au pala 'abi mlau mlinu*
1s belly=1s.POSS.INAL irritate INTS then 1s bake cassava DIR garden
'I was very hungry (lit: My stomach was very irritating), then I baked some cassava in the garden.'

Does the construction type in (59) have a bearing on how (or whether) the enclitic construction type illustrated in (57) and (58) might be disambiguated?

There are similar constructions expressing physical and emotional states in Allang, which may help shed some light on the situation, exemplified in (60) through (62).

(60) *Paliri=au.*
cold=1s.U
'I am cold.'

(61) *Aku hause.*
1s.POSS thirst
'I am thirsty' (lit: I have thirst)

(62) *Pinu aku tia.*
full 1s.POSS stomach
'I'm full' (lit: My stomach is full)

These constructions include the following examples: (60) in which the experiencer is expressed as an undergoer and the experience as a predicate, (61) in which the experiencer is expressed as possessor of the experience and (62) in which the experiencer is expressed as the possessor of a body part which in turn is undergoer of the experience predicate. The third construction is grammatically analogous to the Alune example (59). The second and third of these Allang examples illustrate two contrasting types of experiential construction which Klamer (2001) has noted in other languages of East Nusantara. In (61) the enclitic form attaches to a word representing the experience, while in (62) the enlcitic attaches to a body part of the experiencer.

The Allang examples demonstrate that both the undergoer-experience and possessive-experience constructions can co-occur in a language and that both can co-occur with a possessed body part experience construction. This does not disambiguate the Alune forms, but rather provides evidence that either interpretation of their structure is plausible, thus suggesting that synchronically in Alune the grammatical structure of these clauses is in fact indeterminate between undergoer and possessor. Indeed, this indeterminacy makes these constructions in Alune a potential locus of reanalysis. We may be seeing a window on the possibility of an alienable-inalienable possessive system being a potential source of an agentive system. A close connection between possessive forms and argument coding has been noted in languages that show some form of agentive alignment, including Magey Matbat (Remijsen this volume) and the non-Austronesian languages of the East Nusantara region (Reesink this volume). Of the languages surveyed in this chapter, Alune shows the closest connection between possessive and argument marking, while Allang, with one of the most robust agentive systems discussed here, synchronically shows little relationship between these subsystems of its grammar.

5.2 Sou Amana Teru

Sou Amana Teru is the name currently in use among some inhabitants of Tulehu and certain neighbouring villages for the language they share. It is, together with Allang, part of the Piru Bay group of East Central Maluku languages, but a member of the East Piru Bay subgroup (Collins 1983). The following table of Sou Amana Teru pronominal forms is based on Musgrave (2005).

Table 3: Sou Amana Teru pronoun paradigm

Person Number	**free pronouns**	**clitic pronouns A, S / O**	**Inalienable possessive enclitics**
1s	*yau, au*	*u*	*=u*
2s	*yare, are, ar*	*m(u)*	*=m(u)*
3sh	*ire, ei*	*i*	*=i*
3nh	-	*- / =r(e)*	*=i*
1pi	*ike, ke, ka*	*ke, ka*	*=ke, =ka*
1pe	*yami, ami*	*mi*	*=mi*
2p	*imi*	*mu*	*=mu*

Sou Amana Teru clitic pronouns can function as actors when they occur preverbally and as undergoers when they occur post-verbally. The alienable-inalienable distinction still exists among more conservative speakers of Sou Amana Teru, although the system as it is currently used is not as robust as that found in Alune. For most speakers, possessives are marked by juxtaposition of possessor followed by the possessed. This is the common pattern across all speakers for alienable possession (63). This same pattern of juxtaposition can also be found to indicate semantically inalienable possession (64). More conservative speakers still retain inalienable marking by a possessive enclitic that is co-referential with a pre-head possessor, as in (65).

(63) *Yau lopu-'e na wa'a pe'e?*
1s knife-LM PRED LOC where
'Where is my knife?'

(64) *Au ina taha=i nau'e au a'a mahina sala*
1s mother NEG-3sh see 1s older.sibling female yet
'My mother hasn't seen my older brother's wife yet.

(65) *Bapa Acang eng ina=i*
Mr Acang 3s.POSS mother=3sh.INAL
'Mr Acang's mother'

Sou Amana Teru also has an agentive alignment system. Most intransitive verbs in Sou Amana Teru mark S like A, with an optional pre-verbal clitic, as in (66). Note that the preverbal bound pronoun is actually encliticized to the preceding negative particle.

(66) *Yau tah=u reu sala.*
1s not=1s go.home still
'I haven't gone home yet.'

A small class of intransitive verbs whose S argument is somehow affected obligatorily code the single argument with a post-verbal pronoun in addition to the optional preverbal coding.

(67) *Bapa Acang i-maruhu-i*
Mr Acang 3sh=hungry=3sh
'Bapa Acang is hungry'

(68) *Ar amuri'a=mu.*
2s tired=2s
'You're tired.'

This construction is similar to the so-called stative clause construction in Nuaulu, discussed below. Sou Amana Teru also has another type of double marking. In certain clauses which tend to refer to motion, S is optionally coded with pre-verbal clitic, but may also be optionally coded with a post-verbal clitic, as in (69).

(69) *Ike reu=ka.*
1pi go.home=1pi
'We went home.'

Such constructions are similar to the intradirective clauses (a term following Pawley 1973), also discussed below for Nuaulu.

5.3 Nuaulu

Nuaulu is a member of the Patakai subgroup of East Central Maluku languages (Collins 1983). Table 4 showing pronominal forms in Nuaulu is based on Bolton (1990), which is also the source of the example sentences.

Table 4: Nuaulu pronoun paradigm

Person Number	**Free Pronouns**	**V proclitic**	**V enclitic**	**N proclitic – alienable possession**	**N enclitic – inalienable possession**
1s	*au*	*u=*	*=ku*	*we(a)*	*=ku*
2s	*ano*	*a=*	*=(y)a*	*me(a)*	*=m*
3sh	*ia*	*i=*	*=(k)i*	*ne(a)*	*=i, =(n)i*
3sn	-	*(e)re=*	-	*nene*	*=e, =te*
1i	*ita*	*i=*	*=ta*	*re(a)*	*=ri*
1e	*ami*	*a=*	*=ma*	*mani(a)*	*=ma*
2p	*omi*	*o=*	*=mo*	*mo(a)*	*=mo*
3ph	*sia*	*o=*	*=so*	*no(a)*	*=o*
3pnh	-	*(e)ra=*	*=re*	*?*	*=a, =ta*

Like Alune, Nuaulu makes a clear distinction between alienable possession, marked by proclitics on the possessed, and inalienable possession, marked by enclitics, as in the following examples. Note that the alienable possessive forms, while phonologically proclitics, are written as separate words in Nuaulu orthography.

(70) *Au we topi rei-mo*
1s 1s.AL hat this-TOP
'This is my hat'

(71) *Ruka re=hoka unu=e*
monkey 3snh.ACT=come head=3sn.INAL
'The monkey stuck out his head.'

Nuaulu also has actor proclitics and undergoer enclitics, which attach to the verb. While there are a few sporadic correspondences between the clitics coding verbal arguments and possessive clitics (notably among plural enclitic forms), these paradigms are generally dissimilar. This is quite unlike the situation in Sou Amana Teru and Alune, where the sets of verbal and nominal clitics are nearly identical within each language.

Bolton (1990) presents five types of basic Nuaulu clause. The distinctions are based on possible argument structure, which is realized in part through patterns of clitic use. Her classification includes two types of multi-argument clauses, transitive and bitransitive, and three types of single-argument clauses, intransitive, intradirective, and stative. In all cases Bolton analyses an argument expressed pre-verbally as a subject. This subject can be expressed by free form (NP or pronoun), a proclitic, both or neither. For transitive and bitransitive clauses, a direct object can be expressed as a free form, an enclitic or neither; it is thus implied that the free form and enclitic cannot co-occur (Bolton 1990:97–98).

The single-argument clauses are most relevant to the present discussion of agentive systems. Like multi-argument clauses, all three types of single-argument clauses can have subjects expressed as free forms, proclitics, both or neither. The various types of single-argument clauses are differentiated semantically by the thematic role of the single argument and structurally by how enclitic marking is used. The single argument of an intransitive clause can function as an 'agent, experiencer if it is animate, or patient if it is inanimate' (Bolton 1990:99). Intransitive clauses cannot have an enclitic on the verb; the subject can only be marked preverbally as in (72).

(72) *Ranie re=sita tewasi.*
sun 3snh.ACT=shine not.yet
'The sun didn't shine yet.'

The single argument of an intradirective clause is marked by a proclitic and at the same time can be optionally marked with a co-referential enclitic. The single argument of an intradirective clause is thus 'a patient and is also usually an agent' (Bolton 1990:100). Intradirective verbs usually express a sense of motion.

(73) *Ia pina onate i=sipu=i, i=eu=i*
3sh female big 3s.ACT=get.down=3s.U 3s.ACT=go=3s.U

ria manahane.
inland outside
'The old woman got down and went outside.'

Stative clauses are obligatorily marked with an enclitic that is co-referential with the subject. As with other subjects, the subject of a stative clause may also be optionally marked with proclitic as well.

(74) *Unu=ku* *re=mpeta=i*
head=1S.ACTL= 3snh.ACT=wet=3S.ACT
'My head is wet.'

While examples (73) and (74) appear structurally similar to double marking of the subject by both proclitic and enclitic, the difference is in the optionality of the enclitic. In the intradirective clause the enclitic is optional, while in the stative clause the enclitic is obligatory. Semantically the subject of a stative verb can be an experiencer or a patient. The examples of stative verbs given include 'be dead', ' be sick', 'feel faint' and 'be wet'. While Bolton (1990) does not state it explicitly, this suggests that stative verbs express non-controlled and affected situations similar to those expressed by undergoer intransitives in Allang and Sou Amana Teru. The obligatory use of enclitics to mark the single argument of these verbs also suggests a similarity to undergoer intransitives. Correspondingly, the wider range of semantic functions of Bolton's class of intransitive subjects and their ability only to be marked preverbally suggests a similarity with the actor intransitives of Allang and Sou Amana Teru. The intradirective clauses are similar in structure and semantics to double-marked intransitives found in Sou Amana Teru. Neither Allang nor Alune have intradirectives; similar meanings of motion are expressed by actor intransitive constructions.

5.4 Rutah

The following table contains the pronoun paradigm for Rutah (Florey p.c.).

Table 5: Rutah pronoun paradigm

Person Number	**Free Pronoun**	**V proclitic**	**V enclitic**	**N proclitic – alienable possession**	**N enclitic – inalienable possession**
1s	*au*	*wa=*	*=u, ='u*	*u=, ='u*	*=u, ='u*
2s	*ale*	*u=, a=*	*=mo*	*m=, mu=, =m*	*=mo*
3sh	*iri*	*i=*	*=en*	*eni=, i=, n=*	*=(n)i*
3snh	*elo*	*e=*	*=esi, =lo*	-	-
1pi	*ama, ito*	*ma=*	*=ma*	*ma=*	*=ma*
1pe	*eko*	*ko=*	*=ko*	*ko=*	*=ko*
2p	*imi*	*mi=*	*=mi*	*mi=*	*=mi*
3p	*siri*	*si=*	*=si*	*si=*	*=si*

As with Allang, argument marking in Rutah forms a robust agentive system. Actor prefixes and undergoer suffixes are used to mark the A and P arguments of transitive classes. These same affixes mark the single arguments of intransitive clauses: proclitics are used with actor intransitives and enclitics with undergoer intransitives. Undergoer intransitive verbs are much more prevalent in Rutah than they are Sou Amana Teru or Alune (Florey p.c.). Nonetheless, within Rutah undergoer intransitives are still less common than actor intransitives, which, as in Allang, form the default intransitive verb type.

The following set of examples of first person singular forms illustrate this system with tokens of actor and undergoer clitics in transitive clauses (75) and (76), and intransitive clauses (77) and (78), (all Rutah examples from Flory, p.c.)

(75) *Heterue ne wa=ane hala-lo mo kurang e=m'ele nya.*
earlier this 1s.ACT=eat rice-LM but less.MAL 3sn.ACT=tasty PFV
'Earlier I ate rice but it wasn't tasty.'

(76) *Ale amariri-e=u nai kaalo?*
2s look.at-APPL=1s.U for what
'Why are you staring at me?'

(77) *Bisa wa=turi-e lee.*
able.MAL 1s.ACT=write-APPL also
'I'm able to write.'

(78) *Piri=u le amamuai='u.*
tired=1s.U therefore yawn=1s.U
'I'm tired so I'm yawning.'

Rutah makes a distinction between alienable and inalienable possession, marked by prefixes and suffixes respectively. This system is similar to that found in Alune, and contrasts with the free possessive pronoun system of Allang.

(79) *Pela ku=kepeng se oi nai seri.*
finish 1s.POSS.AL=money.MAL at where for pay
'My money's all gone just on paying the fare.'

(80) *Kalapesi-o=u aka.*
beard-LM-1s.POSS.INAL long
'My beard is long.'

We can see in the pronoun paradigm in Table 5 that many, but not all, possessive clitics have the same form as their corresponding core argument clitics. In this case, there can be specific examples of enclitic pronoun usage that are potentially ambiguous as to whether they code undergoers or inalienable possessive forms similar to the case discussed above for Alune. The statement in (81) has been glossed as an undergoer-intransitive clause, but it could also be understood as a possessive construction meaning 'my thirst' or 'I have thirst'.

(81) *Amalae='u.*
thirsty=1s.U
'I'm thirsty'

In Allang the same concept is unambiguously expressed as the possessive form as seen previously in (61), while numerous other expressions of physical sensation in Allang are clearly marked with undergoer forms. Such physically sensations are also marked with pronominal suffixes in Rutah as in the following:

(82) *Akakele-'e-m?* *Elo,* *akakele-'e-u.*
frighten-APPL-2s.U yes frighten-APPL=1sU
'Are you frightened? Yes, I'm frightened.'

5.5 Banda and Buru

All the languages discussed above are of the Nunusaku branch of East Central Maluku languages. Two non-Nunusaku Central Maluku languages for which data are available are Banda and Buru. Banda is a member of the Proto-Banda branch of East Central Maluku language (Collins 1982), and it also has an agentive alignment system. Among conservative speakers undergoer S is double marked with both a preverbal and a postverbal pronoun. Among younger speakers this paradigm is shifting and undergoer S is now often marked with a single pronoun: a preverbal pronoun by some speakers and a post verbal pronoun by others (Collins and Kaartinen 1998). Thus, in this shifting situation, the language is being pulled in two directions. If postverbal marking were to predominate, Banda would still have an agentive alignment system, but one which begins to look more like the present systems in other Central Maluku languages. However, if preverbal coding were to prevail, the alignment system will have shifted closer to an accusative system, with all S being coded like A. The existence of a semantic alignment system, and the current potential shift away from such a system, is interesting, because, based on other aspects of phonology and morphology, Banda retains many archaic properties which have been reconstructed for proto-Central Maluku, but which are reduced or absent from other languages in the group (Collins 1982, Collins and Kaartinen 1998).

Buru is a West Central Maluku language, and thus the most distantly related of the languages discussed here. Grimes (1991:93–108) classifies the verbs of Buru in three categories. The class of 'non-active' verbs are intransitive verbs where the semantic role of the single argument is the undergoer. Morphologically, these include verbs that are unmarked, or marked with one of the stative affixes *-em*, *-eb*, or *-t*. Syntactically, the S argument precedes non-active verbs, as it does for the S argument of intransitive active verbs, whose single argument is semantically an actor, and the A of transitive active verbs. Thus argument alignment in Buru is syntactically accusative. Morphologically all S arguments and A arguments can be coded with either a clitic or a pre-verbal free pronoun, while P arguments are coded by post-verbal free pronouns (but never by clitics). This pattern also shows an accusative alignment in the availability of clitic forms. (It should be noted however that Grimes (1991) prefers not to classify Buru as an accusative language on the grounds that choice of free pronouns and clitic forms is pragmatically driven, rather than based on grammatical roles.) Yet interestingly, Grimes (1991:156–159) also reports evidence from archaic and stylized speech forms that indicate that Buru once had a semantic alignment system that included both undergoer intransitive verbs marked with undergoer clitics as well as double marked intradirective constructions like those found in Sou Amana Teru and Nuaulu. Thus, Buru appears to be an example of a language which once had an agentive alignment system, but which now most closely resembles an accusative system.

6 Discussion

Allang has a clear and robust agentive alignment system. In addition, the actor-undergoer distinction that is the basis of this system is fundamental to Allang clause

morphosyntax, not only being involved in constituent order and classification of intransitive verbs types, but also being the driving force between verb morphology and valency changing processes. We have also seen that agentive alignment systems make some sort of appearance in many but not all of the Central Maluku languages surveyed here. Allang, Rutah, Banda and Nuaulu appear to have semantic alignment systems that are (or recently were) fairly robust, at least among conservative speakers. The semantic alignment system of Sou Amana Teru is not nearly as robust as that in these other languages, with only a small number of undergoer intransitive verbs. Alune has only a handful of possible undergoer intransitives, which in fact could be analysed as possessive constructions, and Alune might be better characterized generally as having an accusative alignment system. Buru clearly does not have a semantic alignment system, but rather one which is much closer to an accusative system. The frequency of agentive systems in these languages would appear to suggest a strong historical connection. However, it should not be assumed that these grammatical systems can be used to establish these genetic relationships. A taxonomy of the languages discussed here, based on the robustness of agentive alignment, does not correspond with the genetic relationships established by Collins (1983). The two most clearly accusative languages, Alune and Buru, are distant relatives, while more closely related language, for example Rutah and Sou Amana Teru, show very different degrees of development (or rentention) of agentive systems.

Agentive case marking occurs to some degree in many of these Central Maluku languages, but it is not presently a robust or unifying characteristic of all Central Maluku languages. This is consistent with Klamer's (2006) claim that agentive systems are sporadic across Indonesia, and do not define any areal subgroup. It is also consistent with Mithun's (2008) claim that Split-S systems are not a good diagnostic of genetic relationships. That is, here we have very closely related and geographically proximate languages, but in the area of argument coding we find both accusative and agentive systems, as well as languages that tend toward one or the other system, and may be in the process of shifting. There is evidence that the system in Banda is moving away from an older agentive alignment system and possibly towards accusative alignment. Similarly, in Buru, archaic forms suggest the existence of an earlier agentive alignment system, which no longer operates in the language. We do not at present have evidence for whether the alignment systems in either Sou Amana Teru or Alune are showing signs of incipient agentive alignment or the vestiges of a older system; that is, we cannot at this point make claims about a direction of change in these languages. The genetic relationship between these languages, as demonstrated by other means (Collins 1983), is certainly a likely contributing factor to the observed frequent occurrence, synchronically and in some cases historically, of agentive alignment. At the same time, the variation and change seen in the data are also consistent with Mithun's (2008) and Klamer's (2006) claims, and this suggests that these systems cannot be taken as themselves diagnostic of such a relationship.

References

Bolton, Rosemary Ann, 1990. A preliminary description of Nuaulu phonology and grammar. Arlington: University Texas at Arlington. MA thesis.

Collins, James T., 1982. Linguistic research in Maluku: a report on recent fieldwork. *Oceanic Linguistics*, 21:73–149.

— 1983. *The historical relationships of the languages of Central Maluku, Indonesia*. Canberra: Pacific Linguistics.

Collins, James T. and Timo Kaarkenin, 1998. Preliminary notes on Bandanese: language maintenance and change in Kei. *Bijdragen tot de Land-, Taal- en Volkenkunde*, 154:521–570.

Dixon, R. M. W., 1979. Ergativity. *Language*, 55: 59–138.

Donohue, Mark, 2004. Typology and linguistic areas. *Oceanic Linguistics*, 43:221–239.

Durie, Mark, 1985. *A grammar of Acehnese on the basis of a dialect of North Aceh*. Dordrecht: Foris Publications.

Ewing, Michael C., 2005a. Diverging and Converging Patterns of Possession: Allang in its Central Maluku Context. *Monash University Linguistic Papers*, 4:45–64.

— 2005b. *Grammar and inference in conversation: identifying clause structure in Javanese conversation*. John Benjamins: Amsterdam.

Florey, Margaret, 1997. Skewed performance and structural variation in the process of language obsolescence. In Cecilia Odé and Wim Stokhof, eds, *Proceedings of the Seventh International Conference on Austronesian Linguistics*, 639–660. Amsterdam: Editions Rodopi B.V.

— 2001. Verb and valence in Alune. *La Trobe Papers in Linguistics*, 11:73–120.

— 2005. A cross-linguistic perspective on emergent possessive constructions in Central Moluccan languages. *Australian Journal of Linguistics*, 25:59–84.

Foley, William A., 2005. Semantic parameters and the unaccusative split in the Austronesian language family. *Studies in Language*, 29:385–430.

Grimes, Charles, 1991. The Buru language of eastern Indonesia. Canberra: Australian National University. PhD Thesis.

Hoëvell, G.W.W.C. van, 1875. Iets over de vijf voornaamste dialecten der Ambonsche landtaal (bahasa tanah). *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch Indië*, 24:1–136.

Klamer, Marian, 2001. Phrasal emotion predicates in three languages of Eastern Indonesia. In Geert Booij and Saap van Marle, eds, *Yearbook of morphology 2000*, 97–121. Amsterdam: Kluwer Academic Publishers.

Klamer, Marian, 2006. Split S in the Indonesian area: forms, semantics, geography. Paper presented at the Tenth International Conference on Austronesian Linguistics. 17–20 January 2006. Puerto Princesa City, Palawan, Philippines. <http://www.sil.org/asia/Philippines/ical/papers.html>

— 2008. The Semantics of Semantic Alignment in Eastern Indonesia: Forms, Semantics, Geography, Possible Diffusion. In Mark Donohue and Søren Wichmann (ed.) *Typology of semantic alignment*, 221–252. Oxford: Oxford University Press.

Laidig, Wyn D. and Carol J. Laidig, 1991. *Tarus Sou Rikedu - Tata Bahasa Larike - Larike Grammar*. Ambon: Summer Institute of Linguistics and Center for Moluccan Studies and Development, Pattimura University,

— 1995. A Synopsis of Larike Phonology and Syntax. In Wyn D. Laidig, ed., *Descriptive Studies in Languages of Maluku*, 19–42. NUSA 38. Jakarta: Badan Penyelenggara Seri NUSA.

Merlan, Francesca, 1985. Split intransitivity: functional oppositions in intransitive inflection. In Johanna Nichols and Anthony C. Woodbury, eds, *Grammar inside and outside the clause*, 324–362. Cambridge: Cambridge University Press.

Mithun, Marianne, 1991. Active/agentive case marking and its motivations. *Language,* 67: 510–546.

— 2008. The Emergence of Agentive Patient Systems in Core Argument Marking. In Mark Donohue and Søren Wichmann (ed.) *Typology of semantic alignment*, 297–334. Oxford: Oxford University Press.

Musgrave, Simon, 2005. Possession in Souw Amana Teru: internally and externally motivated change. *Monash University Linguistics Papers* 4:65–81.

Musgrave, Simon, Michael Ewing and Margaret Florey, in press. -CV suffixes in Central Maluku languages: problems of form, distribution and function. *Australian Journal of Linguistics*.

Nichols, Johanna, 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago.

Pawley, Andrew K., 1973. Some problems in Proto-Oceanic grammar. *Oceanic Linguistics*, 12:103–188.

7 *Reduced pronouns and arguments in Sou Amana Teru, Ambon*

SIMON MUSGRAVE

1 Introduction

This paper examines the use of reduced pronouns in the language Sou Amana Teru, spoken at the eastern end of Ambon Island. The language has two series of reduced pronouns; one series of proclitics and another series of enclitics. The two series are almost identical, and are both formally derived from free pronouns. The reduced pronouns can appear in various functions, including as the arguments of verbs, as the complements of prepositions and as markers of inalienable possession. My discussion will focus on the use of reduced pronouns in argument positions, especially their use as the sole argument of an intransitive verb or as the more agentive argument of a transitive verb.[1] The evidence which I present shows that reduced pronouns used in this function have properties which are characteristic of a cross-referencing system. Such systems are usually considered to be a part of the morphological component of grammar, but evidence from Sou Amana Teru suggests that the use of reduced pronouns in S and A function largely follows principles of the syntactic component of the language. I suggest that this situation is best understood as part of a process of change from a diachronic state of the language which has true morphological cross-referencing towards a state which is organized more analytically.

The discussion also covers the use of reduced pronouns to mark patient-like arguments. This occurs with transitive verbs (in O function), where reduced pronouns are one possible realization of the argument but do not have a cross-referencing function, and with two classes of intransitive verbs, where the reduced pronoun references the S argument which also appears elsewhere in the clause. The latter case provides evidence that the language has split S properties in at least a small part of its grammar.

The examples used in this paper represent three types of data, and codes are given following the free translation of each example to indicate the source of each example. Many examples are taken from spontaneous narrative or conversation, and such examples have the code SD (= spontaneous discourse). Other examples come from translations of a standard elicitation narrative, used to obtain basic data on a number of languages in several

[1] In what follows, I use the abbreviations S, A and O (after Dixon, 1972) to mean respectively: the single argument of an intransitive verb, the more agent-like argument of a transitive verb, and the more patient-like argument of a transitive verb.

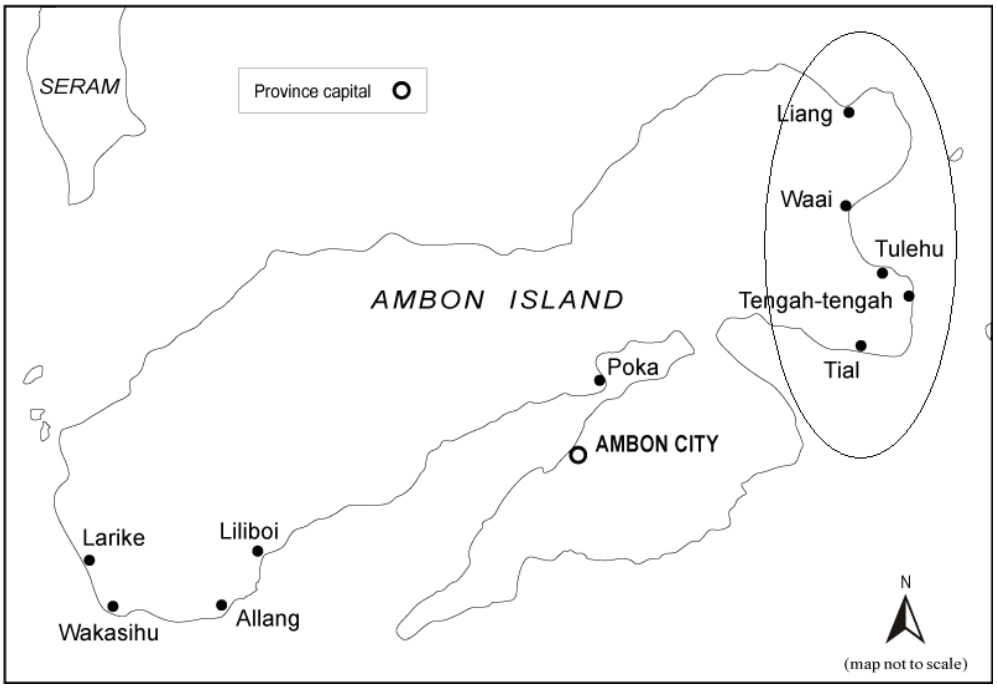
Michael C. Ewing and Marian Klamer, editors.
Typological and areal analyses: contributions from East Nusantara, 143–164. Pacific Linguistics, 2010
© This edition vested with Pacific Linguistics.

locations. Such examples have the code GS (= Garden Story) and the examples cited here come from two different speakers. Finally, some examples were elicited as isolated sentences, and these examples have the code E.

The following section of the paper gives brief background information on the language. Section 3 presents data on the pronoun forms and on the distribution of the different form classes, while section 4 concentrates on the use of reduced pronouns in argument positions. Section 5 examines the question of what properties of the system for marking S and A arguments can be considered morphological, and which can be treated as syntactic properties. A summary of my conclusions is presented in section 6.

2 Background

Sou Amana Teru is a Central Malayo-Polynesian language spoken at the eastern end of Ambon Island (see Map 1). The name means 'language of the three villages', and the primary speech community is located in the villages of Tulehu, Tengah-tengah and Tial. A dialect of Sou Amana Teru is spoken in the village of Liang to the north of Tulehu, and closely related varieties are spoken in villages on the north coast of Ambon Island (see Musgrave 2006 for discussion). The language was also formerly spoken in the village of Waai, immediately to the north of Tulehu, but only a few elderly rememberers remain in that location. In all villages, the indigenous language is used in conjunction with Ambonese Malay (Minde 1997), with standard Indonesian also present via education and the media. There are currently approximately 10,000 active speakers of the language, but the majority of these are over 30 years of age. In younger age groups, linguistic vitality drops rapidly (see Musgrave and Ewing 2006 for preliminary results of the testing of linguistic vitality in Tulehu), but another 6,000 people can be counted as having some knowledge of the language.



Map 1: Ambon Island showing area in which Sou Amana Teru is spoken.

3 Pronouns in Sou Amana Teru

Sou Amana Teru has three sets of pronouns, one set of free pronouns which are phonologically independent forms, and two sets of reduced pronouns which are phonologically dependent on another word. Of the reduced pronouns, one set are proclitics, that is they appear preposed to the host word, and the other set are enclitics, which appear postposed to the host. Table 1 lists these forms.

Table 1: Pronoun forms in Sou Amana Teru

		Free pronouns	**Reduced pronouns**	
			Proclitics	Enclitics
1s		*yau*	*au=*	*=u*
2s		*yare*	*ar=*	*=(a)r, =m(u)*
3s	h	*ire (eng)*	*e=*	*=i*
	nh			*=re*
1p	i	*ike*	*ka=*	*=ka*
	e	*yami*	*m=*	*=m(i)*
2p		*yimi*	*mu=*	*=m(u)*
3p		*sire/isi*	*si=*	*=si*

When a vowel-initial enclitic attaches to a vowel-final host, a glottal stop is predictably inserted in most cases. The two enclitic forms given for second person singular are functionally differentiated, and will be discussed below. The different series of pronouns are used in different functions, and these functions are discussed in the following sections.

3.1 Free Pronouns

Free pronouns can be used in at least the following functions: as the S or A argument of a clause, as the O argument of a clause, as the complement of a preposition, and as a possessor. Examples (1) – (7) illustrate these possibilities, and exemplify the free forms for all person and number combinations, including both of the third person plural forms.[2]

(1) First person singular in O function
Nina e=pa'anusi ***yau****: "Oi apai=si"*
mother 3s=order 1s go call=3p
'Mother told me: "Go and call them." (GS)

(2) Second person singular in complement of preposition function
Bisa *yau puna wa'a* ***yare***
be.able 1s do LOC 2s
'I can do it for you.' (E)

[2] ' indicates a glottal stop. Malay words used as loans, and not phonologically integrated into the language of the example, appear in regular (non-italic) font in examples.

(3) Third person singular in A function
Ire *weuta* *e=pamana*
3s not.want 3s=eat
'S/he doesn't want to eat.' (SD)

(4) First person plural inclusive in A function
Ian *ahia=ma* *taha* boleh ***ike*** *lata'a=re*
fish rotten=that NEG permitted 1pi sell=3nh
'It is not allowed that we sell rotten fish.' (E)

(5) First person plural exclusive in S function
Yami *lai* *wa'a* *ruma=ma* *atatoru* *ea*
1pe come LOC house=that evening already
'When we got home, it was evening already.' (GS)

(6) Second person plural in possessor function
Yau *rarehu=si:* *'**Imi*** *hala'usai* *na* *wa'a* *pe'e?'*
1s ask=3p 2p clothing AUX LOC where
'I asked them: "Where are your clothes?"' (GS)

(7) Third person plural in A function (*sire*)
Sire *taha=si* *supu* *ian-e*
3p NEG=3p catch fish-LM
'They didn't catch any fish.' (E)

(8) Third person plural in A function (*isi*)
Uma ***isi*** *paha-nalar-ma* *aman* *Tui-rehu-i*
then 3p CAUS-name-EMPH village tui.bird-below-3s.INAL
'Then they called the village "Tuirehui".' (SD)

The free pronoun form shown bracketed in Table 1 is a third person singular pronoun form used only for possession. The use of this form is illustrated in example (9).

(9) *Au* *sau='u* *e=tana* ***eng*** *ana-e*
1s in.law=1s.INAL 3s=take 3s.POSS child-LM
'My sister-in-law took her baby' (GS)

This form is an innovation, based on the proclitic form (Musgrave 2005).

3.2 Proclitics

Proclitics appear in S and A function where there is no other non-nominal element preceding the verb, as seen in examples (10) to (12).

(10) ***Ar***=*oi* *tula* *ar-ang* *nina*
2s=go with 2s-LNK mother
'You go with your mother.' (GS)

(11) Setelah *ela-e* ***ka****=hala=r* *wa'a* *rehit-e*
after sago.pulp-LM 1pi=carry=3nh LOC sago.trough-LM
'After it is pulp, we carry it to the trough.' (SD)

(12) ***E****=hose:* *"Mai palamana."*
3s=say let's tour
'He said: "Let's take a trip."' (SD)

Where a full NP occurs as S or A argument, a third person proclitic still appears adjacent to the verb.

(13) *Isi* *ana* *ko'in-e* ***i****=na'e*
3p child small-LM 3s=sleep
'Their baby was sleeping.' (GS)

(14) *Tahina-n* *laha* *matua-na=re* ***si****=jaga* *duren=re*
old.woman-PL with old.man-PL=this 3p=guard durian=this
'These old people were guarding the durian.' (SD)

In several languages of Central Maluku, proclitic pronouns mark inalienable possession (Florey 2005). I have argued elsewhere (Musgrave 2005) that this structure has almost vanished from Sou Amana Teru, but occasional examples can be found in the usage of older speakers. Example (15) was produced by an elderly speaker in the Dutch diaspora, who has had limited contact with the main speech community for a period of 50 years.

(15) *Au* *a'a='u* *malona-e* *tula* ***i****=mahina-e* *isi*
1s older.sibling=1s male-LM with 3s=female-LM 3p

wa'-ene *marinu-e*
LOC-DIST garden-LM
'My big brother and his wife were there at the garden.' (GS)

3.3 Enclitics

Enclitics appear in five functions: as S or A where a non-nominal element precedes the verb, as O in transitive clauses, as the complement of prepositions, as markers of inalienable possession and with quantifiers and numerals.

3.3.1 Enclitics as S or A

Three non-nominal elements host enclitics preceding the verb. Two of these are *taha*, the primary negator in Sou Amana Teru, and *hare* which marks continuing action, exemplified in (16) to (18).

(16) *Tah=****u*** *iming* *ane* *ambelan-e*
NEG=1s wish eat mango-LM
'I don't like to eat mango.' (E)

(17) *Tah=**i** pa-manahu* tapi *e=ane*
NEG=3s CAUS-fall but 3s=eat
'He didn't send [them] down but he ate.' (SD)

(18) *Hare=**si** turi'e*
CONT=3p write
'They are writing.' (E)

The third preverbal element which hosts reduced pronouns, *na*, is more problematic. The most common occurrence of this element is in clauses with non-verbal predicates expressing location, as seen in example (6) (repeated here) and example (19).

(6) *Yau rarehu=si: 'Imi hala'usai **na** wa'a pe'e?'*
1s ask=3p 2p clothing AUX LOC where
'I asked them: "Where are your clothes?"' (GS)

(19) Foto=*ma=pi nana-e sala mama Ceha **na***
photo=that=EMPH exist-LM still PN AUX

wa'=ena-ke
LOC=MED=EMPH
'That photo still exists, mama Ceha, it's over there.' (SD)

Such examples suggest that an analysis of *na* as a verb meaning 'be there' would be possible. There are two arguments against this approach, however. Firstly, *na* can host an enclitic when it occurs in locational clauses, and if such an enclitic occurs, it definitely refers to the S of the clause.

(20) *Nina e=larehu: "Are-ng ana-e kako'ini na=**i***
mother 3s=ask 2s-LNK child-LM small AUX=3s

wa'a pe'e?"
LOC where
'Mother asked: "Where is your baby?"' (GS)

As will be discussed in more detail below, enclitics attach to lexical verbs in two cases. In one case, an enclitic can attach to a transitive verb representing the O argument. It is clear that example (20) is not of this type. In the other case, certain types of intransitive verbs can appear with an enclitic. In such cases, a full NP S argument, as in example (20), is accompanied by a proclitic, as in examples (13) and (14). Example (20) is not of this type either. The second argument against analysing *na* as a verb is that it can co-occur with an unambiguous verb, as in example (21).

(21) ***Na**=isi **lai** mena ea*
AUX=3p come first already
'They came first.' (GS)

In such examples, the behaviour of *na* is parallel to that of *taha* and *hare*, seen in examples (16) to (18). On the basis of these considerations, I analyse *na* as belonging to

the same lexical class as the other two elements which occur preverbally. However, it is not clear what semantic contribution *na* makes to clauses in which it occurs, and I therefore gloss it as 'AUX' where it appears.

The negator *taha* can also appear in negative existential clauses, and in such contexts it can host an enclitic in the same way that *na* does in example (20). Florey (this volume, see especially example (20) and associated discussion) analyses *taha* as a predicate in such contexts, glossing it as 'Neg.Exist'. The arguments presented above as to the status of *na* may also apply to *taha*, but current data possibly is insufficient to resolve the question.

For all the examples discussed in this section, it is intuitively clear that the clitic referencing the S or A argument appears immediately before the verb (if there is one). Given that third person clitics are never hosted by a preceding noun phrase (see examples (13) and (14)), I suggest that the best statement of the position of the clitic is that it attaches to the first element in the predicate constituent of the clause, whatever syntactic category might be assigned to that constituent, and that it must precede the verb if there is one. This statement assumes that auxiliaries and the negator are the first element of a predicate if they occur. Therefore, when an auxiliary or negator is first in the predicate, the clitic follows it and is hosted by it, whether or not a verb follows. On the other hand, when there is no auxiliary or negator, the clitic precedes the verb and is hosted by it.

A second person singular S or A argument which is realized as an enclitic takes the form *-r(e)* (example (22)).

(22) *Taha=**r** supu ian-e?*
NEG=2s catch fish-LM
'Didn't you catch any fish?' (E)

The other second person singular form, *-m(u)*, is used to mark inalienable possession and is shown in example (40) below.

3.3.2 Enclitics as O

An O argument in Sou Amana Teru can be realized as an enclitic to the verb. For human third person O arguments, there is a distinction between a singular clitic form (example (23)) and a plural form (example (24)):

(23) *Jadi ru'a e=supu=**i***
then monkey 3s=catch=3s
'Then Monkey caught him.' (SD)

(24) *Yau rarehu=**si**: "Imi hala'usai na wa'a pe'e?"*
1s ask=3p 2p clothes AUX LOC where
'I asked them: "Where are your clothes?"' (GS)

For a non-human O argument, a single reduced pronoun, *=re/=r/=ar* (afterwards referred to as *=re*), serves for both singular and plural (examples (25) and (26)).

(25) *Ka=tala=**r** usi-e ike ha'a=**re***
1pi=fell=3nh all-PFV 1pi split=3nh
'Having felled it, we split it.' (SD)

(26) Jadi *rua=isi* *pa-tanane'e=**re***
then two=3p CAUS-plant=3nh
'Then the two of them planted them.' (SD)

The clause which precedes example (26) describes how the two protagonists have split a banana tree between them, one taking the trunk and one taking the leaves, and it is clear in context that both parts of the tree are being planted, that is, that *=re* has plural reference.

Another use of the clitic *=re* occurs where the valence of the verb requires a second argument, but its referential content is not important. In this use, the clitic has no specific reference.

(27) *Taha=u* *tea=**re***
NEG=1s know=3nh
'I don't know.' (E)

(28) *Uma* *e*=ambil kesimpulan-*e* *i=oi* *lohi=**re*** menulusuri
then 3s=take conclusion-LM 3s=go seek=3nh explore

waer-ma *lo'o* *rete* *eng* sumber
water-that toward DIR 3s.POSS source
'After reaching a conclusion, he went looking, exploring the river towards its source.' (SD)

In one example, the clitic is used in this way with a verb which is a Malay loan:

(29) *Yami* tunjukkan=***ar*** *wa'a* dunia bahwa *yami* *aman*
1pe indicate=3nh LOC world that 1pe village

Tuirehui-e *taha=mi* *iming* *ena* *ia* kacau *uru* *ekai'i*
Tulehu-LM NEG=1pe wish for good trouble head single
'We show the world that our village of Tulehu does not wish for trouble for anyone.' (SD)

This example suggests a high level of integration of at least some Malay loans into the grammatical system of Sou Amana Teru.

3.3.3 Prepositional complements

The complements of prepositions allow the same range of possibilities as O arguments of verbs for the use of reduced pronouns. Both human (example (30)) and non-human entities (example (31)) can be coded using a reduced pronoun.

(30) *Yau* *lope'e* *ian* *hahua-i* *wa'a=**i***
1s give fish piece-3s.INAL LOC=3s
'I gave him a piece of fish.' (GS)

(31) *Yau* *lope* *ian* *hahua-ri* *wa'a=**r***
1s give fish piece-one LOC=3nh
'I gave it a piece of fish.' (GS)

Example (30) and example (31) are translations of the same prompt given by different speakers. The entity to whom the fish is being given is a dog in each case. In example (30), the speaker treats the domestic animal as quasi-human, while in example (31), the dog is seen as non-human. The fact that example (31) is possible establishes that the relevant semantic distinction is human versus non-human, rather than animate versus inanimate.

3.3.4 Enclitics with intransitive verbs

There are three classes of intransitive verbs in Sou Amana Teru, two of which can occur with an enclitic in the position which would represent an O argument with a transitive verb. The third class is not discussed here, and contains all intransitive verbs which are not included in the first two classes.

3.3.4.1 Undergoer-oriented intransitive verbs

The first class consists of undergoer-oriented verbs, denoting states which affect the S argument. Such verbs, for conservative speakers at least, obligatorily occur with an enclitic which repeats the S argument.

(32) *Yami maruhu'=**amu***
1pi hungry=1pi
'We are hungry.' (GS)

(33) *Ar=amuri'=**amu***
2s=tired=2s
'You are tired.' (GS)

Several of the verbs which fall into this class in Sou Amana Teru allow a second oblique argument to occur in the clause, coded as a PP (for an extensive discussion of such verbs in western Austronesian languages, see Musgrave 2002, Ch. 4). The repetition of the S argument can still occur where an undergoer-oriented verb takes an oblique second argument.

(34) *Yau kere'=**u** ena asu*
1s afraid=1s for dog
'I am afraid of dogs.' (E)

Examples such as (34) suggest that the enclitic which occurs with undergoer-oriented intransitive verbs does not represent a semantic argument distinct from the experiencer. The preverbal S argument of an undergoer-oriented intransitive verb cannot be omitted; in this respect, Sou Amana Teru differs from Allang, another Central Maluku language from Ambon Island (see Ewing, this volume).

3.3.4.2 Intradirective intransitive verbs

The second class of intransitive verbs consists of verbs denoting an action which is initiated by a conscious agent and which affects that same agent. Such verbs are classed as intradirective verbs by Pawley (1973:176), and defined as ones whose 'subject or actor is both the one who causes and who experiences the action'. As Pawley notes, most such

verbs involve either movement or posture. Although these verbs are normally used with a single argument, there are examples in which they occur with an enclitic which doubles the S argument.

(35) *Sori bombonu e*=kecewa *e=oi=**'i***
then turtle 3s=disappointed 3s=go=3s
'Then turtle was sad and he took himself off.' (SD)

(36) *Ike reu=**ka***
1pi go.home=1pi
'We want to go home!' (GS)

This possibility seems to be used to emphasize the commitment of the S argument to the action. Bolton (1990:100) gives similar examples for Nuaulu, a Central Moluccan language from south-central Seram island. All the examples of this construction in my database have a human S argument (or an animal treated as a human), except for one:

(37) Karena *mei-si* *taha* *lai=**re*** *uma* *isi* *apa'=ar*
because tongue-3p.INAL NEG come=3nh then 3p call=3nh

Tulehu
PN
'Because their tongues couldn't come, they called it Tulehu.' (SD)

This example is an explanation of why the name by which the village is usually called, *Tulehu,* differs from the traditional name in Sou Amana Teru, which is *Tuirehui*. The accepted account of this change in the village is that the Dutch colonialists were unable to pronounce /r/ and therefore changed the name. Although the non-human pronoun form is used in example (37), it could be argued that the S argument in this instance is a human body part, and that the construction is therefore being used with a human referent via a semantic extension. Such considerations suggest that this construction may indeed be constrained semantically to human entities in S function, even if a formal distinction between humans and non-humans is still evident.

3.3.5 Inalienable possession

In Sou Amana Teru, enclitic pronouns attached to nouns code inalienable possession. The semantic category of inalienability includes some kin terms (example (38)) and part/whole relations (example (39)), including parts of the body (example (40)). Example (40) also shows the second person singular form used to mark inalienable possession. This is the only case in which the inalienable possession marker has a different form from the enclitic pronoun form used in other functions (that form can be seen in example (22)), and it suggests that the use of reduced pronouns for this function may be of greater antiquity than the use of reduced pronouns for argument marking. The form *-mu* is a reflex of the form reconstructed for the Proto-Austronesian second person singular genitive pronoun (Blust 1977), while both of the forms used to mark arguments, *ar-* and *-(a)r*, are reduced forms of the relevant free pronoun. That form, *yare*, has cognates throughout central Maluku, but is certainly more recent than the *-mu* form.

(38) *Yau oi laha au ama=**'u** laha au ina=**'u***
1s go with 1s father=1s.INAL with 1s mother=1s.INAL
'I went with my father and my mother.' (GS)

(39) *Yau a'a malona-e i=tane kaki-e wa'a marinu*
1s older.sibling male-LM 3s=plant tuber-LM LOC garden

*usu=**'i***
edge=3s.INAL
'My big brother planted sweet potato at the edge of the garden.' (GS)

(40) *Yare mata=**m** ma'e'u*
2s eye=2s.INAL itchy
'Your eyes are itchy' (E)

Examples (9), (15) and (30) also include inalienable possession structures, and Musgrave (2005) has extensive discussion of this possibility.

3.3.6 Numbers and quantifiers

Enclitics also attach to quantifiers and numbers to indicate the group from which the selection is made (examples (41) and (42)).

(41) *Ami haiti=m ami tiau nela ea*
1pe all=1pe 1pe stomach big already
'We were all full.' (GS)

(42) *Nina e=larehu yau: 'Rua=isi na=isi wa'a pe'e?'*
mother 3s=ask 1s two=3p AUX=3p LOC where
'Mother asked me: 'Those two, where are they?'' (GS)

In these examples, it is clear that the enclitic forms a unit with the quantifying element. However, in other cases, the enclitic which attaches to the quantifier could also be interpreted as being the S/A argument clitic (example (43)).

(43) *Uma rua=isi laka kula hutai='ene*
then two=3p take banana plant=MED
'Then the two of them took the banana plant' (SD)

3.4 Summary

Table 2 summarizes the distribution of the different pronoun forms in the various constructions which have been discussed.

Table 2: Distribution of pronoun forms in Sou Amana Teru

	S/A	**O**	**P-Comp**	**Poss(INAL)**	**Poss(AL)**	**Quant**
Free pronoun	√	√	√		√	
Proclitic	√				?	
Enclitic	√	√	√	√		√

This table gives information only about the possibilities for formal realization of the different functions; the actual constructions are more finely differentiated, particularly with regard to the hosts to which clitic pronouns attach. The next section of the paper clarifies these different possibilities, focussing first on the realization of arguments in O function, and then on the more complex issue of the distribution of pronoun forms in S and A functions.

4 Pronouns and argument positions

It can be seen from Table 2 that there is more than one way of realising an argument as a pronoun. And it has also been shown in previous examples that at least some reduced pronouns can appear in the presence of other nominals which also realize the same argument. In this section, I will set out the restrictions on the multiple exponence of arguments, before attempting to place the system in a typological perspective. The analysis of O arguments is more straightforward, and I therefore begin with them.

4.1 O arguments

There are two possible realizations of an O argument in Sou Amana Teru. Firstly, a full NP can occur, as in example (9), and this possibility includes the use of a free pronoun, as in example (44).

(44) *Yami a'a-ng mahina-e eng ana e=ainala*
1pe older.sibling-LNK female-LM 3s.POSS child 3s=wait

yami *wa'a ruma-e*
1pe LOC house-LM
'Our big sister's child waited for us at the house.' (GS)

Secondly, a reduced pronoun can be used, as seen in examples (23) to (29). Crucially, however, these two possibilities are in complementary distribution. It is not possible for a single clause to contain both a free NP in O function and an enclitic referencing that NP. That is, there is no clause corresponding to example (9) in which a reduced pronoun also occurs.

(45) *Au sau'u etana(*=**i**) eng anae*.

This evidence shows that enclitics attached to transitive verbs in Sou Amana Teru are not a cross-referencing phenomenon. The reduced pronoun is the sole exponent of the O argument in these cases, and it retains its status as a true pronoun.

However, Sou Amana Teru does have structures where two exponents of an O argument do seem to occur in the same clause. In such cases, a full noun phrase which seems to be in O function appears at the left edge of a clause, with a pragmatically marked effect, and an enclitic also occurs attached to the verb.

(46) *Isi* pakaian *isi taru=**r*** *wa'a pe'e*
3p clothes 3p place=3nh LOC where
'Their clothes, where had they put them?' (SD)

(47) *Uma selendang-ma Lambi Ulan e=tana=**re** e=pahunia=**re***
then shawl-that PN 3s=take=3nh 3s=hide=3nh

*e=pa-nusu=**ar** wa'a ute wake-i*
3s=CAUS-enter=3nh LOC bamboo joint-3s.INAL
'Then, that shawl, Lambi Ulan took it, he hid it, he put it into the piece of bamboo.' (SD)

As was seen in examples (27) to (29) above, the non-human clitic can satisfy the valence requirements of a transitive verb even where the second argument has no significant reference. This fact, together with the co-occurrence restriction just described, suggests that the clitic in examples (46) and (47) should be interpreted as the true syntactic argument of the verb, with the left-dislocated noun phrase being outside the clausal nucleus. Schematically, the structure suggested for these examples is:

(48) $[_{S}NP_i\ [_{Cl}\ NP\ V=Pro_i]]$

This analysis is supported by the fact that the clitic is used in similar fashion in cases in which it co-refers with an NP that is unambiguously outside the immediate clause, for example when the O of a purpose clause is co-referent with the O of the matrix clause.

(49) *Uma ute wake-ma e=tana=re [ena e=pa'-ia=**re]***
then bamboo joint=DIST 3s=take=3nh for 3s=CAUS-good=3nh
'Then that bamboo joint, she took it to fix it.' (SD)

In example (49), the full noun phrase O of the matrix clause has been left-dislocated, but it is clear that both the clitic in the matrix clause and that in the purpose clause refer to the same entity, the piece of bamboo. These arguments show that the possibility of an enclitic co-referring with a left-dislocated NP does not contradict the analysis that reduced pronouns attached to transitive verbs fill the O argument position.[3]

Sou Amana Teru has no obvious voice system, that is, there are no dedicated morphological markers which signal changes in the diathesis of verbs. This is a property which is expected according to Himmelmann's (2005) account of the typology of non-Oceanic Austronesian languages. Sou Amana Teru belongs to the type which Himmelmann terms 'preposed possessor languages', as can be seen in examples such as (6), (9), (40) and (44) among others, and the lack of a voice system is one of the characteristics of this type. However, the left-dislocation structure just described can be

[3] The intonation of utterances with a left-dislocated NP is inconclusive with regard to this analysis. In many cases, there is an intonation break after the initial NP, but a break does not occur in every case.

used in a way which is functionally very similar to the use of passive in other languages. A salient O argument can be moved to the clause initial position with a co-referent clitic on the verb, and a non-specific pronoun occurring in A function.

(50) Jadi *lapia-re si heta=re*
then sago-this 3p cut=3nh
'So this sago, they cut it.' (SD)

Example (50) is the second clause of a procedural text. No human actors were introduced in the first clause, and the third person plural pronoun in the example therefore does not have anaphoric reference. Its reading is non-specific or generic, and the example could be translated appropriately with a passive clause in English: 'The sago is cut'. Such structures most commonly have a non-human O argument. It is possible to have an enclitic representing a human O, but it is very uncommon for such pronouns to be anteceded by a left-dislocated NP. Where a noun phrase is followed by a proclitic and then a verb, the default option is to interpret the proclitic as co-referential with the noun phrase, that is, it is interpreted as an example of the repetition of A previously discussed. The structure represented schematically in example (51) is preferentially interpreted as indicated in (52), not as an instance of a quasi-passive clause.

(51) NP_O Pro_A Verb=Pro_O

(52) Jadi *ru'a e=supu=i*
then monkey 3s=catch=3s
'Then Monkey$_i$, he$_i$ caught him$_j$.' (SD) NOT: 'Then Monkey$_j$, he$_i$ caught him$_j$.'

Speakers will accept left-dislocation of human O arguments in elicited examples, particularly when there is a difference in number between the A and O arguments to disambiguate as in example (53), and I have recorded a single spontaneous example of this type, example (54).

(53) *Hasan si nau=i*
PN 3p see=3sh
'Hasan they saw.' (E)

(54) *Ana haing-api si=halata=i*
child that-EMPH 3p=strike=3s
'That child was struck.' (SD) (lit: 'That child, they struck him.')

But even examples with both A and O being singular are acceptable, as in example (55).

(55) *Hasan e=nau=i*
PN 3s=see=3sh
'Hasan$_i$, he$_j$ saw.' (E)

Languages can tolerate ambiguity in the reference of verbal arguments (see Ewing, 2005 for discussion), but the absence of examples such as (53) and (55) in natural discourse suggests that Sou Amana Teru speakers prefer to avoid ambiguity in this case.

4.2 S and A arguments

All three types of pronoun can realize an S or A argument in Sou Amana Teru. One possible realization of S, discussed in section 3.3.4, is as an enclitic attached to the verb. But this possibility is restricted to two small classes of intransitive verbs, and I will not discuss it further here. Instead, I concentrate on preverbal realizations of S and A arguments, and the restrictions on the appearance of the different pronoun types.

Three factors determine the distribution of pronouns in preverbal positions and there are relations of logical dependency between these factors. The first factor is whether the S/A argument is human or non-human. Dependent on this factor are the questions of whether or not one of the preverbal elements *taha*, *hare* and *na* occurs in the clause, and whether the S/A argument is third person and can be represented by a non-pronominal NP.

In the case that the S/A argument is non-human, no pronoun occurs preverbally, because no pronoun exists in Sou Amana Teru which can refer to a non-human entity in that function. It will be recalled that the enclitic *=re* refers to non-human entities (cf. §3.3.2), and the third person singular possessive pronoun *eng* can also have a non-human antecedent, as seen in example (56) (and also in example (60)).

(56) Tapi ***eng*** *nalar=ma,* ***eng*** *nalar* *ha'er?*
but 3S.POSS name=that 3S.POSS name what
'But what was its name?' (SD)

This example is taken from a narrative about how the village of Tulehu was named, and in this clause the first inhabitants, having just arrived at the site of the village, are asking what the name of the place might be. But a third-person, non-human entity occurring as an S/A argument can never be represented by a preverbal pronoun. In such cases, the noun phrase denoting the entity stands alone. This is the case when the noun phrase occurs immediately before the verb, example (57).

(57) ***Manu*** *ane* *kula-e*
bird eat banana-LM
'The bird is eating bananas' (SD)

It is also the case when the clause is negated (example (58)) or when some other element intervenes between the S/A argument and the verb (example (59)), a situation where repetition with a pronoun would be expected with a human entity.

(58) Lampu ***taha*** *kina*
lamp NEG bright
'The lamp has gone out.' (E)

(59) *Duren* *ain=ma* **tiap** ***taun=ma*** *pahua*
durian tree=that each year=that bear.fruit
'That durian tree bore fruit every year.' (SD)

It is possible to omit the noun phrase when it is recoverable from the context, as in example (60), but this strategy is very rare in my data. In the procedural text from which example (60) is taken, which describes the harvesting and processing of sago, the sago is

almost always kept in O function where it can be referred to with a pronoun. Example (60) is the only exception to this pattern.

(60) *Lapia=re-na eng isinar=ma-ne masehu aiy-i*
sago=this-EMPH 3s.POSS contents=that-EMPH drop.down 3-3sh

lo'o-ne rehit lare-i
to-EMPH sago.trough inside-3s.INAL

Masehu usie ike pareta waer-e
drop.down all 1pi make.dry water-LM
'The contents of the sago drop down into the trough. After [it] drops down, we dry the water out.' (SD)

It should be noted that some other languages of Central Maluku do have a third person singular non-human pronoun, distinct from the pronoun which refers to third person humans. Examples (61) and (62) from Alune show this contrast (data from Florey 2001).

(61) *Ela' inai-je **i**=ombe….*
elder CLF-DET 3sh-say
'My parent, s/he said…..'

(62) *Apa-le **e**=betu bei au*
pig-LM 3snh-get.up ABL 1s
'The pig got up from me.'

When the S/A argument is a human entity, then a pronoun is possible. Which type of pronoun can appear then depends on the presence or absence of a preverbal element. If such an element is present, an enclitic must attach to it, as seen in the examples in section 3.3.1. If there is no preverbal element, then a proclitic is possible. Such a pronoun is obligatory in the case of a third person S/A argument realized as a non-pronominal NP (examples in section 3.2), but it cannot appear when a free pronoun occurs before the verb, unless a preverbal element intervenes.

(63) *Isi pahai wa'-ene*
3p play LOC-DIST
'They played there.' (GS)

(64) *Yau taha=u pahai-'e*
1s NEG=1s play-LM
'I didn't play.' (GS)

Thus example (63) is possible with a pronoun adjacent to the verb, and example (64) is possible with a free pronoun and a reduced pronoun separated by *taha*. But example (65) is not possible where two pronouns are adjacent.

(65) **Yau u=pahai'e.*

Two generalizations can be made about the patterns just described. Firstly, unless the S/A argument is non-human, a pronoun must immediately precede the verb. Secondly, two exponents of the S/A argument are permitted, but if they are both pronouns, they must be separated by one of the preverbal elements. At first sight, it appears that there is a single position available for a pronoun immediately before the verb which can be filled by a free pronoun, or by a proclitic attached to the verb when no non-nominal element precedes the verb, or by an enclitic attached to a non-nominal preverbal element. I will argue below that this appearance is incorrect and that preverbal free pronouns are in a different position from proclitics.

5 S/A pronouns – morphology or syntax?

The various possibilities for the realization of S/A arguments described above constitute a system which has some of the properties of a cross-referencing system. Such systems are defined by Andrews (1985:75) as having 'various grammatical properties of an NP … registered on an element bearing some specific syntactic relation to the NP'. In Sou Amana Teru, material appears attached to the verb (or at least adjacent to it) which registers the person and number properties of the S/A argument. That such elements are closely related to pronouns, or actually are pronouns, is consistent with the view of such systems developed by Givón (1976 and elsewhere). In addition to referencing a free NP which is an exponent of the S/A argument, these clitics can fill the argument position themselves. This again is a common property of cross-reference systems. Therefore one analysis might be that Sou Amana Teru has a system of S/A agreement or cross-referencing. Such a system would be a morphological system, but different aspects of the system in Sou Amana Teru suggest an interpretation which is either more morphological or more syntactic. The possibility of free pronouns appearing as sole exponent of the S/A argument supports a morphological view of the system, while the instability in the positioning of the clitics, and the absence of cross-referencing for non-human S/A arguments suggest a more syntactic view.

As seen previously, a free pronoun can occur immediately preverbally, and in this case it must be the sole exponent of the S/A argument. This distribution might be taken to suggest that the position occupied by clitics is available to any pronoun, other things being equal. But, if that was the case, it would be predicted that free pronouns could occur between a preverbal element and the verb, and this pattern is not attested. Therefore, I suggest that a free pronoun as sole exponent of the S/A argument must be in the same position as a full NP argument such as those in examples (13) and (14) (human S/A with reduced pronouns following) or in example (57) (non-human S/A without a following reduced pronoun), and that the immediately preverbal slot is restricted to clitics. This implies a more morphological view of the cross-referencing phenomenon: the material which can appear in the position in question is restricted according to form class, rather than a syntactic class.

On the other hand, the positioning of the clitics seems to be determined syntactically. As seen in several examples above, a reduced pronoun representing the S/A argument can appear either as a proclitic attached to the verb or as an enclitic attached to one of the three preverbal particles (and possibly as an enclitic attached to a quantifying expression). Intuitively, the correct generalization is that the two possibilities are really the same: there is a single position in which a reduced pronoun can occur (immediately before the verb) and whether it attaches to the left or to the right depends on the surrounding environment.

But such an account fits better with an analysis in which the syntactic position of the pronoun and its phonological position are independent factors. In a fully grammaticized morphological system, we would expect that the cross-referencing elements all to appear in the same morphological slot. Therefore, this evidence suggests that viewing the clitics as independent syntactic elements is preferable.

The second case in which the S/A argument is not represented by a clitic is that where the S/A argument is non-human. The paradigm of forms in Table 1 showed that Sou Amana Teru has no pronominal forms specific for third person non-human reference except for the enclitic *-re* which cannot be used in S/A function. One approach to this gap in the paradigm would be to posit the existence of a non-human pronoun which is formally realized as zero. Such an account would remove the paradigmatic gap, and the requirement to have a pronoun immediately preceding the verb would then be without exception. Analyses which include clitics with no overt realization have been proposed for languages from another part of Maluku, the island of Aru (Hughes, 2000, Nivens, 1998). For example, Hughes proposes that the Dobel language has a zero clitic for third person inanimate O arguments. However, it is clear from many preceding examples that a preverbal clitic alone is sufficient to fill the S/A argument slot in Sou Amana Teru. Therefore such an account would predict that the zero pronoun could also be an argument in its own right, and that clauses with no overt S/A argument should be acceptable. But as noted in the discussion of example (60), this is not in fact the case. Although other morphological analyses are possible which do not posit a zero form in the pronoun paradigm, the argument just presented tends to weaken the case for a morphological treatment of the preverbal clitics.

The arguments above suggest that the S/A pronoun system of Sou Amana Teru has morphological aspects and syntactic aspects. Collins (1983:24–27) shows that the languages of Central Maluku had a verbal conjugation system at some point in their development which involved variation in the initial consonant of verb forms. Sou Amana Teru shows little trace of such a system today, although closely related varieties spoken at Hitu and Mamala on the north coast of Ambon Island have retained some conjugated forms as free variants. This can be seen in example (66), from Hitu, where the S argument of the two clauses is identical, but different verb forms appear.

(66) *Ite* *kolo* *wa'ale* *kula* *ite* *tolo* *ahasame*
1pe sit LOC-PROX and 1pe sit rest
'We sat down here and we rested.' (GS)

The system described in this paper is formally close to the one which Collins gives for the 'neutral conjugation', where some material which varies precedes the verb stem, but the verb stem itself does not vary.

(67) Conjugated forms of *nanu* 'swim' from Asilulu (north coast, Ambon Island) (Collins 1983:26)
1s *unanu*
2s *ananu*
3s *inanu*
1pe *mananu*
1pi *nanu*
2p *(i)nanu*
3p *sinanu*

The pattern presented by Collins is more clearly morphological, especially in that different paradigms occur depending on the phonological properties of the base form. A similar system almost certainly was part of Sou Amana Teru at an earlier period, and traces of verb conjugation, appearing as free variation, can still be observed in the closely related variety spoken at Hitu (Musgrave 2006). In comparison, the current system in Sou Amana Teru appears more analytic. One set of forms of the cross-referencing pronouns has been generalized across all verbs, and the pronouns are well short of being fully grammaticalized as bound morphemes.

It seems likely that the language will move towards an even more analytic structure in the near future. Morphological simplification is already evident in the language use of younger speakers, with both suffixal markers of inalienable possession and S clitics on undergoer-oriented verbs tending to disappear (see Musgrave 2005 for discussion). The use of S/A cross-referencing is also less common for such speakers. The following examples use data taken from the program of linguistic vitality testing carried out in Tulehu village. For each pair of examples, the first example is the target translation of a Malay prompt provided by a fluent male speaker aged over 60, while the second example is the translation of the same Malay prompt given by a female speaker aged between 20 and 30.

(68) a. *Yau* *ai=’u* *masere*
1s leg=1s.INAL sick

b. *Au* kaki *masere-’e*
1s leg(Malay) sick-LM
‘My leg hurts.’ (E)

(69) a. *Yami* *taha=mi* *maruhu=’amu*
1pe NEG=1pe hungry=1pe

b. *Ki* *taha* *maruhu-‘e*
1p(i) NEG hungry-LM
‘We are not hungry.’ (E)

Example (68) shows the loss of the enclitic marking of inalienable possession, at least partly due to the use of a Malay loan word for the possessed body part. Example (69) shows the loss of the preverbal clitic referencing the S argument and the loss of the enclitic on the verb referencing an affected undergoer S. This example also shows the collapse of the inclusive/exclusive distinction for first person plural. In both examples, the unexpected use of the lexical marker suffix *-e* is a characteristic strategy of younger people establishing their identity as speakers of the language.

As for the system described here, that still used by conservative speakers, it can best be characterized as a syntactic system albeit one which has a morphological look to it. Representing the structure as a flat one, Sou Amana Teru has a position for NPs at the left edge of the clause, followed by auxiliaries and negators, which are followed in turn by a position for a cross-referencing pronoun. The first two positions need not be filled, while the clitic position must be filled except when there is no suitable pronoun form (non-human S/A argument) or when filling it would result in two adjacent pronouns.[4]

[4] I assume a general requirement that there be some overt realisation of the S/A argument.

The preceding discussion has made it clear that S and A are treated as a single category in Sou Amana Teru as opposed to O. The only exception to this documented here is that some S arguments have O-like characteristics. But this does not invalidate the generalization: the O-like properties (being represented by an enclitic on the verb) are in addition to the S-like properties for such arguments, and it is very common to find some aspects of the grammar of a language which display split S properties without that fact calling into question the basic alignment pattern of the language (Verhaar 1990). The argument-marking system of Sou Amana Teru therefore aligns S and A together against O in a nominative-accusative pattern. Despite this, I have avoided the use of the term 'subject' to refer to the S/A argument in Sou Amana Teru, because it is not yet clear to me whether anything is to be gained by assuming the existence of grammatical relations in this language. The most important evidence which might be taken as support for analysing Sou Amana Teru without grammatical relations is the almost complete lack of constructions which use pivots (see Ross 2004 for related discussion of Oceanic languages). The preverbal pronoun representing S or A is retained in most contexts where it might be expected to be omitted, such as with dependent verbs (examples (70) and (71), see also example (3)) and in purpose clauses (example (49) repeated here).

(70) *Isi* asik ***isi*** *pahoi=si*
3p busy 3p wash=3p
'They were busy washing themselves.' (SD)

(71) *Aman-de* pertama-*ma* *tahina-na* *matua-na* *si=turu*
village=this first=that old.woman-PL old.man-PL 3p=descend

si*=lohi* *tampa* *a'-upa-t*
3p=seek place NMLZ-sit-NMLZ
'The ancestors from the first village went down to look for a place to live.' (SD)

(49) *Uma* *ute* *wake=ma* *e=tana=re* *ena* *e=pa'-ia=re*
then bamboo joint=that 3s=take=3nh for 3s=CAUS-good=3nh
'Then that bamboo joint, she took it to fix it.' (SD)

The only exception to this generalization is that the verb of a complement clause following the verb *oi* 'go' cannot have its own S or A pronoun:

(72) *Yau* *oi* *sahe* roti-*'e*
1s go buy bread-LM
'I went to buy bread.' (E)

This exception could be accounted for in various ways without invoking the notion of same-subject deletion. For example, it is semantically almost impossible for the S or A argument of a second verb following 'go' to not be co-referential with the goer. Or, alternatively, examples such as (72) might be analysed as serial verb constructions.

It is not yet clear whether Sou Amana Teru has true relative clauses, but in the closest corresponding construction there is also no omission of arguments. Example (73) is the type of structure given as translation of prompts which contain relative clauses, and it is possible that such structures should be analysed as simple juxtaposition of clauses.

(73) *Yau nau mansia-e **si**=upa wa'a rete Waai*
1s see person-LM 3p=sit LOC direction PN
'I saw the people who live at Waai.' (E)

However, in some cases, speakers insert the Malay relative clause marker *yang* into sentences, but argument omission still does not occur in the relative clause; omission of the argument is obligatory in the corresponding Malay structure (example (74)).

(74) *Mansia* yang ***isi*** *ane ian-e si=upa wa'a rete Waai*
person REL 3p eat fish-LM 3p=sit LOC direction PN
'The people who eat fish live at Waai.'(E)

Orang-orang yang makan ikan tinggal di Waai
person-RDP REL eat fish live LOC Waai
'The people who eat fish live at Waai.'

The evidence presented here does not amount to a conclusive demonstration that Sou Amana Teru lacks the grammatical relations subject and object, but it does suggest that there little is to be gained by assuming their existence in Sou Amana Teru.

6 Conclusion

This paper has presented data on the use of various types of pronouns in Sou Amana Teru. The major part of the analysis has been concerned with the use of clitic pronouns in argument positions. I have argued that enclitic pronouns are one possible realization of O arguments, but that such pronouns do not cross-reference O arguments. In contrast, clitic pronouns do serve a cross-referencing function for S and A arguments. This system is not clearly morphological, with some of its properties being more amenable to a syntactic analysis. I suggest that the system described here should be seen as one point in a process of change which is moving from a morphological system of argument marking to a more analytic system, and that this is part of a more general decline in the use of morphology in the language.

References

Andrews, Avery, 1985. The major function of the noun phrase. In Timothy Shopen, ed., *Language typology and syntactic description: 1 Clause Structure*, 62–154. Cambridge: Cambridge University Press.

Blust, Robert A., 1977. The Proto-Austronesian pronouns and Austronesian sub-grouping. *Working Papers in Linguistics, University of Hawai'i*, 9:1–15.

Bolton, Rosemary, 1990. A preliminary description of Nuaulu phonology and grammar, Arlington: University of Texas at Arlington. MA Dissertation.

Collins, James T., 1983. *The Historical Relationships of the Languages of Central Maluku, Indonesia.* Canberra: Pacific Linguistics.

Dixon, R.M.W., 1972. *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.

Ewing, Michael C., 2005. *Grammar and Inference in Conversation: Identifying clause structure in spoken Javanese.* Amsterdam: John Benjamins.

Florey, Margaret, 2001. Verb and Valence in Alune. Explorations in valency in Austronesian languages. *La Trobe Papers in Linguistics*,11:73–120.

— 2005. A Cross-linguistic Perspective on Emergent Possessive Constructions in Central Moluccan Languages. *Australian Journal of Linguistics*, 25:59–84.

Givón, Talmy, 1976. Topic, pronoun and agreement. In Charles N. Li, ed., *Subject and Topic*, 149–188. New York: Academic Press.

Himmelmann, Nikolaus P., 2005. The Austronesian languages of Asia and Madagascar: typological characteristics . In K. Alexander Adelaar and Nikolaus P. Himmelmann, eds, *The Austronesian Languages of Asia and Madagascar*, 110–181. London: Routledge.

Hughes, Jock, 2000. The morphology of Dobel, Aru, with special reference to reduplication. In Charles E. Grimes, ed., *Spices from the East: papers in languages of esatern Indonesia*, 131–180. Canberra: Pacific Linguistics.

Minde, Don van, 1997. *Malayu Ambong: Phonology, Morphology, Syntax*. Leiden: CNWS.

Musgrave, Simon, 2002. Non-subject Arguments in Indonesian. Melbourne: University of Melbourne. PhD Dissertation.

— 2005. Possession in Souw Amana Teru: internally and externally motivated change. *Monash University Linguistics Papers* 4:65–81.

— 2006. Dialects and varieties in a situation of language endangerment. *Selected Papers from the 2005 Conference of the Australian Linguistics Society.* View at http://www.als.asn.au/proceedings/als2005/musgrave-dialects.pdf (accessed 19 November 2007)).

Musgrave, Simon, and Ewing, Michael C., 2006. Language and religion: a case study of two Ambonese communities. *International Journal of the Sociology of Language*, 179:179–194.

Nivens, Richard J., 1998. Borrowing vs. code-switching: Malay insertions in the conversations of West Tarangan speakers of the Aru Islands of Maluku, eastern Indonesia. Manoa: University of Hawai'i. PhD Dissertation.

Pawley, Andrew., 1973. Some problems in Proto Oceanic Grammar. *Oceanic Linguistics*, 12:103–188.

Ross, Malcolm, 2004. The Morphosyntactic Typology of Oceanic Languages. *Language and Linguistics*, 5:491–541.

Verhaar, John, 1990. How transitive is intransitive? *Studies in Language*, 14:93–168.

8 *Voice and valency in Tetun Dili*

CATHARINA WILLIAMS-VAN KLINKEN

1 Introduction

Tetun Dili, one of the two official languages of the newly independent nation of East Timor, is an Austronesian language derived from the vernacular Tetun Terik, but showing strong Portuguese influence and creole-like simplification of morphology and syntax. It has a fixed word order, traditionally has no passive voice, and has few morphological markers of syntactic structure.

The fixed word order (as SV(O) or topicalized O(S)V) makes the syntactic analysis of the argument structure of most clauses straightforward. This chapter looks at the residual, non-straightforward cases, which consist of an undergoer noun phrase followed by a potentially transitive verb. Such sequences are illustrated by *bee fakar* 'water pour' in example (1) and *DIT harii* 'DIT set.up' in (2). Examples (3) and (4) show the same verbs *fakar* 'pour' and *harii* 'set up' in clearly transitive constructions.

(1) *Bee matan hotu-hotu iha rai okos nak-fera too bee*
water source all-all LOC earth under INTR-split until water

fakar sai.
pour exit
[In Noah's flood:] 'All the underground springs burst open with the result that water poured out.'

(2) *DIT harii iha tinan rihun rua rua liu ba.*
DIT set.up LOC year thousand two two pass go
'DIT (Dili Institute of Technology) (we) set up in the year 2002.' or: 'DIT was set up in the year 2002.'

(3) *Ohin João fakar bee iha sementi leten.*
earlier.today João pour water LOC cement top
'Today João poured water onto the cement.'

(4) *Ami harii ami nia fundasaun nee iha*
1pe set.up 1pe POSS foundation this LOC

Michael C. Ewing and Marian Klamer, editors.
Typological and areal analyses: contributions from East Nusantara, 165-183. Pacific Linguistics, 2010
© This edition vested with Pacific Linguistics.

tinan *rihun* *rua.*
year thousand two
'We set up our foundation in the year 2000.'

Most of this chapter addresses the syntactic question of whether it is possible to determine the argument structure of such clauses. One alternative is an analysis as a subject followed by an intransitive verb (comparable to middle voice 'The water poured out'), as suggested by the translation of (2). The other is an analysis as a topicalized object followed by a transitive verb, with the subject being elided, as suggested by the first translation given for (2). A sub-question is whether there is indeed support for the notions of 'subject' and 'object' as independent grammatical categories in Tetun Dili. The final section of the chapter looks at a new construction now being used in the media, in which undergoer-verb constructions are expanded to include an actor, presented as a peripheral argument. This construction appears to be an incipient agentive passive.

An issue noted at several points throughout the discussion is the effect of other languages on Tetun Dili. The history of East Timor makes such influence almost inevitable. Until 1999, Tetun Dili was primarily an oral language of intercultural communication, with home and local community life carried out in local vernaculars, and writing and public communication carried out in colonial languages. The half-island was ruled by Portugal for several centuries, with Portuguese as the language of education and public life. After the Indonesian invasion in 1975, the Indonesian language took over these functions. The only major exception was that the influential Catholic Church replaced Portuguese with Tetun rather than Indonesian as its liturgical language. The role of Tetun has expanded greatly since 1999, when East Timorese voted for independence, with Tetun now playing a major role in government, media, and public life. The vast majority of new words entering Tetun's vocabulary to meet these new demands are from Portuguese. As a result of the presence of the United Nations and many international organizations, English too has played an increasing role in the country since 1999, in particular as the source of much translation into Tetun.

2 Tetun Dili clause structure

2.1 Default clause structure

Intransitive clauses in Tetun Dili follow strict subject-verb order, with a few lexically-defined exceptions (discussed in Williams-van Klinken *et al.* 2002:54–56). Subject-verb order applies regardless of whether the subject is semantically actor or undergoer, as illustrated by the two clauses in the following example.

(5) *Milísia* *moos* *tauk,* *sira* *halai* *sees* *hotu.*
militia.man also afraid 3p run move.aside all
[We drove fast towards the militia roadblock.] 'The militia men were consequently afraid, (and) they all ran out of the way.'

The default order for transitive clauses is subject-verb-object, as in (6).[1] As will be seen in §4.1, the main characteristic shared by subjects of transitive and intransitive verbs is in fact this default preverbal position.

(6) *Ita foo tua ba apaa.*
1pi give palm.wine to dad
'We'll give palm wine to dad.'

For transitive clauses, there is a one-to-one match between syntactic structure and semantic structure, in that the subject is always actor, and the object is always undergoer. The macro-role of actor is defined, following Foley and van Valin (1984:29), as 'the argument of a predicate which expresses the participant which performs, effects, instigates, or controls the situation denoted by the predicate.' It includes roles such as agent and experiencer. The undergoer is 'the argument which expresses the participant which does not perform, initiate, or control any situation but rather is affected by it in some way.' It includes roles such as theme, patient, and goal.

That is, in Tetun Dili the subject is always more agentive than the object. Thus a sentence like *Oan kiak nee foo osan* 'child orphan/poor this give money' could only mean that the orphan gave money, not that he or she was given it. This strict ordering is discussed further in section §2.3, and is in marked contrast to languages such as Riau Indonesian, where semantic roles are underspecified (Gil 2001).

2.2 Object-fronting

Although the default clause order is SVO, objects can also be fronted to before subject position. Such fronting has two primary functions (Williams-van Klinken *et al.* 2002:52–53). One contrasts the fronted constituent with other members of a set of entities which have either been explicitly mentioned, or which have been implicitly invoked. In example (7), for instance, the fronted object refers to the (cows') milk which the infant does not drink, in contrast to the breast milk which she does take. The fact that she drinks something is taken for granted.

(7) *Maibee susubeen nia laduun hemu. Nia susu deit.*
but milk 3s not.very drink 3s suck just
[Every day we feed our infant daughter porridge, then give her milk.] 'But she doesn't drink much milk. She just breastfeeds.'

The other major function of fronting an object is to mark it as the topic or theme of the sentence. Such topics are often, though not necessarily, marked by anaphoric *nee* 'this'. This is illustrated by the fronted *kuda nee* 'horse this', which precedes the subject in the final clause of example (8), and which continues the topic introduced in the first clause. Example (9) similarly introduces *ai-moruk* 'medicine' in the initial clause, and refers to it again in the fronted object in the final sentence.

[1] SOV word order is possible in informal speech under very restricted conditions (Williams-van Klinken *et al.* 2002:54). However, this order, which appears to amount to object-incorporation, is not relevant here.

(8) *Ami lori kuda, ami hanesan hahaan la iha,*
1pe take horse 1pe like food not have
[When we fled, we took a horse.] 'We took a horse, [and when] like we had no food,

kuda nee ami baa troka fali hahaan.
horse this 1pe go exchange instead food
we went and swapped this horse for food.'

(9) *Kuandu ami foo ai-moruk, loron ida-rua maka nia*
if 1pe give plant-bitter day one-two FOC 3S.POSS

moras la liu,
sick not pass
'If we give [our daughter] medicine, and after a day or two the illness hasn't passed,

foin ami lori nia ba ospitál.
only.then 1pe take 3s to hospital
only then do we take her to hospital.

Ai-moruk nee foo ba la pasa, ami lori nia baa konsulta.
plant-bitter this give to not heal 1pe take 3s go consult
[If we] give [her] this medicine, [and the illness] isn't cured, we take her to see a doctor/nurse.'

2.3 Grammatical role of the single preverbal argument of a transitive verb

Both subjects and objects can readily be omitted if the referent is clear from the context, or is irrelevant. The single preverbal argument of a transitive verb can thus either be its object, such as with *foo* 'give' in the final sentence of (9), or its subject, as with *ita foo* 'we give' in example (10).

(10) *Buat nebee mak nia hakarak, se ita bele foo, ita foo, e.*
thing REL FOC 3s want if 1pi can give 1pi give TAG
'The things which she wants, if we can give [them to her], we give [them], don't we.'

Despite theoretical ambiguity, in practice it is always possible to determine whether the single argument preceding a transitive verb is its subject or a fronted object, simply because subjects of transitive verbs are actors, while objects are undergoers. The animacy hierarchy alone is sufficient to disambiguate the role: If the argument is human or higher animate, it is an actor, such as *ita* 'we' in (10). If it is inanimate, such as *ai-moruk nee* 'this medicine' in (9), it is an undergoer.

There are of course many transitive verbs which take human undergoer objects, such as *hadomi* 'love'. It would thus in principle be possible to have a sentence such as *Paulo hadomi* 'Paulo love', in which it was impossible to determine whether Paulo was loving (actor) or loved (undergoer). However, not surprisingly, it is virtually impossible to use such verbs in undergoer-verb sequences. Indeed, I have noted only one example – a

newspaper headline[2] – in the corpus on which this study is based. Invented examples were invariably rejected in elicitation as being 'forced', even with the help of context and intonation; e.g. **Nia oan moos baku* '3s.POSS child also beat', to mean 'Even his children (he) beats.' As assistants pointed out, an initial human noun phrase would automatically be interpreted as actor.

2.4 Syntactic structure of undergoer argument followed by ambi-valent verb

As noted above, the syntactic structure of the single initial argument of a verb is almost always clear in practice. If the verb is intransitive, then the argument is the subject. If it is transitive, then semantic actors are subjects, and semantic undergoers are fronted objects.

There remains, however, one class of verbs for which the syntactic argument structure is not clear. These are verbs which can either function transitively, or intransitively with an undergoer subject. That is, the subject of the intransitive verb corresponds semantically to the object of the transitive verb, and there is no morphological distinction between the transitive and intransitive uses of the verb. For ease of reference, these will be referred to as 'ambi-valent' in the remainder of this chapter. For these verbs, undergoer-verb constructions can in principle be analysed in three ways:

1. The argument is the subject, and the verb is intransitive. This analysis intuitively fits *bee fakar sai* 'water poured out' in example (1).
2. The argument is the object, the verb is transitive, and the subject has been omitted. This analysis fits *ai-moruk nee foo* 'this medicine (we) give' in example (9).
3. It is impossible to tell, and the question may indeed be irrelevant. In this case, the notion of 'subject' may be unhelpful in the analysis of Tetun Dili, or, at a minimum, the grammatical function is underspecified in such constructions (as proposed by Gil (2001) for similar constructions in Riau Indonesian).

3 Methodology

3.1 Verbs investigated

All potentially ambi-valent verbs which came to my attention were investigated. These included:

1. The 21 verbs (out of almost 1200 verbs) in my dictionary file which had already been tentatively classified as being both transitive and intransitive, either by myself or in Hull's (2002) published Tetun dictionary, and which I suspected could take undergoer subjects in their intransitive use.
2. Two other transitive verbs which came to my attention as commonly occurring in undergoer-verb constructions, namely *harii* 'set up' and *halao* 'conduct'.

There were thus 23 verbs under investigation.

---

[2] The headline read *Marí Sei Investiga Kinta.Feira Semana Ne'e* 'Marí will investigate Thursday week this' = 'Marí will be investigated on Thursday this week' (Nacional Diario, 18/7/06, page 1). This referred to the widely anticipated court appearance of the then prime minister, in which everybody knew that he would be the one being questioned, rather than the one doing the questioning. Students whom I directed to this headline immediately noted its ambiguity.

Two other ambi-valent verbs were excluded from the investigation because they have unique properties and have already been described elsewhere (Williams-van Klinken *et al.* 2002:45–46, 54–55). One is *iha*, which transitively means 'have' and intransitively means 'exist, be present'. The other is *liu*, which transitively means 'pass', and can also be used intransitively of a time period or an event passing.

3.2 Data sources

The data came from both a corpus and elicitation. The corpus consisted of over 40,000 words of transcribed oral texts, and about 200,000 words each of written texts and of translations into Tetun. Using the Toolbox program (2001–2004 SIL International), I listed every example of each of the verbs under examination, and was also able to readily find all examples of certain constructions if these involved a particular lexical item (e.g. by looking for *halo* 'make' in causatives, or certain conjunctions). Interlinearized examples in this chapter are from the corpus, except for (20), which was given by assistants during elicitation.

I first sought examples of all the selected verbs in these texts, looking in particular for all undergoer-verb sequences. If a verb occurred frequently in a particular construction in oral texts, I presumed this was 'normal' correct Tetun. Some, however, occurred only in written or translated texts. Since the languages of education in East Timor have been Portuguese and Indonesian, and the source of much translation is English, I treated any written and translated texts with caution, as perhaps reflecting the influence of one of these three languages.

The many gaps in the data were filled by monolingual Tetun elicitation with four tertiary educated residents of Dili who have taught Tetun to foreigners, and nine students of the Dili Institute of Technology.[3] They worked in pairs or small groups, usually directly with me, on two main tasks:

1. They were asked to orally create contexts and sentences for particular verbs in particular constructions. For instance, when working on causative constructions, I asked them to make a sentence saying that something *taka* 'closed' (to which a response was *Biblioteka taka tuku lima* ' [The] library closes at 5 o'clock'], then, for example, asked them to try to say that some problem caused this to happen yesterday. Occasionally, discussion of one context brought about spontaneous sentences illustrating some other word or construction in which I was interested. The assistants were on the whole creative in seeking possible contexts.
2. They judged the grammaticality of sentences created by themselves or by myself (usually based directly on their examples) and of sentences found in the corpus which represented low-frequency constructions. Although the assistants were generally able to come up with an example for each construction, as being 'the best I can do if I am forced to say this', they were also able to reject their own examples as nevertheless not being correct Tetun.

[3] These were Jorge de Orleans Alberto Magalhães, Simeão Brites Seixas, Terezinha Araujo Cardoso Gusmão, Guilhermina Mouzinho, Alberto Correia, Marito da Silva Alves, Almeida Boavida, Isaias Nivio H.F. Pereira, Dámaso A. Vereira, Carvarinho Bento, Domingos Locatelli, Lídio Inácio Freitas and Maia Faria.

3.3 Tests

The following tests were considered as possible means for distinguishing between an SV and an OV analysis of undergoer-verb constructions, where the verb is potentially transitive.

1. **Use distinguishing properties of subjects and objects** to determine the grammatical function of the single preverbal argument: The results of this test are reported in §4.1.
2. **Use acceptability in dependent clauses** to determine whether the argument is fronted: Topicalization is normally considered to be a sentence-level phenomenon, which does not occur in dependent clauses. On this assumption, if undergoer-verb constructions occur in dependent clauses, they should be analysed as SV rather than as topicalized OV. This test is further discussed in §4.2.
3. **Test whether the verb can function intransitively:** There are a number of tests which can be used to prove that certain verbs are (or at least can be) intransitive with an undergoer subject. The results of these tests are reported in §4.3.
4. **Use the presence or absence of implicit actors** to help determine verb transitivity: Semantically, undergoer-verb clauses with SV structure can be expected to have no implicit actor. OV clauses, on the other hand, have an omitted subject, so an actor is likely to be understood. This test is discussed in §4.4.

The overall results for each test are presented in Table 1 within §4.5.

4 Results

4.1 Properties of subjects and objects

In Tetun Dili, actors of transitive verbs and the single argument of intransitive verbs are syntactically subjects, while undergoers of transitive verbs are objects. The main syntactic feature uniting subjects of transitive and intransitive verbs is their default preverbal position, in contrast to the default post-verbal position for objects. (Peripheral arguments, such as recipients and addressees, also occur after the verb, and are further distinguished by requiring prepositions.)

Subject properties which would distinguish between an SV and an OV analysis of undergoer-verb constructions proved to be elusive, for various reasons. Some of the criteria that distinguish subjects in other languages are not applicable. For instance:

1. Tetun Dili has no subject marking on verbs. In this it is unlike its source language Tetun Terik (van Klinken 1999:172–178).
2. There is no case marking.
3. There is no marking of voice.
4. Both subjects and objects are readily omissible.

Some criteria which in other languages distinguish subjects are in Tetun Dili shared between subjects and fronted objects. These thus distinguish between preverbal core arguments and post-verbal objects but not between subjects and objects as such. In particular:

1. Subjects and fronted objects can both be relativized (as can time phrases).

(11)

sira	*nebee*	*hili*	*Tetun*	*inisiativa*	*nebee*	*sira*	*hala'o*
3p	REL	choose	Tetun	initiative	REL	3p	carry.out
S: 'those who chose Tetun'				O: 'the initiative which they took'			

2. Both subjects and fronted objects can float the quantifier *hotu* 'all'. When an object is fronted, *hotu* quantifies the object rather than the subject.[4]

(12)

Ami	*sae*	*hotu*	*mikrolet.*	*Sasaan*	*ema*	*naok*	*hotu.*
1pe	ascend	all	minibus	goods	person	steal	all
S: 'We all got into the minibus.'				O: 'Someone/people stole all (our) goods.'			

3. Both can be focused by *mak*, as can many other preverbal constituents (such as time, reason prepositional phrases, and fronted question adverbials) (Williams-van Klinken *et al.* 2002:68).
4. Both are prototypically marked as definite, although not obligatorily so.

There are a few subject properties in Tetun which, by virtue of the semantics of the constructions in which they occur, apply only to actor subjects. These are thus not useful in determining whether undergoer-verb constructions could be interpreted as a subject followed by an intransitive verb. These properties include the following:

1. Subjects control reflexives. In reflexive constructions, the subject referent acts on himself/herself; the subject is thus always semantically an actor (e.g. *Timor ukun-an.* 'Timor rule-REFL' = 'Timor is independent.').
2. In reciprocals, it is always the subject referent that is involved in the reciprocal action or relationship. Except in certain idiomatic expressions (e.g. *diak malu* 'good RECP' = 'get on well with each other'), this construction involves transitive verbs (e.g. *Ami hasoru malu.* '1pe meet RECP' = 'We met each other.'), or intransitive verbs with actor subjects and prepositional phrase (e.g. *Sira koalia ba malu.* '3p speak to RECP' = 'They spoke to each other.').
3. The subject is obligatorily omitted before the second verb in same-subject serial verb constructions (Williams-van Klinken *et al.* 2002:92–96), such as *halai sai* 'run exit' = 'run outside', and *mai haan* 'come eat' = 'come and eat'. The subject is semantically the actor of both verbs.

To date I have found only one property which is genuinely restricted to subjects regardless of whether these are actor or undergoer, namely that subjects are obligatorily omitted in reduced complement clauses. For instance, in *Hau hakarak baa* 'I want (to) go', the subject of *baa* 'go' is obligatorily omitted. The categories of verbs which take such complements include verbs of thinking, knowing, wanting, liking, and trying. By virtue of their semantics, these take human subjects, which will normally be interpreted as actor of both verbs. An exception is that the subject of *lakohi* 'not want, refuse' can be the undergoer subject of an intransitive complement verb, as in *Ami lakohi mate* 'We

[4] Donohue (2005) similarly notes that it is the fronted undergoer rather than the actor that is quantified in the Austronesian language Palu'e. He uses this fact, along with a range of others, to argue that in Palu'e, the fronted undergoer is in fact a passive subject, and the actor an adjunct NP.

don't.want (to) die'. After all, one does not necessarily instigate or have control over the things which one does not want to happen. Nevertheless, even for *lakohi*, the subject is, in the corpus, nearly always actor of the complement verb.

In conclusion, support for an independent notion of 'subject' in Tetun Dili is very weak. Apart from default position within the clause and subject deletion of complements of *lakohi* 'don't want', all the properties of subjects are in fact shared properties of preverbal arguments (whether these be subjects or fronted objects), or restricted to semantic actors.

If a clause has only a single preverbal undergoer argument, then there appear to be no syntactic properties of subjects or objects in Tetun Dili which would allow one to determine which syntactic role this argument fulfils.

4.2 Acceptability in dependent clauses

Topicalization, including object-fronting, is normally considered to be a sentence-level phenomenon. Therefore one would expect that object-fronting would not occur within dependent clauses. If this expectation is borne out, then undergoer-verb sequences in dependent clauses must be analysed as SV, with an intransitive verb, rather than as OV, with a transitive verb.

As anticipated, object-fronting appears to be totally impossible for analytic causatives, which consist of *halo* 'make' followed by a causee noun phrase and caused predicate (e.g. *halo nia terus* 'make him/her suffer'). The corpus contains no instances in which objects of the caused predicate verb were fronted to causee position. Attempts to front the object during elicitation were either rejected or led to assistants disagreeing as to the sentence's acceptability (for example ?*Reitór nee sempre halo enkontru halao lalais* 'vice-chancellor this always makes meeting conduct quickly' to mean 'This vice-chancellor always ensures meetings are conducted quickly', where *enkontru* 'meeting' is the fronted object of *halao* 'conduct').

Nevertheless, in adverbial clauses, which are more loosely bonded to the main clause, it is somewhat easier to accept OV order. For complement and adverbial clauses, object fronting occurs (albeit rarely) in the corpus, and some of these corpus sentences were accepted during elicitation. However in all cases, even when an object-fronted dependent clause was accepted, the assistants judged VO order to be better. It appears to be easier to allow OV structure in dependent clauses when the focus is on the result of an action, rather than on the action itself. In these cases there is no interest in the actor. Example (13), from a translation of electoral rules, and accepted by assistants, illustrates this.

(13) *Se loron moris troka tiha, bolu responsavel.*
if day born exchange PFV call head.person
[Regarding rules for registering voters:] 'If the date of birth has been changed [on the identity card], call the person in charge.'

The corpus examples of OV within dependent clauses are from written and translated texts. It seems likely that this construction is being pushed by the influence of the main languages in which current writers of Tetun have been educated, namely Portuguese and Indonesian, as well as English, from which many texts are currently being translated into Tetun. All three of these languages have passive constructions, and thus readily allow dependent clauses to include undergoer subjects of passivized transitive verbs, as in the English translation of example (13).

Although there is some acceptance of OV word order in adverbial clauses, OSV constituent order was not found in any non-main clauses in the corpus, and was not accepted in elicitation. For instance, adding a subject to the conditional clause in example (13) led to its total rejection (**Se loron moris João troka* 'if day [of] birth João change' to mean 'if the date of birth was modified by João'). That is, if the actor is specified in a dependent clause, one must use SVO order (*se João troka loron moris* 'if João modified date [of] birth'), not OSV order.

Since fronted objects in dependent clauses were never unreservedly approved of, I conclude that this test can be used (though with some margin of uncertainty) to distinguish subjects from fronted objects. That is, any potentially transitive verb which readily occurs in undergoer-verb constructions within a dependent clause is actually intransitive with an undergoer subject in this instance. This test shows 12 transitive verbs to also be classifiable as intransitive with undergoer subject. An example is *remata* 'finish', illustrated as a transitive verb in (14), and as intransitive within an adverbial clause in (15). The full test results are presented in Table 1 in section 4.5.

(14) *Molok atu remata hau nia liafuan, dala ida tan hau hakarak ...*
before IRR finish 1s POSS word time one more 1s want
'Before finishing my talk, once more I would like to [express my greetings to your family.]'

(15) *Bainhira João nia servisu atu remata ona, ...*
when João POSS work IRR finish ANT
'When João's work was about to finish, ...'

4.3 Tests for intransitivity

4.3.1 Introduction

All the verbs under consideration can readily be proved to be transitive, simply because they can take both a subject and an object, in an SVO clause. Testing whether they should also be classified as intransitive with an undergoer subject is not so simple. There are two tests involving causative constructions which can be used to prove that some of these verbs are indeed intransitive. However, both constructions accept only lexically restricted sets of verbs. Hence, while they can be used to prove that certain verbs *are* intransitive, they cannot be used to prove that the remainder of the verbs are not.

Nevertheless, if a verb can be shown to function intransitively in some contexts, then when it occurs in a potentially ambiguous undergoer-verb construction, the most parsimonious conclusion is that it is intransitive in that context as well. If, however, there is no independent evidence that a verb can be used intransitively, an object-verb analysis is more convincing, with the verb being transitive.

4.3.2 Morphological causatives

In morphological causatives, the prefix *ha-* is added to an intransitive verb; e.g. *ha-toba* 'make-lie.down' = 'lay down'. It never applies to unambiguously transitive roots, so any roots of such derivations can safely be presumed to be intransitive. This prefix is however far less productive in Tetun Dili than in Tetun Terik, and there are many intransitive roots to which it does not apply.

This prefix can be added to only four of the verbs in this study. The most commonly occurring derivation of this type is *ha-lakon* 'make-disappear' = 'get rid of', illustrated in example (16). Two others verbs are morphologically both a root and a causative, namely *hakiduk* 'reverse' and *haliis* 'slant'. That is, the intransitive root has this form, and when causative *ha-* is prefixed to the final two syllables according to the morphophonemic rules of Tetun (van Klinken 1999), the resulting form looks exactly the same as the root. The final example, *ha-para* 'make-stop', is used by many speakers as a transitive verb meaning 'stop', even though some others dislike it for mixing a Tetun prefix with a Portuguese root.

(16) *Ita tenki ha-lakon tiha divizionizmu ...*
1pi must make-disappear PFV factionalism
'We must get rid of the factionalism ...'

4.3.3 Serial verb causatives

In causative serial verb constructions, the verb *halo* 'make, cause' is immediately followed by an intransitive verb, which is followed in turn by the object, as per *halo lakon* 'make disappear' in example (17).

(17) *... susar tebes atu halo lakon buat hirak nee.*
difficult indeed IRR make disappear thing certain this
'... it will be very difficult to get rid of these things (corruption and crime).'

This test identifies four verbs in the study as intransitive, namely *lakon* 'lose, disappear' and *haliis* 'slant' (both of which can alternatively take a causative prefix), and two Portuguese loans *remata* 'finish' and *estraga* 'ruin' (which as trisyllabic borrowings cannot take a causative prefix). *Remata* 'finish' is illustrated in example (18).

(18) *Nia halo remata tiha nia servisu.*
3s make finished PFV 3s.POSS work
'He has completed his work.'

4.3.4 Summary

Together, these two tests identify seven of the verbs investigated as being intransitive with undergoer subject, with two verbs, *haliis* 'slant' and *lakon* 'lose, disappear', passing both tests. The full list of test results is presented in Table 1 in Section 4.5 below.

4.4 Implicit actors

In object-verb (OV) clauses, the subject is omitted, usually because the actor is either understood from context, or because it is deemed irrelevant. It seems reasonable to expect that in many cases, people would be able to identify who the unspecified actor subject of an OV clause would be.

Assistants clearly understood the concept of implicit actor, and were often able to state whether there was one. For instance, in example (19), the party is run by someone, in particular by the speaker and her extended family, mentioned earlier. In such cases, it is syntactically possible, though perhaps textually odd, to add a subject noun phrase to the clause.

(19) *Ajuda hotu para hanesa festa nee bele ha-lao ho diak.*
help all so.that like party this can make-walk with good
[The food for the wedding is from us two, and from our parents and brothers and sisters. They all contribute.] '[They] all help, so that the party will be run well.'

However, not surprisingly, assistants at times found such judgments impossible to make with certainty. This was no doubt exacerbated by the fact that, in the absence of textual examples, some of the examples given were invented sentences without genuine context. Since the results were not sufficiently clear, this approach was not continued.

4.5 Overall results

The results of the tests for all verbs determined to be ambi-valent are listed in Table 1 below, with bold font indicating that the test supports an analysis as intransitive verb with undergoer subject. The remainder of the verbs tested are discussed in section 4.6.

Since the varying sources of information are not all of equal weight, the table specifies the source, as follows:

- **Text** introduces the total number of examples found in the corpus for this word in this construction. This includes oral, written and translated examples. When this number is reasonably high (e.g. over 5), I assume that the construction is not only grammatical, but also 'normal'.
- **Bible/Translate/Written**: if all corpus examples are from the Bible, other translations or other written texts, the count is introduced accordingly. These examples may well reflect the influence of other languages.
- **Spontan** introduces the number of times that this construction was used spontaneously during elicitation about the verb. If the construction was used spontaneously and subsequently accepted as correct, I assume this construction is more likely to be genuinely acceptable than if it was only accepted during elicitation.
- **Elicit** indicates whether the construction was accepted in elicitation, with 'Y' meaning 'yes', 'N' meaning 'no', '?' indicating that assistants were unsure, and 'YN' showing that some accepted it, while others rejected it. If a verb occurred in a particular construction more than a few times in texts, I did not verify its acceptability through elicitation.

Table 1: Results of tests for verbs found to be both transitive and intransitive with undergoer subject.

		Tests for intransitivity		**Preverbal in dependent clause**	
Word	Gloss	Morphological causative: *ha*-V	Serial verb causative: *halo* VU	Analytic causative: *halo* UV	UV in complement or adverbial clause
aumenta	increase	not applicable	**Elicit: Y**	**Text: 3, Elicit: Y**	**Text: 4, Elicit: Y**
estraga	ruin	not applicable	**Text: 2, Elicit: Y**	**Elicit: Y**	Elicit: ?
fakar	spill	Elicit: N	Elicit: Y?	Elicit: YN?	**Translate: 4, Elicit: Y**

		Tests for intransitivity		**Preverbal in dependent clause**	
Word	Gloss	Morphological causative: *ha*-V	Serial verb causative: *halo* VU	Analytic causative: *halo* UV	UV in complement or adverbial clause
hahuu	start	not applicable	Elicit: N	**Elicit: Y**	**Texts: 2, Elicit: Y**
hakiduk	reverse	**Elicit: Y**	Elicit: YN	**Elicit: Y**	**Elicit: Y**
haliis	slant	**Text: 3, Elicit: Y**	**Elicit: Y**	**Elicit: Y**	**Spontan: 1, Elicit: Y**
komesa	start	not applicable	Elicit: N	**Elicit: Y**	**Texts: 3, Elicit: Y**
kontínua	continue	not applicable	Elicit: ?	**Elicit: Y**	**Elicit: Y**
lakon	lose, disappear	**Text: 29**	**Text: 11**	**Text: 1, Elicit: Y**	**Text: 3**
loke	open	Elicit: N	Elicit: N	**Elicit: Y**	Translate: 3, Elicit: ?
muda	move, change	Elicit: N	Elicit: N	**Elicit: Y**	**Translate: 3, Spontan: 1**
para	stop	**Translate: 13**	Elicit: YN	**Elicit: Y**	**Elicit: Y**
remata	finish	not applicable	**Text: 8**	Elicit: N?	**Text: 17**
taka	close	Elicit: N	Elicit: N	**Elicit :Y**	**Elicit: Y**

For comparison, table 2 shows the results of the tests for a selection of verbs which are only intransitive (with undergoer subject) or only transitive.

Table 2: Results of tests for verbs which are only intransitive with undergoer subject, or only transitive.

		Tests for intransitivity		**Preverbal in dependent clause**	
Word	Gloss	Morphological causative: ha-V	Serial verb causative: halo VU	Analytic causative: halo UV	UV in complement or adverbial clause
Intransitive only					
kole	tired	Bible: 3,[5] Elicit: N	Text: 3	Bible: 1	Elicit: Y
mate	dead, die	Text: 18	Text: 9	Text: 9	Text: 75
monu	fall	Text: 8	Text: 1, Elicit: Y	Text: 9	Text: 32
nakfera	split	Elicit: N	Elicit: N	Elicit: Y	Text: 2
terus	suffer	Bible: 2, Elicit: N	Bible: 25	Text: 6	Text: 5

[5] The Bible translation (*Liafuan diak ba imi*, 2000) is written in liturgical Tetun, which is significantly influenced by Tetun Dili's primary source language, Tetun Terik. It follows Tetun Terik in making more generous use of the *ha*- prefix than does non-liturgical Tetun Dili (the primary variety of Tetun analysed in this article).

Word	Gloss	**Tests for intransitivity**		**Preverbal in dependent clause**	
		Morphological causative: ha-V	Serial verb causative: halo VU	Analytic causative: halo UV	UV in complement or adverbial clause
Transitive only					
ha-lao	carry out (make-walk)	Elicit: N	Elicit: N?	Elicit: N?	Text: 1, Translate: 2, Elicit: YN?
harii	set up	Elicit: N	Elicit: N	Elicit: YN?	Translate: 4, Elicit: Y
ha-tama	insert (make-enter)	Elicit: N	Elicit: N	Elicit: YN	Text: 1, Elicit: Y?
sobu	demolish	Elicit: N	Elicit: ?	Elicit: N	Elicit: ?
troka	replace, exchange	Elicit: N	Elicit: N?	Elicit: N	Translate: 1, Elicit: Y?

4.6 Short-listed verbs which proved not to be ambi-valent

A number of the verbs which were short-listed as possibly being ambi-valent turned out to not be unambiguously so. For some verbs, unambiguous classification proved impossible, due to disagreement amongst Tetun speakers over the basic uses of the verb, reflecting the lack of standardization of the language. (These are not listed in the table above.) For instance, two of the verbs in the short-list were universally accepted as intransitive, but accepted by only some people as transitive. These were *hariis* 'bathe' and *sarani* 'baptized' (with most using *foo hariis* 'give bathe' and *foo sarani* 'give baptized/baptism' as the transitive equivalents). Another verb, *dezenvolve* 'develop' is listed in Hull's (2002) Tetun dictionary as both transitive and intransitive, and in Whitlam *et al.*'s (2001) Portuguese dictionary as both transitive and reflexive, yet it passed none of the tests for intransitive verbs with undergoer subjects during elicitation with student assistants. It is, of course, possible that more Portuguese-influenced people would respond differently to the elicitation questions. Similarly, *aselera* 'accelerate' is listed as both transitive and intransitive by both Hull (2002) for Tetun and Whitlam *et al.* (2001) for Portuguese (with neither specifying whether it takes an actor or undergoer subject), yet for the student assistants it passed only one test for undergoer-subject status (namely the loosest one, of being able to occur in an adverbial clause), and it was strongly felt to have an understood actor.

Some short-listed verbs turned out to be both transitive and intransitive, but to require an actor subject in their intransitive use. These included *halibur* 'gather', *falta* 'absent; lack' and *subar* 'hide'.[6] This reflects a shortcoming in the short-list rather than a problem in the data.

[6] *Subar* is used intransitively with undergoer subject to mean 'be hidden' in some verses of the Bible translation *Liafuan diak ba imi* (2000), for instance in Matthew 14:4. These were not accepted by assistants, and could well reflect influence of the source language.

5 Discussion of ambi-valent verbs

The results of the various tests prove to be moderately consistent. Half of the ambi-valent verbs pass at least one of the intransitivity tests. Given that both tests are lexically restricted, one cannot expect that all intransitive verbs would pass them. All of these verbs are also recognized as intransitive with undergoer subject by one (usually both) of the dependent clause tests.

This study identified only 14 verbs which can reliably be categorized as ambi-valent. This is in addition to *iha* 'have; exist, be present' and *liu* 'pass', mentioned earlier. While the list is not claimed to be complete, it should represent a relatively large proportion of the ambi-valent verbs in the everyday language.

The low number of ambi-valent verbs is in marked contrast to the 'rampant lexical diathesis' which Markey and Fodale (1980, quoted in Bickerton 1981:72) describe as the general pattern for true creoles.

Half of these ambi-valent verbs are Portuguese loans. This is not surprising given that that 44% of the verbs in my dictionary file are of Portuguese origin. All the loans classified as ambi-valent in Tetun are also ambi-valent in Portuguese, being categorized as both transitive and either intransitive or reflexive in Portuguese dictionaries (e.g. Whitlam *et al.* 2001).

6 Pressure on undergoer-verb constructions

There are some transitive verbs which quite often occur in undergoer-verb constructions, but which are not conclusively intransitive. The most obvious are *harii* 'set up' and *halao* 'conduct'. The results for the tests of these two verbs are presented in Table 2 above.

The verb *harii* 'set up' occurs in undergoer-verb constructions in 25% of the 71 examples in the corpus. In the corpus such sequences are found in several dependent clauses, all from rather literal translations. Assistants could readily come up with undergoer-verb sequences within dependent clauses, and accepted such examples unreservedly as correct Tetun. One example produced by them is (20), which they judged to be analogous to Indonesian passives, without an understood actor.

(20) *Kuandu insituisaun PNTL harii fila fali, lei tránzitu*
when institution East.Timor.police set.up return again law traffic

bele lao fali.
can walk again
'When the East Timorese National Police force is re-established [after the current crisis], traffic law will again be operational.'

Nevertheless, *harii* does not pass any other tests for intransitivity, and most undergoer-verb sequences were judged to have an understood actor.

The verb *halao* 'carry out' (from *ha-* 'make' and *lao* 'walk') occurs in undergoer-verb constructions in 9% of the 202 examples in the corpus, including (21), from a translated manual on running the 2002 presidential election.

(21) *Eleisaun nee sei ha-lao hosi tuku 7:00 dadeer*
election this will make-walk from o'clock 7 morning

too tuku 4:00 lokraik.
until o'clock 4 afternoon
'The election will run from 7am to 4pm.'

Halao occurs in several written and translated examples in the corpus in an undergoer-verb construction within a dependent clause, suggesting that it was (at least under the pressure of writing) reanalysed by those authors as an intransitive verb with undergoer subject; however none of these examples received general acceptance by the assistants.

Both *harii* 'set up' and *halao* 'carry out' are common words in bureaucratic texts, where people speak of setting up systems and carrying out tasks. Until 1999, Timorese were accustomed to reading and writing bureaucratic texts only in other languages, first in Portuguese and later in Indonesian. Both languages make heavy use of passives in bureaucratic texts. So too does English, which has been the source for much translation in recent years.

Thus, given the high frequency of passives in the bureaucratic languages with which Timorese are familiar, it is not surprising that these words occur in functionally passive-like undergoer-verb clauses within written and translated texts. It remains to be seen whether sustained pressure in favour of a passive will result in these two verbs being generally reanalysed as intransitive with undergoer subject.

7 A new passive?

Tetun Dili does not (or at least did not) have a passive construction. In this it reflects its source language Tetun Terik (van Klinken 1999), and indeed the general pattern of central-eastern Austronesian languages (Klamer 2002).

However, in newspapers it is now standard to use the formula *Liafuan hirak nee hatoo hosi ...* 'word certain this deliver from' to introduce the source for a preceding statement, as in (22). This has the functional characteristics of a passive, with the undergoer foregrounded to subject position, and the actor backgrounded into a peripheral constituent (Foley and Van Valin 1984:149ff), introduced by the preposition *hosi* or *husi* 'from'. Unlike passives in most languages (Keenan 1985:255), however, there is no morphological marking on the verb and no auxiliary to indicate that this is passive voice.

(22) *Liafuan hirak nee ha-too hosi Senyora Maria de Jesus...*
word certain this make-reach from Mrs Maria de Jesus
'This (the preceding information) was given by Mrs Maria de Jesus ...'

The construction (undergoer verb *hosi*-actor) occurs most frequently in this one set expression, but is also increasingly being used creatively in the media, both written and oral. A small sample of current newspapers contained at least one example per edition, mostly occurring within long complicated sentences. In particular, many occur within relative clauses.[7] An example is (23), from the Timor Post.

[7] It is unclear why this construction should be used in relative clauses. The traditional means of relativising on an object involves subject-verb word order within the clause; e.g. *Relatoriu nebee PNUD halo* 'report REL UNDP make' = 'the report which UNDP produced'.

(23) *... Relatóriu ba Dezenvolvimentu Umanu Nasionál 2006 nian*
... report to development human national 2006 POSS
'... the National Human Development Report of 2006

nebee mak halo hosi PNUD ...
REL FOC make from UNDP
which was produced by UNDP ...'

Usually the construction uses an unmarked form of the verb, whether this be a native Tetun verb, such as *halo* 'make' in (23), or a Portuguese loan verb such as *asisti* 'attend, participate' in the fourth line of example (24).[8]

(24) *Hola parte iha serimónia entrega ekipamentu*
take part LOC ceremony hand.over equipment

nee mak hanesan
this FOC like
'Participating in the handing-over of equipment ceremony, were

funsionáriu balun embaixada Indonézia nian no moos
official some embassy Indonesia POSS and also
some officials of the Indonesian embassy, and (it) was also

asisti direitamente husi Xefe Kanselir ...
attend directly from head legation.chair (name)
attended in person by the legation head (name)

no segundu sekretáriu polítika ....
and second secretary political (name).
and the second political secretary (name).'

Occasionally, however, the Portuguese passive form of the verb is used, such as *preparadu* 'prepared' in (25). (The corpus includes five examples of this). This passive form is used primarily by highly Portuguese-influenced people, and reinforces the suspicion that this passive construction is motivated by the existence of passives in the languages in which Timorese people have been educated (primarily Portuguese and Indonesian). Note that the high percentage of Portuguese words in all these passive examples reflects the high formal register of the examples.

(25) *Nia haktuir katak, orsamentu jerál du estadu nebee prepara-du*
3s relate that budget general of state REL prepare-PASS
'He said that the state budget prepared

[8] Arka and Kosmas (2005) report a very similar construction for Manggarai, on the nearby island of Flores. The Manggarai passive too has no morphological marking on the verb. There, however, the agent phrase (introduced by an oblique marker) is usually obligatory, and the construction is well-established within the language.

husi *primeiru* *ministru* *demisionáriu* *...* *ida* *nee*
from first minister outgoing (name) one this

inválidu *tamba* *...*
invalid because ...
by the outgoing prime minister (name) is invalid because ...'

The passive construction is rarely used in spoken language. Nevertheless, the corpus does contain five spoken examples, including two from media interviews and three from formal meetings. Example (26) is from a music show on television.

(26) *I* *nunee* *ho* *múzika* *nebee* *mak* *sei* *oferese*
and so with music REL FOC will offer

husi *grupu* *Esperiénsia* *Stik*
from group experience drumstick
'And so, [we turn to] the music which will be presented by the band Drumstick Experience.'

In addition to being found in the media, this construction is also typically given as an initial response by students and inexperienced translators when they are asked for a Tetun translation of a passive in another language, regardless of whether the source language is Indonesian, Portuguese, or English.

Despite its increasing use, Tetun Dili speakers disagree as to whether this construction is acceptable. In my large informal sampling of students and writers, many more vote against the passive than in favour. Those disapproving include many who themselves wrote such sentences when given translation tasks. Since the construction is both relatively uncommon and rejected by most Tetun Dili speakers, it was not possible to do elicitation to determine any grammatical or lexical restrictions on it.

Development of a passive in a language which formerly had none is of course not a new phenomenon. Indeed, Heine and Kuteva (2002) list eight paths by which new passives have been observed to develop in the languages of the world. The structure used in Tetun Dili, is, however, not listed there, perhaps because, unlike prototypical passives, it has no morphological or auxiliary marking of voice. Rather, this passive is derived from middle voice undergoer-verb constructions, by the addition of an actor, in a pattern noted for some other languages by Shibatani (1999:409). It remains to be seen whether this incipient passive will expand to become a fully acceptable and widely used construction within Tetun Dili.

References

Arka, Wayan and Jeladu Kosmas, 2005. Passive without passive morphology? Evidence from Manggarai. In I. Wayan Arka and Malcolm Ross, eds, *The Many Faces of Austronesian Voice Systems: Some new empirical studies*, 87–117. Canberra: Pacific Linguistics.

Bickerton, Derek, 1981. *Roots of language*. Ann Arbor: Karoma Publishers, Inc.

Donohue, Mark, 2005. The Palu'e passive: from pragmatic construction to grammatical device. In I Wayan Arka Arka and Malcolm Ross, eds, *The Many Faces of*

Austronesian Voice Systems: Some new empirical studies, 59–85. Canberra: Pacific Linguistics.

Foley, William A. and Robert D. Van Valin Jr., 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Gil, David, 2001. Escaping Eurocentrism: Fieldwork as a process of unlearning. In Paul Newman and Martha Ratliff, eds, *Linguistic fieldwork*, 101–132. Cambridge: Cambridge University Press.

Heine, Bernd and Tania Kuteva, 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

Hull, Geoffrey, 2002. *Standard Tetum-English dictionary* (3rd edition). Winston Hills: Sebastião Aparício da Silva Project in association with Instituto Nacional de Linguística.

Liafuan diak ba imi [Good news for you], 2000. Dili: Catholic Church.

Keenan, Edward L., 1985. Passive in the world's languages. In Timothy Shopen, ed. *Language typology and syntactic description, volume 1: Clause structure*, 243–281. Cambridge: Cambridge University Press.

Klamer, Marian, 2002. Typical features of Austronesian languages in central/eastern Indonesia. *Oceanic Linguistics*, 41:363–383.

Shibatani, Masayoshi, 1999. Voice. In Keith Brown and Jim Miller, eds, *Concise encyclopedia of grammatical categories*, 406–412. Amsterdam: Elsevier.

van Klinken, Catharina, 1999. *A grammar of the Fehan dialect of Tetun, an Austronesian language of West Timor*. Canberra: Pacific Linguistics.

Whitlam, John, Vitoria Davies and Mike Harland, 2001. *Collins English-Portuguese Portugûes-Inglês dictionary* (2nd edition). Glasgow: HarperCollins Publishers.

Williams-van Klinken, Catharina, John Hajek and Rachel Nordlinger, 2002. *Tetun Dili: a grammar of an East Timorese language*. Canberra: Pacific Linguistics.

9 *Grammaticalisation of asymmetrical serial verb constructions in Klon*

LOUISE BAIRD

1 Introduction

Klon is a Non-Austronesian language spoken in south-west Alor by approximately 5000 speakers. Prior to Baird (2008) it was a previously undocumented and undescribed language, and it can be regarded as being endangered due to the increasing number of Klon children being brought up monolingually in Malay. Two dialects are identified by native speakers – Klon Bring and Klon Paneia. The boundary between the two is blurred, but clear distinctions exist in the centre and farthest extremities of each dialectal region. The main differences between the two dialects are phonological and lexical; there are also different pronominal paradigms. All of the data presented in this chapter comes from a corpus of Klon Bring texts, unless otherwise indicated.[1]

Klon is a predominantly verb-final language, with modifiers following the head and final negation. Intransitive clauses always have the structure SV.[2] Transitive clauses may have the structure AOV, OAV or AVO. The choice of constituent order is pragmatically determined, but it is possible to identify AOV as being more basic. S arguments are sometimes morpho-syntactically treated the same way as A arguments and sometimes treated the same way as O arguments. Consequently, the grammatical relations of Actor and Undergoer are identified for Klon. Although there are coordinating conjunctions speakers prefer to make extensive use of verb serialisation and paratactic coordination of

[1] The data used in this chapter was collected by the author on two fieldwork trips of approximately 8 weeks duration each in 2003 and 2004 to Alor, as part of 'Linguistic Variation in Eastern Indonesia: the Alor and Pantar Project' (www.let.leidenuniv.nl/aapp). The corpus consists of approximately 15 hours of digitally recorded, transcribed, glossed and translated texts, with additional elicited sentences and notes.

I gratefully acknowledge all of the assistance I received from the Klon in Probur village, the Nederlandse Organisatie voor Wetenschappelijk Onderzoek, which funded the project 'Linguistic Variation in Eastern Indonesia: the Alor and Pantar Project', and the Endangered Languages Documentation Programme which funded my fieldwork in 2004. Thanks is also heartily given to participants in the 2005 East Nusantara conference who shared their data and insights with regards to an earlier version of this chapter, and to Marian Klamer, Alexandra Aikhenvald and Ger Reesink who read and provided useful comments on the pre-final version of this chapter.

[2] S is used to represent the single argument of an intransitive clause, A represents the most agent-like argument of a transitive clause and O represents the most patient-like argument of a transitive clause.

Michael C. Ewing and Marian Klamer, editors.
Typological and areal analyses: contributions from East Nusantara, 185-201. Pacific Linguistics, 2010
© This edition vested with Pacific Linguistics.

clauses. See Baird (2008) for a more complete explanation and analysis of these and other grammatical phenomena.

In this chapter Aikhenvald's (2006) framework is used to describe Klon serial verb constructions (SVCs). Aikhenvald initially identifies symmetrical and asymmetrical SVCs, defining symmetrical SVCs as those that contain components from unrestricted classes, all being of equal status, whereas asymmetrical SVCs contain a verb from a closed class, which provides some kind of 'modificational specification' of the non-restricted verb(s) (Aikhenvald 2006:29). Aikhenvald then identifies certain semantic types of symmetrical and asymmetrical SVCs.

Three symmetrical and five asymmetrical types of SVCs are semantically identified for Klon. The semantic type of a specific serial verb construction is not always clear-cut. In some cases it is possible to semantically label a single construction in more than one way, and there may be cases for which none of the labels may be appropriate. The Klon SVCs presented in this chapter have been characterised semantically based on their most salient features.

All Klon SVCs can be identified based on the following defining morpho-syntactic characteristics (Baird 2008:136–139).

1. No verbs within Klon SVCs are syntactically subordinate to one another. We know this because of the SVCs' other characteristics.
2. Klon SVCs have a single illocutionary force, and belong to a single utterance type. It is not possible, for example, for some of the verbs in the complex to be declarative, and others to be interrogative.
3. Klon SVCs can only have a single intonation contour, with no stress or intonation breaks typical of those found at the edges of clauses occurring between the verbs.[3]
4. Klon SVCs share a single Actor argument, which is only marked once by either a full NP or a free pronoun.[4]
5. Verbs within Klon SVCs may share Undergoer arguments, or they may take different Undergoer arguments. If an Undergoer argument is marked by a pronominal prefix, all verbs in the verb complex sharing that Undergoer will take the pronominal prefix. Likewise, if a reciprocal prefix is used it will be used on all of the verbs within the construction.
6. Peripheral constituents, such as adverbs, cannot intervene between the verbs. Neither can the coordinating conjunctions.
7. All of the verbs in a Klon SVC have shared aspect and mood, with a single aspect or mood marker having scope over the entire serial complex.
8. A Klon SVC can only take one negator that has scope over the entire verbal complex. The individual verbs within a SVC cannot be separately negated, nor can some be negated and others not.

In this chapter an overview of the symmetrical and asymmetrical SVC types found in Klon will be provided, followed by a description of the grammaticalisation of

---

[3] Auditory analysis of Klon texts confirms this criterion, but acoustic analysis to confirm it remains to be carried out.

[4] In some cases the Actor argument may be ellipsed altogether when co-referential with an Actor argument in the previous clause. This does not alter the fact that the SVC has a single Actor argument.

asymmetrical SVCs. The grammaticalisation processes found in Klon are either aspectual in nature or involve the item *mi*. In all cases it is the syntactic positioning and semantic structure of the verbs that have enabled them to be reanalysed into grammatical items. The verb *agai* 'go, reach' has grammaticalised to mark perfect aspect, and the verb *yeh* 'exist' has grammaticalised to mark continuative aspect (§4.1). The verb *mi* 'to be at, to place' occurs in two types of asymmetrical SVCs: placement SVCs and locational SVCs. From its different functions and positioning within each of the SVCs *mi* is grammaticalising in two different directions: firstly, to become an applicative prefix (§4.2.4) and secondly to occur in temporal expressions (§4.2.5).

2 Symmetrical SVC Types

Symmetrical SVCs consist of verbs that are not restricted to a particular semantic or syntactic sub-type. The three main types of symmetrical SVCs found in Klon are sequential SVCs, manner SVCs, and parallel SVCs. Cross-linguistically symmetrical SVCs are prone to lexicalisation, as is the case in Klon (see Baird 2008:143 for a discussion of this phenomenon).

In sequential SVCs the event is divided into sub-events denoted by separate verbs. The order of the verbs is iconic, following the temporal sequence of sub-events. In example (1) there are two sequential SVCs; in the first SVC the sub-event of 'going home' occurred before the sub-event of 'money being taken', and in the second SVC the sub-event of 'going to the market' occurred before the sub-event of 'buying fish'. When the verbs are non-contiguous, separated by the Undergoer argument of the second verb, sequential SVCs typically only contain two verbs, as in example (1), whereas in contiguous sequential SVCs it is possible to have more verbs. Non-contiguous sequential SVCs are more common.

(1) *Ga alah agai doi méd*[5] *koih*[6] *ga unu agai ibiq qel.*
3ACT home go money take finish 3ACT market go fish buy
'He went home and got money then he went to the market and bought fish.'

In manner SVCs the first verb in the serial construction describes the manner in which the other verb(s) is/are executed, as can be seen in example (2), which contains the serial verbs *kdad adapu* 'quickly cook'.[7] The verbs in this type of SVC are always contiguous.

---

[5] Orthographically <é> is used to represent a close-mid front vowel phoneme, <ò> is used to represent a open-mid back vowel phoneme, and double vowels (eg. <aa>) are used to represent long vowels.

[6] The form *koih* 'finish' is from the Klon Paneia dialect, while the form *koh* 'finish' is from the Klon Bring dialect. In this example, and in other examples in this chapter, *koih* 'finish' does not form a part of the SVC, but is used as a discourse marker, intonationally separate from the SVC.

[7] Although the verb *kdad* can be used meaning 'be quick' when used in mono-verbal clauses, it is more typically used with the meaning 'to shock' in mono-verbal clauses and 'be quick' in SVCs.

(2) *Gen i koh ini awa kdad adapu bo g-eh.*
until DUR finish 3ns[8] again be.quick cook SEQ 3U-feed
'After that they quickly cooked then fed him.'

Parallel SVCs contain pairs of verbs that have a special semantic relationship, such as being (near) synonyms, antonyms, or activities that are somehow seen as typically co-occurring. This is an extremely common type of SVC in all types of Klon speech genres. Parallel serialisations consist of two verbs, which are always contiguous to one another. The verbs in the verb complex are strictly ordered, so, for example, it would not be possible to substitute '*tinghal toara*' for '*toara tinghal*' in example (3). The unalterable ordering of the verbs is typical of all parallel SVCs, which indicates that this SVC type has lexicalised (see Baird 2008: 143).

(3) *Hok pi yo ara tin-ghal yej nang,* kalo[9]
IRR 1nsi.ACT that issue REC-wrong able NEG if

to-ara tin-ghal yo
REC- issue REC- wrong that
'We're unable to make problems with each other, if we make problems with each other (make issue with each other, wrong each other)'

ho pe pa t-ebeer elel.
SIM 1nsi.POSS 1nsi.HORT 1nsi.U-die search
'we're searching for suicide.'

3 Asymmetrical SVC Types

One of the verbs found in Klon asymmetrical SVCs must be of a certain semantic or syntactic sub-type. The five main types of asymmetrical SVCs found in Klon are: directional SVCs, modal SVCs, instrumental SVCs, placement SVCs and locational SVCs. Cross-linguistically asymmetrical SVCs are prone to grammaticalisation. The grammaticalisation of Klon asymmetrical SVCs will be addressed below.

Directional SVCs are SVCs that contain at least one verb which is used to indicate the direction of the event denoted by the SVC. This can be seen in example (4), in which the verb *agai* 'go' indicates motion away from the deictic centre, and in example (5), in which the verb *ma* 'come' indicates motion towards the deictic centre.

(4) *Nok bo, gi-odoin orok ini ge kuur g-oj*
good SEQ 3POSS-brother two 3ns 3POSS dog 3U-call.dog
'Right, so her two brothers called their dog

[8] When a language has dual forms the term 'dual' should be used for reference to two people and 'plural' then implies a number greater than two people. In Klon, the non-singular forms co-occur with the dual forms. Hence, they may indicate either two or more people. Therefore, the term 'non-singular' is a more appropriate label than 'plural' in Klon.

[9] Words in normal font rather than italics indicate words not of Klon origin. Typically such words (like *kalo* 'if' in this example) are from Malay.

bo ini a lam agai koor.
SEQ 3ns 3RES walk go hunting
then they walked going hunting.'

(5) *Buwembui u-huih:*[10] *'Ah hok ngan hok nang', ho borkak ool*
B. VAL-say ah IRR thing IRR NEG SIM greedy woman

ting ma:
jump come
'Buwembui said "Ah it's alright" when the greedy woman jumped out:

'Ah Yap Umemenem e-m~mnem=e ne-m~mnem?'
ah Yap Umemenem 2S.U-RDP~perfumed=DIS 1S.U-RDP~perfumed
' "Ah Yap Umememen are you sweetly perfumed or am I sweetly perfumed?"'

Modal SVCs typically consist of two verbs, the final verb in the serial complex being one of the modal verbs *yaah* 'unable' or *inok* 'able'. Example (6) contains the modal SVC *gemod yaah* 'unable to climb it'. In example (7) *inok* 'able' is used in two modal SVCs – *taan inok* 'able to sell' and *mihok inok* 'able to become' – and is also used in a mono-verbal clause.

(6) *Bo wed i qad, o tok yong qada i mteh ong,*
SEQ now DUR come ah palm this IPFV DUR stand this
'So up until now this palm is still standing here,

ge-mod yaah.
3U-climb unable
unable to be climbed.'

(7) *Ongo lui me~meh ge ih ho taan inok,*
this chilli RDP~betel.vine 3POSS fruit SIM sell able
'This is wild betel vine, its fruit can be sold,

ga doi *mihok inok,*
3ACT money become able
we can turn it into money,

ge irik wei ong onon pi daqan lel gten
3POSS root leaves this PL 1nsi.ACT medicine medicine make
we make medicine from its roots and leaves

de pi naaq kde di inok.
CONJ 1nsi.ACT drink eat also able
and we can also eat them.'

[10] The form *huih* 'say' is from the Klon Paneia dialect, and the form *huh* 'say' is from the Klon Bring dialect, as used, for example, in (15).

Instrumental SVCs contain the verb *puin* 'hold', which adds an instrumental argument to a clause.[11] These SVCs may be either contiguous or non-contiguous, depending on the argument structure of the other verb(s) (not *puin*) in the verbal complex. Positionally *puin* 'hold' always precedes the other verbs in the serial complex, as can be seen in example (8).

(8) *Yo bui we puin g-tet, ata leng,*
that betel.nut spit hold 3U-massage coconut coconut.cream
'(In) that (case) I use betel-nut spit to massage, coconut cream,

kook tbok bisa *bui we bo puin g-tet di inok.*
break break able betel.nut spit SEQ hold 3U-massage also able
breaks can also be massaged with betel-nut spit.'

The other way of adding instrumental arguments to a clause is by using applicative *mi-* (§4.2.4).

There are two other types of asymmetrical SVCs in Klon, both of which contain *mi* 'be at' or 'place'. Placement SVCs contain the verbs *ma* 'come' and *mi* 'be at, place', which always occur in that order. Each verb takes a different Undergoer argument, with the Undergoer argument of *mi* positioned between the two verbs. In locational SVCs the verb *mi* 'be at' always precedes the other verbs in the verbal complex. These types of SVCs will be discussed together with the other verbal uses of *mi* in §4.2.3.

4 Grammaticalisation of asymmetrical SVCs

As mentioned, components of asymmetrical SVCs tend to grammaticalise over time. There are four examples of this in Klon:

1. the perfect aspect marker *agai* being grammaticalised from the verb *agai* 'to go, reach';
2. the continuative aspect marker *yeh* being grammaticalised from the verb *yeh* 'exist';
3. the instrumental prefix *mi-* being grammaticalised from the verb *mi* 'be at, place'; and
4. the verb *mi* 'be at, place' is in the process of grammaticalising to become a postpositional-type constituent in temporal expressions.

4.1 Klon aspect markers *agai* and *yeh*

Perfect aspect '…indicates the continuing present relevance of a past situation' (Comrie 1976:52). In Klon the aspect marker *agai* expresses this aspect, as can be seen in (9) and (10). The perfect aspect marker immediately follows the predicate it has scope over.

[11] Currently the verb *puin* 'hold' retains its meaning within SVCs and has not (yet) grammaticalised to become an instrumental marker. When used in mono-verbal clauses *puin* 'hold' can take theme or patient arguments, but when used in SVCs it typically takes instrumental arguments. There is one example in the corpus of *puin* taking a theme argument in a SVC, which can be seen in example (12) below.

(9) *Bo gan ehek yo igi n-en agai*
SEQ 3ACT place that 2ns.ACT 1s.U-give PFV
'So you have given me a place

de bo nab araa bo na naaq?
CONJ SEQ what water SEQ 1s.ACT drink
but what water will I drink?'

(10) *Tkin! Himbur* kot *breh agai, aa tbal hah ik, bo tkin!*
run H. city tear PFV fence collapse fall COMPL SEQ run
'Run! Himbur city has been torn down, the fence has fallen so run!'

The perfect aspect marker *agai* has transparently been derived from the verb *agai* 'go, reach'. When used verbally *agai* 'go, reach' can be used in either mono-verbal clauses or SVCs. In example (11) *agai* is being used as the sole verb in a transitive clause, and in example (12) it is used in a directional SVC.

(11) *Ini dger agai, ge-lam.*
3ns border.between.coast.&.farmland go 3U-walk
'They went to the area between the coast and the farmland, walking there.'

(12) *Ini ge ininok nuk g-ebeer,*
3ns 3POSS person one 3U-die
'They killed one of their people,

gi-to g-tak non ini puin agai atal o
3POSS-head 3POSS-leg p 3ns hold go above that

Totomibak eben.
Totomibak village
they went up with his head and legs to Totomibak village.'

It is probable that the use of *agai* in certain SVC constructions lead to its reanalysis as an aspect marker. Synchronically, there are still contexts in which, despite the meaning of the utterance being clear, it is not always apparent whether *agai* is being used as a verb in a directional SVC or as an aspect marker. In example (13) *agai* is used three times. In the first instance *agai* is clearly being used as a verb meaning 'go', and in the third instance it is clearly being used as a perfect aspect marker. However, the second occurrence of *agai* in this example could be the verbal *agai*, forming a directional SVC with *gelam* 'walk there' meaning 'went walking there'; or it may be the perfect aspect marker marking that the 'walking there' has been done. Although not indisputable, it is more likely in this instance that *agai* is being used as a perfect aspect marker, because, presumably, if it was part of a SVC *agai* 'go' would share the Undergoer argument (3rd person *ge-* 'it, there') with *lam* 'walk', yet it is not marked (see defining characteristic 6. in §1).

(13) *Bo ni lam, ni agai u-hiid,*
SEQ 1nse.ACT walk 1nse. ACT go VAL-reach
'So we walked reaching (there)

ho jam *nuk ge-lam* *agai* *yo, eteq yo ini* sengsor *agai,*
SIM time one 3U-walk ??? that wood that 3ns chainsaw PFV
in an hour, walked there, they had already chain-sawed the wood,

eteq kak o a=ubei nah, bo hos yeh.
wood board that INTS=many very SEQ place exist
there were very many planks of wood placed there.'

In summary, *agai* can be used:

1. as a verb, in mono-verbal clauses;
2. as a verb, in directional SVCs to indicate direction away from the deictic centre; and
3. as a perfect aspect marker, following predicates.

There is a tendency for *agai* to precede other verbs when used in a serial verb complex. When following verbs it appears *agai* has begun to grammaticalise from verb to aspect marker. However, this process is incomplete and ambiguous cases remain in which it is unclear whether *agai* functions as a verb or aspect marker.

The Klon continuative aspect marker *yeh* has been derived from the existential verb with the same form. The grammaticalisation path from existential verb to continuative or progressive aspect marker is cross-linguistically not uncommon, particularly in East Nusantara. For example this phenomenon is also found in Pidgin Malay Derived varieties [such as Alor Malay, (Baird, *et al* 2004)], in which the existential *ada* is used to indicate progressive aspect (Adelaar and Prentice 1996).

When used as an existential verb, *yeh* behaves the same way as other verbs in mono-verbal clauses. For example, *yeh* 'exist' is negated in the same way as other verbs, as can be seen in (14).

(14) *Ga u-huih: 'Labegei ge kdeh ong hok* *yeh* *nang ong.'*
that VAL-say Labegei 3POSS head this IRR exist NEG this
'He said: "Labegei's head isn't here."'

Continuative aspect indicates an ongoing state or situation. *Yeh* is continuative, rather than progressive, because unlike progressive aspect markers, which can only be used with non-stative situations (Comrie 1976:51), *yeh*, when used aspectually, can follow both verbs denoting states, as in (15) and dynamic situations, as in (16).

(15) *Om qad u-huh abang: 'E ul ool kaklok agai,*
husband come VAL-say say 2S.POSS CLF woman give.birth PFV
'The husband came saying "Your woman has given birth,

ul okne aal nuk, lale *taa yeh.'*
child woman big one below sleep CONT
[to] a large girl, [she's] sleeping down there."'

(16) *Mteh dgim di,* *lem yeh* *yo, wa* bapa, *he-eh nok bo*
stand strong first shake CONT that wow father ha-ha good SEQ
'Stand strong first, [it's] shaking, wow Dad, ha-ha okay so

her bapa, *Meos oyor Meos.*
descend father Meos move.aside Meos
come down Dad, Meos move aside.'

It is frequently not altogether clear whether *yeh* is an aspect marker or the existential verb occurring in SVCs. Synchronically, it is analysable as both. There are many instances in which it is clear that *yeh* is being used as an aspect marker, as in examples (15) to (16) above. However, there are also ambiguous instances, as can be seen in example (17).

(17) *Hu-u ele lam yeh.*
uh-huh 3d walk ???
'Uh-huh those two (are there) walking.'

Although there are ambiguous occurrences of both *agai* and *yeh* in the corpus, in which it is not always clear whether they are functioning as verbs in a serial complex or as aspect markers, in both cases the one form only has two functions and the path from verb to aspect marker is transparent. The case of the form *mi* is not as clear, as will be discussed in the next section.

4.2 Grammaticalisation of *mi*

4.2.1 Overview

Klon *mi* has cognates in many of the languages spoken in the Alor archipelago. Cognate forms typically fill the roles of being a verb or a postposition, but may have multiple functions. For example: in Blagar (on Pura) *mi* is used as both a postposition and a verb (Steinhauer p.c.); in Adang (on Alor) the form *mi* is used as both a comparative marker and as a locative verb meaning 'in, at' (Haan 2001); in Abui (on Alor) *mi* is a verb used to refer to location or motion towards a deictic centre (Kratochvíl p.c.); in Kaera (on Pantar) *mi* is a locative postposition which has a cognate verbal form *ming* 'be located at X' (Klamer this volume); and in Western Pantar (on Pantar) the form *me* is a postposition. Although *me* can be used verbally this use is considered 'weird' (Holton p.c.). The form *mi* in Klon is used as a comparative, as a verb in both mono-verbal clauses and two types of serial verb constructions, as an applicative prefix and in temporal expressions.

4.2.2 Comparative *mi*

The use of *mi* as a comparative prefix attached to adjectives, can be seen in example (18).[12] This comparative use of *mi* is synchronically unrelated to the other uses of *mi*.

(18) *Mleng ni ge eteq aan,*
yesterday 1nse.ACT 3POSS wood carry
'Yesterday we carried his wood,

[12] This example contains reduced NPs, in which the head noun has been ellipsed leaving behind modifiers – in this case adjectives – that refer to it. This more commonly occurs with demonstratives or numeral and classifier combinations than adjectives (see Baird 2008:88).

mi-tuang	*ni*	*aan*	*ik,*	*mi-ubei*	*qada.*
COMP-a.little	1nse.ACT	carry	COMPL	COMP-much	IPFV

we've carried the lesser (of it), (there's) still the larger (amount).'

4.2.3 Verbal mi

Verbal *mi* can occur in mono-verbal clauses, and as a part of placement and locational SVCs. When used verbally *mi* means 'be at' or 'to place'. The difference in meaning is based on the animacy of its arguments. Semantically, verbal *mi* always takes a THEME and a LOCATION as its arguments. The THEME may be either animate or inanimate. When the THEME is animate, *mi* occurs in transitive clauses which have the structure '(THEME) Y is at (LOCATION) Z', the THEME is realised as an Actor argument and the LOCATION is realised as an Undergoer argument. When used in this way *mi* is translated as 'be at'. An example of this can be seen in (19).

(19)

Ni	*qad*	*ho*	*gan*	*alah*	*mi.*
1nse.ACT	come	SIM	3ACT	house	be.at
			A/THEME	U/LOCATION	V

'We came and he was at the house.'

If the THEME argument is inanimate a third animate argument (an AGENT) is semantically required by *mi*, resulting in: '(AGENT) X makes (THEME) Y be at (LOCATION) Z'. There are two ways in which Klon speakers syntactically deal with the semantic argument structure of *mi* when the THEME argument is inanimate: firstly, they may create ditransitive clauses; or secondly, they may use SVCs.

Although a predicate may semantically require three arguments, Klon has very few syntactically ditransitive clauses.[13] I found only six instances in the corpus of *mi* being used in ditransitive clauses.[14] In three of these the bare form *mi* was used, while in the other three *mi* was prefixed by valence-increasing prefix *u-*. When occurring in the ditransitive clauses with bare *mi*, the AGENT argument is realised as an Actor, the LOCATION is realised as a Primary Undergoer (PU), positionally immediately preceding the verb, and the THEME argument is realised as a Secondary Undergoer (SU).[15] This can be seen in example (20). This contrasts with example (21) (taken from example (32) below) in which the valence increasing prefix *u-* is used to rearrange the mapping relationship between semantic roles and Undergoers. The AGENT argument is again realised as an Actor, but the THEME is realised as a Primary Undergoer (PU), positionally immediately preceding the verb, and the LOCATION argument is realised as a Secondary Undergoer (SU).

(20)

Ge	*huud*	*yo*	*di*	*ga*	*gi*-bad	*hok*	*mi,*
3POSS	spoon	that	also	3ACT	3POSS-clothes	pocket	place
	SU/THEME			A/AGENT	PU/LOCATION		V

'He also put his spoon in his clothes' pocket,

---

[13] There is only one verb *en* 'to give' which is consistently realised ditransitively.

[14] Note that three of these instances were produced by a non-native Klon speaker (originally from the nearby island of Pura). This fact, coupled with the low incidence of such clauses perhaps brings into question the acceptability of such utterances.

[15] The terms Primary and Secondary Undergoer are adapted from Dryer's terms Primary and Secondary Object (Dryer 1986). The Primary Undergoer in a ditransitive clause is morpho-syntactically treated in the same way as an Undergoer in a transitive clause (see Baird 2008:34).

ge	*kol*	*lab*	*yo*	*di*
3POSS	tube.for.crushing.betel.nut	chisel	that	also
	SU/THEME			

he also put his tube for crushing betel nut and chisel

ga	*gi*-bad	*hok*	*mi.*
3ACT	3POSS-clothes	pocket	place
A/AGENT	PU/LOCATION		V

into his clothes' pocket.'

(21)

Ga	*ge-uur*	*kwet*	*yo*	*ihi=e*	*u-mi.*
3ACT	3U-see	basket	that	faeces=FOC	VAL-place
A/AGENT	SU/LOCATION	PU/THEME	V		

'She saw the basket was full of faeces.'

The second, more commonly employed, construction used when the THEME argument of verbal *mi* is inanimate, is a placement SVC. An example of this can be seen in (22), with an ellipsed Actor / AGENT argument. Placement SVCs[16] contain the verbs *ma* 'come' and *mi* 'be at, place'.[17] Syntactically, each of these verbs within the SVC is transitive, taking an Actor argument and an Undergoer argument. The verb *ma* 'come' takes the THEME as its Undergoer argument and the verb *mi* 'be at, place' takes the LOCATION as its Undergoer argument. Both Undergoer arguments precede their verbs, and so the verbs are non-contiguous. As with all Klon SVCs, the Actor argument is shared. Examples of placement SVCs can be seen in (22) to (24). Note the Actor argument has been ellipsed in (22).

(22)

Go-ma	*kwet*	*mi.*
3U-come	basket	place
U/THEME	U/LOCATION	

'Come with it [and] put [it] in the basket.'

(23)

Pi	*hol*	*tak,*
1nsi.ACT	split	pieces

'We split pieces

bo	*pi*	*meh*	*ma*	*t-ad*	*mi,*
SEQ	1nsi.ACT	betel.vine	come	1nsi.POSS-mouth	place
	A	U/THEME		U /LOCATION	

and we bring the betel vine and put it in our mouths,

---

[16] Placement SVCs could potentially be regarded as being symmetrical because of the sequential nature of their sub-events. They are classified here as asymmetrical, because, as with the other Klon asymmetrical SVC types, they contain verbs 'of a certain semantic or syntactic sub-type'. Whereas there are no 'rules' regarding the verbs that may occur in sequential SVCs, placement SVCs always contain particular verbs (*ma* 'come' and *mi* 'place'), which, as noted, is a feature typical of asymmetrical SVCs.

[17] See Klamer (this volume) for a discussion of the various functions and grammaticalisation of *ma* 'come' in the two Pantar languages Teiwa and Kaera. Note that in Klon *ma* 'come' is only used as a verb, and is not used as an oblique marker.

koh ewer ma tang ho i kiik yo.
finish chalk come above SIM DUR red that
then comes the chalk and it becomes red.'

Placement SVCs are not restricted to containing just the verbs *ma* 'come' and *mi* 'be at, place'. Although these verbs are always present, other verbs may also occur in the verbal complex, as in (24), which contains the four verbs *gtal* 'lift', *méd* 'take', *ma* 'come' and *mi* 'be at, place'.

(24) *Ini at gtal méd ma hok*[18] *mi.*
3ns bamboo.spikes lift take come small.basket place
A U /THEME U /LOCATION
'They took bamboo spikes and put them in a small basket.'

Mi 'be at, place' also occurs in locational SVCs. The use of *mi* 'be at, place' in these SVCs allows for a LOCATION argument to be added to an otherwise intransitive clause as an Undergoer. *Mi*'s behaviour in locational SVCs is quite different to that found in placement SVCs. The verbs in locational SVCs are always contiguous, with *mi* preceding the other verb(s). In (25) the verb *kukui* 'play' occurs alone in an intransitive imperative clause, while in (26) it occurs in a locational SVC containing the verb *mi* 'to be at', which licenses the LOCATION Undergoer argument *lapang ong* 'this field'.

(25) Lapang *ong pi al go-aran bo kukui, tinggen nang.*
field this 1nsi.ACT all 3U-cut SEQ play fight NEG
'We all cleared this field, so play, don't fight.'

(26) *Eh yo yo nang=e!* Lapang *ong pi al mi kukui.*
hey that that NEG=DIS field this 1nsi.ACT all be.at play
'Hey don't be like that! This field is for all of us to play on.'

In example (27) the locational argument *wòr* 'rock' is introduced by the use of *mi* in a locational SVC.

(27) *Nabe nuk gten mlir ong awa iqes, nang bo*[19] *lega wòr*
what one do eel this again live NEG SEQ 3S.TOP rock

mi obon bo a imih.
be.at block SEQ 3RES stay
'Something made the eel come back to life, so he stayed hidden behind a rock.'

Based on its locative-like semantics and the configuration of constituents in mono-verbal clauses and locational SVCs, *mi* may initially appear to be a postposition forming a postpositional phrase with the preceding NP. However, *mi* is analysed as a verb rather than a postposition in these contexts, because the argument plus *mi* do not behave as a

[18] The form *hok* is heterosemous, meaning 'pocket', 'small basket' and 'half'; it is also the irrealis marker.

[19] The negator and sequential marker (*nang bo*) or sometimes the negator alone (*nang*) are used as discourse markers, separating different sections within discourse, and are typically translated with 'so' or 'then' (Baird 2008:161).

single syntactic constituent or form a phonological phrase, rather *mi* forms a syntactic unit with the following verb. Examples (28) to (30) illustrate how *mi* and its argument *lale Hwak weer* 'below Hwak river' do not form a single syntactic constituent. In (29) the argument of *mi* can be moved, like other Undergoer arguments, while the verb complex remains clause-final. Example (30), in which *mi* is moved together with the NP and separated from the verb *taa* 'sleep' can only be grammatically read as containing two clauses. *Mi* is the sole verb in the first mono-verbal clause, without an overtly stated Actor argument. This first clause would mean '(they) were below at Hwak river'. The second clause contains the new SVC *gen agai taa*, translatable as '(they) eventually went to sleep'.[20]

(28) *Ini gen agai lale Hwak weer mi taa.*
3ns until go below H. river be.at sleep
'They went until below Hwak river sleeping there.'

(29) *Lale Hwak weer ini gen agai mi taa.*
below H. river 3ns reach go be.at sleep
'They went until below Hwak river sleeping there.'

(30) *Lale Hwak weer mi. Ini gen agai taa.*
below H. river be.at 3ns reach go sleep
'[They] were at below Hwak river. They eventually went to sleep.' (*They went until below Hwak river sleeping there.)

Further evidence to support the claim that *mi* forms a part of a serial verb complex is that both placement SVCs and locational SVCs fulfil all of the characteristics of Klon SVCs, listed in §1 above.

4.2.4 Applicative *mi-*

Klon contains two valence-increasing operators, both of which are used to introduce an Undergoer argument into a clause. Applicative *mi-* typically introduces Undergoer arguments with the semantic role of INSTRUMENT, while the general valence-increasing prefix *u-* introduces Undergoer arguments with a much wider range of potential semantic roles, such as PATIENT, RECIPIENT, GOAL or THEME. Applicative *mi-* only prefixes to verbal bases, while the general valence increasing prefix *u-* may occur on a variety of bases.

When occurring in locational SVCs, *mi* always precedes the other verbs in the construction. It is probably due to this positioning that it has grammaticalised, becoming a verbal prefix. However, unlike when used verbally, the prefix *mi-* typically introduces instrumental arguments or arguments with other typically oblique semantic relations, rather than locational ones.

Mi, because it is a phonetically small unit, attaches prosodically to the constituent that follows it. The only instances in which *mi* does not occur in the same phonological phrase as the following constituent, and hence does not prosodically attach to it, is when *mi* occurs in temporal expressions (see §4.2.5). In locational SVCs *mi* is treated as a separate unbound constituent, because its behaviour is verbal, and there is no phonological reason

[20] The presence of the verb *gen* 'reach' indicates that there was some effort involved in the endeavour, hence it is translated with 'eventually'.

to believe that it might be bound. However, when schwa-initial verbs are prefixed by applicative *mi*, the schwa is deleted, providing evidence that *mi* in this context is indeed bound. For example *mi-eweel* 'APPL-bathe' is realised as [miwe:l]. This contrasts with the verb's phonetic realisation when preceded by a phonetically similar unbound item: *ni eweel* '1pe bathe' realised as [ni əwe:l].

Syntactically, the applicative prefix *mi-* introduces an Undergoer argument into the clause (just as verbal *mi* 'be at' does in locational SVCs). In some instances this results in transitive clauses, as in (31), while in other cases it results in ditransitive clauses, with the introduced argument taking the grammatical relation of Secondary Undergoer, as in (32). As with all Undergoer arguments, the Undergoer argument of an applicative verb may be ellipsed when understood from previous discourse or shared knowledge, as in (33).

(31) *Ni* *nger* *elel* *knai* *elel,*
1nse.ACT candle.nut search cenari.nuts search
'We search for candle nuts and cenari nuts

ni *<u>mi</u>-gbon* *<u>mi</u>-ghek,*
1nse.ACT APPL-roast APPL-dry.in.sun
we roast them, dry them in the sun,

ni *eben* *agai* *taan* *kde,*
1nse.ACT village go sell eat
we go sell them in the village,

doi *<u>mi</u>-tel* seng *<u>mi</u>-tel.*
money APPL-exchange money APPL-exchange
exchange money.'

(32) *Bo* *kwet* *op* *ga* *kbak* *<u>mi</u>-tpan,* *bo* *ga* *u-ilin,*
SEQ basket that 3ACT spear APPL-stab SEQ 3ACT VAL-lick

ho *mkal,*
SIM bitter
'Then she stabbed the basket with a spear and licked it, it was bitter

bo *ga* *ge-uur,* *kwet* *yo* *ihi=e* *u-mi.*
SEQ 3ACT 3U-see basket that faeces=FOC VAL-place
then she saw the basket was full of faeces.'

(33) *A* *naaq* *a* *kde=we* *a* *<u>mi</u>-eweel* *a* *<u>mi</u>-ruh.*
2s.ACT drink 2s.ACT eat=DIS 2s.ACT APPL-bathe 2s APPL-massage
'You drink (it)you eat (it), you use (it) to bathe, you use (it) to massage.'

We can see the difference between verbal *mi* in a locational SVC and applicative *mi-* in the following examples, using the verb *uur* 'to see'. In (34) the verb *uur* 'to see' is used intransitively, and in (35) it is used transitively with an Undergoer prefix.

(34) *Bo rung iri bo ni uur, ho doqom=e mung,*
SEQ bang sound SEQ 1nse.ACT see SIM grandfather=FOC fall
'Then bang and we looked and grandfather had fallen

nang bo ni tkin ma, bo ni go-puin.
NEG SEQ 1nse.ACT run come SEQ 1nse.ACT 3U-hold
so we ran and we held him.'

(35) *Nang ogol ple adaq ge-uur.*
NEG beginning 1di still 3U-see
'So to start we'll still look at it (leaves).'

In the following two examples, (36)[21] and (37), the verb *uur* 'to see' is preceded by *mi*. When co-occurring with *uur* 'to see' we know that one use of *mi* is verbal and the other applicative. When used as an applicative prefix attached to the verb *uur* 'to see' an instrumental argument is introduced, as in example (36). This contrasts with the use of *mi* in the locational SVC in (37), in which *mi* is verbal with a locational argument.

(36) *Na* kacamata *mi-uur.*
1s.ACT glasses APPL-see
'I see with glasses.'

(37) *Nang bo ga ge-mod agai, hek po go- hkek,*
NEG SEQ 3ACT 3U-climb go door that 2U- open
'So she climbed up and opened the door

bo ga oot mi uur, ho Buwembui=we qad,
SEQ 3ACT room be.at see SIM B.=FOC come
and looked into the room, Buwembui had come

bo ga Yap Umemenem ong a g-lain yeh.
SEQ 3ACT Yap Umemenem this 3RES 3U-seduce CONT
and was seducing Yap Umemenem.'

4.2.5 Temporal expressions with mi

In all of the uses of *mi* that we have seen thus far, *mi* forms a part of a phonological phrase with the following constituent (a verb), rather than any preceding constituents, such as an NP. However, *mi* can form a phrase – both phonologically and syntactically – with a preceding constituent, when that constituent is a temporal adverb, forming temporal expressions. The precise status of *mi* in this context is unclear. It behaves in a similar way to a postposition, but instead of having a nominal argument takes an adverbial complement. Temporal expressions containing *mi* may occur in multiple slots within a clause, typically either clause-initially or pre-verbally.

The meaning of *mi* in temporal expressions can be considered a semantic extension from its verbal use. As seen, when used verbally *mi* means 'be at' taking a locational

[21] This is an elicited sentence.

argument. This meaning 'be at' appears to have been extended from referring solely to locations ('at location Z') to temporal concepts ('at time A').

Examples (38) to (40) illustrate temporal expressions containing *mi*: *makna mi* 'in the past', *minuk mi* 'at one time' and *nuk mi* 'once' respectively, in different positions within clauses.

(38) *Ni-man* *n-oi* *non* *ah*
1s.POSS-father 1s.POSS-mother PL ah
'Ladies and gentlemen ah

ul *òm* *gi-eqeben,* *ah* *mi-bong* *heb,*
CLF older.sibling 3POSS-eldest ah COMP-youth young
the eldest sibling is now younger,

ah *wed* *pi* *qad* *ma* *ongo,*
ah now 1nsi.ACT come come this
we come here

makna mi *lahtal* *ta* *mi* *tengtang* *mi~mih* *t~t-ebeer.*
past MI God above be.at fate RDP~sit RDP~1nsi.U-die
from the past God above decides our fate.'

(39) *Minuk mi* *ini* *a* *weer* *agai,*
one.moment MI 3ns 3RES river go
'One time they went to a river,

ini *her* *gi-tbur* *elel* *gi-ahkol* *elel.*
3ns descend 3POSS-crab search 3POSS-shrimp search
to search for their shrimp and crabs.'

(40) *Ge-agai* *nuk mi* *heng-heng* *bo* *qad.*
3U-go one MI rushed SEQ come
'Once someone approached them in a hurry.'

In summary, from use of the verb *mi* 'to be at' in locational SVCs, *mi* appears to have grammaticalised into an applicative prefix and a postposition-like element marking temporal expressions. As an applicative prefix, semantically the concept of location has been extended to instruments and other oblique-like semantic relations. By prosodically attaching to the preceding item rather than the following item *mi* is taking on postpositional-like qualities, forming temporal expressions.

As we have seen, there are frequent instances in the corpus where there is either no Actor argument in a clause containing verbal *mi* or the Actor argument is shared with another verb, thus leaving only the locational argument. Currently *mi* remains verbal in such clauses (and indeed they are clauses and not phrases), but it could conceivably grammaticalise into a postposition in this context, as it apparently has done so in other languages in the Alor archipelago.

5 Summary and conclusion

As seen, eight types of productively used SVCs can be identified in Klon: the three symmetrical types of sequential SVCs, manner SVCs and parallel SVCs; and the five asymmetrical types of directional SVCs, modal SVCs, instrumental SVCs, placement SVCs and locational SVCs. Although synchronically productive, some symmetrical SVCs are being lexicalised, while some asymmetrical SVCs are being grammaticalised. The process of grammaticalisation of the verbs *agai* 'go, reach' into a perfect aspect marker and *yeh* 'exist' into a continuative aspect marker, although incomplete, is a relatively transparent process. This contrasts with the grammaticalisation of the verb *mi* 'be at, place', which is simultaneously grammaticalizing in two different directions. The first grammaticalisation path from verbal *mi* 'be at' to applicative *mi-* has arisen through the use of verbal *mi* 'be at' in asymmetrical locational SVCs. The second grammaticalisation path from verbal *mi* 'be at' to postpositional-like *mi* 'at' in temporal expressions has arisen through semantic extension of *mi*'s meaning: from 'at a location' to 'at a time'.

References

Adelaar, K. Alexander and David J. Prentice, 1996. Malay: Its history, role and spread. In Stephen A. Wurm and Peter Mühlhäusler, eds, *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*, Vol. II, 1:673–693. Berlin/New York: Mouton de Gruyter.

Aikhenvald, Alexandra Y., 2006. Serial verb construction in typological perspective. In Robert M.W. Dixon, ed., *Serial verb constructions: A cross-linguistic typology*, 1–68. Oxford: Oxford University Press.

Baird, Louise, 2008. *A grammar of Klon: A non-Austronesian language of Alor, Indonesia*, Canberra: Pacific Linguistics.

Baird, Louise, Marian Klamer and František Kratochvíl, 2004. Alor Malay as a distinct variety of Malay. Paper presented at *ISMIL 8*, Penang, Malaysia.

Comrie, Bernard, 1976. *Aspect: an introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

Dryer, Matthew, 1986. 'Primary objects, Secondary objects and antidative, *Language*, 62:808–845.

Haan, Johnson Welem, 2001. The grammar of Adang: A Papuan language spoken in the island of Alor East Nusa Tenggara – Indonesia. Sydney: The University of Sydney. PhD dissertation.

Klamer, Marian, this volume. 'One item, many faces: 'come' in Teiwa and Kaera. In Michael Ewing and Marian Klamer, eds, *Typological and areal analysis: Contributions from east Nusantara*. Canberra: Pacific Linguistics.

10 *One item, many faces: 'come' in Teiwa and Kaera*

MARIAN KLAMER

1 Introduction

This chapter describes the many faces of the item *ma* 'come' in Teiwa (T) and Kaera (K), two non-Austronesian languages spoken on Pantar island.[1] We will see that *ma* functions as a deictic verb, as a change-of-state verb, marker of intentions or future tense, as a marker of hortatives and imperatives, as a conjunction that indicates that time has elapsed between subsequent events, and as an oblique marker that introduces secondary objects into the clause.

It will be argued that all these synchronic functions of *ma* are manifestations of one and the same lexical item. In the analysis of the various functions of *ma*, I distinguish between firstly, the level of lexical semantics; and secondly, the level of contextualised meanings, or pragmatics. At the level of lexical semantics, *ma* is analysed as a predicate with a single, semantically unspecified argument. Following general insights on the semantics of deictic verbs, I will argue that the lexical meaning of *ma* is actually not 'come', but rather something like 'move towards deictic centre'. At the lexical semantic level, *ma* consists of two basic semantic components: a motion, and a deictic component (cf. Talmy 1985, Wilkins and Hill 1995). The deictic component contains information on the motion's 'path' (towards/from) and its 'ground' (deictic centre).

We will see that the variable functions of *ma* mentioned above are all contextualised meanings that depend on the grammatical context for *ma*. They can be divided into three broad types, depending on the grammatical context of *ma* and the animacy value of its argument. The first type is found when *ma* is used as an independent predicate of a clause.

[1] I would like to thank the people of the village of Madar on Pantar island who hosted me and helped me collecting and analysing a corpus of Teiwa texts in 2003, 2004 and 2007. In particular, I would like to thank the following Teiwa speakers (in alphabetical order): Paulus Kay, Aser Pering, Seprianus Pering, Amos Sir and Florens Titing, whose narratives I recorded, and who assisted in the transcription and analysis of these texts. I also thank Marianus Waang, native speaker of Kaera, for contributing the Kaera data and for inspiring discussions on his language. Parts of this chapter were presented at the East Nusantara Conference held at Leiden University in the summer of 2005, and I would like to thank the audience for stimulating questions and discussion. Wayan Arka, Louise Baird, Bill Foley, Gary Holton, Beth Levin, Andrej Malchukov, and Bernhard Waelchli read and commented on drafts of this chapter, and their comments and suggestions have helped to shape it into its current form.

Michael C. Ewing and Marian Klamer, editors.
Typological and areal analyses: contributions from East Nusantara, 203-225. Pacific Linguistics, 2010
© This edition vested with Pacific Linguistics.

In such contexts, it is interpreted as a deictic verb when the argument is animate, and as a change-of-state verb when the argument is inanimate. However, as soon as it occurs in serial verb constructions (SVCs) *ma* is found in other functions. One class of such functions is when *ma* has an **animate** argument and occurs in an SVC: it is then interpreted as a tense/mood marker, or as a conjunction. When *ma* has an **inanimate** argument and occurs in an SVC, however, *ma* has another type of function altogether: it now marks secondary objects as obliques. Semantically, these obliques are varied, and include locations, instruments, and displaced themes. Their interpretation depends on the semantics of the major verb in the SVC.

In general, I will argue that although *ma* has many different faces, we are dealing with one and the same 'animal' all the time. The many faces of this animal reflect how its interpretation can shift when its grammatical context changes. The following two contextual factors will be identified as crucial in the shifting interpretations of *ma*. Firstly, the animacy value of its argument, and secondly, the fact whether or not *ma* appears in an SVC, that is is followed by a major verb. In addition, Teiwa and Kaera also have certain particular grammatical characteristics which enable the reinterpretation of *ma*. (See §3.2 and §3.3).

The structure of this chapter is as follows. To understand the grammatical factors involved in the interpretations of *ma*, §2 presents an overview of the grammar of Teiwa and Kaera. After discussing clausal structure (§2.1), I describe how primary objects (§2.2) and secondary objects (§2.3) are encoded and introduce the deictic verbs and serial verb constructions (§2.4). Readers who prefer to move straight to the core of the chapter can just refer to §2.5 which contains a summary of the relevant grammatical observations.

In §3 the various functions of the item *ma* are described and analysed. First it is proposed how the lexical semantics of *ma* may be represented, and how they are to be distinguished from its contextualised meanings (§3.1). Then the various functions of *ma* in relation to the animacy of its argument are discussed: when the argument is animate (§3.2), and when it is inanimate (§3.3). In both cases, I suggest an analysis where the lexical semantics of *ma* are reinterpreted under influence of its structural context. Section 4 sums up the conclusions.

2 Grammatical overview of Teiwa and Kaera

2.1 Clausal structure

Teiwa (T) and Kaera (K) are two closely related non-Austronesian languages, spoken on Pantar island, which is located just north of Timor island, in Eastern Indonesia. Teiwa is spoken in the north-west of Pantar by approximately 4,000 speakers, Kaera is spoken on the eastern coast by approximately 10,000 speakers. The Teiwa data were collected on site between 2003 and 2007, and the Kaera data come from research with a native speaker residing in the Netherlands in 2005 to 2006. At the time of writing, the Teiwa corpus contained approximately 1200 records of utterances, and the Kaera corpus about 400 records. A reference grammar of Teiwa is Klamer (2010).

Both languages have clause-final verbs, final conjunctions and final negations. Subjects and objects are preverbal. Both languages make extensive use of serial verb constructions. Nominal predicates do not occur with copulas. Neither Teiwa nor Kaera distinguish

between finite/non-finite verb forms, and tense and aspect are not grammatically marked.[2] These grammatical features are illustrated for Teiwa in the utterance in (1) and for Kaera in the utterance in (2). Example (1a) illustrates a subject with an intransitive verb, (1b) a verb with a subject and a patient marked as an oblique and a final conjunction, (1c) a serial verb construction (SVC) consisting of an intransitive and a transitive verb with a shared subject and a final conjunction and (1d) a negated nominal predicate.[3]

(1) a. *Karian pati sin amidan a wa*
work PROG first what 3s say
'[She] is just working, what..., she thinks

b. *haliwai la ga-ti*[4] *ma sii-n ba*
black.ant TOP 3s-bottom OBL bite-RLS SEQ
(it's) black ants that bite in her bum so

c. *a tup-an ga-nuan kiqax si maan,*
3s get.up-RLS 3s-cloth shake.out SIM NEG
she gets up, shakes out her sarong and no,

d. *haliwai dan axa'a maan*
black.ant part this.one NEG
these are not black ants.' [Teiwa]

Example (2a) contains an SVC with four verbs sharing one subject, the first three intransitive, the last one transitive, followed by a final conjunction. Example (2b) illustrates a transitive verb with a subject and object constituent and a final negation.

(2) a. *Gang wa-t urung mid ui gu lal-i asi*
3s go-MOD look.up ascend person that see-MOD but
'He looks up to watch the other person climb up,

b. *ui gu gang lal-i bino.*
person that s/he see-MOD NEG
but that person does not see him.' [Kaera]

The following examples are two Teiwa clauses with adverbs for time and manner, followed by three Kaera clauses. Time and manner adverbs occur as topicalised

---

2 Teiwa verbs may be inflected for realis, taking a suffix glossed as 'RLS' (see Klamer, 2010, for a description of the 'realis state' marking functions of this suffix, as well as its discourse functions). Kaera verbs can take inflectional suffixes which are provisionally glossed here as Modality 'MOD' suffixes.

3 The translations of the examples use present tense, unless the context indicates otherwise, though tense is not specified in Teiwa and Kaera. Orthographical conventions Teiwa: <q> represents a uvular stop, <x> a pharyngeal fricative /ħ/, and <'> a glottal stop. In Kaera, *x* represents a uvular fricative /x/ and <'> a glottal stop.

4 Possessor prefixes and object marking prefixes are homophonous in Teiwa, in Kaera, prefixes marking alienable possessors and object prefixes are identical. Similarity of object and possessor prefixes is found in all the languages of Alor and Pantar; Klamer and Kratochvíl (2006) discuss this for Teiwa and Abui.

constituents, as in (3), or precede the object and verb as in (4) to (7); they never occur in a clause-final position.

(3) *Iliar ga'an*[5] *u a mulai*[6] *gi te-tei wraak*
daybreak 3s DIST 3s begin go RDP-wood search
'That daybreak he begins to search for wood.' [Teiwa]

(4) *Iman kal-kalan wa ma palan si, ...*
they RDP-slowly go come inspect SIM
'They slowly came to inspect and...' [Teiwa]

(5) *Ui umux gang miaag la ma*
person woman 3s yesterday FOC come
'That woman came yesterday' [Kaera]

(6) *Ging kali-kali tei baxi gu wang ekeng...*
3p RDP-slowly wood branch that be.at climb
'Slowly they climb on that tree branch...' [Kaera]

(7) *Ilwang gang user-user bir bleling g-om*
Ilwang 3s RDP-quickly run outside 3s-inside

mi eser-it,...
OBL exit-MOD
'Ilwang quickly runs outside...' (lit. '...runs exiting to outside's inside') [Kaera]

Note that sentences (6) and (7) also contain a location. In (7) it is expressed as an oblique constituent marked *mi.* (More examples of oblique constituents are discussed in §3.3.) Obliques always precede the verb, just like time and manner adjuncts. In (6), the location is encoded as the object of the transitive locative verb *wang* 'be at/with/near X'. Both Teiwa and Kaera have several such verbs; additional examples are Teiwa *me'* and Kaera *ming* 'be at location X' (cf. §3.3). In (8) and (9) these verbs are illustrated with the location *hafan/abang* 'village' as primary object:

(8) *Uy ga'an hafan me'-an ba aria'.*
person 3s village be.at-RLS SEQ arrive
'That person arrives from the village' (lit. '...is at the village then arrives')
[Teiwa]

(9) *Ui gu abang ming la da.*
person that village be.at CONJ ascend
'That person comes up here from the village' (lit. '...is at the village then comes up here') [Kaera]

---

5 The third person object pronoun *ga'an* has a secondary function as demonstrative pronoun. In adnominal function it marks a known/previously introduced entity.

6 From Indonesian/Malay *mulai* 'begin'.

2.2 The encoding of core objects

In order to compare the object marking patterns of Teiwa and Kaera, I analysed 54 transitive constructions in Teiwa, and 49 in Kaera. In both languages, transitives can have maximally one object. There are two transitive verb classes: one class indexes the object with an object marking prefix, the other class expresses the object as a separate nominal constituent and does not allow it to be marked with a prefix. Verbs taking prefixes are illustrated in (10); verbs not taking prefixes are given in (11).[7]

The verbs with object marking prefixes in (10) proto-typically[8] have animate objects. Note that 'someone' refers to animates (people and animals), in contrast to 'something'.

(10) Some verbs with object-marking prefixes

Teiwa		**Kaera**	
sas	'feed someone'	*as*	'feed someone'
rian	'look after someone'	*rian*	'look after someone'
wei	'bathe'	*wei*	'bathe'
pak	'call someone'	*pek*	'call someone'
tiar/tir	'chase someone'	*ter*	'chase someone'
lal	'show to someone'	*taring*	'show to someone'
soi	'order someone'	*iling*	'order someone'
ayas	'throw at someone'	*od*	'throw at someone'
an	'give someone'	*eng*	'give someone'

The verbs that do not take object prefixes are given in (11) and these proto-typically have inanimate objects. In Teiwa, the absence of an object prefix correlates strongly with the inanimate character of the object referent. I investigated 32 transitive verbs that occurred without an object prefix, for 28 of these the object was inanimate (an entity or a place). In Kaera, however, the verbs of the prefixless class can generally take both types of objects.

(11) Some verbs without P-marking prefixes

Teiwa		**Kaera**	
na	'eat something'	*na*	'eat something'
pai	'cut something in many small pieces'	*pay*	'cut someting/someone'
ol	'buy something'	*wal*	'buy something/someone'
qas	'split something'	*xas*	'split something'
kiri	'pull something'	*ker*	'pull something/someone'
mat	'take something'	*met*	'take s.th (marry someone)'
bali	'see something'	*lal*	'see something/someone'

Comparing the verbs in (10) and (11), we find a correlation between the semantics of the verbs and their object marking properties: the prefixing verbs in (10) denote events that proto-typically involve an animate object (a benefactive/recipient/goal); while the

---

[7] Of the 49 verbs investigated for Kaera, 28 belong to the first group, 20 to the second group, and one is unclear. Of the 54 verbs investigated for Teiwa 18 belong to the first group, 32 to the second group, and for four verbs the group is unclear.

[8] In Teiwa, 3 out of 19 verbs with an object prefix can also have an inanimate object, in Kaera 2 out of 25.

prefixless verbs in (11) express events that typically involve an inanimate object (a patient/theme).

In sum, despite some differences in detail, the following characterisations of Teiwa and Kaera object marking patterns apply: firstly, prefixes on transitive verbs crossreference animate objects, and secondly, inanimate objects are encoded by independent nominal constituents. Illustrations are given in (12) and (13). In (12a-b), the animate object of 'feed' is marked with a verbal prefix, in (13a-b), the inanimate object of 'eat' is expressed as an independent constituent.

(12) a. *Na bif ga-sas*
I child 3s-feed
'I feed the child' [Teiwa]

b. *Uxai gu gi-as-o.*
child that 3p-feed-RLS
'That child feeds them' [Kaera]

(13) a. *Uy ga'an qar na*
person 3s rice eat
'That person eats rice' [Teiwa]

b. *Gang naxar na*
3s rice eat
'S/he eats rice' [Kaera]

This is the general pattern of object encoding in Teiwa and Kaera. There are a few verbs that show exceptional behaviour, however, which are mentioned here for the sake of completeness. For example, Teiwa *mar* 'take' belongs to both class (10) and class (11): in (14a) it has an inanimate, free object, in (14b) an animate, prefixed one. The interpretation of *mar* changes with the animacy of the object.

(14) a. *Na ga'an mar*
1s 3s take
'I take/get it' [Teiwa]

b. *Na ga-mar*
I 3s-take
'I follow him/her' [Teiwa]

Further, Teiwa also has a few verbs that cross-reference their object both when it is animate and when it is inanimate. In such cases the animacy distinction is marked by the shape of the prefix: inanimate objects are marked with the prefix that normally marks animate objects, animate objects are marked with a prefix ending in a glottal stop. Examples are *ga-* [ga] vs. *ga'-* [gaʔ] in (15).

(15)

wulul		'speak, talk, tell'
ga-	*wulul*	'talk about it, tell it'
ga'-	*wulul*	'talk with/tell him/her'
wultag		'talk'
ga-	*wultag*[9]	'talk about it'
ga'-	*wultag*	'talk to/about him/her, tell him/her'

Apart from these exceptional cases, the overall system of Teiwa and Kaera is that animate objects are cross-referenced by verbal prefixes and that inanimate objects are expressed as free constituents.[10] Asymmetrical object marking patterns like these are in line with cross-linguistic observations on how animacy can affect agreement patterns (see Comrie 1989, Croft 2003, Siewierska 2004 for discussion and exemplification).[11] As agreement is sensitive to the discourse prominence of arguments, and as animate objects have more discourse prominence than inanimate ones, animate participants are more eligible to be indexed on the verb.

2.3 The oblique marking of themes

Having considered the encoding of the objects of two-place predicates, let us briefly consider how Teiwa and Kaera express events with three participants (agent, patient/theme, benefactive/recipient). As mentioned above, Teiwa and Kaera lack ditransitive verbs (see also Klamer, forthcoming a). Both languages distinguish between core and oblique arguments: a core object is expressed as the single object of a monotransitive verb (semantically a patient or a recipient/benefactive), while an additional argument, semantically a displaced theme, is introduced with an oblique marker. Transitive verbs with a benefactive/recipient core object cross-reference this object on the verb. As mono-valent verbs, they cannot have an additional 'displaced theme' (or 'object of transfer') as core object. Therefore, such displaced themes are either not expressed, as in (16), or they are introduced by a separate predicate as in (17), or they are marked as an oblique argument. Examples of themes introduced by a verb are given in (17) and (18). Themes introduced as oblique arguments are illustrated in (19) and (20). In Teiwa the oblique maker is *ma*, in Kaera it is *mi*.

(16)

Na	*ha-mian*
1s	2s-put.at

'I give (something) to you.' [Teiwa]

(17)

Na	*meet*	*mar-an*	*ha-mian*
1s	betelvine	take-RLS	2s-put.at

'I give you some betelvine' [Teiwa]

---

[9] This verb is usually pronounced in an abbreviated form, as [*g-ultag*].

[10] This is not unusual in Papuan languages. Examples of Trans New Guinea languages with animate object marking are Usan (Reesink 1987:108–109) and Nggem (Etherington 2002). Furthermore, in all the examples of verbal object agreement in Papuan languages presented by Foley (2000:378–379) the object affixes have animate referents.

[11] Differential encoding of animate and inanimate objects is part of a broader pattern also known as 'differential object marking', see Aissen (2003), Malchukov (2008), among others.

(18) *Gang ge-topi gu met, a*
s/he 3s-hat that take CONJ
'He takes that hat of his,

xabi mampelei utug met mi kunang masik namung gu gi-eng
then mango three take OBL children male many that 3p-give
then takes three mangoes to give to those boys. [Kaera]

(19) *Uy ga'an u sen*[12] *ma n-oma' g-an*
person 3s DIST money OBL 1s-father 3s-give
'That person gave money to my father.' [Teiwa]

(20) *Ui gu gang doi mi na-mam g-eng.*
person that 3s money OBL 1s-father 3s-give
'That person gave money to my father.' [Kaera]

In §3.3 below, we shall consider the oblique constituents marked with *ma/mi* in more detail. Not only do *ma/mi* mark displaced themes as obliques, but also locations and instruments.

2.4 Deictic verbs and serial verb constructions

The Teiwa verb that is the topic of this chapter, *ma*, belongs to a class of deictic motion verbs. Examples of such verbs are given in the (non-exhaustive) list in (21).

(21) Some deictic verbs in Teiwa and Kaera

Teiwa	**Kaera**	
ma	*ma-*	come (towards deictic centre)
gi	*gi-*	go (from deictic centre)[13]
daa	*da-*	ascend (towards deictic centre)
mir	*mid*	ascend (from deictic centre)
yaa	*ya*	descend (towards deictic centre)
yix	*ip*	descend (from deictic centre)
wa	*wa-*	go (from deictic centre; not far)

Talmy (1985) understands a 'motion' scene to constitute of a 'figure' in 'motion' along a 'path' oriented towards one or more 'grounds'. The 'figure' is expressed as the argument of the deictic verb. The deictic verb itself can be analysed into two basic components: a motion component, and a deictic component. As the translations of the verbs in (21) suggest, the deictic component of the verbs in Teiwa and Kaera contains information on the 'path' (towards/from), and the 'ground' (deictic centre) of the motion (cf. Wilkins and Hill 1995:249). In §3.1 below, I will come back to the semantic components of the deictic verb *ma*.

The origin of *ma* is unclear. Apparent cognates of it are found in both the non-Austronesian and the Austronesian languages spoken in the area. Some examples are

---

12 Teiwa *sen* and Kaera *doi* (cf. the following example) are loans from Malay (< Dutch *cent* 'cent', *duit* 'coin').

13 Probably from Indonesian *pergi*/ Malay *pigi,* see below.

listed in (22). The reconstructed Austronesian (Proto Malayo-Polynesian) form for 'to come' is **maRi*. It is thus possible that the verb *ma/man/mai* in the non-Austronesian languages of Alor-Pantar-Timor has an Austronesian origin, but the formal similarity may also be accidental. There is no evidence that *ma* is related to proto Austronesian **ma* 'and' (cf. fn 16).

(22) Possible cognates of *ma*

a. In neighbouring non-Austronesian languages
West Pantar (Lamma): *ma* 'come' (Gary Holton, p.c. 2006)
Adang: *ma* 'come (from a short distance)' (Haan 2001: 248)
Klon**:** *ma* 'come' (Baird 2008)
Abui: *me* 'come' (Kratochvíl 2007)
Kafoa: *mai* 'come', Hamap: *ma* 'come' (Baird 2003)
Bunaq (Timor): *man* 'come' (Klamer 2002)

b. In neighbouring Austronesian languages
Alorese (Alor) *mene* 'come (here)' (Klamer, forthcoming b)
Tetun Terik (C Timor) *mai* 'come' (Van Klinken 1999:262)
Mambai (E Timor) *ma,* Tokodede (E Timor) *mai* (Klamer 2002)
Kemak (E Timor) *mai,* Lakalei (E Timor) *man* (Klamer 2002)
Idate (E Timor) *ma* (Klamer 2002)

Deictic verbs like *ma* often occur in serial verb constructions. Serial verb constructions are here defined (pre-theoretically) as two or more separate verbs occurring in a single intonation contour, which share their subject and/or their object, and are within the scope of a single negation and/or coordinating conjunction, if such an item is present.[14] One of the ways to classify serial verb constructions is their composition along a continuum of 'symmetrical' and 'asymmetrical' combinations of verbs (Aikhenvald 2006). Symmetrical serial verb constructions consists of two or more verbs from semantically and grammatically unrestricted classes (so-called 'major' verbs). Asymmetrical serial verb constructions include at least one verb from a grammatically or semantically restricted class (a 'minor' verb). The deictic verbs in Teiwa and Kaera are typically the minor verbs in asymmetrical serial verb constructions.

2.5 Grammatical overview: summary

We have seen that Teiwa and Kaera are head-final, with final verbs, final negations and final (coordinating) conjunctions. Lacking syntactically subordinated clauses, Teiwa and Kaera make extensive use of serial verb constructions. Subjects and objects occur pre-verbally, as do time and manner adjuncts. Locations are expressed in two ways: as obliques or as object of transitive location verbs. In both cases, they are pre-verbal.

Teiwa and Kaera lack ditransitive verbs. A core argument is expressed as the single object of a (mono)transitive verb, and is semantically a recipient/benefactive or a patient. Oblique arguments are marked with *ma/mi*, and have a variety of semantic functions (cf. §3.2 for details). Teiwa and Kaera have a grammatically asymmetrical system where only

[14] Serial verbs are notoriously difficult to define, see Crowley (2002:8–19). It is beyond the scope of this chapter to present a full analysis of serialisation patterns in Teiwa and Kaera. See Klamer (2010) for additional information on Teiwa serial verb constructions.

animate objects are cross-referenced by verbal prefixes. Inanimate objects are expressed as independent constituents.

The Teiwa verb that is the topic of this chapter, the deictic verb *ma,* belongs to a set of deictic verbs whose lexical semantics are formally analysable into two elements: a motion component, and a deictic component. Deictic verbs like *ma* can occur as independent verbs and in asymmetrical serial verb constructions. In the second context, they are always the minor verb.

3 The item *ma*

3.1 Lexical semantics versus contextualised meanings of *ma*

I analyse *ma* using a representational model of lexical semantic representation in which meanings of verbs are decomposed into more basic elements. Current theories of predicate decomposition usually have one or more primitive predicates that represent the meaning of the verb and act as argument-taking functions (see Levin and Rappaport Hovav 2005 for an overview). In such models, the number of argument positions associated with predicates may vary from one to three: arguments may be hierarchically organised (for example, as internal vs. external argument), and have different syntatic categories and/or different semantic features.

Analysis of *ma* assumes that *ma* is a one-place verb, and that the semantics of its single argument are left unspecified. In the analysis below I distinguish between:

(i) a level that is concerned with the stored communicable information associated with conventional signs, often referred to as lexical semantics; and

(ii) another level which is concerned with the final interpretation of utterances and their parts in particular contexts, also referred to as contextualised meanings or pragmatics (cf. Wilkins and Hill 1995 and the references cited there).

Under this view, it is possible to characterise a lexical item in lexical semantic terms independently of other lexical items. But its functions and the way it is pragmatically interpreted, that is, its contextualised meaning, depend on the grammatical context of the item, and may therefore be variable.

The distinction between the lexical semantics and the contextualised meanings of *ma* will become relevant in the sections below, when the various kinds of synchronic functions that *ma* has developed are discussed. My hypothesis is that all of the grammaticalised functions of *ma* are derived from is conceptual sturucture as the deictic verb *ma*. In the present section, I propose a lexical conceptual representation of *ma,* and I will explain how its various synchronic functions relate to its lexical conceptual structure in §3.2 and 3.3.

As mentioned above, the verb *ma* can be used as an independent verb, or as part of a serial verb construction. When it is used as an independent verb, that is, as the single verb of a clause, it expresses a motion and an orientation towards the deictic centre. The deictic centre (often pragmatically designated to be the speaker) is the understood endpoint of the motion. This use of *ma* is illustrated in (23) and (24).

(23) *Ha'an la ma le na'an la wa?*
2s FOC come or 1s FOC go
'Are you coming (to me) or am I going (to you)?' [Teiwa]

(24) *Ui gu gang miaag la ma*
person that 3s yesterday FOC come
'That person came (here) yesterday.' [Kaera]

The lexical semantic components of *ma* can be represented schematically as in (25).

(25) *ma:* <x> MOVE [DEICTIC CENTRE (Goal; Endpoint)]

Ma expresses a motion of <x> with the deictic centre as destination or 'goal'; and since the motion ends there, the deictic centre is also the endpoint of it (see Wilkins and Hill 1995 for exemplification of this distinction). As an intransitive verb, *ma* takes a single argument <x>. As mentioned above, I assume that the semantics of this argument are unspecified, that is, the lexical conceptual representation of *ma* does not contain any specification(s) of the semantic role of this argument (as agent, patient, location, instrument, etc.). Instead, I assume that the argument's semantic interpretation is contextually determined.

Because the MOVE component of *ma* forces the figure to be canonically interpreted as a moving entity, a proto-typical argument of *ma* is aminate, as illustrated in (23) and (24), but the argument is not lexically specified as animate. Observe that *ma* can also occur with an inanimate argument as in (26), where the referent of *a'an* is a sword:

(26) *...palan qas maan, a'an **ma** mosan*
split.bamboo NEG, 3s come sword
'...it wasn't a bamboo, it had become a sword.' [Teiwa]

In grammatical contexts like (26), the argument of *ma* is inanimate, and the nominal predicate *mosan* 'sword' is the deictic centre ('goal' and 'endpoint'). As a result, MOVE must be reinterpreted as a change of state predicate, whereby an inanimate figure (metaphorically) moves towards a certain state.

(27) *ma:* <*inanimate*> MOVE [DEICTIC CENTRE = *nominal predicate state, for example 'sword'*]

The reinterpretation of motion verbs into predicates marking change-of-state is cross-linguistically widely attested (cf. English 'come' > 'become'). For *ma* such a reinterpretation does not involve a change in the lexical semantics of *ma,* but is entirely determined by its grammatical context. When *ma* combines with an inanimate, non-moving argument and with a stative nominal predicate, there is no way that MOVE could ever be interpreted as expressing a physical motion, and the closest interpretation that remains available in this type of context is to view *ma* as a change-of-state predicate.

To conclude, *ma* can occur with an animate or inanimate argument, and the animacy value of the argument is one of the contextual factors that determines its interpretation as either a deictic verb or as a change-of-state predicate.

In the following sections, the various grammaticalised functions of *ma* will be discussed. We will see more evidence that the animacy value of its argument and the grammatical context in which the verb appears determine its interpretation.

In Teiwa, *ma* functions as a grammatical tense/mood marker ('venitive', intentions, hortatives, imperatives) and as a conjunction (§3.1), while it is also used as a marker of oblique arguments (§3.2). The former type of function is seen when *ma* has an animate

argument, the latter when it has an inanimate argument. In Kaera, which also has a verb *ma*, the verb only developed functions as a tense/mood marker and conjunction. The oblique marking function was probably blocked for Kaera because of the pre-existence of the oblique marker *mi*.

3.2 *Ma* as marker of movement in time

When *ma* occurs with an animate argument in a serial verb construction, it can be used to mark events that will take place in the future. This is illustrated in (28) and (29). Example (30) shows that *ma* can also be used to express intentions.

(28) a. *Ha ma nili pat-an*
2s come debt pay.back-RLS
'You will pay back the debt.'

Cf. b. *Ha nili pat-an* [Teiwa]
2s debt pay.back-RLS
'You pay back the debt.'

(29) a. *Na la ma lal-o.*
1s TOP come look-MOD
'I will be the one who looks.'

Cf. b. *Na la lal-o.* [Kaera]
1s TOP look-MOD
'I am the one who looks.'

(30) a. *Na ma walas?*
1s come talk
'Shall I tell [the secret]?'

Cf. b. *Na walas?* [Teiwa]
1s talk
'Do I talk? (*Shall I tell?)'

In contexts with an animate argument, *ma* still functions to express a motion, but now the movement may be interpreted as a movement in time; the figure is moving towards a temporal endpoint rather than a physical one. Such a temporal interpretation of *ma* is possible, because the meaning component MOVE in itself does not say anything about the dimension in which the movement takes place. When the grammatical context is such that the component DEICTIC CENTRE is the event depicted by the (major) verb (phrase), clearly the endpoint of the 'movement' is a point in time when that particular event takes place. This is the function of *ma* in (28) to (30). In (31) and (32) below, *ma* functions to mark an imperative and a hortative clause. In these contexts, the verb is interpreted even more broadly as a movement towards a situation that takes place in the future.

(31) a. *Ma na-walas.*
come 1s-talk
'Talk with me!'

b. **Na-walas!* [Teiwa]
1s-talk
(Not good for: 'Talk with me!')

(32) *'Ah Liwang! Ma ping wat igang mo mi rap-o.'*
Oh Liwang come 1pi go place overthere OBL search-MOD
'Oh Liwang! Let's go search over there.' [Kaera]

Note that I have glossed *ma* in (28) to (32) as 'come', to indicate that it is the same item as the one used in (23) and (24). The argument of *ma* is animate, and is thus moving, in space and/or in time. The only difference I see between the function of *ma* in (23) and (24) on the one hand, and in (28) to (32) on the other, that in the former, *ma* is an independent verb, while in the latter set of examples, it is part of a serial verb construction and shares its argument with another verb. This other verb is the major one of the serial verb construction, and is semantically the more important one of the two. In other words, (28) to (32) are about arriving at a **situation** of paying back debt/looking/talking/going, rather than about arriving at a literal geographical **location.** The more abstract temporal interpretation of *ma* is also witnessed when it is interpreted to mark a time lapse between events. Consider (33): both (33b) and (33c) are possible sequences to (33a), but in (33b) *ma* functions as a conjunction and in (33c) it does not.

(33) a. *...qavif ga'an hala ta ga-finan,*
goat 3s others TOP 3s-catch,

qavif ga'an bir-an kuat maan ba
goat 3s run-RLS strong NEG SEQ
'... that goat was caught by them, the goat couldn't run fast, so

b. *hala ta gi er-an gula'* ***ma*** *harabax ma*[15] *gad.*
people TOP go do-RLS finish come stable come put
people went to get [it] then put [it] in a stable'. [Teiwa]

c. *hala ta gi er-an gula' harabax* ***ma*** *gad.*
people TOP go do-RLS finish stable come put
people went to do [that] put [it] in a stable.' [Teiwa]

In the sentence in (33b), two events are expressed: *<people went to do (that)>* and *<(people) put (a goat) in a stable>*. The events are separated by *ma,* which marks a time lapse between the two events ('x then y'). For example, there could have been a time lapse when the people first caught the goat, and then had a smoke before putting it in the stable. By contrast, in (33c) there is no overt marker to indicate that time elapsed between the two events. Note that *ma* in (33b) is grammatically still a verb: it has 'people' as its animate argument, an argument it shares with *gi eran* 'go do', *gula'* 'finish' and *gad* 'put'. Recall that as a head-final language, Teiwa conjunctions are clause final. As *ma* is structurally

[15] Here, *ma* functions as an oblique marker to introduce a location as secondary object, see §3.3 for a discussion of this function of *ma.*

part of the clause preceding it, it may be reinterpreted to function here as a kind of coordinating conjunction ('then'), linking two separate clauses.[16]

In sum, when *ma* is used in a serial verb construction with an animate argument, it behaves as a verb. It combines with a major verb (phrase), which consitutes the event that is the deictic centre, as sketched in (34):

(34) <*animate*> MOVE | DEICTIC CENTRE = event (for example, pay debt, talk, look, put in stable)

In such contexts, the goal/endpoint of the movement is the situation when a particular event takes place. That is, when *ma* occurs in a grammatical context which allows its meaning be generalised to a moment in time, it functions to mark intentions, hortatives and imperatives, as well as time lapses between events.

The functions of *ma* discussed in this section are cross-linguistically quite commonly found as grammaticalisations of the verb 'come'. Heine and Kuteva (2002:68–70, 78) mention that 'come' often develops into a 'venitive', comparable to (24) or a 'hortative', comparable to (32) or a 'consecutive', comparable to (33b), and they also mention 'come to' as a common source for future markers, comparable to (29). Bybee, Perkins and Pagliuca (1994:268–269) present an analysis of motion verbs developing functions as markers of future, intentions, and imperatives in various languages. In the case of *ma,* there is no reason (yet?) to say that synchronically, the verb has developed into a set of different functional items. In the analysis presented here, one and the same lexical item is used in the examples (23) to (32), but the functional interpretation of this item depends on the grammatical context in which it appears.

Pragmatically, *ma* functions as a marker of intentions, hortatives, imperatives, and is similar to a conjunction. However, this type of interpretation is only possible if the argument of *ma* is inanimate and if *ma* is followed by a constituent expressing an event/proposition: a verb, verbal phrase, or a clause. In addition, certain specific grammatical properties of Teiwa and Kaera enable the reanalysis of *ma*. They include:

(i) the productiveness of serialisation;

(ii) the absence of morpho-syntactically marked clause subordination markers and/or a distinction between finite and non-finite verb forms, which enables easy re-interpretation of verb forms into grammatical morphemes;

(iii) the predominantly 'left-branching' character of Teiwa and Kaera (where branching constituents precede non-branching constituents, see Dryer 1992), so that [NP *ma*] phrases preceding a verb may be reinterpreted as branching PP-like constituents.

(iv) the absence of inflection markers for (future) tense or irrealis mood, so that the tense/mood interpretation of *ma* does not need to compete with pre-existing morphemes.

---

[16] The conjoining function of Teiwa and Kaera *ma* is but one of the many contextualised interpretations of the deictic verb *ma*. The question may then be asked if *ma* is related to Proto-Austronesian (PAN) conjunction *ma.* Section 3.1 presents cross-lingustic evidence which suggests that *ma* might be related to the Proto-Austronesia verb **maRi* 'come'. There is no such evidenc to link it with the PAN conjunction **ma* 'and'.

In sum, *ma* as a marker of intentions, hortatives, imperatives and time lapses is not an accidental or random process, but interpretation depends on its grammatical context as well as on more general morpho-syntactic and phrase structure properties of Teiwa and Kaera.

3.3 *Ma* as oblique marker

In this section, I consider constructions in which *ma* is used with an inanimate argument. With an inanimate argument, *ma* can function as an oblique marker when it is a minor verb in the serial verb construction. It then marks adjunct-like participants in the clause. In Kaera, such constructions are not found, since Kaera *ma* cannot have an inanimate argument. Instead, Kaera uses the marker *mi* to mark oblique arguments. Teiwa *ma* and Kaera *mi* have very similar functions, as we will see.

The origin of these two oblique markers is quite different. Example (35) presents a hypothetical scenario of how *ma* and *mi* may each be related to a verbal source. On the left, the original item is the intransitive verb *ma* 'come (here)', on the right, the source is the locative verb *me'* (Teiwa) 'be at (location) X' or *ming* (Kaera) 'be at location X'. Given their similar semantics, *me'* and *ming* may be cognates. The direction of grammaticalisation is indicated with arrows; no arrows means that the grammaticalisation of an oblique marker was blocked because such an item already existed.

(35) Possible diachronic relations between Teiwa *ma*, Kaera *mi*, Teiwa *me'* and Kaera *ming*

Intransitive verb		Oblique		Transitive verb
Teiwa *ma*	→	Teiwa *ma*		Teiwa *me'* 'be at location X'
Kaera *ma*		Kaera *mi*	←	Kaera *ming* 'be at location X'

In other words, the function of Teiwa *ma* as oblique marker must have blocked the development of an oblique from the Teiwa verb *me',* while the existence of the oblique marker Kaera *mi* must have blocked the development of an oblique from the Kaera verb *ma*. An oblique marker *mi* also occurs in West Pantar (Lamma) (Holton p.c. 2006), and in Klon (Alor) (Baird, this volume), and these languages are also like Kaera in using *ma* as a deictic verb only. Since the oblique marking function of *ma* is not attested in languages closely related to Teiwa, I assume that the original function of Teiwa *ma* is verbal, and its oblique marking is a derived function. It is the oblique marking function of Teiwa *ma* that will be further examined in this section.

Recall that Teiwa and Kaera have no ditransitive verbs, so that when three participants are involved in a single event, the third participant is either intorduced by its own predicate (one of the mechanisms giving rise to serival verb constructions in these languages), or with an oblique marker. When Teiwa *ma* occurs with an inanimate argument, it functions as an oblique marker to introduce various semantic types of participants; locations, instruments as well as displaced themes.

In what follows, I compare Teiwa constructions with *ma* to minimally contrasting Kaera sentences containing *mi*.[17] In the examples, *ma* and *mi* will both be glossed as 'OBL(ique)' for expository reasons. It is not necessary to assume that *ma* has actually changed its

[17] Sentences (36) to (43) are narrative data elicited with Mayer's (1969) picture book 'Frog where are you?'. Sentences (45) and (46) are from a set of some twenty sentences of an initial survey which used Indonesian as language of elicitation.

category (for example from verb to its case marker): the lexical semantics of *ma* remain unchanged, and its various interpretations are derived by its grammatical context.

In (36) and (37) *ma* and *mi* introduce a locational adjunct: the place where the object (a frog) is kept. In both examples, the 3sg prefix on the final verb refers to an animate object, the frog.

(36) *...bif g-oqai nuk mauqubar ga-fin-an gula' pin aria'*
child 3s-child one frog 3s-catch-RLS finish hold arrive
'a child caught a frog, brought it home,

toples[18] *g-om ma ga-rian*
jar 3s-inside OBL 3s-look.after
kept it in a jar.' [Teiwa]

(37) *Ging gang toples nuk mi ga-dag*
3p 3s jar one OBL 3s-leave
'They put it in a jar.' [Kaera]

In the following two sentences, *ma* and *mi* also encode locations, in this case it is the place where the subject is standing (38), or hiding (39).

(38) ...[*war uwaad nuk ga-he'en*] *ma tas...*
rock big one 3s-close OBL stand
'...stands close to a big rock...' (lit. 'stands at the close of a big rock') [Teiwa]

(39) *Liwang ula* [*war er gu*] *mi ong, gang gang lal-i bino.*
Liwang also rock close that OBL hide 3s 3s see-MOD NEG
'Liwang hides close to that rock [lit. 'at that rock's close'], [but] he did not see him.' [Kaera]

In (40) and (41), the obliques with *mi/ma* introduce the goal of the event of 'falling':

(40) *... ba'-an yaa saf her ma yaa,*
fall-RLS descend river.bank close OBL descend

rus waal ta bir-an gi...
deer that TOP run-RLS go
'...(after) falling down from the riverbank, that deer ran away...' [Teiwa]

(41) *Ilwang mi Liwang unang ir boi mi ba*
Ilwang OBL Liwang be.together water river OBL fall.
'Liwang fell into the river together with Ilwang.' [Kaera]

In (42) and (43) the semantic role of the participant marked by *ma/mi* is ambiguous. It may be the location/goal or the inanimate patient of 'bite'. If 'bite' has an animate object, it is marked as a core object with a verbal prefix, as shown in (44).

---

[18] From Alor Malay *toples* < Dutch *stopfles* 'jar'.

(42) *...banaq g-om bali wa katak ma palan si,*
puddle 3-inside see go frog come inspect SIM
'...goes to inspect the frogs (in the) puddle

katak waal ta daa g-et bag ma sii.
frog that TOP ascend 3-eye seed OBL bite
and that frog bites his eyes.' [Teiwa]

(43) *'Ooh Liwang! Dur nuk n-iming mi si.*
Oh Liwang mouse one 1s-nose OBL bite

Aaach... iit yas-o,"
aaah hurt bad-MOD
'Ooooh Liwang! A mouse bit my nose. Aaah, it hurts so much' [Kaera]

(44) *Katak ga-sii.*
frog 3-bite
'A frog bites him.' [Teiwa]

In (45) and (46), the constituent marked with *ma/mi* introduces a participant that is the instrument of 'cut':

(45) *Uy nuk ped ma tei taxar*
person one machete OBL wood cut
'Someone cuts wood with a machete.' [Teiwa]

(46) *Ui gu gang ped mi tei ptak-o pati*
person that 3s machete OBL wood cut-MOD while
'That person is cutting wood with a machete.' [Kaera]

Instruments cannot be core objects and must be introduced by *ma*, as shown by the grammaticality contrast between (47a, b).

(47) a. *Na ma* [[*tei bun ma*] [*yivar g-ua'*]]
I come wood piece OBL dog 3s-hit
'I'll hit the dog with a stick.' [Teiwa]

b.* *Na ma tei bun yivar g-ua'.*
I OBL wood piece dog 3s-hit

Finally, the oblique marker *ma/mi* also marks displaced themes (also referred to as 'objects of transfer'). Verbs like Teiwa *-an* and Kaera *-eng* 'to give to someone' are mono-transitive, with a recipient core object. In (19) and (20) above it was illustrated how objects of transfer are marked as oblique arguments with the markers *ma/mi*. Additional illustrations are (48) and (49). Example (48) contains the verb *mian* 'to give', literally 'to put at someone', a transitive verb with a benefactive/recipient object. The displaced theme must be marked as oblique, as shown in (48b); the recipient cannot be oblique, as shown in (48c).

(48) a. *Na-xala' yir ma bif ga-mian hufa'.*
1s-mother water OBL child 3s-put.at.s.o. drink
'My mum gives water to the child to drink.' [Teiwa]

b. *Naxala' yir bif ga-mian hufa'.*
1s-mother water small 3s-put.at.s.o. drink

c. *Naxala' yir bif ma ga-mian hufa'*
1s-mother water child OBL 3s-put.at.s.o. drink

The sentence in (49) also contains a transitive location verb, but in this case the verb takes an inanimate locational object: *tanat* 'to place/put on something'. The NP ***afat ki'*** 'his big toe' is thus the locational core object, and the thing placed, *tab ga'an* 'that spear', is marked as oblique.

(49) *A tab ga'an ma ga-fat ki' uwaad tanat olaxhamar*
3s spear 3s OBL 3s-foot toe big place.on.s.t. recite.poetry
'He places [the point of] that spear on his big toe [and] recites poetry.' [Teiwa]

Additional illustrations of Kaera clauses with oblique displaced themes are (50) to (52):

(50) *Gang buku mi n-eng*
3s book OBL 1s-give
'He gave me a book.' [Kaera]

(51) *Gang naxar mi n-aso*
3s rice OBL 1s-feed
'He fed me rice.' [Kaera]

(52) *A ta war upar ma ga-ayas*
3s TOP rock pebble OBL 3s-throw.at
'He throws pebbles at him.' [Teiwa]

Note that entities introduced by *ma/mi* are not only displaced themes, but also include themes that are pointed out or shown to someone, as in (53) to (54), as well as themes that are e.g. bought for the benefit of someone, as in (55).

(53) *Yitar ga-qau ma na-lal-an*
road 3s-good OBL 1s-show.to-RLS
'[You] show me the right way.' [Teiwa]

(54) *Gang foto*[19] *mi ne-taring*
s/he picture OBL 1s-show
'He shows me a picture.' [Kaera]

---

19 From Malay/Indonesian *foto* < Duch *foto* 'photograph'.

(55) *Nang semering nuk mi gang pay-o*
I knife one OBL 3s buy-MOD
'I buy him a knife.' [Kaera]

Observe once again that the objects marked by *ma/mi* are all inanimate. How, then, are animate displaced themes marked? For example, how would the theme *the child* in *she gave me the child* be expressed? The answer is, not as an oblique argument, since animate theme objects can only be introduced through a separate serial verb construction, as illustrated in (56) where *pin* 'hold' introduces the displaced theme *biar kriman* 'small children'. (The subject of *pin* 'hold' is *hala* 'others'.)

(56) a. *Jadi hala biar kriman la pin aria' ma ni-mian...*
so others children small FOC hold arrive come 1p.e-put.at
'So others brought small children here and gave them to us...'
(lit. 'So others took small children, arrived, and put [them] at us.') [Teiwa]

b. *Jadi hala biar kriman ma ni-mian...*
so others children small come 1p.e-put.at

The displaced theme is the core object of the verb *pin,* and cannot be marked as oblique with *ma.* The subject *hala* is shared by all the verbs in the serial verb construction, including the pre-final verb *ma* (since this *ma* has an animate argument, it functions as a conjunction that marks a time lapse between events (cf. its similar function in 33b above).

Serial verb constructions can, of course, also introduce inanimate displaced themes, as illustrated in (57), where *bag nuk* 'one seed' is the core object of *mar-an* 'take-RLS'.

(57) *... mar-an bag nuk, mar-an gula', a ma ga-sas gula'...*
take-RLS seed one take-RLS finish 3s come 3s-feed finish
'... one seed (he) takes, having taken it, he then came and fed [it] to him...'
[Teiwa]

The Kaera oblique marker *mi* normally takes an inanimate argument. It does not occur with an animate argument, except for one particular grammatical context, illustrated in (58). In (58a), *geumux* 'his wife' is the single argument of the intransitive serial verb construction *unang gi* 'go together', and the oblique headed by *mi* contains the comitative object *ui gu* 'that person'. In (58b), the argument of *unung gi* is *ui gu* 'that person', and now the oblique contains the comitative object *geumux* 'his wife'. Note that in (58a) the possessor prefix *ge-* '3sg' of 'wife' is interpreted to refer to 'that person', that is, the previously mentioned participant. In (58b), however, there is no preceding nominal to which the possessor prefix can refer.

(58) a. [*Ui gu mi*]$_{PP}$ *ge-umux unang gi*
person that OBL 3s-wife be.together go
'That person$_j$ goes together with his$_j$ wife.' (lit. 'With that person his wife goes together').

b. [*Ge-umux mi*] *ui gu unang gi*
3s-wife OBL person that be.together go
'His$_j$ wife goes together with that person$_k$.' [Kaera]

In conclusion, we have seen that Teiwa *ma* and Kaera *mi* mark identical types of oblique participants. The semantic roles of these participants vary, and are determined by the meaning of the major verb in the clause. For example, when an oblique argument combines with the major verbs 'leave', 'stand', 'hide', 'descend' or 'fall', it marks a location or goal (as in (37) to (41)); when it combines with the verb 'bite', it marks a location or (inanimate) patient (as in (42) and (43)); when it combines with the verb 'to cut something' it marks an instrument (as in (45) and (46)), when it combines with the verb 'to put at someone' it marks a displaced theme (as in (48) to (50)), with the verb 'to feed someone', it marks the food (as in (51)), with the verb 'to show someone' it marks the thing shown (as in (53) and (54)), et cetera. In short, oblique markers productively introduce inanimate participants of all kinds into events, and the semantic interpretation of these participants is largely determined by the major verb of the clause.

The oblique marking function of *ma* is related to its lexical semantics as follows. *Ma* combines with an inanimate argument as represented in (59). Because the argument is inanimate, it is impossible to interpret *ma* here as a predicate of motion. That is, the semantic component MOVE cannot be part of the contextualised interpretation of *ma* here, and the deictic component is the one that prevails. Since there is no motion involved, the deictic centre is not interpreted as goal or endpoint but simply as a static location.

(59) <*inanimate*> ~~MOVE~~ DEICTIC CENTRE (static location)

As a result, the interpretation of an oblique construction with *ma* is something like 'inanimate argument <x> is located at deictic centre'. What is the deictic centre here? When *ma* is followed by a major verb, the deictic centre is part of the discourse: the event expressed by that major verb (phrase). The oblique marking of *ma* is thus interpreted as something like '<x> is located at the event reported here'. As an illustration, consider (60). The 'deictic centre' is the scene depicted by *a ga-ayas* 'he throws at him'. At this deictic centre, an additional participant is located, namely *war upar* 'pebble', the inanimate argument of *ma*. This additional participant is pragmatically interpreted as the secondary object of *ayas* 'throw'.

(60) *A ta war upar ma ga-ayas*
3s TOP rock pebble OBL 3s-throw.at
'He throws pebbles at him.' [Teiwa]

We thus distinguish between the lexical semantics of *ma* and the contextual interpretation it can get in particular grammatical contexts. In this way, it is possible to account for the fact that a single item is used as a deictic verb, a change-of-state verb, a tense/aspect marker, a conjunction and as a marker of oblique constituents, indicating static locations. I have suggested how these functions are related to the two lexical semantic component *ma,* and thus to each other, and how they can shift depending on grammatical context.

Because the analysis presented here refers to the grammatical context of *ma,* it is implied that *ma* does not get to function as an oblique marker by accident, or that it is a function that developed randomly. Firstly, this is because it can only happen when *ma* occurs with an inanimate argument. Secondly, *ma* must be followed by a major verb (phrase) in a serial verb construction. And thirdly, the oblique marking interpretation of

ma is enabled by the following, more general structural properties of the grammar of Teiwa:[20]

(i) the productivity of serialisation;

(ii) the absence of morpho-syntactic clause subordination markers and/or a distinction between finite and non-finite verb forms, which enables easy re-interpretation of a verb form into a grammatical morpheme;

(iii) the lack of ditransitive verbs, so that there is a communicative need to create structures which introduce additional arguments;

(iv) the fact that adjunct constituents expressing time and manner are pre-verbal constituents, so that it is natural to interpret a preverbal constituent with *ma* as a preverbal adjunct as well.

4 Conclusions

In conclusion, we have seen that *ma* functions as a deictic verb and as a change-of-state verb, as a marker of intentions, hortatives, and imperatives, as a conjunction that indicates a time lapse between subsequent events and as an oblique marker that introduces additional participants. All these synchronic functions of *ma* are surface manifestations of a single lexical item.

In the analysis of the functions of *ma*, a distinction between the level of lexical semantics and of contextualised meanings is relevant. *Ma* is an intransitive predicate with an unspecified argument position, and it contains two basic semantic components: a motion, and a deictic centre. The motion component is represented in the lexical semantics of *ma* as MOVE. The deictic component, which contains information on path (towards/from) and ground (deictic centre) of the motion, is represented as DEICTIC CENTRE (Goal; Endpoint).

The many different synchronic functions of *ma* are contextualised meanings of a single item, that is, they depend on the grammatical context in which *ma* appears. Two contextual factors play a crucial role in the variable interpretations of *ma*: firstly, the animacy value of its argument, and secondly, whether or not *ma* appears in an serial construction, that is is followed by a major verb.

The functions of *ma* are of three broad types. First, *ma* functions as a deictic or change-of-state verb when it is the single verb of a clause. Second, in serial constructions it takes on other functions: with an animate argument in an SVC, *ma* is interpreted as a tense/mood marker or as a conjunction marking a time lapse. These interpretations of *ma* are available when the semantic component MOVE is interpreted as movement in time (as well as in space). The third type of function of *ma* is found when it occurs with an inanimate argument in SVC's. In such contexts, it functions as an oblique marker introducing arguments of various semantic types, including locations, instruments and displaced themes. In contexts with an inanimate argument, the component MOVE is not available for interpretation, and the deictic component prevails. Since there is no motion involved, it expresses a static location.

---

[20] Durie (1988:20) mentions factors very similar to the ones suggested here as factors which 'constrain' the development of prepositions out of verbs in Oceanic languages.

Comparing the functions of *ma* in Teiwa and Kaera, we find that in both languages *ma* functions to mark 'movement in time', but that only in Teiwa does *ma* function as an oblique marker. For the oblique marking function to be possible, *ma* must occur with an inanimate argument, which is not possible in Kaera. Instead, Kaera marks obliques with *mi,* a marker that is possibly derived from the transitive location verb *ming.*

Regarding the overall grammatical structure of Teiwa and Kaera, we observed that both languages lack ditransitive verbs, and distinguish between core and oblique arguments. Within the group of (core) objects, only the animate ones are cross-referenced by verbal prefixes, the inanimate ones are expressed as independent constituents. Additional arguments are marked as obliques with *ma/mi* if they are inanimate. Animate arguments cannot be marked with *ma/mi*, but are introduced with a separate verb. The animacy value of participants thus plays a crucial role in how they are encoded.[21] In addition, the animacy of the argument of *ma* is one of the crucial determinants of the variable interpretation of *ma*.

References

Aikhenvald, Alexandra, 2006. Serial verb constructions in typological perspective. In Alexanda Y. Aikhenvald and R.M.W. Dixon, eds, *Serial verb constructions: A cross-linguistic typology*, 1–68. Oxford: Oxford University Press.

Aissen, Judith, 2003. Differential object marking: iconicity vs. economy, *Natural Language and Linguistic Theory*, 21: 435–483.

Aristar, Anthony R., 1997. Marking and hierarchy types, and the grammaticalisation of case-markers, *Studies in Language*, 21:313–368.

Baird, Louise, 2003. Alor Survey. Unpublished database. Leiden University.

— 2008. *A grammar of Klon: A non-Austronesian language of Alor, Indonesia*. Canberra: Pacific Linguistics.

Bybee, Joan, Revere Perkins and William Pagliuca, 1994. *The evolution of grammar*. Chicago/London: The University of Chicago Press.

Comrie, Bernard, 1989. *Language universals and linguistic typology.* Chicago/London: The University of Chicago Press.

Croft, William, 2003. *Typology and universals* [2nd edition]. Cambridge: Cambridge University Press.

Crowley, Terry, 2002. *Serial verbs in Oceanic.* Cambridge: Cambridge University Press.

Dryer, Matthew, 1992. The Greenbergian word order correlations, *Language*, 68:81–138.

Durie, Mark, 1988. Verb serialization and 'verbal prepositions' in Oceanic languages, *Oceanic Linguistics*, 27,1–2: 1–23.

Etherington, P. A., 2002. Nggem Morphology and Syntax. Alice Springs: The Northern Territory University, Australia. PhD dissertation.

Foley, William A., 2000. The Languages of New Guinea, *Annual review of Anthropology*, 29:357–404.

---

[21] This suggest that the Animacy Hierarchy (Croft 1990:111ff.) not only influences the way 'grammatical' cases (such as nominative and accusative) develop, but also how 'non-grammatical' cases (such as locative and instrumental) come into being (cf. Aristar 1997).

Haan, John, 2001. The grammar of Adang. Sydney: The University of Sydney. PhD dissertation.

Heine, Bernd and Tania Kuteva, 2002. *World Lexicon of Grammaticalization.* Cambridge: Cambridge University Press.

Klamer, Marian, 2002. *East Timor survey.* Database. Leiden University.

— 2003. *Bahasa Alor survey data.* Database. Leiden University.

— 2010. *A grammar of Teiwa*. Berlin/New York: Mouton de Gruyter.

— Forthcoming a. Ditransitive constructions in Teiwa. In Bernard Comrie, Martin Haspelmath and Andrej Malchukov, eds, *Studies in Ditransitive Constructions.* Berlin/New York: Mouton de Gruyter.

— Forthcoming b. A short grammar of Alorese. MS. Leiden University.

Klamer, Marian and František Kratochvíl, 2006. The role of Animacy in Teiwa and Abui (Papuan). Paper presented at the 32nd Annual Meeting of the Berkeley Linguistics Society (BLS 32), February 10–12, 2006. Forthcoming in *BLS 32 Proceedings*.

Kratochvíl, František, 2007. *A grammar of Abui.* Leiden University. Utrecht: LOT Publications. PhD dissertation.

Levin, Beth and Malka Rappaport Hovav, 2005. *Argument realization.* Cambridge: Cambridge University Press.

Malchukov, Andrej L., 2008. Animacy and asymmetries in differential case marking. *Lingua*, 118:203–221.

Reesink, Ger P., 1987. *Structures and their functions in Usan, a Papuan language of Papua New Guinea.* Amsterdam/Philadelphia, Benjamins.

— 2002. *Languages of the eastern Bird's Head*. Canberra: Pacific Linguistics.

Siewierska, Anna, 2004. *Person*. Cambridge: Cambridge University Press.

Talmy, Leonard, 1985. 'Lexicalization patterns'. In Timothy Shopen, ed., *Language typology and syntactic description*, 57–149. Cambridge: Cambridge University Press.

van Klinken, Catharina, 1999. *A grammar of the Fehan dialect of Tetun, an Austronesian language of West Timor*. Canberra: Pacific Linguistics.

Wilkins, David P. and Deborah Hill, 1995. When 'go' means 'come': Questioning the basicness of basic motion verbs. *Cognitive Linguistics,* 6.2/3: 209-259.

11 *Negation in Moluccan languages*

MARGARET FLOREY

1 Introduction

There has been a long-standing interest in negation amongst typologists (compare Croft 1991, Dryer 1989, Hovdhaugen and Mosel 1999, Kahrel and van den Berg 1994, Payne 1985), but until recently scant information on the languages of the Asia-Pacific region has been available to contribute to these investigations. Attention of late has turned to this region, with analyses of negation emerging from the increasing number of descriptions of the Austronesian and Papuan languages of East Nusantara (broadly encompassing eastern Indonesia, East Timor and Papua New Guinea) and Oceania. Central amongst these investigations are Reesink's (2002) analysis of negation in eastern Austronesian and Papuan languages and Mosel's (1999) typological overview of negation in Oceanic languages from New Caledonia, Papua New Guinea, and Polynesia. Tryon (2006) has recently expanded available information on negation in Oceanic languages with a survey of forty languages of Vanuatu. Included amongst the key typological issues in relation to typologies of negation in this region are word order and placement of the negator, the functions of negation, and whether or not a distinction is drawn between negation in declarative and imperative clauses.

Payne defines standard negation as that which applies to 'the most minimal and basic sentences' (1985:198). His analysis indicates that in all languages standard negation is marked with the addition of at least one of negative verbs, particles and derivational morphemes (1985:228). Mosel (1999:5) found that negatives in Oceanic languages most frequently were particles and verbs which filled five basic functions: negative answer to questions, negative existential, negation of non-verbal assertive predicates, negation of verbal assertive predicates and negation of imperatives. Tryon's analysis added a sixth negative function, *not yet*, to Mosel's five basic functions.

Dahl (1979) analysed the position of the negator in a sample of 240 languages and concluded that the canonical place of the negative adverb is pre-predicate. Payne (1985:224) also notes that typologically with negative particles (as with negative verbs) the common word order for SVO, VSO and VOS languages is for the particle to precede the verb, while with SOV languages 'it is equally likely to precede or follow'. Reesink asserts that the canonical pattern does not hold for the Austronesian and Papuan languages of the Moluccas and West Papua, stating that 'The distribution of clause-final NEG is widespread in the[se] languages' (2002:245). Throughout his paper, Reesink refers to the 'rigid

Michael C. Ewing and Marian Klamer, editors.
Typological and areal analyses: contributions from East Nusantara, 227-250 . Pacific Linguistics, 2010
©. This edition vested with Pacific Linguistics.

clause-final position of NEG in Austronesian languages of the Moluccas' (see, for example, 2002:263, 264)

Himmelmann (2005) includes clause-final negation amongst the eight characteristic features which he suggests distinguish two major types within a typological grouping for non-Oceanic[1] Austronesian languages: symmetrical voice and preposed possessor languages. Symmetrical voice languages account for some 60% of the approximately 800 non-Oceanic Austronesian languages and the preposed possessor languages comprise a further 25%. The remaining 15% are transitional: neither preposed possessor nor symmetrical voice languages (Himmelmann 2005:114)

The languages of central Maluku (the Moluccas) fall typologically within Himmelmann's preposed possessor group. These SVO languages largely share the following seven of the eight proposed characteristic features (2005:175):

1. no voice alternations,
2. preposed possessor,
3. alienable/inalienable distinction,
4. clear-cut differences between narrative and equational clauses,
5. person marking prefixes or proclitics for S/A arguments,
6. numerals/quantifiers follow head,
7. V-second or -final.

In regard to the eighth feature, negation, Himmelmann suggests that amongst the non-Oceanic Austronesian languages, 'negators also generally precede the negated constituent, with the exception of most preposed possessor languages, where negators usually occur in clause-final position' (2005:141-2)

Himmelmann's and Reesink's assertions about clause-final negation are of interest because they potentially identify a region of variance from the typologically more common word order of negator preceding the verb in SVO languages. However, in a linguistically diverse region that remains under-documented, both analyses relied on the small sample of Moluccan languages that was then available. Reesink's analysis drew on just four Austronesian languages from north (Taba), central (Buru) and south (Leti, Kei) Maluku (that is, 3.5% of the languages of Maluku). Reesink (2002:245) notes that the extension of his conclusions to other Austronesian languages was hampered by the absence of data. Himmelmann's analysis included eleven preposed possessor languages, three from East Timor, two from West Papua, and six which are drawn from north (Taba), central (Buru, Alune) and south (Selaru, Dobel, and Leti) Maluku. Himmelmann notes that

> '...the present survey is somewhat biased towards symmetrical voice languages. One reason for this bias is the fact that the author is most familiar with these languages. There are also far fewer publications about preposed possessor languages' (2005:115).

Himmelmann excepts one of the Moluccan languages, Leti, from the typological characterization of clause-final negation in preposed possessor languages.

---

1 Himmelmann (2005) uses the term 'western Austronesian' to refer to the non-Oceanic Austronesian languages.

In the light of these recent papers it is timely to expand on the analysis of negation in Moluccan languages, drawing on a wider database that is becoming available through new research.[2] The present analysis focuses on seven central and one north Moluccan language (see Map 1). It draws on the documentation in progress of four languages in central Maluku – Kouro (Ethnologue code AMQ), Allang (ALO), Sou Amana Teru (TLU) and Haruku (HRK). Further data are drawn from available descriptions of three central Moluccan languages – Alune (ALP) (Florey 2001; 2005; Florey and Kelly 2002; and additional unpublished data), Nuaulu (NXL) (Bolton 1990), and Buru (MHS) (Grimes 1991). One north Moluccan language, Taba (MKY) (Bowden 1998), is also included in this analysis.

Map 1: Indonesia and the islands of Maluku and North Maluku Provinces

---

2 This chapter draws on data from the research project *Endangered Maluku languages: Eastern Indonesia and the Dutch diaspora*, which was funded by a Major Documentation Project grant (MDP0009) from the Hans Rausing Endangered Languages Project, SOAS, UK, and by an Australian Research Council Discovery Project grant (DP0343379). I am indebted to my colleagues Michael Ewing and Simon Musgrave for sharing their data for Allang and Sou Amana Teru respectively, and for numerous discussions of the analyses. We are very grateful to the many people who have assisted us in gathering the data reported here. I also thank Tania Strahan and two anonymous reviewers for valuable feedback. This chapter was completed while the author was a visiting scholar at the Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands, and I thank Prof. Gunter Senft and the MPI for providing a stimulating and convivial work environment.

Language naming is complex in the Moluccan region and Ethnologue codes (Gordon 2005) are included here to give a general point of reference for the languages. Few ethnolinguistic groups have a composite name for their language and there is often no agreement amongst members of speech communities about composite names. Indigenous languages may be referred to by the local word for 'language'. For example, *Kouro* 'language' is used here to refer to the language spoken in south central Seram Island in the five villages of Amahei, Rutah, Makariki, Soahuku and Haruru,[3] which is most commonly called *Amahai* in the literature. Other groups have adopted a new name to refer to their language: for example, *Sou Amana Teru* 'the language of three villages' is used to refer to the language spoken in eastern Ambon Island in the villages of Tulehu, Tial and Tengah-tengah. *Allang* is the name of a village in western Ambon Island that shares a language with the neighbouring villages of Wakasihu and Larike.[4]

The seven central Moluccan languages that are analysed in this chapter are subgrouped within the Central Malayo-Polynesian branch of the Austronesian language family. At a lower level, Collins (1983) subgroups Buru within Proto-West Central Maluku and Nuaulu within the East Seram branch of Proto-East Central Maluku. Alune, Kouro, Haruku, Allang, and Sou Amana Teru are all members of the Nunusaku branch of Proto-East Central Maluku: Alune in the Three Rivers branch and Kouro, Haruku, Allang, and Sou Amana Teru in Proto-Piru Bay. Taba is a north Moluccan language.

This chapter seeks to examine the functional range of negative constructions and the position of negative particles in these constructions, and whether a distinction is drawn between negation in declarative and imperative clauses. It addresses the question of whether clause-final negation is indeed widespread amongst the languages of the (central) Maluku and whether this feature is useful as an eighth typological characteristic of preposed possessor languages. The data analysed here indicate that negation in this region is more complex than indicated by earlier analyses. Cross-linguistically, Moluccan languages exhibit a number of different negative constructions, including pre-predicate negation, post-predicate and clause-final negation, and double particle (Givón 1984:66) or the so-called 'embracing' negation (exemplified by French *ne* V *pas*) (Reesink 2002:241). Further, language internally, different word orders for the placement of negative particles may be used for different negative functions. However, rather than limiting the analysis to the primary negation functions (declaratives and imperatives), this chapter also examines negation in existential clauses as well as in two constructions *not yet* and *no longer* formed through the interaction between negative particles and aspectual and modal particles. This broader analysis suggests that an investigation of negation focused solely on primary negators is not sufficiently nuanced and that further evidence concerning clause-final (or post-predicate) negation may be found through a cross-linguistic examination of a wider range of negative constructions.

---

3 The inclusion of Haruru in this language group is controversial. Local people from these five villages insist that they are all one ethnolinguistic group. Collins (1983) separates Haruru from the Amahai language and subgroups it closely with Yalahatan and Awaiya. C. Beckley (SIL) is undertaking research in Yalahatan, and notes (p.c.) that Yalahatan people and her own research identify Haruru and Yalahatan as members of one ethnolinguistic group.

4 Most of the *Kouro* data presented here were recorded in Rutah by the author, with some earlier data collection in Soahuku and Makariki. Most of the Sou Amana Teru data were recorded by Musgrave in Tulehu, while Allang was the primary site for Ewing's data collection.

2 Primary negators

This section examines primary negative constructions in declarative and imperative clauses in the languages, and also in existential clauses. Reesink (2002) notes that the results of Kahrel's (1996) survey of negation in declarative and imperative clauses in 40 languages indicate that a distinction is drawn between these in approximately 50% of the world's languages. The eight Moluccan languages described here share that characteristic and use distinct strategies to negate declarative and imperative clauses (as also observed by Reesink 2002:244). The most frequently occurring order of elements within transitive clauses for the eight languages in this sample is SVO. All of the languages use particles to mark negation. A summary of the data is tabulated at the end of each section for cross-linguistic reference.

2.1 Negation in declarative clauses

Amongst the eight sample languages, there are three different word orders for placement of the negative particle. Three central Moluccan languages (Alune, Buru, Nuaulu) and the north Moluccan language Taba display post-predicate negation of declaratives. In the central Moluccan languages Kouro, Sou Amana Teru and Allang, the negative particle precedes the predicate, and Haruku uses an embracing construction.

Person-marking on negative declarative particles

In negative declarative clauses, three of the languages – Kouro, Sou Amana Teru and Haruku – share the feature that the negative particle optionally takes a pronominal subject-marking enclitic, the presence of which suggests a verbal origin for the negator. The enclitic may co-occur with a full lexical noun phrase and/or a proclitic on the verb, exemplified in (5) below in Kouro with the 3s pronominal enclitic *-ñ* cross-referencing the subject 'my father'. Note that the subject is further cross-referenced with a 3s proclitic on the verb in this clause.[5] In contrast, in (6) the 3s pronominal enclitic on the negator *aya* is the sole subject marker in the negated clause, whilst in (7) the subject is not overtly expressed, and can be traced only through the context of an earlier question. The contrast between (8) and (9) illustrates the optionality of person marking on the negator in Sou Amana Teru. In the Haruku embracing construction the second element of the negative optionally takes a subject-marking enclitic pronoun which may co-occur with a full subject lexical noun phrase, as illustrated in (11). Examples (12) and (13) illustrate the optionality of person marking on the negator in Haruku and demonstrate that all other elements of the verb phrase follow the negator, including prepositional phrase adjuncts (as in 13).

Post-predicate negation

Declaratives in Alune, Buru, Nuaulu and Taba are negated with a particle which follows other elements of the clause except for the modal and aspectual particles (Bolton 1990:129; Bowden 1998:399; Grimes 1991:232f)

---

5 In central Moluccan languages more widely, subjects may occur as a full noun phrase, a pronoun, or a proclitic, or NP plus cross-referencing proclitic, or pronoun plus cross-referencing proclitic (described in Florey 2001, 2005). These patterns are exemplified in (1) for Alune, in (30) for Kouro, in (31) for Sou Amana Teru, in (44) for Nuaulu, in (45) for Taba, and in (50) for Haruku. Argument marking in Allang is described in Ewing (this volume).

(1) Alune
i-mei *i-lepa* *lo'o* *Lisona* *be* *su'a* *i-'ai*
3sh-LOC 3sh-talk ALL Lisona COMP want 3sh-marry

po *Lisona* *i-su'a* ***mo***
but Lisona 3sh-want NEG
'He said to Lisona that he wanted to marry her but Lisona didn't want to.'

(2) Buru (Grimes 1991:232–166)
sira *hapu* *lafa-t* *la* *yako* *langina* ***moo***
3p tie food-NM for 1s.BEN earlier NEG
'They didn't tie up trailfood for me earlier.'

(3) Nuaulu (Bolton 1990:126–14)
ne *munata* *i-amanaku* ***tewa,*** *au* *tentene* ***tewa***
but if 3sh-agree NEG 1s force NEG
'But if she doesn't agree, I won't force her.'

(4) Taba (Bowden 1998:400–19)
nik *calana* *kuda-k* *asfal* ***te***
1s.POSS trousers be.black-APPL bitumen NEG
'My trousers are not blackened with bitumen.'

Pre-predicate negation

Negation of declaratives in Kouro, Sou Amana Teru and Allang is marked with a particle which precedes the predicate. However, unlike Kouro and Sou Amana Teru, negators in Allang do not take person marking.

(5) Kouro
ama-'u ***aya-ñ*** *i-supu* *ia-no*
father-1s.POSS.INAL NEG-3s 3sh-get fish-NM
'My father didn't catch any fish.'

(6) Kouro
heterue *wa-pe'e* *ri'u-'u* *wa'uwe* *uma Haji*
earlier 1s.ACT-give voice-1s.POSS.INAL BEN elder Haji

tapi ***aya-ñ*** *ponno*
but.MAL NEG-3s hear
'Yesterday I greeted Mr Haji but he didn't hear.'

(7) Kouro
aya *na* *iyane* *manuwo-lo*
NEG IRR eat hen-NM
'[I] don't eat chicken.'

(8) Sou Amana Teru
ire hose ***taha****-u berani*
3s say NEG-1s brave
'She said "I'm not brave".'

(9) Sou Amana Teru
motor-re ***taha*** *hola*
motorbike-PROX NEG sound
'The motorbike didn't start.'

(10) Allang
Nalisa ite ***ta*** *supu luma-nu*
tomorrow 1pi NEG meet RECP-NM
'We won't meet tomorrow.'

Embracing construction for negation

Payne (1985:224) notes that in the languages of the world there is a 'strong tendency for particle negatives to be emphasized and reinforced … by the addition of a further particle elsewhere in the sentence, forming a pair of linked negatives'. In the sample analysed here, Haruku is the only language which uses an embracing strategy to negate declaratives, with a pre-predicate negator *ta'* and the post-predicate negator *sa.* In other contexts, the morpheme *sa* means 'still, yet' and marks continuative aspect and, in these functions, *sa* is discussed further in §3.1 on the negator *not yet*. In the negation of declaratives, *sa* has been bleached of its aspectual function and serves to reinforce the negative particle *ta'*.

(11) Haruku
mahina ***ta'*** *na'e* ***sa****-i*
female NEG sleep NEG-3sh
'The woman is not sleeping.'

(12) Haruku
au ***ta'*** *ane* ***sa*** *hahu*
1s NEG eat NEG pig
'I don't eat pork.'

(13) Haruku
au ***ta'*** *oi* ***sa*** *heke wa'asia*
1s NEG go NEG DIR forest
'I'm not going to the forest.'

Summary

The negative particles and the three strategies used by the sample languages to mark negation of declarative clauses are summarized in the following table.

Table 1: Summary of negation of declaratives

Negation strategy	**Languages**			
Post-predicate	Alune *mo*	Buru *moo*	Nuaulu *tewa*	Taba *te*
Pre-predicate	Kouro *aya*	Sou Amana Teru *taha*	Allang *ta*	
Pre- and post-predicate (embracing construction)	Haruku *ta' V sa*			

2.2. Negation in existential clauses

In five central Moluccan languages – Alune, Buru, Nuaulu, Kouro and Sou Amana Teru – and the north Moluccan language Taba, existential clauses are negated with the same particles and placement of the negative particle as is found with negation of declaratives. Allang and Haruku differ in having dedicated negative existential particles. In the case of Haruku, the word order in negative existentials also differs to that in negation of declaratives.

Person-marking on negative existential particles

Negative existentials in Kouro, Sou Amana Teru and Haruku take a subject-marking clitic which, in the current corpus, appears to be obligatory for Kouro and Sou Amana Teru and optional in Haruku. Kouro and Sou Amana Teru follow the pattern shown in 2.1 for the negator of declaratives, with an enclitic on the negator which may co-occur with a full lexical noun phrase to cross-reference the subject, exemplified in (18) and (19) below for Kouro and (20) for Sou Amana Teru. In Haruku, subject-marking on negative existentials differs from the pattern discussed in 2.1 with respect to negation of declarative clauses. Instead of an enclitic pronoun, the single negative existential particle optionally takes a subject-marking proclitic, as shown in (23). The use of a proclitic accords with subject-marking for other predicate types in Haruku and other languages of this region, and highlights the verbal properties and likely verbal origin of negators in Haruku.

Existential formed with lexeme 'one'

Grimes (1991:376) and Bolton (1990:127) note that an existential predicate in Buru and Nuaulu respectively is formed with the lexeme ‘one’, which also functions as an indefinite article in Buru. In Alune, (*e*)*sa* ‘one’ has been grammaticized as an existential predicate *sae*. In each of these languages, existential predicates are negated with the same placement of the negative particle as occurs with negation of declaratives: that is, with the negator following all other elements in the clause.

(14) Buru (Grimes 1991:376–23)
geba ***sa*** ***moo***
person EXIST NEG
‘There was nobody there.’

(15) Nuaulu (Bolton 1990:127)
ia ***isa*** *eu mai* ***tewa***
3sh EXIST go here NEG
'There was no-one (who) came here.'

(16) Alune (Florey 2001:87–27)
'ane ele mere ne'a, ***sae*** *nasu* ***mo***
eat DEI DIST just EXIST anger NEG
'Meals (Lit. to eat) were like that; there wasn't any anger.'

(17) Alune (Florey 2001:87–28)
due mei hena mise le'we po ***sae*** *supu 'epene* ***mo***
stay LOC village good also but EXIST get money NEG
'Staying in the village is good too, but there isn't any money to be had.'

Negative existential predicate

In Kouro, Sou Amana Teru and Taba, the negative particle functions as a predicator in a negative existential clause. In these three languages, the negative particle has the same form and position as it does in negation of declaratives. Bowden (1998:401, 473) notes for Taba that the discourse marker *mai* (glossed 'well, but') commonly precedes *te* in negative existential constructions to mark, for example, that the situation is counter to one's expectations (as in 22)

(18) Kouro
heterue wa-oi mai tapi Bapak Rahaban ***aya-ñ***
earlier 1s-go PROX but.MAL Mr.MAL Rahaban NEG.EXIST-3s
'I came here earlier but Mr Rahaban wasn't here.'

(19) Kouro
ma-reu nya mo ana-yo ***aya****-si*
1p-return.home RLS but child-PL NEG.EXIST-3p
'We wanted to go home but the children weren't there.'

(20) Sou Amana Teru
ana ko'i-e ***taha****-yisi*
child small-NM NEG.EXIST-3p
'The little children were not there.'

(21) Taba (Bowden 1998:401–21)
nik dalawat ***te***
1s.POSS girlfriend NEG.EXIST
'I don't have a girlfriend.'

(22) Taba (Bowden 1998:401–22)
te mai ***te*** *gula mai* ***te*** *kofi mai* ***te***
tea but NEG.EXIST sugar but NEG.EXIST coffee but NEG.EXIST
'There's no tea; there's no sugar; there's no coffee.'

Dedicated negative existential particle

Unlike the other languages in this sample, Allang and Haruku each have a dedicated existential negative particle which differs in form to the negator of declaratives. Whereas Allang *ta* negates declaratives, *tahi* is the existential negative particle. In both cases, the negative particle precedes the negated constituent. As discussed in §2.1, Haruku uses an embracing construction *ta'* V *sa* to negate declaratives. In contrast, a single existential negative particle *taha* precedes the negated constituent.

(23) Haruku
*esi-**taha*** *eni-asu*
3p-NEG.EXIST 3s.POSS.AL-dog
'They don't have any dogs.'

(24) Haruku
taha *pa'u* *uwei*
NEG.EXIST sago.biscuits source
'There aren't any sago biscuits.'

(25) Allang
tahi *untunu*
NEG.EXIST advantage
'There's no advantage.'

(26) Allang
nalia *hima* *ma* *tau* *agama* *Kristen* *sala* ***tahi***
year DIST DIST NEG.PFV religion.MAL Christian still NEG.EXIST

agama *mati* *parcaya* *ria* *upu* *lani-ta* *mana*
religion.MAL 3p.ACT believe.MAL exist TITLE sky-NM just
'In those years there wasn't yet Christianity. There wasn't religion. They just believed in the god of the skies.'

Summary

The strategies used by the sample languages to mark negation of existential clauses are summarized in the following table.

Table 2: Summary of negation of existential clauses

Negation strategy	**Languages**		
Existential predicate + post-predicate NEG	Buru *sa(a) X moo*	Nuaulu *isa X tewa*	Alune *sae X mo*
NEG particle = NEG existential predicate	Taba *te*	Sou Amana Teru *taha*	Kouro *aya*
Dedicated (pre-predicate) NEG existential particle	Allang *tahi*	Haruku *taha*	

2.3 Negation of imperatives

Clause-initial or pre-predicate negative particles are used to mark negation of imperatives in six languages, and only Alune and Taba display post-predicate (and clause-final) negation of imperatives. Note, however, that a further four of the sample languages, Kouro, Sou Amana Teru, Buru and Allang also have an emphatic imperative, and in three of these emphatic imperative constructions one of the negative particles is clause-final. Negative imperative particles do not take subject-marking in any of the languages. This is not surprising, given that subjects are commonly not expressed in imperative constructions.

Post-predicate negation

In Alune and Taba, negation of an imperative clause is marked in clause-final position with Alune *ya'e* and Taba *oik*.

(27) Alune
'ane pia ***ya'e***
eat papeda NEG.IMP
'Don't eat sago porridge!'

(28) Alune
lalei-mu *nete* *lo-ri* ***ya'e***
emotions-2s.POSS.INAL remember ALL-zone.2.LOC NEG.IMP
'Don't think of home!'

(29) Taba (Bowden 1998:401–23)
h=momas meu komo mai h=momas-ak meu calana ***oik***
2p=wipe 2p.POSS hand but 2p=wipe-APPL 2p.POSS trousers NEG.IMP
'Wipe your hands, but don't wipe them with your trousers.'

Pre-predicate negation

Negation of an imperative clause is marked with a pre-predicate negative imperative particle in Kouro, Sou Amana Teru, Haruku, Buru, Nuaulu, and Allang**.**

(30) Kouro
Preti, ***hakai*** *u-hanu-'e* *ari-mo* *ena*
Preti NEG.IMP 2s-awaken-APPL younger.sibling-2s.POSS.INAL first
'Preti, don't wake your younger sibling.'

(31) Sou Amana Teru
nina i-hose ***ehe'e***
mother 3s-say NEG.IMP
'Mother said "Don't!".'

(32) Sou Amana Teru
masasusu ***ehe*** *loru-r mena, maela* ***ehe*** *rua'a-r mena*
sweet NEG.IMP swallow-3nh first bitter NEG.IMP spit-3nh first
'Don't immediately swallow the sweet, don't immediately discard the bitter.'

(33) Haruku
ehe'e *puri* *au.* *au* *ta'* *kewa* *sa*
NEG.IMP ask 1s 1s NEG know NEG
'Don't ask me. I don't know.'

(34) Buru (Grimes 1991:214–53)
bara *e(p)-mata* *kono*
NEG.IMP CAUS-die 1SU
'Don't kill me!'

(35) Nuaulu (Bolton 1990:127–18)
pene *o-ai* *ana* *rei-mo*
NEG.IMP 2p-eat rice this-PROX
'Don't eat that rice.'

The Allang negative imperative is marked by clause initial *naka*, regardless of whether the second person subject is explicitly expressed, as in (36) or not, as in (37).

(36) Allang
naka *ane* *lepa*
NEG.IMP 2S.ACT speak
'Don't talk!'

(37) Allang
naka *lepa* *sou* *Belanda*
NEG.IMP speak language Dutch
'Don't speak Dutch!'

In addition to the negative imperative *naka* illustrated above, Allang has an admonitive discourse particle *ya'a*, which can stand alone, as in (38). This particle is cognate with the Alune negative imperative *ya'e*. Allang *ya'a* commonly precedes negative imperatives yielding by far the most common construction with double marking of the imperative (Michael Ewing, p.c.). This is exemplified in (39).

(38) Allang
ina *hete-nu* ***ya'a***
mother say-NM ADMON
'Mother said "Don't".'

(39) Allang
ya'a ***naka*** *snapau* *kakei*
ADMON NEG.IMP walk fast
'Don't walk fast!'

Emphatic negative imperative

Emphatic negative imperative constructions are also found in Kouro (40) and reportedly in Sou Amana Teru (Simon Musgrave, p.c.), with the negative particle optionally co-occuring in post-predicate position. The Buru emphatic imperative construction is formed

with *bara* pre-predicate and the negative particle *moo* in post-predicate (and clause-final) position (41).

(40) Kouro
hakai *u'u-mi* ***hakai***
NEG.IMP noisy-2p.U NEG.IMP
'Don't you all be so noisy!'

(41) Buru (Grimes 1991:214–54)
ma ***bara*** *em-tako* ***moo***
1p NEG.IMP STAT-fear NEG
'We shouldn't be afraid.'

Summary

The particles and strategies used by the sample languages to mark negation of imperatives are summarized in the following table.

Table 3: Summary of negation of imperatives

Negation strategy	**Languages**					
Post-predicate	Alune *ya'e*	Taba *oik*				
Pre-predicate	Kouro *hakai*	Sou Amana Teru *ehe('e)*	Haruku *ehe'e*	Buru *bara*	Nuaulu *pene*	Allang (1) *naka*
Emphatic: Pre-predicate						Allang (2) *ya'a naka*
Emphatic: Pre- and post-pred. (embracing)	Kouro *hakai X hakai*	Sou Amana Teru *ehe('e) X ehe('e)*		Buru *bara X moo*		

3 Complex constructions

Two negative constructions are formed in the eight sample languages through interaction between negative particles and aspectual and modal particles, giving meanings such as 'not yet, still not yet' (Indonesian *belum*) and 'no longer, no more, not/nothing at all' (Indonesian *tidak lagi*)

Person-marking on negative particles

In the two complex negative constructions, both Kouro and Sou Amana Teru again display subject-marking on the negator. In each case, the languages use an embracing construction, and the first (pre-predicate) element of the negative optionally takes a subject-marking enclitic pronoun which may co-occur with a full lexical noun phrase and/or a proclitic on the verb. This is illustrated for the *not yet* construction in (46) and

(47) for Kouro (with a 3s human and non-human subject respectively), and (48) for Sou Amana Teru (with a 3p subject), and for the *no longer* construction in (60) for Sou Amana Teru (with a 1s subject). There is no evidence in the data that Haruku permits subject-marking on the negator in these two complex constructions.

3.1 Not yet

In all the sample languages, a *not yet* construction is formed with the negative particle in combination with a morpheme that has one or more functions in the languages, commonly meaning 'still, yet' and marking continuative aspect, and, in some languages, marking irrealis. In seven of the eight languages, the second operator in the *not yet* construction occurs in post-predicate position. In Haruku alone both elements occur in pre-predicate position.

Post-predicate negation

In Alune, Buru, Nuaulu and Taba, *not yet* is post-predicate. In each case, the negative particle is the same as that used in the negation of declaratives and existential clauses. The second particle of the *not yet* construction encliticizes to the negator and the resultant complex construction is clause-final.

(42) Alune
au 'eu po au bete-' 'ena ina-'u
1s go but 1s say-APPL to mother-1s.POSS.INAL

'ai ama-'u ***mo-sa***
and father-1s.POSS.INAL NEG-still
'I'd go but I haven't yet told my mother and my father.'

(43) Buru (Grimes 1991:236–185)
da kaa ***mo-hede***
3s eat NEG-still
'He hasn't eaten yet.'

(44) Nuaulu (Bolton 1990:129–22)
au u-apahia-i ***tewa-si***
1s 1s-open-3s NEG-still
'I haven't opened it yet.'

(45) Taba (Bowden 1998:403–29)
karna ta-plod ***te-hu*** *manusia l=oas-do*
because DET-erupt NEG-still people 3p=flee-RLS
'Because the mountain had still not erupted when everyone fled.'

Pre- and post-predicate negation

In Kouro and Sou Amana Teru, the *not yet* negative is formed with an embracing construction with the negative particle preceding the negated constituent and the second element in post-predicate position. The negative particle is the same as that used in the

negation of declaratives and existential clauses. *Kala* in Kouro and *sala* in Sou Amana Teru encompass the meanings 'still, yet' and mark continuative aspect.

(46) Kouro
ina-'u *Ona* ***aya-ñ*** *i-reu*
mother-1s.POSS.INAL Ona NEG-3s 3sh-return.home

heri'e *mirim-o* ***kala***
from garden-NM still
'Mrs Ona hasn't come home from the garden yet.'

(47) Kouro
mabulaa-no *ne* ***aya-ñ*** *e-puhua* ***kala***
mango-NM this.(near) NEG-3s 3snh-bear.fruit still
'This mango tree hasn't yet fruited.'

(48) Sou Amana Teru
yau *a'a* *mahina-e* *tula* *yau* *ina-u*
1s older.sibling female-NM with 1s mother-1s

taha*-yisi* *upa* ***sala***
NEG-3p sit still
'My big sister and mother didn't sit yet.'

Pre-predicate negation

As discussed in §2 above, Haruku negative declaratives are formed with the embracing construction *ta' X sa* in Haruku. *Not yet* constructions in Haruku are also formed with two particles – in this case, both are placed before the predicate. *Ta'u* is followed by *sa*, which is cognate with the operators found in other languages (Alune, Nuaulu, Allang, Sou Amana Teru, Kouro) that mark continuative aspect and encompass the meanings 'still, yet'.

(49) Haruku
au ***ta'u*** ***sa*** *ninu* *kopi* *ke* *wae-le* *putui*
1s NEG still drink coffee or water-NM warm
'I haven't yet drunk either coffee or tea.'

(50) Haruku
asu *ito* ***ta'u*** ***sa*** *e-ane*
dog DIST NEG still 3snh-eat
'That dog still hadn't eaten.'

Dedicated* not yet *particle

In §2 we saw that Allang *ta* negates declaratives and that *tahi* is a dedicated negator for existential clauses. Allang also has a dedicated negator *tau* which occurs pre-predicate to mark *not yet*, as exemplified in (51). It is the only language in this database which can mark *not yet* with a single morpheme. However, a second *not yet* construction also occurs in Allang – in this case an embracing construction mirroring the Kouro and Sou Amana Teru pattern with the negator *tau* in pre-predicate position and *sala* 'still, yet' in post-

predicate position. The construction *tau* X *sala* is illustrated in (26) above and (52) below. Michael Ewing (p.c.) estimates that approximately 50% of *not yet* constructions are formed with the two elements *tau* and *sala*, and suggests that *sala* is grammaticizing to become part of this negative construction.

(51) Allang

ina	***tau***	*kari'i*	*a'a*	*undana*	*mana*	*ana*
mother	NEG.PFV	see	older.sibling	male	3s.POSS	child

'Mother hadn't yet seen my older brother's child.'

(52) Allang

mane	***tau***	*wito*	*mana*	*ulo*	***sala***
3sh.ACT	NEG.PFV	tie.in.bun	3s.POSS	head	still

'She hadn't put her hair up yet.'

Summary

The particles and strategies used by the sample languages to mark *not yet* are summarized in the following table.

Table 4: Summary of *not yet*

Position of NEG	**Languages**			
Post-predicate	Alune *mo-sa*	Buru *mo-hede*	Nuaulu *tewa-si*	Taba *te-hu*
Pre- and post-predicate (embracing construction)	Kouro *aya X kala*	Sou Amana Teru *taha X sala*	Allang (2) *tau X sala*	
Pre-predicate	Haruku *ta'u sa X*			
Dedicated (pre-predicate) *not yet* particle	Allang (1) *tau*			

3.2 No longer

A second complex negative construction occurs in the sample languages to encode 'no longer, no more, not/nothing at all'. A negator combines with a morpheme that marks perfective and/or, in some languages, realis, and encompasses the meanings 'any more, any longer, just, simply'.

Post-predicate negation

In Alune, Nuaulu and Taba, the two particles in the *no longer* construction are both post-predicate. The same particle used in negation of declaratives and existentials is in penultimate position in the clause, and the second particle of the construction follows, encliticized to the negator in Alune and Taba. Alune *ne('a)* encompasses meanings such as 'any more, any longer, just, simply', and forms part of *pene'a*, perfective 'already'. Bolton (1990) analyses *nea* in Nuaulu as a tense marker for completed events or perfective 'already'. Realis in Taba is marked with *do*, which can combine with the negator *te* to form what Bowden (1998:402) analyses as a realis negative.

(53) Alune

ele'i	*meme-i*	*Si'ua*	*meije*	*e-supu*	*ma-'ane*	***mo-ne'a***
then	uncle-3sh.POSS.INAL	Si'ua	PROX	3snh-able	NMLZ-eat	NEG-longer

supu	*'eu*	*'atu*	***mo-ne'a***
able	go	defecate	NEG-longer

'Then her uncle Si'ua he was no longer able to eat, no longer able to shit.'

(54) Nuaulu (Bolton 1990:129–25)

i-rori	*muaine*	*osi-ki*	***tewa***	***nea***
3sh-bring	food	to-3s	NEG	PFV

'He didn't bring food to him any more.'

(55) Taba (Bowden 1998:402–27)

mai	*a-ne*	*l=pe-ik*	*saguer*	***te-do***
but	DEM-PROX	3p=make-APPL	palm.wine	NEG-RLS

'But here they don't make palm wine with it any more.'

In Buru, the morpheme *tehuk* 'longer' combines with the negator *moo* to form the complex negative *tehuk moo* 'no longer', in which both elements are post-predicate. Grimes (1991:236) notes that *tehuk* can occur in three positions in the clause, as illustrated below, including the (reportedly rare) occurrence in two slots, but it always precedes the negative particle, which is clause-final. Buru thus contrasts with Alune, Nuaulu and Taba in the order of elements in this negative construction.

(56) Buru (Grimes 1991:236–187)

da	*kaa*	*gehu-t*	***tehuk***	***moo***
3s	eat	taro-NM	longer	NEG

'She doesn't eat taro any more.'

(57) Buru (Grimes 1991:236–188)

da	*kaa*	***tehuk***	*gehu-t*	***moo***
3s	eat	longer	taro-NM	NEG

'She doesn't eat taro any more.'

(58) Buru (Grimes 1991:236–189)

da	*kaa*	***tehuk***	*gehu-t*	***tehuk***	***moo***
3s	eat	longer	taro-NM	longer	NEG

'She doesn't eat taro any more.'

Pre- and post-predicate negation

In Kouro, Sou Amana Teru and Haruku, *no longer* is formed with the negator of declaratives in pre-predicate position. The aspectual morpheme is in clause-final position.

(59) Kouro

tapi	***aya***	*ne*	*bisa*	*na*	*ko-hi'a-no*	***nya***
but.MAL	not	PROX	able.MAL	IRR	1pe-pound-NM	PFV

'But it's not possible for us to pound sago any more.'

(60) Sou Amana Teru
ire hose sa'a eeee
3s say climb [hes]

***taha**-u sa'a **ea**; yau kere'e-u*
NEG-1s climb PFV 1s afraid-1s
'He said climb on, I didn't climb on then; I was afraid.'

(61) Haruku
*ie-le ti **ta'** makahina **seiya***
thing-NM this.(near) NEG beautiful PFV
'This thing is no good any more.'

In Allang, *no longer* is also formed with an embracing construction with *tahi*, the negator of existential clauses in pre-predicate position and the second morpheme in clause-final position.

(62) Allang
*au **tahi** palime-nu **mana***
1s.ACT NEG play-NM PFV
'I didn't play at all.'

Summary

The particles and strategies used by the sample languages to mark *not yet* are summarized in the following table.

Table 5: Summary of *no longer*

Position of NEG	**Languages**			
Post-predicate	Alune *mo ne('a)*	Nuaulu *tewa nea*	Taba *te do*	Buru *tehuk moo*
Pre-and post-predicate (embracing construction)	Kouro *aya X nya*	Sou Amana Teru *taha X ea*	Haruku *ta' X seiya*	Allang *tahi X mana*

4 Discussion

Data from eight SVO Moluccan languages (including seven central Moluccan and one north Moluccan language) have been analysed in this areal study of negation. Widening the database of central Moluccan languages from the two (Buru and Alune) used in previous analyses of negation in this region has revealed a range of negation constructions, both in what are commonly regarded as the primary negators (negating declaratives and imperatives) as well as in existential constructions and two negative constructions marking *not yet* and *no longer*. A cross-linguistic summary of the range of word orders in placement of negative particles in six negative constructions is provided in Tables 6 and 7. These tables demonstrate that the languages fall largely into two clusters. In Group 1 (Alune, Taba, Nuaulu and Buru) there is a strong preference for post-predicate negation.

The negative particle is also clause-final in the group in the negation of declaratives and existentials, and imperatives for Alune and Taba. In *not yet* and *no longer* constructions, the complex negative is clause-final in this group, with the negative particle in penultimate position.

Table 6: Summary of word order in negation – Group 1

	NEG declarative	**NEG existential**	**NEG imperative**	**emphatic NEG. imperative**	***not yet* position of NEG**	***no longer* position of NEG**
Alune	post-predicate	post-predicate	post-predicate	-	post-predicate	post-predicate
Buru	post-predicate	post-predicate	pre-predicate	pre- and post-predicate	post-predicate	post-predicate
Nuaulu	post-predicate	post-predicate	pre-predicate	-	post-predicate	post-predicate
Taba	post-predicate	NEG as predicator	post-predicate	-	post-predicate	post-predicate

In contrast, the four languages in Group 2 (Allang, Sou Amana Teru, Kouro and Haruku) – all of which subgroup within Proto-Piru Bay – show a strong preference for the negative particle to be positioned before the predicate and/or to use an embracing strategy with negative particles both pre- and post-predicate. This group is further differentiated from Group 1 by the feature that three of its languages – Sou Amana Teru, Kouro and Haruku – permit subject-marking on the negative particle in the negation of declaratives, existentials, *not yet* and *no longer*. The fact that the two complex negatives share this feature with the declaratives and existentials supports their inclusion in this analysis as it indicates that there is a cluster of features affecting negation more widely.

Table 7: Summary of word order in negation – Group 2

	NEG declarative	**NEG existential**	**NEG imperative**	**emphatic NEG. imperative**	***not yet* position of NEG**	***no longer* position of NEG**
Allang	pre-predicate	dedicated NEG existential particle	pre-predicate	pre-predicate	1) pre-predicate 2) pre- and post-predicate	pre-predicate
Sou Amana Teru	pre-predicate	NEG as predicator	pre-predicate	pre- and post-predicate	pre-predicate	pre-predicate
Kouro	pre-predicate	NEG as predicator	pre-predicate	pre- and post-predicate	pre-predicate	pre-predicate

Haruku	pre- and post-predicate	dedicated NEG existential particle	pre-predicate	-	pre-predicate	pre-predicate

As noted in §2, Kahrel (1996) found that a distinction is drawn between negation in declarative and imperative clauses in approximately 50% of the world's languages. The cross-linguistic comparison of the negative particles provided in Table 8 shows that all eight languages accord with that finding. Different negative particles are used to negate declaratives and imperatives. Five of the languages use the same word order for placement of the negative particle in both declarative and imperative clauses: post-predicate in Alune and Taba, and pre-predicate in Allang, Sou Amana Teru and Kouro (see Tables 6 and 7). In three of the languages, different word orders are used for negation in declarative and imperative clauses. In Buru and Nuaulu, the negative particle is post-predicate in negation of declaratives, and pre-predicate with imperatives. Haruku accords with the Group 2 languages in having pre-predicate negation of imperatives. However, it is the only language that uses the embracing strategy in the negation of declaratives. It is also worth noting that the second negative particle in the emphatic negative imperative in Buru, Kouro and Sou Amana Teru is post-predicate.

Table 8: Summary of negative particles

	NEG declarative	**NEG existential**	**NEG imperative**	**emphatic NEG imperative**	***not yet***	***no longer***
Alune	*mo*	*sae X mo*	*ya'e*	-	*mo sa*	*mo ne'a*
Buru	*moo*	*sa(a) X moo*	*bara*	*bara X moo*	*mo hede*	*tehuk moo*
Nuaulu	*tewa*	*isa X tewa*	*pene*	-	*tewa si*	*tewa nea*
Taba	*te*	*te*	*oik*	-	*te hu*	*te do*
Allang	*ta*	*tahi*	*naka*	*ya'a naka*	*tau (sala)*	*tahi mana*
Sou Amana Teru	*taha*	*taha*	*ehe'e*	*ehe'e X ehe'e*	*taha X sala*	*taha X ea*
Kouro	*aya*	*aya*	*hakai*	*hakai X hakai*	*aya X kala*	*aya X nya*
Haruku	*ta' V sa*	*taha*	*ehe'e*	-	*ta'u sa*	*ta' seiya*

Table 8 indicates that Allang and Haruku show the widest range in form of negative particles, contrasting with the other six languages which consistently use the same negative particles in the negation of declaratives and existentials and in the *not yet* and *no longer* constructions. In these two moribund languages it is difficult to find good evidence for the historical path of development of this wider range of negative particles.

The data analysed in this chapter indicate that an investigation of the clausal operators (aspect and mood) that interact with negative particles highlights further interesting issues. Word order in the complex negatives *not yet* and *no longer* is summarized in Tables 9 and

10. It is in the interaction between the two elements in these negatives that we see two striking patterns emerge: firstly, a strong preference for negative particles to occur at the boundaries of the verb phrase; and secondly, the noteworthy involvement of post-predicate position in negation. Table 9 demonstrates that, amongst the Group 1 languages, both elements in the *not yet* and *no longer* negatives cluster in post-predicate and clause-final position. Amongst the Group 1 languages, the second element in the *not yet* construction cliticizes to the negator. This also occurs in the *no longer* construction in Alune and Taba. In Alune (and possibly in the other Group 1 languages) *mosa* (not yet) and *mone'a* (no longer) have become unanalysable single particle negatives for younger speakers.

Three of the Group 2 languages in Table 10 (Allang, Sou Amana Teru and Kouro) position the negator pre-predicate whilst the second element in the negative is in post-predicate position, as is the case with Group 1. Haruku follows this pattern for the *no longer* construction. In contrast, it is the only language in either group in which the second element in the *not yet* construction is also pre-predicate.

Table 9: Position of negator and second particle in *not yet* and *no longer* – Group 1

	not yet	***no longer***
Alune	both particles post-predicate and clause-final	both particles post-predicate and clause-final
Buru	both particles post-predicate and clause-final	both particles post-predicate and clause-final
Nuaulu	both particles post-predicate and clause-final	both particles post-predicate and clause-final
Taba	both particles post-predicate and clause-final	both particles post-predicate and clause-final

Table 10: Position of negator and second particle in *not yet* and *no longer* – Group 2

	not yet		***no longer***	
	NEG	**2nd particle**	**NEG**	**2nd particle**
Allang (1)	pre-predicate	-		
Allang (2)	pre-predicate	post-predicate and clause-final	pre-predicate	post-predicate and clause-final
Sou Amana Teru	pre-predicate	post-predicate and clause-final	pre-predicate	post-predicate and clause-final
Kouro	pre-predicate	post-predicate and clause-final	pre-predicate	post-predicate and clause-final
Haruku	pre-predicate	pre-predicate	pre-predicate	post-predicate and clause-final

Allang and Haruku provide useful hypotheses for ongoing grammaticization in the negative constructions *not yet* and *no longer*. There are two pieces of evidence which indicate that Haruku *sa* has been bleached of its aspectual meaning 'still, yet' and

reanalysed and grammaticized as one of two negative particles (as postulated in §2.1). Firstly, the two particles *ta'(u)* and *sa* are used to form negation both of declaratives and the *not yet* construction in Haruku. The use of *sa* does not contribute any aspectual meaning in the negation of Haruku declaratives (see examples 11 to 13). Word order alone differentiates the two negatives, with an embracing construction for the former and a pre-predicate construction for the latter. Secondly, we have seen that in Sou Amana Teru and Kouro, a subject-marker optionally enclitices to the negative particle. In Haruku, the optional enclitic attaches to *sa* – the second element of the negative (see example 11). Aspectual markers do not take person marking in any of the languages, providing further evidence that *sa* has been reanalysed as a negative particle in Haruku. A similar process of bleaching and reanalysis may be implicated in the apparent grammaticization in progress of the optional Allang *not yet* construction *tau X sala*, which is seeing the development of an embracing construction.

The data analysed in this chapter and summarized in the tables in this section permit the questions raised in the introduction to be addressed. First, these data show that some Moluccan languages do provide a contrast with the typologically more common word order of negator preceding the verb in SVO languages, noted by Payne (1985). The four Group 1 languages have post-predicate (and clause final) negation. However 50% of this sample – the four Group 2 languages – accord with Payne's finding.

Second, the more substantial question concerns whether the claims can be substantiated that negation is rigidly clause-final in Moluccan languages (as suggested by Reesink 2002) and that clause-final negation is a characteristic typological feature of preposed possessor languages (Himmelmann 2005). The data analysed in this chapter have broadened the sample of SVO languages in which the negator does not precede the verb. Of the languages surveyed here, only central Moluccan Alune and north Moluccan Taba consistently use clause-final position across the range of the negative functions examined in this chapter. Nuaulu and Buru also show a strong preference for clause-final negation with only the negative imperative marked pre-predicate. However, the inclusion of the two complex negatives in this chapter demonstrates that, although negation is not rigidly clause-final in Moluccan languages, post-predicate position certainly plays an important role in the formation of negative constructions. The two complex negatives lend some support to the argument for clause-final negation as a feature of this region. In the *not yet* construction (Tables 9 and 10), at least one of the two negative particles is post-predicate and clause-final in seven of the eight languages, with Haruku being the exception. In the *no longer* construction, at least one of the two particles is post-predicate and clause-final in all eight languages.

Reesink has argued that clause-final negation in eastern Austronesian languages (which is primarily evidenced here in Group 1) is the result of change induced by contact with adjacent Papuan languages. However, this argument is contentious for the central Moluccan region in which there are no extant Papuan languages, and little synchronic evidence can be found to support that hypothesis. It may more precisely be applied to the South Halmahera West New Guinea (SHWNG) languages of north Maluku and West Papua, which are the primary focus of Reesink's analysis. Clause-final negation in some of the languages of central Maluku may be an ancient change resulting from contact, but could also be an independent development.

This areal analysis has shown the complexity of negative constructions, and may suggest areas of further typological research. The range of negative constructions which is found in Maluku suggests that clause-final negation may not be useful as an eighth typological characteristic of preposed possessor languages. These questions will continue

to benefit from data emerging from the growing number of languages that are currently being documented throughout the East Nusantara region.

References

Bolton, Rosemary A., 1990. A preliminary description of Nuaulu phonology and grammar. Arlington: The University of Texas at Arlington. MA thesis.

Bowden, John, 1998. Taba (Makian Dalam); description of an Austronesian language from eastern Indonesia, unpublished ms. University of Melbourne.

Collins, James T., 1983. *The historical relationships of the languages of Central Maluku, Indonesia*. Canberra: Pacific Linguistics.

Croft, William, 1991. The evolution of negation, *Journal of Linguistics*, 27:1–27.

Dahl, Östen, 1979. Typology of sentence negation, *Linguistics*, 17:79–106.

Dryer, Matthew S., 1989. Universals of syntactic typology. In Edith Moravcsik, Jessica Wirth and Michael Hammond, eds, *Studies in syntactic typology*, 93–124. Amsterdam: John Benjamins.

Florey, Margaret, 2001. Verb and valence in Alune. In Peter Austin, Barry J. Blake and Margaret Florey, eds, *Explorations in valency in Austronesian languages*, 73–120. Melbourne: La Trobe Papers in Linguistics.

– 2005. A cross-linguistic perspective on emergent possessive constructions in Central Moluccan languages. In Margaret Florey and Patrick McConvell, eds, *Language shift, code-mixing and variation* (Special volume of) *Australian Journal of Linguistics,* 25:59–84.

Florey, Margaret and Barbara F. Kelly, 2002. Spatial reference in Alune. In Giovanni Bennardo, ed., *Representing space in Oceania; Culture in language and mind*, 11–46. Canberra: Pacific Linguistics.

Givón, Talmy, 1984. *Syntax: a functional-typological introduction.* Vol. 1. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Gordon, Raymond G., Jr., 2005. *Ethnologue: Languages of the World*. SIL International. Dallas, Texas: SIL International.

Grimes, Charles E., 1991. The Buru language of eastern Indonesia. Canberra: The Australian National University. PhD dissertation.

Himmelmann, Nikolaus P., 2005. The Austronesian languages of Asia and Madagascar: typological characteristics. In Alexander Adelaar and Nikolaus P. Himmelmann, eds, *The Austronesian Languages of Asia and Madagascar*, pp. 110–81. London: Routledge.

Hovdhaugen, Even and Ulrike Mosel, eds, 1999. *Negation in Oceanic languages*. Muenchen: Lincom Europa.

Kahrel, Peter, 1996. Aspects of negation. Amsterdam: University of Amsterdam. PhD dissertation.

Kahrel, Peter and René van den Berg, eds, 1994. *Typological studies in negation*. Amsterdam: John Benjamins.

Mosel, Ulrike, 1999. Towards a typology of negation in Oceanic languages. In Even Hovdhaugen and Ulrike Mosel, eds, *Negation in Oceanic languages*, 1–19. München: Lincom Europa.

Payne, John R., 1985. Negation. In Timothy Shopen, ed., *Language typology and syntactic description. Volume 1: Clause structure*, 197–242. Cambridge: Cambridge University Press.

Reesink, Ger P., 2002. Clause-final negation: structure and interpretation, *Functions of Language*, 9:239–68.

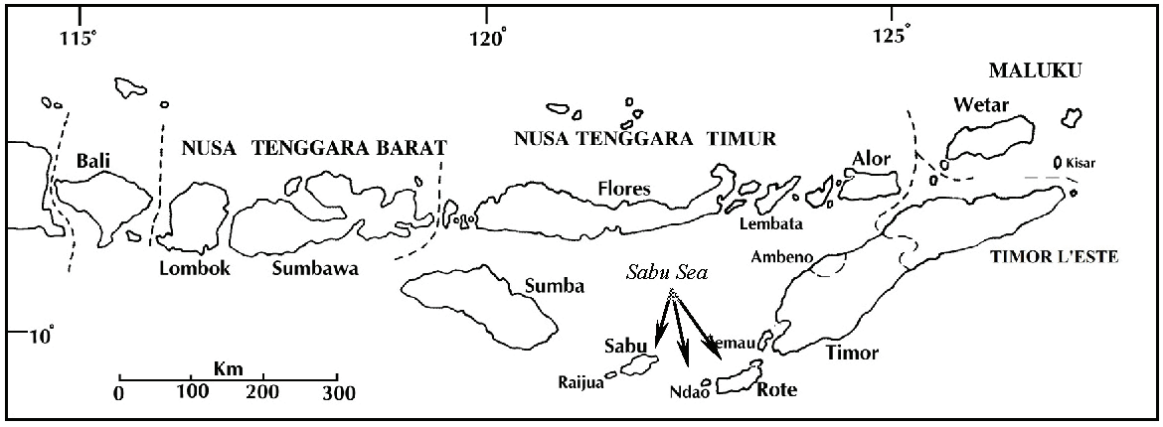
Tryon, Darrell, 2006. Negation in the languages of Vanuatu. Paper presented at the *Tenth International Conference on Austronesian Linguistics*, Puerto Princesa City, The Philippines.

12 *Hawu and Dhao in eastern Indonesia: revisiting their relationship*[*]

CHARLES E. GRIMES

1 Introduction

How the languages of eastern Indonesia fit into the bigger picture of Austronesian languages has long been the subject of debate (see C. Grimes 2000 for a summary). The debate has been based on sketchy, sporadic, and often unreliable data. One recurring question has been whether the Austronesian languages surrounding the Sabu Sea group with the western Austronesian languages or with the Austronesian languages to the east (see Map 1). Arthur Capell (1944–45/15:19) groups the languages to the north and east of the Sabu Sea with the 'Timor subgroup' which he associates with the eastern Austronesian languages, but then notes 'Curiously enough, Savu is very different.' This idea of Hawu being 'different' occurs repeatedly in the linguistic literature over the past 150 years.



Map 1: The section of eastern Indonesia with Sabu, Ndao, and Rote islands (indicated by arrows)

[*] An earlier version of this paper was presented at the 10th International Conference on Austronesian Linguistics, Puerto Princessa, Philippines, 17–20 January 2006. I am grateful to Joseph Grimes and Michael Ewing for helpful comments.

Michael C. Ewing and Marian Klamer, editors.
Typological and areal analyses: contributions from East Nusantara, 251–280. Pacific Linguistics, 2010
©. This edition vested with Pacific Linguistics.

The Hawu language and culture have a relatively long history of descriptions (for example, Donselaar 1872, Riedel 1889, Kern 1892; Wijngaarden 1896, Jonker 1919, Radja Haba 1958, Walker 1982, Fox 1972, 1977). Material has also been published in the vernacular (for example, YPPII 1985, and Ly, *et al* 2006).

In contrast, very little has been written specifically about Dhao (for example, Jonker 1903). Some works focusing primarily on other topics also include sketches of Dhao (for example, Walker 1982 and Grimes 1999). Vernacular materials have only recently begun to be published (for example, Ranoh *et al*, 2000, 2004, 2005, 2006).

The obvious similarities in the lexicons between Hawu and Dhao have led some observers to claim that Dhao is a dialect of Hawu. However, after looking at more than just the lexicon, Walker (1982) tentatively suggested they could be considered separate languages. This study examines the question more closely by comparing a number of subsystems, and concludes that in their current forms the two speech varieties cannot reasonably be considered dialects of the same language. What becomes obvious from the data in this present study are that the vast majority of the grammatical functors are different between the two languages, and so it should come as no surprise that intelligibility is blocked.

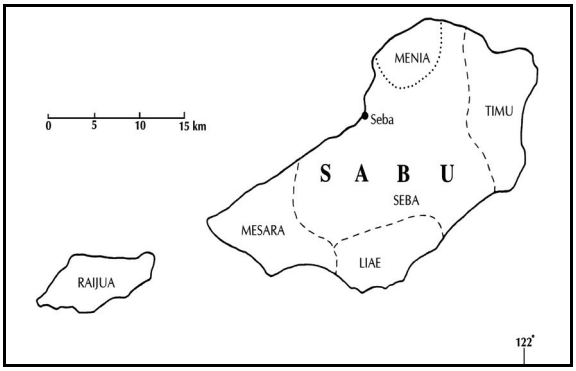
This study also raises some interesting questions that are beyond its scope to address. On methodology, it shows severe limitations to drawing definitive conclusions simply based on comparing word lists. Assuming the obvious similarities imply the two speech varieties diverged from a common source, in the area of typology and classification, if the lexicons appear so similar, but the grammars and typologies are so different, how did they get that way? On language contact and change; what happened (assuming multiple contributing factors) to cause the two speech varieties to become so different at such a fundamental level in what appears must have been a relatively short period of time? Does one variety represent a conservative, relatively unchanged language, and the other reflect radical change, or did they both undergo significant change from an earlier form?

2 A note on language and island names[1]

The Hawu language and its dialects have no /s/. Versions of the language name written with /s/ trace back to the Portuguese era, and commonly occur through the Dutch era. They continue to be preserved in the Indonesian era, apparently preserving names on maps from the Dutch without further investigation. There is an /s/ → /h/ correspondence in loans. The sound written /w/ in Hawu ranges across dialects and various speakers from a light labiodental fricative (upper teeth to inner lower lip), to a light bilabial fricative, to a semivowel. The light labiodental fricative seems most prevalent, but the older generation insists it be written as a /w/, with written Dutch as their reference point, whereas many of the younger generation prefer the symbol /v/ with reference to Indonesian and English

[1] My Hawu data are primarily from Bernadus Lado (Seba dialect, deceased May 2007), Rev. Simon Tari (Dimu dialect, deceased March 2007), and Rev. Thomas Ly (Dimu dialect). I have worked with approximately 15 other speakers, male and female, young and old, from the Seba, Dimu, and Raijua dialects. My exposure to the Liae and Mehara dialects has been limited to date. Menia is subsumed within the Seba dialect. An online resource is C. Grimes (2008). My Dhao data are primarily from Lazarus Aplugi, Michael Sina, Rev. Dr. Ayub Ranoh, Paul Ledo, Helena Aplugi, and Lazarus Lusi. I have worked with approximately 25 other speakers, male and female, young and old. My contact with Dhao is significantly greater than my contact with Hawu, ranging into a few months. An online resource is C. Grimes, *et al* (2008).

writing conventions. The name of the island continues to be written as Sabu within the political system under which it currently functions. Outsiders in the region do not fricativise the intervocalic consonant and pronounce it as [sabu]. Hawu is most fully described in Walker (1982). Map 2 shows the location of Hawu dialects.



Map 2: The princedoms (and corresponding dialects) on Sabu Island[2]

The Dhao language has no /nd/ sound or sequence. The spelling and pronunciation as 'Ndao' come from the dominant Rote languages on near-by Rote Island (described in Fox and Grimes 1995). The /dh/ digraph in the practical orthography represents a slightly retroflexed and slightly affricated voiced dental obstruent represented for convenience as [ɖ̪]. The name of the tiny island continues to be Ndao within the political system under which it currently functions. A sketch of Dhao phonology is found in C. Grimes (1999). Notes on Dhao serial verb constructions are found in Jacob and Grimes (2005).

I write the name of the two languages in accord with discussions and requests from influential speakers of those languages who have examined the options, discussed the issues among themselves, and made informed decisions.[3] Speakers of both languages have expressed dismay that they are known to the outside world by names given to them by outsiders, rather than by the way they refer to themselves.

3 Lexical similarity gives the impression that Hawu and Dhao are dialects of the same language

When compared with languages spoken on the other islands around them, the obvious similarities between the vocabularies of Hawu and Dhao have led some writers (for example, Jonker 1903:85-89; Fox 1977:268) to conclude that Dhao (with around 7,000 speakers) is a dialect of Hawu (with 60,000-80,000 speakers in five dialects).[4]

Table 1 gives a sample of words that are identical in form and meaning. Comparison of my data processed to date shows only three percent of vocabulary to be identical in form and meaning.

---

[2] From C. Grimes *et al* (1997).

[3] Left to my own assessment, I would write the language name as Havu.

[4] A print dictionary of Hawu (C. Grimes *et al* 2003) has 1,250 entries, while a more recent online dictionary (C. Grimes 2008) has over 1,650 entries. A small Hawu text corpus for concordance searches has over 1,000 entries. A print dictionary of Dhao (Aplugi *et al* 2000) has 1,586 entries, and a slightly larger online online dictionary (C. Grimes *et al* 2008) is also available. A growing Dhao text corpus for concordance searches has over 10,000 sentences.

Table 1: Sample of vocabulary that is identical in form and meaning

Dhao Form	**Meaning**	**Hawu Form**	**Meaning**
afa	teach (Malay loan)	*afa*	teach (Malay loan)
afu	tree, wood	*afu*	tree, wood
ama	father; classificatory father	*ama*	father; classificatory father
amo	root (of plant)	*amo*	root (of plant)
are	rice (on stalk)	*are*	rice (on stalk)
aru	eight	*aru*	eight
ɗara	1) inside, 2) character[5]	*ɗara*	1) inside, 2) character

Table 2 gives a sample of vocabulary that is fairly similar in form and identical in meaning.

Table 2: Sample of vocabulary that is similar in form and identical in meaning

Dhao Form	**Meaning**	**Hawu Form**	**Meaning**
ana talora	middle child	*ana telora*	middle child
bhəfi	sleep, lay down	*bəfi*	sleep, lay down
bhəlu	forget	*bəlu*	forget
bəsi	iron, steel	*ɓəhi*	iron, steel
dhasi	sea, body of salt water	*dahi*	sea, body of salt water
dhu	relativiser 'which'	*do*	relativiser 'which'
dedha	above	*dĩda*	above
dua	two	*dũe*	two
ətʃi	one	*əhi*	one
subha	swear an oath	*hupa*	swear an oath
jiʔi	we (exclusive)	*ʃĩi*	we (exclusive)
madhutu	1) follow, 2) obey	*pedute*	1) follow, 2) obey

Table 3 shows vocabulary that is similar in form, and only slightly different in meaning or usage. Comparison of my data processed to date shows I have marked only ten percent of vocabulary to be similar in form or only slightly different meaning or use.[6]

[5] This is actually a complex entry as a noun (2 senses), preposition, and TAM marker. The structure and functions appears identical in both languages.

[6] I am aware that this is not yet systematic or complete, and thus am hesitant to make any claims that it is representative.

Table 3: Sample of vocabulary that is similar in form and slightly different in meaning or use

Dhao Form	**Meaning**	**Hawu Form**	**Meaning**
ae	many, much	*ae*	many, much[7]
aʔa	elder sibling, same sex	*aʔa*	elder sibling, same sex[8]
ana	1) offspring, 2) person[9]	*ana*	1) offspring, 2) person, 3) small
ari	younger sibling, same sex	*ari*	younger sibling, same sex[10]
bani	bold (intrasitive)[11]	*ɓani*	bold (intransitive), scold (transitive)
bhədo	1) closed, 2) prison	*ɓədo*	enclosed
bətʃu	full, satiated	*ɓəhu*	1) satiated, 2) trailfood[12]
dhəu	person	*dəu*	person[13]
kabodho	back of, behind s.o.	*keɓodo*	anus (vulgar)
məu-məda	day and night	*mədȁ-nilodȏ*	night and day
pa-məu	1) clean, flatten, 2) clarify, explain	*pe-məu*	clean something (CAUS-clean)
tadəŋi	hear	*rəŋe*	hear

The large bulk of my data are vocabulary items in which the Hawu and Dhao forms are unrelated or are used quite differently. Table 4 presents a sample of such items.

Table 4: Sample of vocabulary that is unrelated in form

Dhao Form	**Meaning**	**Hawu Form**	**Meaning**
məu	daytime, daylight	*nilodȏ*	daytime, daylight
leo, paleo, kaɓi	marry, wed	*baŋa əmu*	marry, wed
rəŋu	3p free pronoun	*roo*	3p free pronoun
cee	who	*naduu*	who
aaʔi-aaʔi	all, every	*hari-əle*	all, every

[7] Different order in NP. Different collocations. Different intensifiers.

[8] Phrasal modifications for specifying gender and other relationships are quite different.

[9] The Dhao entry for *ana* also has several subentries for which there are no close parallels in Hawu.

[10] Phrasal modifications for specifying gender and other relationships are quite different.

[11] Dhao does not have the transitive use with the sense of 'scold, rebuke'.

[12] Dhao does not use this form as a noun meaning 'trailfood'. It uses a different lexeme altogether.

[13] The Hawu form becomes *do* when modified. The Dhao form remains constant.

Dhao Form	**Meaning**	**Hawu Form**	**Meaning**
nidhu	1) spirit of dead pers; 2) evil spirit	*vaŋo*	evil spirit
te ŋaa	but	*tapulara*	but

While the impression that Hawu and Dhao are both dialects of the same language is not at all unreasonable, I argue here that a careful examination of various subsystems of the languages indicates that they are so different in their current forms they **must** be considered separate languages. Since language is a complex system comprised of many complex subsystems, looking at only one or two subsystems for comparison should be inadequate by definition for understanding the relationship between two related speech varieties that have both clear similarities and clear differences.

4 The sound systems (phonologies)

In comparing the sound inventories of the two languages it becomes quickly apparent that, while there are clear similarities, there are also significant systemic differences.

4.1 Dhao phonology

Dhao consonants are displayed in Table 5.

Table 5: Dhao consonants (with practical orthography symbols in <>)[14]

	Labial	**Apical**	**Laminal**	**Dorsal**	**Glottal**
Stop VOICELESS	p	t	ʧ <c>	k	ʔ <'>
Stop VOICED	b	d	ʤ <j>	g	
Implosive	ɓ <b'>	ɗ <d'>	ʄ <j'>	ɠ <g'>	
Affricate/Retroflex	b͡ß <bh>	ɖ <dh>			ʁ <#VV>
Fricative	(f)	s			h
Nasal	m	n	ɲ <ny>	ŋ <ng>	
Lateral		l			
Trill		r			
Semivowel	(w)		(y)		

Dhao has three contrastive grades of voiced obstruents.[15] The /bh/ is a lightly articulated voiced bilabial affricate. As mentioned above, the /dh/ a slightly retroflexed, lightly articulated alveolar affricate. The voiced glottal sound marked /ʁ/ is articulated by some speakers as a pharyngeal constriction to a vowel onset, and by other speakers as a

[14] Loan phonemes are indicated in brackets. These are found mostly in proper names, and a few high frequency loans that have come in via Malay.

[15] Walker (1982:57ff) only identified two for Dhao. This skewed his comparison of Hawu and Dhao phonologies.

lack of a glottal stop onset (contrasting with a glottal stop onset) to a vowel-initial word in a phrase.[16] Examples 1 to 3 illustrate some of the contrasts.

(1)	[ana bebo]	*ana bebo*	'calf (of leg)'
	[babia]	*babia*	'burden'
	[baɓa]	*bab'a*	'short'
	[b͡ßab͡ßənːi]	*bhabhəni*	'female arrogance'
	[kab͡ßisa]	*kabhisa*	'bag woven from lontar leaves'
	[kabitʃu]	*kabicu*	'corner'
(2)	[madəɖːi]	*madèdhi*	'sit'
	[maɖe]	*madhe*	'die, dead'
	[musi maɖa]	*musi madha*	'eyeball'
	[meda]	*meda*	'yesterday'
	[mədːa]	*mèda*	'night'
	[məɖːa]	*mèdha*	'thing'
	[məɗːu]	*mèd'u*	'2-hold'
(3)	[ʤara]	*jara*	'horse'
	[ƒara]	*j'ara*	'path, way'
	[ƒala]	*j'ala*	'castnet' (loan)
	[ƒərːa]	*j'èra*	'difficulty'

Dhao has a six vowel system as displayed in Table 6.

Table 6: Dhao vowels

	Front	**Central**	**Back**
High	i		u
Mid	e	ə <è>	o
Low		a	

Phonetically long vowels are a sequence of two vowels. There are no single-unit diphthongs. Word stress falls on the penultimate syllable of the word. Each vowel is a syllable nucleus, so stress falls on the penultimate vowel in VV sequences, regardless of whether the two vowels are same or different. A stressed schwa /ə/ in a VCV# sequence results in phonetic lengthening of the following consonant. Because this is predictable it is not written in either the transcription or the practical orthography (as shown in examples 1 to 3 above).

These patterns result in a need to be able to distinguish between single (short) vowels, double (long) like vowels, and intervocalic glottal between identical vowels as illustrated in examples 4 and 5.

[16] When presented with a systematic way of representing these three grades of voiced obstruents, native speakers are able to not only read them easily, but also identify and write them consistently. The idea to write the contrastive onset in the practical orthography as a double vowel onset was suggested by Lazarus Aplugi and Michael Sina. It works quite brilliantly.

(4)	[ne]	*ne*	‘3s proximal object pronoun (short)’
	[neʔe]	*ne'e*	‘proximal deictic’
	[ŋaŋeː]	*ngangee*	‘thinking’
	[ʧeː]	*cee*	‘who’

(5)	[ra]	*ra*	‘3p pronominal clitic’
	[raː]	*raa*	‘blood’
	[raʔa]	*ra'a*	‘3p-eat’

The double vowel onset in the practical orthography noted in the discussion of Table 5 is contrastive. But it has extremely limited distribution. It seems to function at a lexical or phrasal level, and hence is not a phoneme of the same sort as the others. This is illustrated in example (6).

(6)	[ɖəu ʔae]	*dhèu ae*	‘many people’
	[ɖəu ʁae]	*dhèu aae*	‘king’
	[ʔəlːe]	*èle*	‘complete, finished, done’
	[ʔele]	*ele*	‘disappear’
	[təkːe ʁele]	*tèke eele*	‘abandon s.th., leave s.th. behind’[17]
	[ʔiki]	*iki*	‘small’
	[ʔana ʁiki]	*ana iiki*	‘small child’
	[ʔuru]	*uru*	‘first, prior, preceding’
	[nətːi ʔuru ka mai]	*nèti uru ka mai*	‘from long ago’
	[ʔana ʁuru]	*ana uuru*	‘eldest child’
	[ʁaː]	*aa*	‘and (conjunction)’
	[ʁoː]	*oo*	‘also, as well, additionally (adverb)’

4.2 Hawu phonology

Hawu consonants are displayed in Table 7.

Table 7: Hawu consonants (with practical orthography symbols in <>)

	Labial	**Apical**	**Laminal**	**Dorsal**	**Glottal**
Stop VOICELESS	p	t		k	ʔ <'>
Stop VOICED	b	d	ʤ <j>	g	
Implosive	ɓ <b'>	ɗ <d'>	ʄ <j'>	ɠ <g'>	
Fricative	v <w>				h
Nasal	m	n	ɲ <ny>	ŋ <ng>	
Lateral		l			
Trill		r			
Semivowel			<y>[18]		

[17] The word *eele* here turns an activity verb into an accomplishment verb, see Jacob and Grimes (2005).

[18] The Seba dialect has a [y] onset for some lexical items that have implosive [ʄ] in the Dimu and Raijua dialects.

The labiodental voiced fricative /v/ is most commonly articulated with upper teeth to inner lower lip. It is sometimes realised as a semivowel, but this seems to be linked more to speaker preference than to dialect or complementary distribution. I have heard it fricativised before all vowels and in both stressed and unstressed syllables, although it is more likely to be articulated as a semivowel in the onset of an unstressed syllable. Examples of contrasts of Hawu consonants can be found in Walker (1982).

Hawu has a six-vowel system as displayed in Table 8.

Table 8: Hawu vowels

	Front	**Central**	**Back**
High	i		u
Mid	e	ə <è>	o
Low		a	

Phonetically long vowels are a sequence of two vowels. There are no single unit diphthongs. Word stress falls on the penultimate syllable of the word. Each vowel is a syllable nucleus, so stress falls on the penultimate vowel in VV sequences, regardless of whether the two vowels are the same or different. A stressed schwa /ə/ in a VCV# sequence results in phonetic lengthening of the following consonant. Because this is predictable it is not written in the practical orthography (the same pattern for Dhao is illustrated in examples (2) and (3) and for Hawu in example (7)).

These patterns result in a need to be able to distinguish between single (short) vowels, double (long) identical vowels, and intervocalic glottal between identical vowels.

(7)	[ŋa]	*nga*	'with (preposition)'
	[neŋaː]	*nengaa*	'what?'
	[ŋaʔa]	*nga'a*	'eat, consume (v), food (n)'
	[ŋali]	*ngali*	'senile'
	[ŋəlːu]	*ngèlu*	'wind'

4.3 Comparing the two systems

The similarities and differences between the two languages can be seen more easily when making a composite of the two as in Table 9. Sounds that are unique to only one of the two languages are underscored and identified with (H) or (D).

Dhao has a whole consonant grade that Hawu does not have. Dhao has /ʧ/ and /s/ which Hawu does not have. Hawu has /v/ which Dhao does not have. The contrastive vowel onset found in Dhao (marked in Table 9 as ʁ) has no parallel in Hawu.

The six-vowel systems and word stress are virtually identical between the two languages. This includes phonetically long vowels functioning phonemically as two syllables (widespread in Central Malayo-Polynesian languages), and the stressed schwa /ə/ triggering phonetic lengthening of the following consonant.[19]

[19] Similar but slightly different patterns with schwa /ə/ triggering lengthening of the following consonant are found in other languages in the region, such as Ngad'a on the island of Flores to the north of Sabu (see Djawanai and Grimes 1995).

Table 9: Dhao consonant system compared with Hawu consonant system

	Labial	**Apical**	**Laminal**	**Dorsal**	**Glottal**
Stop VOICELESS	p	t	ʧ (D)	k	ʔ
Stop VOICED	b	d	ʤ	g	
Implosive	ɓ	ɗ	ʄ	ɠ	
Affricate/Retroflex	b͡ß (D)	ɖ (D)			ʁ (D)
Fricative	v (H)	s (D)			h
Nasal	m	n	ɲ	ŋ	
Lateral		l			
Trill		r			
Semivowel			(y) (H)		

There seems to be a template-driven vowel in the antepenultimate slot for prefixes, even frozen ones, such that Dhao uses /a/, while Hawu uses /e/. Examples are given in Table 10.

Table 10: Antepenultimate vowel (Dhao /a/, Hawu /e/)

Dhao	**Hawu**	**Gloss**
ca-ŋasu	*he-ŋahu*	one hundred
ca-ŋuru	*he-ŋuru*	ten
ca-bəka	*he-ɓəka*	side, part
kacui aai	*kehui ai*	hand, finger // finger
kapai	*kepai*	big, important // big
madhaʔu	*medaʔu*	afraid
madhera	*medera*	long // tall, long
pa-mari	*pe-mari*	RECP-laugh
pa-bhəlu	*pe-bəlu*	CAUS-forget (active, deliberate)
sabaʄa	*heɓaʄa*	pray, worship (loan)

The information laid out above in Table 9 is actually misleading, giving the impression that, because the two consonant SYSTEMS are similar, there is a one-to-one correspondence between the two languages (for example, an implosive /ɓ/ in Dhao is represented by an implosive /ɓ/ in the cognate word in Hawu). The data show such correspondence is not what one would assume. Table 11 shows that there are actually significant differences in the phonologies of the two languages. Note that in this table, 17 out of 47 sound correspondences between Dhao and Hawu are different (36%)

Table 11: Sound correspondences between Dhao and Hawu[20]

Dhao	**Hawu**	**Dhao**	**Hawu**	**Gloss[21]**
p	p	*peka*	*pika*	speak
p	p	*əpa*	*əpa*	four
t	t	*tadhe*	*tade*	know s.o.
t	t	*suti*	*huti*	drip // flow out
k	k	*kako*	*kako*	walk, move, go
k	g**	*j'əke*	*j'ege*	scold, rebuke (??)
ʔ	ʔ	*ngaʔa*	*ngaʔa*	1pe-eat // eat
c	h**	*caʔe*	*haʔe*	ascend, climb
c	h**	*əci*	*əhi*	one
s	h**	*sab'a*	*hab'a*	work // business, effort[22]
s	h**	*risi*	*rihi*	more, excess
h	h	*həba*	*həb'a*	mouth, orifice, door[23]
h	v**	*hadhu*	*vovadu*	rock, stone
h	v**	*hahi*	*vavi*	pig, swine
b	ɓ**	*badha*	*b'ada*	animal
b	ɓ**	*həba*	*həb'a*	mouth, orifice, door
ɓ	(?)	*b'uʔi mari*	*??*	smile (??)
ɓ	ɓ	*sab'a*	*hab'a*	work // business, effort (??)
bh	b**	*bhəni*	*bəni*	female, woman
bh	b**	*sabha*	*haba*	lontar leaf bucket
bh	p**	*subha*	*hupa*	swear an oath (??)
d	ɗ**	*dai*	*ɗ'ai*	arrive, until
d	ɗ**	*məda*	*məɗ'a*	night
ɗ	ɗ	*ɗ'əlu*	*ɗ'əlu*	abdomen, stomach
ɗ	ɗ	*loɗ'o*	*loɗ'o*	1) sun, 2) day[24]
dh	d**	*dhari*	*dari*	rope, cord
dh	d**	*madhe*	*made*	die, dead
j	ʄ/y**	*jaʔa*	*j'aa/yaa*	1s pronoun
j	j	*pajuu*	*pejuu*	order, command (??)

[20] (**) in the Hawu column flags correspondences that are different, in order to aid visual scanning. (??) in the gloss column indicates the correspondence is tentative, since data are sparse. Only zero to two cognates have been found to illustrate the correspondence, so confidence is not high.

[21] In the gloss column, commas represent the semantic range of glosses that work for that lexeme; double-slash // distinguishes different meanings for the similar Dhao (left of //) and Hawu forms (right of //); 1) and 2) represent different senses (polysemy) to that lexeme.

[22] In Hawu the high-frequency verb for 'work' in general is *j'èga*.

[23] In Hawu the form *vuɓa* is more commonly used for a human mouth or a spokesperson.

[24] Dhao has a third sense of 'when'.

Dhao	Hawu	Dhao	Hawu	Gloss[21]
ʃ	ʃ	*j'amo*	*j'amu*	wash (hands)[25]
ʃ	ʃ	*aj'u*	*aj'u*	wood, tree
g	g	*golo*	*golo*	loose (adj) // free (transitive) (??)
g	??	*haga*	*??*	foot, leg (??)
ɠ	ɠ	*g'ətu*	*g'ətu*	pick, pluck (fruit, grain) (??)
ɠ	ɠ	*cahag'e*	*hag'e*	some, several (??)
m	m	*madha əi*	*mada əi*	well // spring
m	m	*ama*	*ama*	father, classificatory father
n	n	*naha*	*nava*	wave (of sea)
n	n	*ina*	*ina*	mother, classificatory mother
ɲ	ɲ	*nyiu*	*runyiu*	coconut // coconut palm (??)
ɲ	ɲ	*manyilu*	*menyilu*	sour
ŋ	ŋ	*ŋara*	*ŋara*	name
ŋ	ŋ	*manəŋi*	*menəŋi*	ask, request[26]
l	l	*lere*	*lere*	escort
l	l	*ele*	*ele*	disappear
r	r	*rai*	*rai*	land, region[27]
r	r	*ari*	*ari*	younger sibling, same sex

4.4 Historical sound correspondences

Differences between Dhao and Hawu are also highlighted when comparing historical sound correspondences as in Table 12. Historical final consonants are lost in both languages.

Table 12: Historical sound correspondences with PAN/PMP[28]

PAN/PMP	Dhao	Hawu	Comment
*p	Ø	Ø	initial and medial
*p	p	p	prefixes and numbers
*(m)p	p	p	
*b	h	v**	initial and medial
*b	b	ɓ**	initial in antepenultimate syll. (??)
*b	bh	b**	reduced CC cluster?
*C	dh	d**	initial and medial

[25] The Dhao word includes 'plates' as a common object, whereas the Hawu word does not.

[26] In Hawu, *huba* and *ami* are used with much greater frequency.

[27] Hawu has a third sense of 'clay'.

[28] This is not exhaustive. It is limited by the patterns observable in my data to date. As before, (**) in the Hawu column flags correspondences that are different, in order to aid visual scanning. (??) in the gloss column indicates the correspondence is tentative, since data are sparse. Only zero to two cognates have been found to illustrate the correspondence, so confidence is not high.

PAN/PMP	Dhao	Hawu	Comment
*t	dh	ɗ**	initial and medial
*T	dh	ɗ**	initial
*d	d	ƒ**	initial
*d	r	r	initial and medial
*D	ɗ	ɗ	initial
*D	r	r	medial
*j	r	r	medial
*z	ƒ	ƒ	initial
*k	Ø (ʔ)	Ø (ʔ)	initial
*k	ʔ	ʔ	medial
*(n)k	k	k	
*s	c	h**	initial and medial (high vowels?)
*s	s	h**	medial
*S	Ø	Ø	medial
*m	m	m	initial and medial
*n	n	n	initial and medial
*ɲ	ɲ	ɲ	initial
*ɲ	n	n	medial
*ŋ	ŋ	ŋ	initial and medial
*l	r	r	initial and medial
*R	Ø	Ø	initial and medial
*q	Ø	Ø	initial and medial
*w	Ø	Ø	initial and medial
*i	i	i	
*e	e	e	
*e	ə	ə	
*e	a	a	last syllable in disyllabic root
*a	a	a	
*u	u	u	
*u	o	o	both syllables *u (??)
*-ay	i	i	
*-ay	e	e	harmony of V height with mid V
*-ey	e	e	
*-uy	i	i	
*-aw	o	o	
*-iw	ƒu	ƒu	*kaSiw ‘wood’
*-iw	i	i	*laRiw ‘run’
*i(C)a	ə(C)i	ə(C)i	any consonant // metathesis
*u(C)a	ə(C)u	ə(C)u	any consonant // metathesis

PAN/PMP	Dhao	Hawu	Comment
*a(C)i	əi	əi	where *Consonant is lost
*a(C)u	əu	əu	where *Consonant is lost
*au	əu	əu	
*a + *i	e	e	

So, from Table 12 it can be seen that while many of the historical sound correspondences are the same even down to some fine complexities in the vowels, there are also a number of significant and noticeable differences.

5 Various grammatical subsystems

When one looks beyond the phonologies and lexicons, at various grammatical functors and grammatical subsystems, one is struck by the cumulative effect of the differences.

5.1 Personal deixis (pronominal systems)

The first type of functors to compare are the pronominal sets. Dhao has a more complex pronominal system than Hawu, as illustrated in Table 13. V-initial verb roots are inflected for person and number in Dhao but not in Hawu. Dhao also has a set of pronominal clitics which can function as Actor proclitics, Undergoer enclitics, and Genitive enclitics. Hawu uses the free pronoun for all those functions.

Table 13: Comparing Dhao and Hawu pronominal systems

	Dhao			**Hawu**	
	Vb pref	Clitic	Free Pronoun	Free Pronoun	Variant
1s	*k-*	*ku*	*jaʔa*	*yaa*	*ɟaa, joo*[29]
2s	*m-*	*mu*	*əu*	*əu*	*au, ou*
3s	*n-*	*na*[30]	*nəŋu*	*noo*	
1pi	*t-*	*ti*	*ədhi*	*dii*	
1pe	*ŋ-*	*ŋa*	*jiʔi*	*ɟii*	
2p	*m-*	*mi*	*miu*	*muu*	
3p	*r-*	*ra*[31]	*rəŋu*	*raa*	*naa*[32]

Note that only the 2s free pronouns are identical in form. There is potential confusion between similar forms between Dhao 2s clitic *mu* and Hawu 2p free pronoun *muu*. There is also potential for confusion between Dhao 3s clitic *na* and Hawu (Raijua dialect) 3p free pronoun *naa*.

---

[29] The dominant Seba dialect uses *yaa*. Dimu and Raijua Wawa use *j'aa*. Raijua D'ida uses *joo*.

[30] An alternate form for 3s object enclitic which adds proximal referential deixis is *ne*.

[31] An alternate form for 3p object enclitic which adds proximal referential deixis is *si*. It also doubles as a collective plural.

[32] Raijua dialect.

5.2 Spatial, temporal and referential deixis

The second type of grammatical functors to compare are the deictics. Compared to Dhao, the Hawu system is quite complex.

5.2.1 Dhao deictics

Dhao deictics are based on a three-way system, also marked for number. The Dhao system is illustrated in Table 14.

Table 14: Dhao deictic system

	Singular	**Plural**
Proximal 'here, now, this'	*neʔe, ne*	*seʔe, se*
Distal 'there, then, that'	*əəna*	*səra*
Remote 'yonder'	*nəi*	*səi*

5.2.2 Hawu deictics

The Hawu deictics are obviously complex, and my data are incomplete, so what is presented in Table 15 is preliminary.[33]

Table 15: Hawu deictic system (preliminary)

	Simple		**Comparative**[34]
	Seba dial.[35]	Dimu dial.	
Proximal 'here, now, this'	*dee*	*nee*	*mi (na) hee*
Distal 'there, then, that'	*nee*[36]	*əne, ne*	*mi (na) hare*
	əne[37]		
Remote 'yonder'	*nii*	*nəi*	*mi (na) hii/hei*

It is sufficient to note similarities and differences with Dhao. Overall, the systems are quite different, and a source of confusion for speakers of one language learning the other. Note that the form *nee* in the Seba and Dimu dialects have significantly different functions.

5.3 Negation

The negation systems show significant differences in form, function, and distribution. These different functors are illustrated in Table 16.

---

[33] Walker (1982:11–12) presents an even more complex system. I am not yet in a position to either confirm or dispute his analysis.

[34] For example, like this, like that, in this way, in that way, and so forth.

[35] In addition, in one text I have eight occurrences of a deictic *napune*, which is specific and referential, but unclear how it fits into the larger system of deictics. An example is *Ta əbe ke rovi roo ne na vavi lara napune la əmu ta peloje ri roo.* 'They took that (aforementioned) yellowish piglet home to raise it.' The piglet had been introduced three sentences prior to the use of *napune*.

[36] Distal deictic directly following noun (as modifier) = *nee*. E.g. *dəu nee* 'that person', *əmu nee* 'that house'.

[37] Distal deictic following preposition (as nominal) (e.g. *la, ma, pa, ti, na*) = *əne.* For example, *pa əne* 'at there', *ti əne* 'from there'.

Table 16: Dhao and Hawu negatives

Function	Dhao form	Dhao gloss	Hawu form	Hawu gloss
General NEG	*aaɗo*	no, not	*aɗo, doɗo*[38]	no, not
Verbal	*boe, be*	not	*ɗo*	not
Prohibitive	*baku*	don't, shouldn't	*ɓole*	don't
Temporal	*dhae*	not yet	*aɗo dae*	not yet
	məka	not yet	*(daeɗo, deɗo)*[39]	not yet
	heka[40]	no longer, not any more	*ɗoke*	no longer
	nəbhu boe	not long from now (soon)	*tui ɗo*	not long
Existential	*abhu boe*	there is no	*peʔee ɗo*	there is no
Modal	*ia ka*	so that…not, lest		
	parluu boe	not necessary	*tenge ɗoke*	not nec.
	bisa boe	not allow, can't	*ie ɗo*	not allowed
Tag	*do aaɗo*	or not?	*ve aɗo*	or not?

5.4 Inflection on verbs

Which verbs are marked, and how they are marked are very different between Dhao and Hawu.

5.4.1 Inflection on Dhao verbs and prepositions

Dhao is a dominant head-marking language (Nichols 1986). It inflects vowel-initial verb and preposition roots for person and number as seen in Table 17. Hawu has no such inflection.

Table 17: Subject/Actor inflection on Dhao verb roots

	Pref	**-ad'o 'visit'**	**-a'a 'eat'**	**-are**[41] **'take'**	**-e'a 'know'**
1s	*kU-*	*kad'o*	*ko'a*	*kore*	*ke'a*
2s	*mU-*	*mad'o*	*mo'a*	*more*	*me'a*
3s	*n-*	*nad'o*	*na'a*	*nare*	*ne'a*
1pi	*t-*	*tad'o*	*ta'a*	*tare*	*te'a*
1pe	*ŋ-*	*ŋad'o*	*ŋa'a*	*ŋare*	*ŋe'a*

[38] Walker (1982:47–48) observes that *ad'o* is the form most commonly negating nominal arguments, whereas *d'o* is most commonly found negating verbal constituents.

[39] Both of these forms are from the Raijua dialect.

[40] Also *aad'o heka*.

[41] In serial verb constructions, this can also mark aspect in the sense of 'already (perfective)'. See Jacob & Grimes (2005) for more examples and discussion; *nare* can also function as a temporal preposition meaning 'for (duration)'.

	Pref	**-ad'o ‘visit’**	**-a'a ‘eat’**	**-are**[41] **‘take’**	**-e'a ‘know’**
2p	*mI-*	*mad'o*	*mi'a*	*mere*	*me'a*
3p	*r-*	*rad'o*	*ra'a*	*rare*	*re'a*

	Pref	***-ədhi* ‘see’**	***-əd'u* ‘hold’**	***-əti***[42] **‘carry’**	***-inu* ‘drink’**	***-o'o* ‘want’**
1s	*kU-*	*kədhi*	*kəd'u*	*kəti*	*kinu*	*ko'o*
2s	*mU-*	*mədhi*	*məd'u*	*məti*	*minu*	*mo'o*
3s	*n-*	*nədhi*	*nəd'u*	*nəti*	*ninu*	*no'o*
1pi	*t-*	*tədhi*	*təd'u*	*təti*	*tinu*	*to'o*
1pe	*ŋ-*	*ŋədhi*	*ŋəd'u*	*ŋəti*	*ŋinu*	*ŋo'o*
2p	*mI-*	*mədhi*	*məd'u*	*məti*	*minu*	*mo'o*
3p	*r-*	*rədhi*	*rəd'u*	*rəti*	*rinu*	*ro'o*

Dhao also has one intradirective verb in which the Actor performing the action denoted by the verb is coreferential with the Undergoer, whose location is being changed (Pawley 1973). This verb is inflected with a set of Undergoer suffixes as illustrated in Table 18.

Table 18: Undergoer suffixes on Dhao intradirective verb

	Suff	***la'-* ‘go’**[43]
1s	*-ku*	*laku*
2s	*-mu*	*lamu*
3s	*-e*	*la'e*
1pi	*-ti*	*lati*
1pe	*-a*	*la'a*
2p	*-mi*	*lami*
3p	*-si*	*lasi*

5.4.2 Object marking on Hawu verbs

Hawu has a large class of verbs that mark number agreement with the Object NP.[44] The plural object is the unmarked form. The singular object is marked by a suffix *-e*, which can trigger morphophonemic changes and harmony of vowel height. Examples 8 to 13 present only a sampling. Dhao has no such agreement system.

---

[42] *nəti* also functions as a preposition ‘from, about’, and as a conjunction ‘because’.

[43] From PMP *lakaw ‘go, walk’.

[44] This analysis is the most consistent with the data and least problematic of various hypotheses examined so far. It is entirely possible that a better analysis will be found with further study.

(8) ***a*** + **-e** → ***e***

Plural object	Singular object	Gloss
aj'a	*aj'e*	teach (loan)
bəka	*bəke*	split, divide
b'ata	*b'ate*	1) block, 2) herd
əba	*əbe*	escort
jaga	*jage*	watch, guard (loan)
j'əga	*j'əge*	work, job responsibilities
mata	*mate*	wait for
ŋaʔa	*ŋaʔe*	eat

(9) ***i*** + **-e** → ***e***

Plural object	Singular object	Gloss
b'ui	*b'ue*	sprinkle, spray
b'uki	*b'uke*	write
d'ai	*d'ae*	arrive at
d'ahi	*d'ahe*	1) weigh, 2) evaluate
d'ui	*d'ue*	shoulder carry
həmi	*həme*	receive, get

(10) ***u(C)u*** + **-e** → **u*(C)e***

Plural object	Singular object	Gloss
b'uju	*b'uje*	touch
pedutu	*pedute*	1) follow, 2) obey
pejuu	*pejue*	order, command
pepuru	*pepure*	lower s.th

(11) ***o(C)o*** + **-e** → ***o(C)e***

Plural object	Singular object	Gloss
heroʔo	*heroʔe*	carry while embracing
kelolo	*kelole*	confess
pemoko	*pemoke*	1) prepare s.th, 2) serve

(12) ***i/ə(C)u*** + **-e** → ***i/ə(C)o***

Plural object	Singular object	Gloss
bəlu	*bəlo*	forget
g'ətu	*g'əto*	pick, pluck (fruit, grain)
həku	*həko*	try, attempt
pej'iu	*pej'io*	bathe s.o.

(13) ***i(C)a*** + **-e** → ***e(C)e*** *(harmony of height)*

Plural object	Singular object	Gloss
liba	*lebe*	broadcast (seed)
pepika	*pepeke*	tell, proclaim, report

5.5 Tense-aspect-mood markers and other adverbials

Understanding TAM marking and adverbials is critical for successful communication. It is here where nuances of meaning are finessed. Missing the subtle clues in the TAM

systems can make one miss the point, even if one understands the content words. The 'noise' is too great. Not only are these functors very different between Dhao and Hawu, but the way they are used in sentences is also very different. There is no attempt to be exhaustive with the examples presented in Table 19. Note that only about a quarter of the examples given are obviously similar in both form and meaning.

Table 19: Comparing Dhao and Hawu TAM markers

Dhao **Form**	**Gloss**	**Hawu** **Form**	**Gloss**
ho	Irrealis, purpose	*la*	IRR (away from Actor/speaker)
———	(no equivalent data)	*ma*	IRR (toward Actor/speaker)
nia, abhu, bisa	able to	*iʔa, nara*	can, be able to
parluu	need to (loan)	*teŋe*	need to
səku	try to	*həku*	try to
hudî	must, have to	———	(no equiv. data)
tema, biasa	normally	*tima, biasa*	normally
ore, iia ɗara	pointlessly	———	(no equiv. data)
ku, laa	first, now; imper.	———	(no equiv. data)
era	still, yet	*dae, ko*	still, yet
dhu	durative, -ing	*do*	stative (durative?)[45]
———	(no equiv. data)	*ŋa*	durative, continuative
nare	perfective in SVC	*məke*	already, PFV
əle, le	already, PFV	*əla*	already, PFV
pe	later, in future	*pe*	completive
———	(no equiv. data)	*pee*	keep doing, continuous

The situation is actually even more complex than it seems from the examples in Table 19. Because the form in one language is identical to another word with a different meaning or function in the other language, there is great potential for miscommunication. So in Table 19 Hawu *la* clashes with Dhao *la* 'locative/dative preposition "at, to"'. Hawu *ma* clashes with Dhao *ma* 'human dative preposition "to"'. Hawu *iʔa* clashes with Dhao *iʔa* 'fish'. Hawu *teŋe* clashes with Dhao *teŋe* 'look for'. Dhao *hudî* clashes with Hawu *hudî* 'chase'. Dhao *ore* clashes with Hawu *ore* 'along edge'. Dhao *laa* clashes with Hawu *la* 'irrealis'. Hawu *dae* 'still, yet' potentially clashes with Dhao *dhae* 'not yet' that has the opposite polarity. Hawu *do* clashes with Dhao *do* 'or'. Hawu *ŋa* clashes with Dhao *ŋa* '1pe clitic'. Hawu *məke* 'already, perfective' clashes directly in aspect with Dhao *məka* 'not yet'.

A similar situation is found comparing other adverbials. Note that in the examples in Table 20 only one is obviously similar in both form and meaning. Gaps in my data are included here to reflect the complexity and possible asymmetry of the subsystems.

[45] Walker (1982:28) describes the function of this *do* as 'stative'. I'm still analysing it.

Table 20: Comparing other Dhao and Hawu adverbials

Dhao Form	**Gloss**	**Hawu Form**	**Gloss**
oe	almost, nearly	*higa*[46]	almost, nearly
dai	very	*ɗai*	very
seli, kolane	excessively	———	(no equiv. data)
həi, kahəi	also	*leko, lema, ma*	also
dhoka, dhodhoka	only, exclusively	*tuu ta ma,*	only, exclus.
(dhoka)…dii	just, merely	*heve, neke, ve*	just, merely
hari	again	*rike, rii, keɓali, vari*	again
taruu-taruu	continually	*loro-loro, peloro*	continually
lai-lai, malai	quickly	*məriai*	quickly
karohe-rohe	quickly	———	(no equiv. data)
nee-nee, mau-mau	quietly	———	(no equiv. data)
ako	rather, a bit, -ish	———	(no equiv. data)
asa	increasingly	———	(no equiv. data)
rupa-rupa	various sorts of	———	(no equiv. data)
kəna	emphatic	———	(no equiv. data)
———	(no equiv. data)	*vəri*	anymore
tareʔa-reʔa	really, truly	*petuu-petuu, təra-təra*	really, truly
mema	1) really, 2) directly 3) beforehand	*dəŋe*	1) directly 2) beforehand

Again, the situation is actually more complex than it initially appears from the examples in Table 20. Dhao *dii* clashes with Hawu *dii* '1pi pronoun', Dhao *hari* clashes with Hawu *hari* 'all', and Hawu *dəŋe* clashes with Dhao *dəŋe* 'with'.

When comparing time words and phrases, the potential for confusion and miscommunication is equally great, as illustrated by the examples in Table 21. Note again that less than one-fifth of these functors are obviously similar in both form and meaning.

Table 21: Comparing selected Dhao and Hawu time words and phrases

Dhao Form	**Gloss**	**Hawu Form**	**Gloss**
deo na, deo neʔe	recent, just now	*ngine*	just now, earlier
pe	later, in future	*hine, leto*	later, in future
meda	yesterday	*miɗa*	yesterday
məda	night	*məɗa*	night
bəli	next day	*ɓəli rai*	next day
madae	morning	*madae*	soon, near future

[46] The Dimu form is *ege* 'almost, nearly'.

Dhao Form	**Gloss**	**Hawu Form**	**Gloss**
madae aae	early morning	*məu rai*	morning, day
məu	daytime	*nilodo, məu rai*	daytime
lodo nətu	noon, midday	*nətu lodo*	noon, midday
nihia	afternoon	*horo lodo, mədə lodo*	late afternoon
ropa lodo cəna	at sunset	*ta jəna ne lodo*	at sunset
camədə	day after tomorrow	*nəli (S), heməd̀a (D)*	day after tomorrow[47]
məu-mədə	day and night	*mədə-nilodo*	night and day
nəbhu	long time	*tui*	long span of time
nəbhu-nəbhu	eventually	*pedəka rai*	eventually
nəbhu boe	soon, not long	*tui do*	soon, not long
lodo-lodo	daily, often	*helodo-helodo*	daily, normal
təu	year	*təu*	year
uru	first, prior	*uru*	first, prior
nəti uru ka mai	from long ago	*rai uru (rai telora)*	from long ago
taa, tebho, lodo	time	*ave*	time
lodo babeʔa	holiday	*lodo mone ae*	holiday
toke dai mia-mia	forever and ever	*hape la lodo namii-mii*	forever and ever

As with TAM markers and other adverbials above, the situation with time words and phrases is more complex than just the examples presented in Table 21. Dhao *pe* clashes with Hawu *pe* 'completive' in Table 19. Words like *madae* and *camədə* vs. *heməd̀a* have the same or similar forms, but the differences in meaning can easily cause serious miscommunication. Dhao *məu* clashes with Hawu *məu* 'clean', and Dhao *taa* clashes with Hawu *taa* 'bag'.

5.6 Prepositions

As with other subsystems, there are very few prepositions (only one in this case) that are identical in form and meaning between Dhao and Hawu. The differences with these functors is seen in the examples in Table 22.

Table 22: Comparing Dhao and Hawu prepositions

Dhao Form	**Gloss**	**Hawu Form**	**Gloss**
aku	according to		
asa, sa	to (toward place)	*pa*	at, to, towards
la	to (Dative → person SG)	*la*	away from source

[47] In the Seba dialect *nəli* is 'day after tomorrow', and *heməd̀a* is 'TWO days after tomorrow'. In the Dimu dialect, *heməd̀a* clashes directly, meaning 'day after tomorrow', and there is no **nəli*. Speakers of the two dialects acknowledge that this semantic clash causes communication problems.

Dhao		**Hawu**	
Form	**Gloss**	**Form**	**Gloss**
mi, ma	to (Dative → person PL)	*ma*	toward source
ɗara	in, inside	*ɗara*	in, inside
dedha	on, above	*ɗida*	above, on top
haha	below, under	*ɗaʔi, vava*	below, under
dəŋe	with	*ŋa*	with, and
ətu, tu	at (location)		
hia	for (benefactive)	*natuu, tuu*	to, for (benefactive)
nare	for (duration)		
nəti, ti	from, about, because	*ŋati, ti*	from, about
toke (dai)	until	*hape, ɗai*	until
		ɟara (lua)	about
		rai	since, from (time)

6 Typology and syntax

The way one strings words together in sentences to convey and track role information is quite different between the two languages. Briefly, Hawu tends to be verb-initial and have ergative marking with *ri*,[48] whereas Dhao is verb medial and has no ergative marking. The order of head and modifier is also different in some types of phrases. The differences in active intransitive clauses are illustrated in §6.1. Examples of active transitive clauses are found in §6.2. Examples of non-active clauses are in §6.3. And differences in order within phrases are further exemplified in §6.4. Collectively these fundamental differences interfere significantly with communication.

6.1 Active intransitive clauses

(Subject = macrorole of Actor; Subject NP is boxed; main verb is underlined)

(14) Dhao

Lazarus	*kako*	*madhutu*	*nebhe*	*dhasi.*
(name)	walk	follow/along	shore	sea

Lazarus walked/was walking along the edge of the sea.

[48] The Seba ergative marker is *ri* as in *ri roo* 'by them'. The Dimu ergative marker is *ro* and long form *rovi*, as in *ro roo* 'by them'. The Raijua Ergative marker is *la*, as in *la naa* 'by them'. The differences in this critical functor are a source of continuing miscommunication among speakers of the different dialects.

(15) Hawu
ta *nəru* *ke* *Simo* *oro* *ngidi* *dahi.*
NPST? walk DECL[49] (name) along edge edge sea
Simo was walking along the edge of the sea.

(16) Dhao
həia *ra* *kako* *taruu* *asa* *Baʔa.*
then 3p walk CONT PATH Ba'a (on Rote)
Then they continued walking/traveling towards Ba'a.

(17) Hawu
ta *nəru* *ke* *roo* *teruu* *la* *Həɓa.*
NPST? walk Decl 3p CONT DAT Seba (on Sabu)
They kept walking to Seba.

6.2 Active transitive clauses

(Subject = Actor; Subject NP is boxed; main verb of clause in question is underlined)

(18) Dhao
Ropa *ra* *poro* *rare* *kətu* *na,....*
when 3p cut 3p-PFV head 3s
When they had cut off his head…

(19) Hawu
Ta *la* *əte* *ke* *ri* *roo* *ne* *kətu* *noo.*
NPST? go cut off Decl ERG 3p PROX head 3s
They went and cut off his head.

(20) Dhao
Te ŋaa *ra* *pamadhe* *ne.*
but 3p CAUS-die 3s.OBJ.prox
But they killed him.

(21) Hawu
Tapulara *pemade* *noo* *ri* *roo.*
but CAUS-die 3s ERG 3p
But they killed him.

6.3 Non-Active clauses

(Subject = Undergoer; Subject NP is boxed; main verb is underlined)

[49] The function of *ke* remains unclear. It does not appear to be part of the Ergative-Absolutive marking, since it often co-occurs with the Ergative marker *ri* (152 co-occurrences in my current corpus), as in *ke ri roo*. Walker (1982:32) was also unable to pin down the function of *ke*, although he observed it did not occur in imperative sentences.

(22) Dhao
ladhe *ama* *na* *madhe*,...
if/when father 3s die
When his father dies,…

(23) Hawu
ki *made* *ama* *noo*,...
if/when die father 3s
When his father dies,…

(24) Dhao
na *ərа* *titu kəna.*
3s strong very much
He was incredibly strong.

(25) Hawu
ɗai təra *noo* *ne* *rui.*
very much 3s PROX strong
He was incredibly strong.

6.4 Order of head and modifier in phrases

While most NPs in both languages follow the order of Head + Modifier, there are some differences. In Hawu, the different order may indicate a degree of lexicalisation. Table 23 shows some contrastive examples.

Table 23: Different order in phrases

Dhao		**Hawu**	
Form	**Gloss**	**Form**	**Gloss**
dhəu ae	many people	*ae dəu*	many people
loɗo nətu	noon, midday	*nətu loɗo*	noon, midday

7 Interclausal relations

There are many differences in the markers of interclausal relations. Here I highlight only a few.

7.1 Complementisers

Typical of many languages of eastern Indonesia, complementisers flag realis or irrealis mode, and there are special complementisers associated with speech acts and verbs involving cognition and perception. Table 24 illustrates that the forms of the Dhao and Hawu complementisers are completely different.

Table 24: Comparing Dhao and Hawu complementisers

Dhao Form	**Function**	**Hawu Form**	**Function**
ho	Irrealis, purpose	*la*	IRR (away from Actor/speaker)
———	(no equiv. data)	*ma*	IRR (toward Actor/speaker)
peka na	Realis, speech acts	*mihe ane*	Realis, speech acts
aku nəŋu na	Realis, speech act (3s)	———	(no equiv. data)
aku rəŋu na	Realis, speech act (3p)	———	(no equiv. data)

As with other subsystems such as TAM markers and other adverbials discussed previously, because the form in one language is identical to a word with a different meaning or function in the other language, there is great potential for miscommunication. So Hawu *la* clashes with Dhao *la* 'locative-dative preposition "at, to"'. And Hawu *ma* clashes with Dhao *ma* 'human dative preposition "to s.o."'.

7.2 Connectors/conjunctions

Various paragraph, sentence, and clause-level connectors also illustrate the significant differences between the two languages. Some of these functors are illustrated in Table 25.

Table 25: Comparing a sampling of Dhao and Hawu connectors

Dhao Form	**Gloss**	**Hawu Form**	**Gloss**
lodo	sun, day, when, while	*mudî həku*	at the time, while
ropa, rapa	when, while	*dai*	when, while
həia	then	*fe*	then
ka	then, and		
ladhe	if, when	*ki*	if, when
te ŋaa	but	*tapulara*	but
əle əəna ka	after that, then	*əla pe na əne*	after that
lodo əəna, taa əəna	at that time	*pa awe na əne*	at that time
lodo əci	one time	*(pa dara) he vari*	one time
nəti, taga	because	*pelai, taga*	because
nəti əəna ka	because of that, that is why	*ri mi na həre*	because of that, that is why
sama səmi, nuka səmi	like X	*helaʔu ŋa*	like X
ele boe	perhaps		
malo-malo	fortunately		
karəi, krəi	since		

Dhao Form	Gloss	Hawu Form	Gloss
		helama	as long as (loan)
		mudî həku	actually, in fact
		ŋgee-ŋgee	lest

Hawu *ɗài* potentially causes confusion with Dhao *dai* 'until'.

8 Question words

Question words also show a high degree of difference between the two languages to a degree that one would expect would interfere with successful communication. These functors are illustrated in Table 26.

Table 26: Comparing a sampling of Dhao and Hawu question words

Dhao Form	Gloss	Hawu Form	Gloss
cee	who	*naduu, mi*	who
ətu mia, ka mia	where	*pa mii*	where
ŋa tao ka	why	*ta ŋaa*[50]	why
ŋaa	what	*neŋaa,*[51] *ŋaa*	what
pəri	1) how many, how much, 2) several, a few	*pəri*	1) how many, how much, 2) several, a few
tasa mia, tasa mera mia	how	*mi na mii*	how

Hawu *mi* 'who' potentially causes confusion with Dhao *mi* '2p pronoun', and Hawu *ta ŋaa* 'why' clashes with Dhao *te ŋaa* 'but'. How these words are used in questions can also be quite different, as illustrated in examples (26) and (27). The question words are marked by boxes.

(26) Dhao
Ngara èu *cee*?
name 2s who?
What's your name?

(27) Hawu
Naduu *ne* *ngara* *èu?*
who? PROX name 2s
What's your name?

[50] Raijua dialect uses *ta nyaa*.

[51] Dimu dialect uses *ningaa*. Raijua uses *nenyaa*.

9 Notes on intelligibility

Existence of a cognate does not tell anything about frequency of use or how widely it is known in society. I have encountered many linguistic cognates between Dhao and Hawu, but a significant number of these have one form widely used in one language, but the cognate form in the other language is not the normal form to be used for that meaning, and often it is not widely known either. This presents a mismatch between the picture from historical linguistics and the sociolinguistic facts of synchronic intelligibility.

In several communities on Sabu Island and in several dialects, I have asked people about interaction and comprehension with speakers of Dhao. Few Hawu speakers have interacted with Dhao speakers. Of the ones who have, not a single one claims to understand Dhao beyond recognising a few words.

Of the many Dhao speakers I have queried over several years, I have met only two who claim to understand Hawu. One's mother was ethnically Hawu, so he grew up interacting with Hawu speakers. In testing his knowledge with common Hawu words and phrases, it also turned out his claims of proficiency went far beyond the reality. The other speaker was posted as a teacher on Sabu for several years, and had to learn Hawu for social survival.

At a workshop to train speakers of local languages to keyboard their own material on computer, I encouraged the Dhao speakers and Hawu speakers to compare notes on language and orthography. It was immediately obvious to them that their languages had more similarities with each other than with other languages around. But they all found it impossible to understand each other in both oral communication and in written form. How information is strung together in sentences is simply too different.

10 Conclusion

Agard (1984) presents a detailed overview of the phonological development of eight of the major Romance varieties. In looking at structural differences that do and do not differentiate languages, he notes that simple and systematic phonological differences do not split languages. More distant (on the phone chart), or more complex differences such as a change of environmental conditioning factors do. Milliken (1988) found that when structural differences become too complex, resulting in two sets of varieties that are no longer congruent in their phonological structure, intelligibility breaks down.

Grimes (1988), looking at eleven studies from around the world, also observes that an apparently high degree of lexical similarity in content words can give the false impression that two speech varieties ought to be intelligible, because the evidence shows that intelligibility is likely to be blocked if the functors or grammatical subsystems have significant differences.

When comparing Hawu and Dhao, which show superficial similarities of content words, not only do we find significant differences in the vocabulary and phonologies, but as this study has repeatedly shown, the overwhelming majority of grammatical functors are different, the two languages have whole subsystems of the grammar that are different, and they are **typologically** different – Hawu tends to be verb-initial and marked for the ergative argument, while Dhao is verb-medial and has no ergative marking. How one indicates who is doing what to whom is significantly different. The order of modifiers in relation to the head in NPs is often different as well.

Thus it should be no surprise that I can find no speakers of Hawu who claim to understand Dhao, and the only Dhao speakers who claim to understand Hawu are those

few who have learned it through years of significant contact. This is learned bilingualism, not inherent intelligibility. The cumulative effect of the differences, particularly in the functors and grammatical subsystems, require the conclusion that Hawu and Dhao **must** be considered separate languages.

Comparing Hawu and Dhao highlights potential limitations with various methodologies. In some cases, comparing vocabulary similarity may give an adequate picture of whether two speech varieties are similar or different. In the case of Hawu and Dhao this is not sufficient. Comparing phonological systems may give a reasonable picture of similarities in some cases, but in the case of Hawu and Dhao it does not show the mismatch between similar phonemes. For Hawu and Dhao it is necessary to compare functors and grammatical subsystems in their entirety to get a realistic picture of how great the differences are between the two languages.

Now the question remains: given that there are obvious similarities in parts of their vocabularies, how did the two languages become so different in their grammatical subsystems?

Bibliography

Agard, Frederick B., 1984. *A Course in Romance Linguistics*. *Volume 2: A diachronic view.* Georgetown: Georgetown University Press.

Aplugi, Lazarus, Charles E. Grimes, Ayub Ranoh, and Michael Sina, compilers, 2000. *Kamus Pengantar Lii* Dhao *(Bahasa Ndao).* Kupang: Artha Wacana Press. [Introductory dictionary of Dhao.]

Capell, Arthur, 1944–45. Peoples and languages of Timor, part two. *Oceania* 151:19–48.

Djawanai, Stephanus, and Charles E. Grimes, 1995. Ngada introduction. In Darrell Tryon, ed., *Comparative Austronesian Dictionary; an introduction to Austronesian studies*, Part 1, Fascicle 1:593–599. Trends in Linguistics, Documentation 10. Berlin: Mouton de Gruyter.

Donselaar, W.M., 1872. Aanteekeningen over het eiland Savoe. *Mededeelingen van wege het Nederlandsch Zendeling Genootschap*, 16:281–332.

Fox, James J., 1972. The Savunese. In Frank Lebar, ed., *Ethnic groups of insular Southeast Asia*. New Haven: HRAF Press. Vol. 1:77–80.

— 1977. *Harvest of the palm: ecological change in eastern Indonesia.* Cambridge: Harvard University Press.

Fox, James J. and Charles E. Grimes, 1995. Roti introduction. In Darrell Tryon, ed., *Comparative Austronesian Dictionary; an introduction to Austronesian studies*, Part 1, Fascicle 1:611–622. Trends in Linguistics, Documentation 10. Berlin: Mouton de Gruyter.

Grimes, Charles E., 1999. Implikasi penelitian fonologis untuk cara menulis bahasa-bahasa daerah di Kawasan Timur Indonesia. In Soenjono Dardjowidjojo and Yassir Nasanius, eds, *PELBBA 12: Pertemuan Linguistik (Pusat Kajian) Bahasa dan Budaya Atma Jaya Kedua Belas*, 173–197. Kanisius: Yogyakarta, Indonesia. [Implications from phonological research for ways of writing vernacular languages in eastern Indonesia].

— 2000. Introduction: new information filling old gaps in eastern Indonesia. In Charles E. Grimes, ed. *Spices from the East: Papers in languages of eastern Indonesia*, 1–8. Canberra: Pacific Linguistics.

— 2008. *Lii* Hawu *(Sabu) online dictionary.* Kupang: UBB-GMIT. www.e-kamus2.org.

Grimes, Charles E., Bernadus Lado, Thomas Ly and Simon Tari, compilers, 2003. *Kamus Pengantar Lii* Hawu *(Bahasa Sabu).* Kupang: Artha Wacana Press. [Introductory dictionary of Hawu.]

Grimes, Charles E., Ayub Ranoh and Helena Aplugi, compilers, 2008. *Lii* Dhao *(Ndao) online dictionary.* Kupang: UBB-GMIT. www.e-kamus2.org.

Grimes, Charles E., Tom Therik, Barbara Dix Grimes, and Max Jacob, 1997. *A guide to the people and languages of Nusa Tenggara.* Paradigma B–1. Kupang: Artha Wacana Press.

Grimes, Joseph E., 1988. Correlations between vocabulary similarity and intelligibility. *Notes on Linguistics*, 41:19–33.

Jacob, June and Charles E. Grimes, 2005. Aspect and directionality in Kupang Malay serial verb constructions. Paper presented at the 4th International East Nusantara Linguistics Conference held in Leiden, 30 June & 1 July 2005.

Jonker, Johann C.G., 1903. Iets over de taal van Dao. In *Album Kern*. (Opstellen geschreven ter eere van Dr. H. Kern), 85–89. Leiden: E.J. Brill.

— 1919. Sawoeneesch. In D.G. Stibbe, ed., *Encyclopaedie van Nederlandsch-Indië*, 3:712–713. s-Gravenhage: Martinus Nijhoff.

Kern, Hendrik, 1892. Sawuneesche Bijdragen; Volzinnen, samensprakken en woordenlijst, met eene grammatische inleiding. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië*, 41:157–196, 513–553. s-Gravenhage: Martinus Nijhoff.

Ly, Thomas, Bernadus Lado, Simon Tari, and Charles E. Grimes, translators, 2006. *Lii Haga Dara jara lua Yesus pedutu nga do hure do buke ri Mau.* Kupang: UBB-GMIT. [Gospel of Mark in Hawu, with sketch of phonology in introduction.]

Milliken, Margaret E., 1988. Phonological divergence and intelligibility; A case study of English and Scots. Ithaca, N.Y: Cornell University. PhD dissertation.

Nichols, Johanna, 1986. Head-marking and dependent-marking grammar. *Language*, 62:56–119.

Pawley, Andrew K., 1973. Some problems in Proto-Oceanic grammar. *Oceanic Linguistics*, 12:103–188.

Radja Haba, M.C., 1958. Havunese Phonemes. Bloomington: Indiana University. MA thesis.

Ranoh, Ayub, Lazarus Aplugi, Michael Sina, Charles E. Grimes, translators, 2000. *Li Lolo Bea nèti Lamatua Yesus madhutu sasuri Markus.* Kupang: Artha Wacana Press. [Gospel of Mark in Dhao, with sketch of phonology in introduction].

— 2004. *Lii Lole nèti Ana Pajuu-paleha Lamatua Yesus.* Kupang: Artha Wacana Press. [Acts of the Apostles in Dhao].

— 2005. *Li Lolo dhu uru tèka.* Kupang: Artha Wacana Press. [Genesis in Dhao].

— 2006. *Li Holo Nori mi Dhèu Sarani.* Kupang: UBB-GMIT. [Pastoral epistles in Dhao].

Riedel, J.G.F., 1889. Bijdrage tot de Kennis van het Sawusch Dialect. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië*, 38:10–12.

Walker, Alan T., 1982. *A Grammar of Sawu.* NUSA Linguistic Studies in Indonesian and Languages of Indonesia. Vol. 13. Jakarta: Badan Penyelenggara Seri NUSA.

Wijngaarden, J.K., 1896. *Sawuneesche Woordenlijst.* Uitgegeven door het Koninklijk Instituut voor de Taal-, Land- en Volkenkunde van Nederlandsch-Indië. s-Gravenhage: Martinus Nijhoff.

YPPII, 1985. *Lai Lua Jagga Rasul-Rasul.* Dept. Misi, Yayasan Persekutuan Pekabaran Injil: Batu, Malang. [Acts of the Apostles in Hawu].

13 *Nouns and verbs in Magey Matbat*

BERT REMIJSEN

1 Background information on Magey Matbat

Magey Matbat is spoken on Misol, an island in the Raja Ampat archipelago, about halfway between the western tip of New Guinea and the Moluccan island of Seram. The language has around 500 speakers, spread over four villages: Magey, Kapacol, Aduwey, and Salafen (Map 1). Two or three other languages are used on the island: Tomolol Matbat, Ma'ya and Biga. Tomolol Matbat is closely related with Magey Matbat, and the two may constitute dialects of one language.[1] More work on Tomolol Matbat is required to settle this question. At least in Magey, adults and children alike are fluent in Matbat.

Matbat is a self-referent term, meaning 'people of the land', and distinguishes the language communities of Magey Matbat and Tomolol Matbat jointly from the Biga and the Ma'ya – or Matlow 'people of the sea', as they are called in Matbat villages. In other words, on Misol the term Matbat refers to an ethnic group, encompassing both the Magey Matbat and the Tomolol Matbat communities. Little information on Magey Matbat is available so far. There is a short wordlist in Wallace (1869 – list no. 50); an analysis of the tone system (Remijsen 2001, 2007), and a Swadesh list (Remijsen 2001). Smits and Voorhoeve (1992) present three word lists on Tomolol Matbat. Blust (1978) has classified Matbat within the South Halmahera-West New Guinea (SHWNG) subgroup of Austronesian.

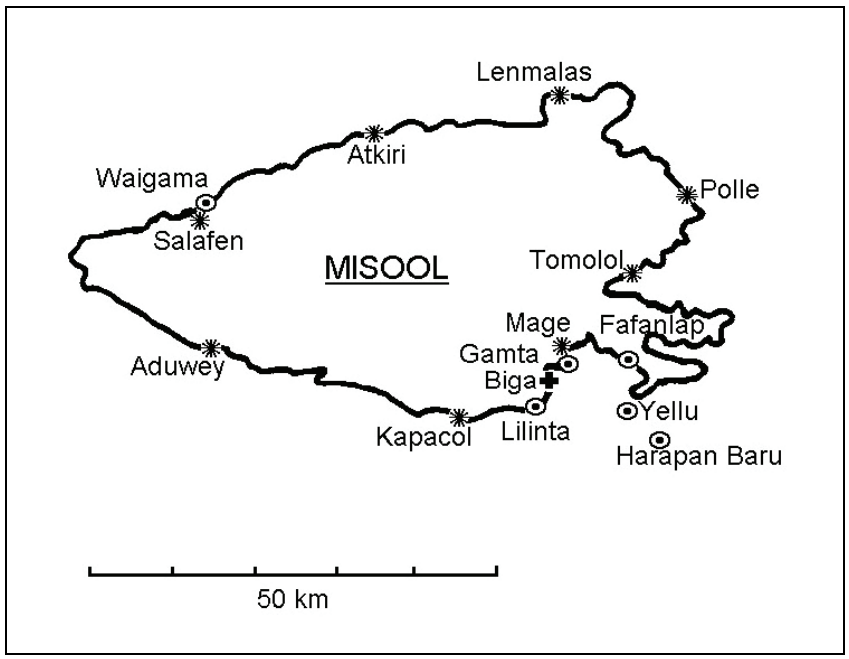
The Magey Matbat system of segmental phonemes comprises the consonants /p,t,k,b,d,g,m,n,ŋ,s,f,l,(r)[2],w,y/, and the vowels /i,e,ɛ,a,ɔ,o,u/. In addition, each syllable can have one of six lexically distinctive tone patterns: Low, High, Extra High Fall, Low Fall, Low Rise, and Rise-Fall. They are transcribed by means of a numeric notation, in which the speaker's pitch range is represented by the range from 1 (low pitch) to 4 (high pitch). Using this system, the tones are transcribed /1, 3, 41, 21, 12, 121/, respectively. Tones are represented after the vowel of the syllable with which they are associated. Syllables may be toneless, although content words have at least one syllable specified for tone. If the

1 In Remijsen (2001) I classified Tomolol Matbat and Magey Matbat as dialects of one language, based on native-speaker judgments. However, data which I have collected on Tomolol Matbat since then show substantial lexical and grammatical differences.

2 The rhotic is only phonemic in shallowly integrated loan words – for example *te^{3}r* 'paint' < Dutch teer 'tar'. In native or fully incorporated words, it is an intervocalic allophone of the phoneme /d/ – for example *mu^{3}ru^{3}* 'easy' < Malay mudah (same meaning).

Michael C. Ewing and Marian Klamer, editors.
Typological and areal analyses: contributions from East Nusantara, 281–311. Pacific Linguistics, 2010
©. This edition vested with Pacific Linguistics.

final syllable of a word has the Low Fall (/21/) associated with it, then a paragoge vowel /-o/ appears in the context of a following intonational phrase boundary. The syllable template is (C)V(C), and a substantial proportion of words is monosyllabic.



Map 1: The island Misol, with indication of the languages:
Tomolol and Magey Matbat, marked by an asterisk;
Ma'ya, marked by an encircled dot; and Biga, marked by a plus.

This chapter presents a description of some basic characteristics of the morphology and syntax of nouns and verbs in Magey Matbat (hereafter referred to as Matbat). It is structured as follows. Sections 2 and 3 describe aspects of the morphosyntax of nouns and verbs in Matbat, respectively. In §4, these findings are related to the structure of non-verbal predicates. The chapter is concluded in §5, with a discussion of the structure of Matbat from a typological and areal perspective.

2 Nouns, pronouns, and noun phrases

2.1 In general

Matbat has little inflectional morphology – it can be categorised best as an isolating language. In line with this characterisation, there is no morphological process that can serve as a heuristic to determine whether a word is a noun. Instead, nouns can better be defined in syntactic terms, based on the modifiers they can take in noun phrases, and on the internal structure of these syntactic constituents. Nouns in Matbat, then, are constituents that can take the following modifiers: demonstratives, possessors, numerals, and adjectives. The syntactic structure of the noun phrase is represented schematically in (A), where the optional modifiers are bracketed. An example is provided in (1). The head noun may be followed by one or more adjectives, a quantifier and a demonstrative. The possessor is the only modifier that regularly precedes the possessor.

A. (possessor) noun(-number) (adjective(s)) (quantifier) (demonstrative)

(1) *ak-dɛ^{12}l si^{3}ni^{3} nu^{3}n-ha-no*
1s-friend Chinese CLF-one-DEM
'This Chinese friend of mine.'

Inflection in the noun phrase is confined to two areas. First, Matbat distinguishes between alienable and inalienable possession. Inalienably possessed nouns show agreement with the possessor. Second, Matbat has a system of numeral classifiers. The numerals from 1 to 10 take affixes selected on the basis of particular semantic properties of the noun referents. The main topics in the morphosyntax of noun phrases are discussed in the following subsections. I begin with outlining the system of personal pronouns (2.2), which is relevant further on throughout this chapter, among others in the discussion of possessive noun phrases in §2.3. Section 2.4 deals with the quantifier slot in the noun phrase, and the related topic of classifiers. Section 2.5 is about compounding.

2.2 Personal pronouns

The paradigm of personal pronouns distinguishes person, number, and, in the first plural, inclusive versus exclusive. In Table 1 the full or non-reduced forms are given in column A, and reduced forms appear in columns B and C. All personal pronouns, both full and reduced forms, are 'anaphoric devices' – elements that can refer to an entity in the absence of a co-referent noun. The same pronominal forms are used across grammatical relations, including the expression of subject, object, and nominal possessor.

Table 1: The paradigm of pronouns. Full, that is, non-reduced forms appear in (A), and reduced forms in (B) proclitics and (C) enclitics.

	A. Full	**B. Proclitic**	**C. Enclitic**
1s	*ya^{21}ka*	*ak*	*ka*, *aka*
2s	*ya^{21}wa*	*aw*	*w*, *awa*
3s	*i^{21}*	*i*	*i*
1pi	*ya^{21}ta*	*at*	*ta*
1pe	*ya^{21}ma*	*am*	*ma*
2p	*mi^{21}na*	*min*	*mina*
3p	*hafɔ^{12}*	*ha*	*ha*

The non-reduced form of pronouns is used mainly when the word has particular emphasis – for example, when uttered in isolation in answer to a question-word question. In other contexts, these full forms are reduced as the pronoun cliticise to adjacent content words. Pronouns appear as enclitics on prepositions and in object position, and as proclitics in subject position and when used as the possessor in a noun phrase. In (2), the pronominal form *ka* '1s' encliticises to a preposition. Similarly in (3), the object-marking pronoun *w* '2s' appears as an enclitic to the verb. But when a pronoun serves as the possessor before a noun (1), or as the subject of a verb (4), it may cliticise at the left side of a stem. Example (4) also illustrates the fact that a subject-marking pronoun separated from the verb by a preverbal adverb is repeated after the intervening constituent.

(2) *ak-s-i^{12}p* *mɛ^{41}-ka*
1s-1-bathe alone-1s
'I bathe by myself.'

(3) *ak-k-e^{21}-w* *mɔ^{41}n* *pa^{21}n* *di^{41}-ha*
1s-1s-give-2s areca_nut half CLF-one
'I give you half an areca nut.'

(4) *ak-nu^{3}mna^{121}n* *ak-tɔ^{12}* *dadi^{3}* *ha^{3}y*
1s-tomorrow 1s-chop cut wood
'I will chop wood tomorrow.'

When the plural forms of the personal pronouns refer to a small number of individuals, the number of individuals tends to be made explicit, by attaching the relevant numeral to the pronoun – see (5) and (6). Here again, the pronouns appear as proclitics.

(5)^[3] *am-lu^{12}* *m-a^{3}p* *bɔ^{3}t* *bɔ^{3}y-pa^{12}n*
1pe-two:CLF 1pe-row reach Boy-entrance
'The two of us rowed up to the entrance of the Boy [name of small stream].'

(6) *ha-lu^{12}* *hi^{12}p* *ha-wa^{3}y*
3p-two:CLF wash 3p-child
'The two of them are washing their child/children.'

Moreover, the identity of third-person referents can be made fully explicit, by specifying them alongside the pronoun. When the pronoun is the first person exclusive, then such a noun phrase, referring to the third party, follows the pronoun, as in (7). When the pronoun is the third plural, both the positions preceding and following the pronoun can be filled by nouns that make explicit the parties referred to by the pronoun, reducing the relevance of the pronoun to that of a conjunction – see example (8).

(7) *am-lu^{12}* *ak-nɔ^{3}ŋ* *nun-wa^{3}y* *faya^{1}w* *yi^{1}n*
1pe-two:CLF 1s-brother:1s CLF-small hunt fish
'Me and my younger brother are hunting fish.'

(8)^ *wa^{3}w* *ha-lu^{12}* *kayɔ^{12}w* *bawo^{21}* *fali^{12}w*
pig 3p-two:CLF tick run win
'The pig and the tick held a running contest.'

2.3 Possessive noun phrases

In a typological overview of Austronesian languages, Himmelmann (2005:164) writes: '[i]n preposed-possessor languages, there are typically two types of possessive constructions, an alienable and an inalienable one.' This holds in Matbat: the possessor precedes the possessed term, and inalienably possessed head nouns are marked.

---

3 The ^ signposts examples that do not result from controlled elicitation.

The possessor is expressed by a pronoun or by another noun. We have already come across several instances of the first scenario (1), (6) and (7). When the possessor is expressed as a noun rather than through a pronoun, then possessor and possessed are separated by the clitic *i-*, as in (9) and (10). This marker is the reduced form of the third singular pronoun. However, its distribution in possessive noun phrases is not specific to the person and number of the possessor. This is evident from (10), where the possessor is neither third person nor singular – it is in the second plural. In §3.1.1 below we will see that the clitic *i-* it fulfills a similar linking function between subject and predicate.

(9)^ *wa^{3}n* *i-pɛ^{21}l* *i-fi^{121}n-a^{12}*
conical_basket 3s-price CLF-how_much-WHQ
'What is the price of the wickerwork basket?'

(10) *mi^{21}n* *matba^{3}t* *i-nu^{3}*
2p Matbat 3s-village
'The village of you, the Matbat.'

For a subset of nouns, the possessor is additionally marked by means of an inflection on the possessed term. This subset includes most but not all words referring to body parts and kinship terms. This inflection is obligatory, that is, words like *fa^{21}* 'husband:3' and *ba^{21}m* 'shoulder:3' are never completely vague as to the possessor. In the case of a body part, the possessor is obviously the person to whose body the part belongs. In the case of a kinship relation, the possessor is the ego, relative to whom the kinship relation holds. While a regular noun such as *wa^{3}ŋ* 'dugout canoe' can be used in a general sense, without specification of a possessor, these inalienably possessed nouns cannot. This inflectional marking for possessor only appears on inalienable body parts and kinship terms. As such, the marking of inalienable possession can be interpreted as a noun categorisation device: it distinguishes a subset of nouns (Aikhenvald 2000:257f).[4]

The paradigms of five inalienably possessed nouns are listed in Table 2. While the system of personal pronouns (see Table 1) distinguishes seven combinations of the factors number, person, and inclusive versus exclusive, fewer are distinguished in the marking of inalienable possession: second singular, second plural, and first plural exclusive are not distinguished from one another; the same goes for third singular versus third plural. As we will see, the situation is similar for agreement inflection on the verb (see §3.1.1). In both cases, finer distinctions are made for the first person than for second and third person. Foley (1986) notes that this phenomenon is found in many Papuan languages.

As seen from the paradigms in Table 2, the third person form constitutes an unmarked base form, which is modified in the other inflections. These modifications invariably involve the nasals /m,n,ŋ/, either in word-medial or in word-final position. In addition, there may be other changes in tone and/or in segments. The tonal suprafix in the marked forms of the paradigm is most often the High Level toneme, but in certain cases it is the Low Rise.

4 I thank Michael Ewing for pointing this out to me.

Table 2: Some paradigms of inalienably possessed nouns

	body	**eye**	**shoulder**	**back**	**husband**	**mother-in-law**
1s	*ni*^{3}*ŋsu*	*mɔ*^{21}*ŋa*	*ba*^{12}*ŋ*	*si*^{3}*ŋbo*^{21}*m*	*fa*^{3}*ŋ*	*imɔ*^{12}*ŋu*^{3}
2s	*ni*^{3}*msu*	*mɔ*^{21}*ma*	*ba*^{12}*m*	*si*^{3}*mbo*^{21}*m*	*fa*^{3}*m*	*imɔ*^{12}*mu*^{3}
3s	*ni*^{21}*su*	*manɔ*^{21}*n*	*ba*^{21}*m*	*sabo*^{21}*m*	*fa*^{21}	*imoyu*^{3}
1pi	*ni*^{3}*nsu*	*mɔ*^{21}*na*	*ba*^{12}*n*	*si*^{3}*nbo*^{21}*m*	*fa*^{3}*n*	*imɔ*^{12}*nu*^{3}
1pe	*ni*^{3}*msu*	*mɔ*^{21}*ma*	*ba*^{12}*m*	*si*^{3}*mbo*^{21}*m*	*fa*^{3}*m*	*imɔ*^{12}*mu*^{3}
2p	*ni*^{3}*msu*	*mɔ*^{21}*ma*	*ba*^{12}*m*	*si*^{3}*mbo*^{21}*m*	*fa*^{3}*m*	*imɔ*^{12}*mu*^{3}
3p	*ni*^{21}*su*	*manɔ*^{21}*n*	*ba*^{21}*m*	*sabo*^{21}*m*	*fa*^{21}	*imoyu*^{3}

Several inalienable nouns referring to body parts are similar to *sabo*^{21}*m* 'back:3', in that they have an initial syllable [sa-] in the third person, which conjugates into [si^{3}ŋ-], [si^{3}m-], [si^{3}n-] in other forms. Other such nouns include *saga*^{21}*w* 'buttocks:3', *sapa*^{21}*w* 'belly:3', *sapɛ*^{21}*ŋ* 'penis:3', *saba*^{21}*ŋ* 'tail:3', *sabɛ*^{121}*mta*^{21} 'chest:3'. Some speakers conjugate such nouns with respect to tone and nasal only, while the vowel remains /a/ throughout the paradigm – so *sa*^{3}*ŋbɔ*^{21}*m*, *sa*^{3}*mbɔ*^{21}*m*, *sa*^{3}*nbɔ*^{21}*m*.

As stated above, not all body parts and kinship terms take this inflection – for example, the body-part noun *te*^{121}*lo* 'head' does not, and neither does *ba*^{12}*p* 'father', probably a loan from *Malay* bapak (same meaning). Inalienable possession is pervasive in the Matbat grammar, in the sense that its marking is not strictly limited to nouns. First, various verb predicates with a meaning related to a physical or emotional state take the inflections of inalienable possession as subject markers (see §3.1.2). Second, body-part nouns are used in the expression of spatial deixis (see footnote 7).

2.4 Quantifiers

2.4.1 Indefinite quantifiers

Indefinite quantifiers include *pu*^{3}*nha* 'some', *tɔ*^{12} 'many', *samu*^{1}*tɔ*^{3}*t* 'a little', *batu*^{3}*p* 'all', and *fasa*^{21}*y* 'another'. They all modify both countable and uncountable nouns, animate and inanimate. Examples are presented in (11) and (12). As seen from (11), indefinite quantifiers follow adjectives, and precede demonstrative. The quantifier *dɛ*^{21}*rɛm* 'every' is exceptional in this respect – it precedes the noun, as in *dɛ*^{21}*rɛm la*^{121} 'every day'.

(11) *wa*^{3}*ŋ* *wa*^{3}*y* *pu*^{3}*nha* *i*^{21}*no*
boat small some DEM
'These few dugout canoes here.'

(12)^ *kayɔ*^{12}*w* *i*^{21} *di*^{1}*n* *fayo*^{21}*l* *i-dɛ*^{12}*l-ha* *batu*^{3}*p* *lɛ*^{12}*y*
tick 3s call invite 3s-friend-PL all real
'The tick invited really all of his friends.'

2.4.2 Numerals or definite quantifiers

Just like most indefinite quantifiers, numerals follow the noun, and in the sequence of noun modifiers they likewise fit between the adjective slot and any demonstrative. Table 3 presents an overview of the numerals. The system is base-10 and abstract, that is, there is

no obvious relation between the counting system and a physical reality (for example body parts). In general, all elements of the Matbat numeral system are derived from Proto-Austronesian or Proto-Malayo-Polynesian forms. They have cognates in closely related Ma'ya, and the organisational structure of numbers is also the same as in Ma'ya.

Table 3: The numeral system of Matbat

1	*i-tɛ^{3}m*	6	*i-nɔ^{12}m*	11	*ya^{41} ma-tɛ^{3}m*	100	*tu^{3}n ha^{3}*
2	*i-lu^{3}*	7	*i-fi^{3}t*	12	*ya^{41} ma-lu^{3}*	200	*tu^{3}n lu^{3}*
3	*i-tɔ^{3}l*	8	*i-wa^{3}l*	13	*ya^{41} ma-tɔ^{3}l*	1000	*yi^{3}p ha^{3}*
4	*i-fa^{3}t*	9	*i-si^{3}w*	20	*ya^{41} lu^{3}*	2000	*yi^{3}p lu^{3}*
5	*i-li^{3}m*	10	*i-ya^{41}*	30	*ya^{41} tɔ^{3}l*	10000	*yi^{3}p ya^{41}*

In the expression of multiples of ten, hundred, or thousand, a number between one and nine follows the word for *ya^{41}* '10', *tu^{3}n* '100', or *yi^{3}p* '1000'. When linked to a higher order number, the units take the prefix *ma-* 'plus'. These principles are illustrated in (13).

(13) *tu^{3}n lu^{3} ya^{41} li^{3}m ma-si^{3}w*
100 2 10 5 plus-9
'259.'

2.4.3 The system of numeral classifiers

Numbers between 1 and 10 take a classifier affix. These affixes on numerals refer to salient semantic properties inherent to the referent of the noun with which the numeral is associated (see Aikhenvald 2000). Numbers above 10 are not inflected in this way. The phenomenon is illustrated in (14). The noun in (14a) – *lɔ^{12}ŋ* 'fishing spear' – refers to a long object, and this salient property is marked by the prefix *di-* on the numeral *lu^{3}* 'two'. Similarly in (14b), *ho^{1}ŋ* 'mango' is a round object, and this is evident from the classifier prefix *pa-*. In (14c), association with an animate noun is marked on the numeral by a tone change, in addition to a prefix classifier. That is, the High tone on *lu^{3}* 'two' is replaced by the Low Rise, as the numeral modifies the animate noun *wa^{3}y* 'child'.

(14) a. *lɔ^{12}ŋ* *di-lu^{3}*
fishing_spear CLF-two
'Two fishing spears.'

b. *ho^{1}ŋ* *pa-lu^{3}*
mango CLF-two
'Two mangos.'

c. *wa^{3}y* *ha-lu^{12}*
child CLF-two:CLF
'Two children.'

The referents of nouns can be divided into six profiles on the basis of the numeral classifier they take. These six profiles are I. Animate; II. Boat and House; III. Long; IV. Round; V. Sago; VI. Rest. In a more general sense, we can distinguish three more basic semantic classes among these profiles: animacy (profile I); physical properties (profiles III and IV), and function (profiles II, V, and, as will be explained shortly, also I). These are the three general classes of meaning that languages use as they carve up the semantic space by means of noun categorisation devices (Aikhenvald 2000:271).

As seen from the examples above, the numeral classifiers are primarily prefixes, and, to a lesser extent, tonal suprafixes. There is no inflectional marking of the classification system on the noun itself nor anywhere else in the noun phrase. Moreover, numerals preceded by classifier prefixes constitute anaphoric expressions, which can come in the place of a nominal constituent. An example can be found in (32). The various numeral classifier affixes – including tonal changes – are illustrated in Table 4.

The 'Animate' affixes are found primarily on numerals associated with animate entities, both human and non-human. Examples include *ma^{3}t* 'person', *no^{3}* 'brother:3', *yɛ^{12}m* 'dog', *pɔ^{3}ntu^{1}* 'ghost', *ipo^{121}* 'fish' and *kasa^{12}s* 'kind of mollusc'. The animate class is distinguished from all other classes in that it takes the prefix *nu^{3}n-* / *nun-* on the numerals *tɛ^{3}m* and *ha* – both meaning 'one'. The distribution between these allomorphs is governed as follows: *nu^{3}n* combines with toneless *ha*; its toneless allomorph *nun-* prefixes to *tɛ^{3}m*, which has a lexical tone itself.

Table 4: The numeral classifiers associated with the six semantically based sets.

	I. Animate	**II. Boat and House**	**III. Long**	**IV. Round**	**V. Sago biscuit**	**VI. Rest**
1	*nun-tɛ^{3}m* *nu^{3}n-ha*	*ha-tɛ^{3}m* *ha^{21}-ha*	*di-tɛ^{3}m* *di^{41}-ha*	*pa-tɛ^{3}m* *pa-di^{41}-ha*	*ta-tɛ^{3}m* *ta-di^{41}-ha*	*i-tɛ^{3}m* *di^{41}-ha*
2	*(ha-)lu^{12}*	*ha-lu^{3}*	*di-lu^{3}*	*pa-lu^{3}*	*ta-lu^{3}*	*i-lu^{3}*
6	*(ha-)unɔ^{121}m*	*ha-unɔ^{12}m*	*di-nɔ^{12}m*	*pa-unɔ^{12}m*	*ta-unɔ^{12}m*	*i-nɔ^{12}m*
10	*(ha-)ya^{41}*	*ha-ya^{41}*	*di-ya^{41}*	*pa-ya^{41}*	*ta-ya^{41}*	*i-ya^{41}*

When two or more tokens are involved, then animacy is marked by a change in tone on the numeral, and optionally also by the prefix *ha-*. The only tonemes occurring on numerals 1-10 are the High /3/, the Rise /12/, and the Extra High Fall /41/. When modifying an animate noun, the High tone of the numeral is replaced by the Rise, the Rise is replaced by the Rise-Fall /121/ – and the Extra High Fall remains unchanged. This can be seen from Table 4, by comparing the tonemes on the numerals marked for Animate with those modifying members of the other five sets. While this tonal change uniquely distinguishes numerals marked for Animate, the prefix *ha-* is also used in the marking of 'Boat and House' nouns. For example, compare *ma^{3}t ha-unɔ^{121}m* 'six people', versus *wa^{3}ŋ ha-unɔ^{12}m* 'six dugout canoes'. It therefore does not come as a surprise that the redundant *ha-* prefix is optional when the tonal classifier is present. For example, both *ma^{3}t si^{12}w* and *ma^{3}t ha-si^{12}w* are both used with the meaning 'nine people'.

There are a few nouns referring to inanimate entities, which nevertheless take the same numeral classifiers as nouns with animate referents. These are *ba^{21}l* 'sago porridge' and *bi^{3}ga^{21}*, *bi^{3}na^{21}s*, *ma^{3}tma^{1}n*, all three of which refer to devices for the storage of uncooked sago. These three terms are also used to express amounts of uncooked sago. The set membership of these nouns is evident from an example such as *lɛ^{41}n bi^{3}na^{21}s ha-lu^{12}* 'two *binas* units of uncooked sago', where Animate class membership is tonally marked on *lu^{3}*. Sago is the source of carbohydrate for the Matbat, and the only food that can be preserved for extensive periods of time. Because these nouns relating to an economically important concept are included in it, the semantic basis of the Animate set is to be interpreted as partly functional, in addition to its primary grounding in animacy.

It is worthwhile noting that the system of numeral classifiers fits together seamlessly with the pronoun system. As noted in §2.2, pronominal reference to a third person plural entity tends to be specific as to the number of individuals involved, as in (15) – repeated from (6). The sequence *ha-lu^{12}* could be analysed as a combination of the numeral with the shortened form of the third plural morpheme *hafɔ^{12}*, just as the latter appears as a possessive pronoun on the object *ha-wa^{3}y*. Alternatively, however, *ha-lu^{12}* can be analysed as a floating quantifier, with the initial syllable being the prefix marking Animate set

membership, rather than as a contracted form of *hafɔ*^{12}. Further research with inanimate subjects and floating quantifiers is needed to settle this issue.

(15) *ha-lu*^{12} *hi*^{12}*p* *ha-wa*^{3}*y*
CLF- two: CLF wash 3p-child
'The two of them are washing their child/children.'

The second set ('Boat and House') comprises only three nouns, to the best of my knowledge. These are *wa*^{3}*ŋ* 'dugout canoe', *de*^{3} 'house', *ka*^{1}*pa*^{3}*l* 'ship'. An example is presented in (16). This paradigm is semantically grounded in function. Both *wa*^{3}*ŋ* and *de*^{3} refer to essential objects in Matbat life. Dugout canoes – usually with two outriggers – are the main medium of transportation. They are used to reach gardens upriver and also to reach other villages along the coast. The membership of the third noun – a Malay loan referring to machine-propelled ships, often with a non-wooden hull – is probably due to the semantic similarity with *wa*^{3}*ŋ*. Like the other classifiers, *ha-* is toneless in most contexts, but it appears with a tone when the following morpheme is toneless itself.

(16)^ *ak-k-o*^{21} *na*^{21}*w* *de*^{3} *ha*^{21}*-ha-no*
1s-1s-go see house CLF-one-DEM
'I am going to visit this house.'

The following two classes are based on physical properties. The prefix *di-* is found on numerals associated with noun referents that are long and often sharp. Examples include *lɔ*^{12}*ŋ* 'spear', *ya*^{21} 'trunk', *yu*^{41}*su* 'knife', *pa*^{1}*ku*^{3} 'nail' (Malay loan) and *wa*^{3}*n* 'conical rectangular carrying basket'. The numeral classifier prefix *pa-* is associated with concepts that are round-shaped. Many fruits are part of this set. Examples include *sapa*^{1}*y* 'papaya', *hɔ*^{1}*ŋ* 'mango', *nu*^{1} 'coconut', *mɔ*^{41}*n* 'areca nut', *na*^{1}*n* 'betel fruit', *pɔ*^{3}*n-so*^{1} 'citrus fruit', *manɔ*^{21}*n* 'eye', *bɛ*^{3}*m* 'plate', *gala*^{3}*s* 'glass' (loan from Malay or Dutch) and *tɔ*^{21}*l* 'egg'.

The word *ni*^{12} 'sago biscuit' is the only noun with which the classifier *ta-* associates. Sago biscuit is made by baking sago flour in a mold. It is the only locally available source of carbohydrate that keeps over longer periods of time. In a hunter-gatherer system without fridges or shops, its importance cannot be overestimated. So whereas the Sago biscuit set comprises one noun only, the difference with the other classes is smaller when the frequency in language use is taken into account.

The numeral classifier *i-* serves as the 'elsewhere' marker. It is found on numerals associated with nouns whose referent does not have one of the characteristics on the basis of which the other classifiers are selected. Examples include: *sate*^{21} 'piece of', *la*^{3}*n* 'song', *fe*^{1} 'shed', *gaŋgɛ*^{12}*ŋ* 'illness', *lu*^{1}*n* 'ladder', *tɛ*^{121}*lo* 'head', *kɔ*^{3}*s* 'T-shirt' (Malay loan), *bu*^{3}*y* 'cloth, sarong', *pata*^{21} 'tooth' and *meŋa*^{21}*wpo* 'hill'. Evidently, all nouns referring to abstractions take this classifier. Also, certain nouns can take either this default classifier or another, marked one. For example, a numeral modifying *tɔ*^{21}*l(o)* 'egg' takes either the 'Round' classifier *pa-* or the default classifier *i-*. Similarly, *pa*^{1}*ku*^{3} 'nail (Malay loan)' varies between 'Long' classifier *di-* and the 'Rest' classifier *i-*. Its association with the latter is probably due to the fact that *pa*^{1}*ku*^{3} is a loan word. Another loan word, *ba*^{3}*l* 'ball', also takes the default numeral classifier, even though the shape of the referent would appear to qualify it for membership of the Round set. The *i-* prefix is also used when the numerals are used outside the context of noun modification: for counting in the abstract, and when used as an adverb. In (17), for example, the numeral modifies the predicate head

na^{41}*ŋ* ‘tall’. It shows no semantically grounded agreement, neither with the subject nor with the predicate attribute.

(17) *Sɛ*^{3}*m* *dɛ*^{21}*l-te* *ak-ma*^{3}*m* *na*^{41}*ŋ* *i-tɛ*^{3}*m*
Sem with-CONJ 1s-father tall CLF-one
‘Sem and my father are the same height.’

In conclusion, it is evident that the numeral classifier system reflects the importance of concepts in the ecological situation. There are classifiers specifically for ecologically important items such as boats and dwellings on the one hand, and sago biscuit on the other. The importance of sago is also evident from the fact that several other words relating to sago take the same markers as nouns referring to animates.

So far we have come across two noun classification devices – first, the difference in marking of possession on alienably versus inalienably possessed nouns; and second, the numeral classifiers. There is a third candidate in this context. Whereas native nouns are predominantly monosyllabic, exceptions to this generalisation in the native vocabulary tend to involve a toneless initial syllable with the segmental structure *i-*, *ka-*, or *sa-*. Word patterns that begin with these syllables occur with conspicuous frequency in the lexicon. The same initial syllable is often found in the words for similar animals. For example, the words *kama*^{1}*y*, *kahɔ*^{12}*m*, *kala*^{12}*ŋ*, *kapi*^{121}, *kapna*^{41}, *karɔ*^{12}*n*, and *katubo*^{41} all refer to ant varieties. Similarly, *ifu*^{41}, *igɔ*^{12}*l*, *ikɔ*^{1}*t*, *iku*^{3}*m*, *impa*^{12}*n*, *imu*^{41}*n*, *imya*^{1}*p*, *ina*^{121}*m*, *iŋa*^{1}*k*, *ipo*^{121}, *isa*^{1}*ŋ*, *iwa*^{3}*nke*^{1}*w*, *iwa*^{1}*y*, *iwe*^{3}*y* and *iya*^{3}*m* all refer to birds. Interestingly, the initial *i-* is also found in several female kinship terms – *ila*^{12}*y* ‘female in-law’, *ima*^{12}*y* ‘sister-in-law’, *imo*^{12}*yu*^{3} ‘mother-in-law’, *ina*^{41} ‘widow’. However, there are also exceptions. For example, several kinds of birds do not have an initial syllable *i* – for example *sabalo*^{12} ‘king bird of paradise’, *kala*^{41}*ŋ* ‘eclectus parrot’, and *wi*^{1}*tna*^{41} ‘palm cockatoo’. The noun-initial *i-* is equally absent in *wayu*^{3} ‘woman’. This means that, if noun-initial *i-* once served as a classificatory marker that was prefixed to a set of nouns including those referring to birds and female animates, this process is no longer productive. More research is needed to establish the origin and the function, if any, of these high-frequency initial syllables.

2.5 Compounding

Nouns combine with nouns and other stems in the formation of compounds. There is a clear difference in compositional semantics between head-initial and head-final compounds. When the first stem is the head, the second stem contributes a more precise specification. In other words, a head-initial compound X-Y tends to have a meaning that can be paraphrased as ‘kind of X, characterised by Y’. For example, *mat-lo*^{3}*w* ‘Ma'ya’, which is made up of the nouns for ‘person’ and ‘sea’, refers to a kind of person. Similarly, *ni*^{12}*-ga*^{41}*n* ‘sago baked in bamboo’, consists of the words for ‘sago cake’ and ‘bamboo’, and refers to a kind of sago cake, and not to a kind of bamboo. More examples are listed in (18). An interesting case is *sakɔ*^{41}*m-ha*^{3}*y* ‘cassava’ in (18e). At first blush, it appears to be a counterexample, on the assumption that the second constituent *ha*^{3}*y* ‘wood, tree’ is the head. But this is not the case. The word *sakɔ*^{41}*m* means ‘sweet potato’, a starch similar to cassava in its function. The compound *sakɔ*^{41}*m-ha*^{3}*y* marks the difference between cassava and sweet potato by referring to a salient difference: while the cassava root projects a

slender tree, sweet potato does not. The correct paraphrase, then, is 'tree-like sweet potato' – compare Dutch *aard-appel* 'potato' from 'earth+apple'.

(18)	a.	*mat-ba^{3}t*	Matbat	'person+land,earth'
	b.	*pɔ^{3}n-so^{1}*	citrus fruit	'INDEF+sour'
	c.	*ni^{12}-ga^{41}n*	sago baked in bamboo	'sagocake+bamboo'
	d.	*ha^{3}y-wa^{1}t*	k.o. mangrove tree	'wood+k.o. mangrove'
	e.	*sakɔ^{41}m-ha^{3}y*	cassava	'sweet potato+wood'
	f.	*ko^{3}la^{3}k-sakɔ^{41}m*	mash of sweet potato	'mash+sweet potato'
	g.	*wo^{41}-napa^{21}t*	westerly wind	'wind+west'
	h.	*ifu^{41}-tarɔ^{1}t*	k.o. fruit dove	'fruit_dove+Ternate'

Head-final noun compounds can be interpreted as partitive or part-whole constructions.[5] That is, when the second stem is the head in the structure X-Y, then Y is a particular part of X. As seen from the examples in (19), this construction is used to refer to sections of the human body and of plants, among others. There is a parallel with possessive noun phrases here, which have the order possessor-possessed. The second noun is the head in both constructions, and the partitive construction and a possessive noun phrase are similar in meaning. The two constructions are still different, though, in that partitive compounds lack the linking clitic *i-*, a characteristic of possessive noun phrases involving a non-pronominal possessor (see §2.3).

(19)	a.	*we^{21}ta-ŋa^{21}l*	toenail	'leg:3 +branch, nail'
	b.	*pata^{21}-yi^{21}*	saliva	'tooth,mouth:3 +water'
	c.	*ha^{21}po-ta^{21}*	glans	'penis:3+front_side_of'
	d.	*na^{1}n-sate^{21}*	piece of betel fruit	'betel+piece of'
	e.	*nu^{1}-gɛ^{21}l*	coconut	'coco+fruit'
	f.	*ma^{1}t-da^{21}n*	guava leaves	'guave+leaf'
	g.	*lɛ^{41}n-bi^{3}na^{21}s*	k.o. unit of raw sago	'sago+k.o._unit_of_sago'
	h.	*mɔ^{41}n-pa^{21}n*	half of an areca nut	'areca_nut+half'

In most cases, compounding does not affect the phonological form: the segments and tones of the constituent word roots are preserved in compounds.[6] As a result, head-initial compounds with an adjective as the second root, such as *maba^{21}y-bu^{3}s* 'pig (nose:3-white)', or like *pɔ^{3}n-so^{1}* in (18b), are identical to noun-modifier constructions. Similarly, a head-final compound headed by a postposition, such as *ha^{21}po-ta^{21}* 'glans' in (19c), is morphologically identical to a postpositional phrase (see footnote 7).

The selection of noun classifiers in compound nouns is determined by the semantic characteristics of the head noun. This is illustrated by the examples in Table 5. For

---

5 I thank Michael Ewing for pointing this out to me, and also the parallel with possessive noun phrases.

6 The only exception I have noted in this respect are compounds with the noun *ma^{3}t* 'person/people'. Its High toneme disappears in the context of a following High toneme, as in *mat-ba^{3}t* 'Matbat' and *mat-lo^{3}w* 'Ma'ya', but *ma^{3}t-ma1t* 'corpse'.

example, *na^{1}n* ‘betel fruit’ takes the ‘Round’ numeral classifier *pa-*, but the compound *na^{1}n sate^{21}* ‘a piece of betel fruit’ takes the default classifier *i-*.

Table 5: Compounding and noun classifiers

Simplex noun		**Compound noun**		
na^{1}n	*pa-lu^{3}*	*na^{1}n*	*sate^{21}*	*i-lu^{3}*
betel_fruit	CLF-two	betel_fruit	piece	CLF-two
two betel fruits		two pieces of betel fruit		
nu^{1}	*pa-tɛ^{3}m*	*nu^{1}*	*ya^{21}*	*di-tɛ^{3}m*
coconut	CLF-one	coconut	trunk	CLF-one
one coconut		one coconut tree		

3 Verbs, verb phrases and main clause syntax

3.1 Verbs

Five classes of verbs can be distinguished on the basis of the subject-marking prefixes they take. One of these is controversial, as its inflectional markings are identical to those of inalienably possessed nouns. This class will be discussed in §3.1.2; the four other classes will be introduced first, in §3.1.1.

3.1.1 Main classes of verbs

The main four classes of verbs are illustrated by the verb paradigms in Table 6. Class I comprises vowel-initial verb stems, such as, *-a^{21}* ‘eat’, *-o^{12}l* ‘stand’, and *-ɛ^{41}n* ‘lie down’. These verbs take an onset consonant prefix. While the pronoun system distinguishes seven combinations of person, number, and inclusive versus exclusive, there are only four different class I prefixes. That is, some levels of the pronominal paradigm are not distinguished from one another in terms of class I prefixes. In particular, second singular and second plural are not distinguished from one another, and the same is true for third singular, third plural, and first plural exclusive. Very similar levels are distinguished in the inflectional marking of inalienable possession (§2.3). The only difference is that, in the marking of inalienable possession, the first person plural exclusive patterns along with the second person forms, rather than with the third person forms.

Table 6: The four Matbat verb classes, as illustrated by one member of each class

	Class I (*-ɛ^{41}n* ‘sleep’)	**Class II (*fu^{21}* ‘say’)**	**Class III (*bawo^{21}* ‘run’)**	**Class IV (*hɔ^{121}l* ‘sit’)**
1s	*k-ɛ^{41}n*	*fu^{21}*	*k-awo^{21}*	*s-ɔ^{121}l*
2s	*m-ɛ^{41}n*	*fu^{21}*	*bawo^{21}*	*hɔ^{121}l*
3s	*n-ɛ^{41}n*	*fu^{21}*	*bawo^{21}*	*hɔ^{121}l*
1pi	*t-ɛ^{41}n*	*fu^{21}*	*t-awo^{21}*	*s-ɔ^{121}l*
1pe	*n-ɛ^{41}n*	*fu^{21}*	*bawo^{21}*	*hɔ^{121}l*
2p	*m-ɛ^{41}n*	*fu^{21}*	*bawo^{21}*	*hɔ^{121}l*
3p	*n-ɛ^{41}n*	*fu^{21}*	*bawo^{21}*	*hɔ^{121}l*

The three other classes of verbs comprise stems that begin with a consonant. If these stems took the same onset prefixes as class I verbs, this would result in complex syllable onsets. This is what we find in related languages such as Taba (Bowden 2001) and Ma'ya (van der Leeden, no date), two other SHWNG languages featuring similar systems of verb inflection. In Matbat, however, complex onsets occur much less frequently, and they are never the result of affixation. Class II comprises onset-initial verbs that do not take any inflection for subject agreement. Examples include *fu*^{21} 'to say', *nɔ*^{41}*ŋ* 'to hear', *la*^{3}*s* 'to swim', and *me*^{21}*s* 'to lift up'.

Classes III and IV also consist of consonant-initial verbs. Here the onset consonant is replaced by a subject-marking prefix consonant in some forms. However, agreement marking on class-II and class-III verbs is more limited than for class I verbs. In the case of class III verbs, the initial consonant of the stem is /b/. These verbs take a prefix *k-* for first singular and *t-* for first plural inclusive. Other combinations of person and number are not marked through inflection. Examples include *bawo*^{21} 'to run', *bo*^{21} 'to go', and *bɔ*^{3}*t* 'to reach'. Class IV consists of verbs beginning with /h/. Here subject marking is even more limited. The first person forms are distinguished from other forms but not from one another (see Table 6). Examples include *hɔ*^{121}*l* 'to sit', *ha*^{3} 'to go up', *hu*^{3}*ŋ* 'to enter' and *ho*^{3}*y* 'to return'. Importantly, not all verbs that begin with /b/ or /h/ are in classes III or IV, respectively. Instead, some of these verbs – for example *bu*^{121} 'to embrace' and *ha*^{41}*l* 'to dig out' – are in class II, taking no subject marking at all.

In summary, the system of verb prefixes makes more detailed distinctions in relation to first person than for second and third person, just as we observed in the case of possessor marking on inalienable nouns. The verbal prefixes function as agreement markers rather than as anaphoric devices. That is, a subject-marking prefix is not a sufficient marking of the subject on its own – either a noun or a pronoun is required to express the subject. The distinction between the verb classes does not reflect a syntactic distinction. In particular, both transitive and intransitive verbs are found in each of the four verb classes. This is illustrated in Table 7, which presents transitive and intransitive members of each of the four verb classes. Within the set of intransitive verbs, however, verbs that are used as noun modifiers invariably belong to class II, which does not show any inflection.

Table 7: The four Matbat verb classes crossed with transitivity

	Class I	**Class II**	**Class III**	**Class IV**
Transitive	*-a*^{21} 'eat'	*na*^{21}*w* 'see'	*be*^{21} 'give'	*ha*^{1}*n* 'cook'
Intransitive	*-o*^{12}*l* 'stand'	*ni*^{12} 'laugh'	*bawo*^{21} 'run'	*hɔ*^{121}*l* 'sit'

When the subject is not a pronoun, then subject and verbal predicate tend to be linked by the clitic *i-*, the reduced form of the third singular pronoun *i*^{21}. This copular element is absent when the subject itself is a pronoun, as in (2). This clitic is also absent when the subject is inflectionally marked for inalienable possession, as in (20). Finally, the clitic *i-* is absent when the subject is relativised, as in (21) – but not in (22), where a different constituent in the subordinate clause is relativised.

(20) *k-ma*^{21}*ŋa* *ba*^{121}
1s-arm:1s stiff
'My arm is stiff.'

(21)^ *be*^{21} *mɔ*^{41}*n* *be-pɔ*^{3}*n-in* *ta* *hɔ*^{121}*l* *le*^{21}*yo*
give areca REC-PRN-DEM REL sit LANDW
'Give an areca nut to the one that is sitting landward.'

(22)^ *a*^{21}*w* *m-u*^{21}*n* *nu*^{3} *t-i-hɔ*^{121}*l-ɛ*^{31}
2s 2s-know place REL-3s-sit-YNQ
'You knew the place where he was sitting, didn't you?'

3.1.2 Inalienable verbs

There is a small set of verbal predicates that derive from inalienable noun stems. Most of the stems involved refer to body parts. The predicates that are based on them tend to refer to emotional or physical states (compare Klamer 2000), as in (23,24). An important characteristic of such verb predicates is that their meaning is not the predictable sum of the parts – the emotion expressed in these examples in conventionally fixed.

(23) *ya*^{21}*k* *lɔ*^{3}*ŋ-de*^{121} *ya*^{21}*wa*
1s want:1s-sick 2s
'I hate you.'

(24) *ya*^{21}*k* *lɔ*^{3}*ŋ* *i-fi*^{3}
1s want:1s 3s-good
'I am happy.'

Table 8 gives the inflectional paradigms of three such verbs. For the sake of comparison, a class I verb and an inalienably possessed noun are presented alongside.

Table 8: The inflectional pattern of inalienable verb predicates, contrasted with the paradigms of class I verbs and of inalienable nouns. The latter are represented by the Class I verb *-a*^{21} 'eat' and the inalienable noun *ma*^{21}*ta* 'arm+hand:3s', respectively.

	Emotional- / physical-state verbs			**Reference data**	
	'feel, want'	'have'	'sleep'	'eat'	'arm+hand'
1s	*lɔ*^{3}*ŋ*	*ni*^{3}*ŋ(a)*	*mɔ*^{3}*ŋyu*^{1}	*k-a*^{21}	*ma*^{21}*ŋa*
2s	*lɔ*^{3}*m*	*ni*^{3}*m(a)*	*mɔ*^{3}*myu*^{1}	*m-a*^{21}	*ma*^{21}*ma*
3s	*lo*^{21}	*ni*^{21}	*mɔyu*^{1}	*n-a*^{21}	*ma*^{21}*ta*
1pi	*lɔ*^{3}*n*	*ni*^{3}*n(a)*	*mɔ*^{3}*nyu*^{1}	*t-a*^{21}	*ma*^{21}*na*
1pe	*lɔ*^{3}*m*	*ni*^{3}*m(a)*	*mɔ*^{3}*myu*^{1}	*n-a*^{21}	*ma*^{21}*ma*
2p	*lɔ*^{3}*m*	*ni*^{3}*m(a)*	*mɔ*^{3}*myu*^{1}	*m-a*^{21}	*ma*^{21}*ma*
3p	*lo*^{21}	*ni*^{21}	*mɔyu*^{1}	*n-a*^{21}	*ma*^{21}*ta*

As seen from these examples, the inflectional marking of the 'inalienable verb predicates' exhibits the same general pattern as found in the inflections on inalienable nouns. Most of the verb predicates at issue here consist of two content word stems, formed by combining *lo*^{21} – inflected as in Table 8 – with a non-inflecting second stem. Eleven of such combinations are given in (25).

(25)	*lo*^{21}*-bɔ*^{3}*t*	'remember'	*lo*^{21}*-i-fi*^{3}	'be happy'	*lo*^{21}*-me*^{21}*y*	'forget'
	lo^{21}*-de*^{121}	'hate'	*lo*^{21}*-sa*^{1}*ma*^{3}	'agree'	*lo*^{21}*-saksa*^{12}	'be sad'
	lo^{21}*-dɛ*^{3}*l*	'love'	*lo*^{21}*-i-tɛ*^{3}*m*	'agree'	*lo*^{21}*-he*^{41}	'be jealous'
	lo^{21}*-sɔ*^{12}*k*	'be stupid'	*lo*^{21}*-kopa*^{21}*l*	'be crazy'	*lo*^{21}*-do*^{21}	'want to have'

The inalienable inflection on *lo*^{21} is indicative of a nominal origin. In related Ma'ya, its cognate *lo*^{3}*n* is a noun, referring to the 'seat of emotions' (van der Leeden 1993:20). In some of these constructions, the syntactic structure is in keeping with a phrasal-predicate analysis whereby some state is predicated of the seat of emotions. For example, in *lo*^{21}*-i-fi*^{3} 'be happy' the two stems are separated by the copular clitic *i-* (24). However, this analysis does not work for transitive predicates such as *lo*^{21}*-bɔ*^{3}*t* 'to remember' and *lo*^{21}*-de*^{121} 'to hate' (23). Both of these predicates take a direct object. On this basis, they can be interpreted best as compound verbs. Another example supporting a verbal interpretation is presented in (26), where *lo*^{2} appears by itself to express volition. Here *lo*^{21} must be the verbal predicate to which the following clause is subordinated.

(26) *ya*^{21}*k* *lɔ*^{3}*ŋ* *t-ak-ni*^{3}*ŋ* *lɔ*^{12}*ŋ*
1s want:1s REL-1s-have:1s spear
'I want to have a spear.'

Similarly, in an analysis of emotion predicates in three languages of Eastern Indonesia (Buru, Kambera, Tetun), Klamer (2000) shows that in each of these languages, emotion predicates can appear either as phrasal predicates or as compound verbs. On the basis of semantic, morphological and syntactic grounds, she argues that such constructions are lexical units, even when they are 'syntactically discontinuous', that is, in phrasal predicates.

Interestingly, the phenomenon of emotion predicates in Matbat is connected to inalienable possession. In inalienably possessed nouns, the possessor is marked by inflection on the body part (see §2.3), just as the subject is marked by inflection in various verb classes (see §3.1.1). This means that, in the process of reinterpretation of a lexicalised phrasal predicate as a compound verb, the marker for inalienable possession can be readily interpreted as subject-agreement marking. Some inalienably possessed noun stems can be interpreted in this way as verbal predicates. The example in (27) was recorded on a hike, as somebody commented on the fact that we were all sweating. This utterance refers to an event, and a verbal interpretation ('we are sweating') is more appropriate than a nominal one ('it is our sweat').

(27)^ *ya*^{21}*t* *si*^{3}*nbɔ*^{12}*t*
1pi sweat:1pi
'We are sweating.'

In some stems, this process of reinterpretation as verbs has progressed to the point where the nominal inflection is optional or absent. In the last case, the derived verb form corresponds to the unmarked (third person) form of the noun. In this way, *sabo*^{21}*m* 'to welcome (somebody)' and *sapa*^{3}*w* 'to be satisfied with food' in (28) are derived from the nouns meaning 'back' and 'belly', respectively.

(28) *ak-sapa^{3}w / ak-sa^{3}ŋpa^{3}w he-do^{21}*
1s-satisfied / 1s-satisfied:1s COMPL-EXIST
'I have already eaten enough.'

Apart from *lo^{21}* and complex predicates based on it, there are two other stems that also behave as predicates, even though they inflect as inalienable nouns: *mɔyu^{1}* 'to sleep' and *ni^{21}* 'to have'. The nominal origin of these forms is clear. The inflected forms of *mɔyu^{1}* 'sleep' (in Table 8) are conspicuously similar to those of *manɔ^{21}n* 'eye' (in Table 2). As for *ni^{21}* 'to have', its Ma'ya cognate is a possession marker in noun phrases. This is illustrated in (29) and (30), from van der Leeden (1993:83) and Remijsen (2001), respectively.

(29) *ˈni^{3}-k ˈu^{3}m-ya pa ˈle^{3} / ˈni-ya ga ˈna^{12}n*
POSS:1s-1s house-SG big POSS:3s-SG small
'My houses are big, those of him are small.'

(30)^ *wa-ka ˈsa^{12}m po ˈga^{3} ˈso^{12}m-ya ˈni kop ˈko^{12}p-si*
3s-distribute PRP tree holy-SG POSS:3s twig-PL
'They hang (them) in the twigs of the holy tree.'

As seen from these examples, Ma'ya *ni-* is a function word that inflects for the possessor, and that appears before the possessed term. It is clearly part of the noun phrase: in the second clause in (29), *ni-* is followed by *-ya*, which marks singular number on noun phrases (van der Leeden, personal communication). Its Matbat cognate *ni^{21}*, by contrast, functions as a verb. Examples are presented in (31,32), and also in (26). Examples (26) and (32) illustrate the use of *ni^{21}* in transitive predicates. In (32), *ni^{21}* is part of a serial verb construction. As we will see below, verb serialisation is common in Matbat.

(31)^ *ya^{21}k ni^{3}ŋa*
1s have-1s
'It is mine (lit. 'I have').'

(32)^ *yak ba^{1} k-ni^{3}ŋa i-tɔ^{3}l*
1s remain 1s-have:1s CLF-three
'I've got three [single-use bags of washing powder] left.'

It is likely that Ma'ya *ni* reflects the historical nature of this morpheme more than Matbat *ni^{21}* does, because the inflection it takes is characteristic of possessive noun phrases in both languages. If Matbat *ni^{21}* 'to have' represents a diachronic instance of the reinterpretation of noun phrase structure as verb phrase structure, then it becomes more plausible to postulate the same analysis synchronically, as I did in the analysis of (27).

3.2 Verb serialisation

It is common in Matbat for two verbs to be serialised, that is, to be combined in a single clause without evidence of either subordination or compounding. In (33), for example, both verbs are inflected, which rules out a compound analysis. Nonetheless, they are not separated by a conjunction or a complementiser, which would be the case if the second

verb would head a separate clause. Instead, I hypothesise that such constructions constitute a single main clause.

(33)^ *ak* *k-a*^{3}*p* *s-i*^{12}*p* *he*^{3}*-do*
1s 1s-row 1-wash COMPL-EXIST
'I have already washed.' [in the river, going out with the boat]

In serialisations, verbs appear with the same agreement prefixes which they take in a single-verb predicate, if any. Exceptions to this generalisation, in which a constituent verb root does not have the prefix which it would take as a single verb, can be attributed to grammaticalisation – they will be discussed at the end of this section. Serialised verbs are not separated by arguments. This suggests that Matbat serialised verbs are nuclear junctures in the framework of Foley and Olson (1985).

The following paragraphs and examples cover the various functions of serial verbs in Matbat. Many cases of serialisation can be considered as instances of co-lexicalisation in the sense of Givón (1991) – that is, the combination of the two verbs gives rise to a more detailed conceptual representation of an event. This is illustrated in (34) to (36), among others. In (34), for example, the presence of *-a*^{21}*t* conveys that the palm nut was peeled using the teeth. This added detail is expressed more succinctly than if the speaker had conjoined two main clauses, or if he had combined a main clause with a subordinated clause. In some cases, such as (36), a co-lexicalised serialisation corresponds to a monomorphemic verb in English and many other languages.

(34)^ *ya*^{21}*k* *k-a*^{21}*t* *wu*^{12}*ŋ* *mɔ*^{41}*n*
1s 1s-bite peel areca_palm_nut
'I peel an areca palm nut using my teeth.'

(35)^ *i-ni*^{12} *yɔ*^{3}*l*
3s-laugh yell
'He roars with laughter.'

(36) *ak* *k-ɛ*^{3}*ŋ* *k-u*^{21}*n-i*
1s 1s-see 1s-know-3s
'I recognise him.'

The component verbs that are juxtaposed in a serial verb construction in Matbat can describe different characteristics of a single event, as in (34) to (36), or a sequence of events, as in (37) and (38). In their typology of serial verb constructions in languages of Eastern Indonesia, van Staden and Reesink (2008) refer to these types as component versus narrative serialisation, respectively. As a communicative device, narrative serialisation is similar in function to a multiclausal packaging of the same information. In fact, the sentence in (38) is equally felicitous with the subordinating operator *te* separating the two verbs.

(37)^ *ak-ba*^{3}*s* *fasa*^{21}*l* *ku*^{3}*rsi*^{3}
1s-lift_up move chair [Malay loan].
'I lift up and move a chair.'

(38) *ak-k-o*^{21} *k-aya*^{21}*y* *hɔ*^{21}
1s-1s-go 1s-hang_out POL
'I am going for a walk.'

While co-lexicalised serialisations yield a more detailed conceptual description of an event, there are other kinds of serialisation, in which one of the two components has a more grammatical function. These more grammatical types of serialisation are well known cross-linguistically (Givón 1991:82–83, Foley and Olson 1985:48ff). In Magey Matbat, serialisations can be used to mark aspect, or to add a peripheral argument to the clause. The use of serialisation to specify the aspectual structure of the clause is illustrated in (39) and (40). Verbs such as *to*^{1}*y* 'to stop', *bɔ*^{3}*t* 'to reach', and *ho*^{3}*y* 'to return' have an inherent aspectual structure. They are frequently found in serial constructions, with the effect of communicating that the event is completed.

(39)^ *m-a*^{3}*l* *to*^{1}*y* *do*
2-take stop EXIST
'Stop taking now.'

(40)^ *ak-tu*^{21}*m* *bɔ*^{3}*t* *paro*^{21}
1s-speak reach IMM
'I have finished speaking.'

A second set of verbs is used in serialisations with the effect of adding arguments to the clause. One of these, *-u*^{21}*t* 'to take', adds an instrument slot to the argument structure of the clause, as in (41). Another, *na*^{21}*w*, has the meaning 'to see', but in serialisation its function is akin to that of a preposition, as it introduces a direction (42). Finally, the use of *bɔ*^{3}*t* in (5) can also be interpreted as a mechanism to add an argument (goal).

(41)^ *ak-k-a*^{3}*p* *k-u*^{21}*t* *lawi*^{12}*n*
1s-1s-row 1s-take sail
'I go rowing, taking the sail with me.'

(42)^ *ya*^{21}*k* *k-o*^{21} *na*^{21}*w* *de*^{3} *ha*^{21}*-ha-no*
1s 1s-go see house CLF-one-DEM
'I am going to visit this house.'

The use of these verbs as function morphemes represents a bleaching process. Both in the case of aspectual and argument-adding verb serialisations, the more grammatical constituent invariably fills the second slot of the verb sequence. This is noteworthy, because most clause-level modifiers appear at the end of the sentence, including those marking aspect, negation, questions, imperative, and politeness. This means that the final positioning of aspectual verbs in serialisations can be interpreted as a sign of their development towards clause-level grammatical morphemes. In the case of *bɔ*^{3}*t*, the process of grammaticalisation is far along: in (40) the verb does not take an agreement prefix, even though it does so when it appears as the single main verb of the clause. Similarly, *be*^{21} 'to give' also takes the class III prefixes when it is the only verb in a predicate, yielding the forms *k-e*^{21} '1s-give' and *t-e*^{21} '1pi-give'. But when it introduces a recipient, as in (43) and (44), this inflection is absent.

(43) *ya^{21}k k-u^{1}t ba^{1}k be-ak-nɔ^{3}ŋ*
1s 1s-take tobacco FOR-1s-brother:1s
'I bring tobacco for my brother.'

(44)^ *ya^{21}p-pɔ^{3}n bɛ-k-ma^{3}t-ha*
cook-PRN FOR-1s-person-PL
'I am cooking for my people.'

In these examples, grammaticalised *be* does not appear immediately adjacent to the other verb stem in the clause. This suggests that it is best interpreted here as a preposition. In other words, while serialisation may have played a role of the grammaticalisation of *be^{21}*, this process is completed in the synchronic grammar. Finally, the verb *na^{21}w* 'to see' is a class II stem. Because class II verbs do not take prefixes, the morphological heuristic to gauge grammaticalisation is not available. To the best of my knowledge, *na^{21}w* introduces an argument expressing a direction or a goal only when it is serialised with another verb stem. This suggests that *na^{21}w* has not yet fully bleached to a preposition.

3.3 Clause-level modifiers

In those non-Oceanic Austronesian languages in which the possessor precedes the possessed term in noun phrases, it is common for negation markers to appear in clause-final position (Himmelmann 2005:141–142). In Magey Matbat, this slot is the docking site for a wide range of clause-level modifiers, including negation, but it also accomodates markers for tense, aspect and modality (TAM).

3.3.1 Tense

In many utterances, the communication of tense is left implicit, to be inferred from context. For instance, while neither of the sentences in (45) and (46) have any marking of tense, (45) refers to an event in the past, while (46) refers to a future state. In these and other examples, a forward slash marks a prosodic phrase boundary.

(45)^ *hi^{12}p he^{3} / am-lu^{12} ho^{3}y*
bathe COMPL / 1pe-two:CLF return
'After bathing we returned.'

(46)^ *ya^{21}k k-o^{21} k-aya^{21}y he^{3} / k-s-o^{3}y k-ɛ^{41}n*
1s 1s-go 1s-hang_out COMPL / 1s-1-return 1s-lie_down
'I will go for a walk, and then I'll return to go to bed.'

Alternatively, tense can be expressed by means of adverbs or function words. The forms *ga^{21}(pe)* 'now' (47) and *beta^{21}* 'FUTURE' (48) both appear before the verb; *paro^{21}* 'now' is clause-final, see (49). Other modifiers are relatively free as to the position in the clause where they appear, and they can be interpreted as adverbs on that criterion. Examples include *nu^{3}mna^{121}n* 'tomorrow', *wo^{21}y* 'earlier on', and *wɔ^{21}l* 'later on'.

(47) *i^{21} ga^{21} i-bɔ^{3}t*
3s just_now 3s-reach
'He has arrived just now.'

(48)^ *ya^{21}k beta^{21} k-na^{3}w ak-wa^{3}y*
1s FUT 1s-see 1s-child
'I will meet my daughter.'

(49)^ *ho^{3}y a^{21}m paro^{21}*
return 1pe IMM
'Coming back already.' [reply to question 'Where are you going?']

3.3.2 Aspect

Aspect is marked syntactically in two ways – by means of sentence-final aspectual markers and by means of verb serialisation. The relevance of serialisation to the marking of aspect was covered in §3.2. In this section, we will consider the role of sentence-final markers in this context. The most frequently used sentence-final aspectual marker is *he^{3}*, which expresses completion. This morpheme does not imply a particular tense structure, which can be seen by comparing its use in (45), set in the past, versus (46), set in the future. In both of these sentences the subordinate clause is marked for completion by means of *he^{3}*. In main clauses, *he^{3}* is always followed by another sentence-final marker. This is illustrated in (50) to (53). Most frequently, the following marker is *do^{21}*, as in (50) and (51). We will consider *do^{21}* in detail in §4.When used in combination with *do^{21}*, *he^{3}* is sometimes realised without its High tone, and the quality of the vowel can be centralised. There are several other sentence-final aspectual markers, such as *pɔ^{121}re* 'not yet' (52), *sado^{21}* 'no longer' (62), and *sa^{41}yde* 'still'.

(50)^ *mat-lo^{3}w da^{1}y yi^{41} he^{3}-do^{21}*
person-sea draw water COMPL-EXIST
'The Ma'ya already draw water.' [re. upriver water collection in the dry season]

(51)^ *ya^{21}w hi^{12}p he^{3}-d-ɛ^{31} – he^{3}-do^{21}*
2s bathe COMPL-EXIST-YNQ COMPL-EXIST
Have you washed yourself already? Yes (lit. already)

(52)^ *i^{21} n-a^{21} pɔ^{3}n he^{3} pɔ^{121}re*
3s 3s-eat PRN COMPL NOTYET
'He has not yet finished eating.'

3.3.3 Markers of modality

There are three sentence-final operators marking modality. They are *fi^{3}*, *-u^{21}n*, and *sa^{21}m*. The former two, *fi^{3}* and *-u^{21}n*, are transparently related to content words. To begin with *fi^{3}*, this is originally an adjective, meaning 'good'. When it appears in sentence-final position, however, it adds a specific modality to the clause, which can usually be translated as 'to feel like', 'to like', 'to tend to be'. This is illustrated in examples (54,55).

(54) *ya^{21}k k-ani^{121}m ba go^{1}l fi^{3} po^{21}w*
1s 1s-drink INS sugar MOD NEG
'I don't like to drink (tea) with sugar.'

(55) *ya^{21}k k-a^{21} lɛ^{41}n ba^{21}l ta du^{3}-a fi^{3}*
1s 1s-eat sago porridge REL hard-BNDRY MOD
'I like to eat sago porridge that has become solid.'

The radical difference between the lexical and grammatical uses of *fi^{3}* is well illustrated in (56), where the lexical interpretation of *fi^{3}* 'good' is out of the question given the lexical meaning of *sɛ^{12}w* 'refuse, be naughty'. Ambiguity between the lexical and grammatical uses of *fi^{3}* is at issue, however, in a sentence like (57). Alternative to the gloss in the example, *fi^{3}* could act as an adverb meaning 'well', in which case the meaning would be 'he plays chess well.' The context is crucial here: there is no difference in word order between adverbial and auxiliary uses of *fi^{3}*.

(56)^ *i-sɛ^{12}w fi^{3}*
3s-naughty MOD
'He likes to be naughty / He is often naughty.'

(57) *i-baya^{21}y ska^{3}k fi^{3}*
3s-play chess MOD
'He likes to play chess.'

The second modal element, *-u^{21}n*, is derived from the verb meaning 'to know'. In addition to this lexical meaning, this morpheme can be used as a sentence-final modal marker, expressing 'to be able to, to have the skill to', as in (58) and (59). Interestingly, when *-u^{21}n* is used as a sentence-final modal operator, it takes agreement prefixes, just as it does when it is used as a main verb. In most cases, the order of constituents reveals how it is used. Whereas modal *-u^{21}n* follows the object, the lexical verb precedes it, as in (36). The third modal operator, *sa^{21}m*, illustrated in (60), also expresses ability, just as *-u^{21}n* does. In contrast with both *-u^{21}n* and *fi^{3}*, *sa^{21}m* is not derived from a content word.

(58) *ya^{21}w m-a^{3}p fara^{21}w m-u^{21}n-ɛ^{31}*
2s 2s-row fast 2s-MOD-YNQ
'Can you row fast?'

(59) *ya^{21}k ha^{41}l wa^{3}ŋ k-u^{21}no*
1s carve canoe 1s-MOD
'I can carve out a canoe.'

(60)^ *ya^{21}t famo^{21}l de^{3} sa^{21}mo*
1pi make house MOD
'We know how to build houses.'

Other modalities are expressed lexically or left to context. As seen from the examples in (61) to (63), for example, conditional clauses are not expressed lexically or morphologically.

(61) *na^{1} ka^{1}m wo^{21}l-ne / ya^{21}k s-i^{12}p po^{21}w*
rain evening next-DEM / 1s 1s-bathe NEG
'If it rains tonight, I won't bathe.'

(62)^ *aw-ma^{21}ŋa pla^{21}yo / aw-kɛ^{12}s sado^{21}*
2s-hand wounded / 2s-write NOLONGER
'If your hand would be hurt, you wouldn't write any longer.'

(63)^ *bafi^{3} do^{21} / aw-dɔ^{12}l do^{21}*
careful EXIST / 2s-fall EXIST
'Be careful, or you will fall.'

3.3.4 Constituent order and the syntax of clause- and sentence-final markers

The constituent order in Matbat is SVO. Object- and verb-initial orderings only appear in pragmatically marked constructions – see example (49). As seen from several of the examples above, the majority of clause- and sentence-level morphemes appear at the right edge of the SVO sequence. This is further illustrated in example (64). Here the negation marker *po^{21}w* is positioned clause-finally, separated from the main verb by a subordinate clause. An ordering in which *po^{21}w* appears earlier is ungrammatical.

(64)^ *ya^{21}k k-u^{21}n Ma^{3}rtɛ^{3}n i-nu^{1} te i-hɔ^{121}l po^{21}w*
1s 1s-know Marten 3s-place REL 3s-sit NEG
'I don't know where Marten is.' [lit.: 'I don't know Marten's place where he sits.']

The following interpretation – different from the one intended by the speaker – is also grammatical: 'I know the place where Marten is not sitting'. In this alternative interpretation, the negation modifies the subordinated verb. In the case of the sentence in (65), the two interpretations – with the negation modifying the main verb versus the subordinated verb – appear more or less equally likely, and context is crucial to determine which is correct.

(65) *ya^{21}k k-u^{21}n Ma^{3}rtɛ^{3}n n-amlɛ^{12} po^{21}w*
1s 1s-know Marten 3s-wash_face NEG
Either: 'I do not know whether Marten has washed his face.'
Or: 'I know that Marten has not washed his face.'

In addition to the TAM markers discussed above, the set of operators that occur in this position includes markers for politeness, questions, and commands. Illustrations of some of these appear in (66) to (69). In the context of the Florey's (this volume) study on negation in languages of the Moluccas, it is worthwhile pointing out that all operators that involve negation appear in clause-final position, see (66) and also (52), (62) and (64).

(66)^ *pe^{1}l ba^{21}ta na^{3}*
open door PROH
'Don't open the door!'

(67)^ *ya^{21}k k-o^{21} hɔ^{21}*
1s 1s-go POL
'I am going.' [with your permission / if you don't mind]

(68)^ *ma*^{3}*n* *ta* *ha*^{3} *nu*^{1} *ina*^{12}
who REL go_up coconut_tr. WHQ
'Who has climbed the coconut tree?' [to harvest this coconut]

(69)^ *ya*^{21}*w* *m-anɔ*^{21}*k* *p-ɛ*^{31}
2s 2-inhale NEG-YNQ
'You don't [want to] smoke, do you?'

The above examples and those in the previous subsections indicate that a wide range of clause-level modifiers are positioned at the end of the clause. When several clause-final markers co-occur in a sentence, their ordering is not random. Instead, a modal marker consistently appears closest to the predicate; *do*^{21} follows the aspect marker *he*^{3}, but precedes other aspect-marking operators, such as *sado*^{21}, etc. The complete ordering of the various clause-final particles is presented in (B).

B. modality *he*^{3} *do*^{21} aspect negation/prohibition politeness/question

Interestingly, *he*^{3} is the only aspectual marker that has been observed at the end of a subordinate clause, see (49) and (50). By contrast, the other aspectual markers, *do*^{21}, and the markers that appear further to the right in B have only been observed in sentence-final position. This suggests that *he*^{3} is clause-final, whereas the other aspect markers are sentence-final, or more precisely, associated with the right edge of the constituent consisting of main clause plus any subordinate clauses.

4 Non-verbal predicates

4.1 Existential, locational and possessive predicates

Existential predicates are constructions that introduce a concept without anything being predicated about it – compare 'there is/are X' in English. Existentials are formed using the function word *do*^{21}. In §3.3, this sentence-final marker already appeared in various examples involving a verbal predicate. One more of these appears in (70), and an adjective predicate appears in (71). In these sentences *do*^{21} communicates a pragmatic meaning akin to assertion or affirmation. Its independent contribution will become clearer as we examine existential predicates and related constructions.

(70) *wa*^{3}*y* *su*^{3}*m* *do*^{21}
child grow EXIST
'Children grow up.'

(71) *ha*^{1}*n* *fa*^{3}*s* *ina* *m*^{3}*uru*^{3} *wɔ*^{21}*t* *do*^{21}
cook rice PART easy just EXIST
'Cooking rice is easy.'

The use of *do*^{21} in non-verbal predicates is illustrated in (72) to (74). Example (72) presents a noun phrase with nothing being predicated of it beyond the statement of its existence, either in a general sense or in the more limited sense of availability. In the absence of a verbal predicate, *do*^{21} merely asserts existence or availability. Context is required to choose between a general existential interpretation, and an expression of

availability. This ambiguity is common in non-Oceanic Austronesian languages (Himmelmann 2005:138). Example (74) shows that *do*^{21} can equally act as an existential by itself, in which case it functions as an affirmation, mirroring the use of Malay *ada*.

(72) *ya*^{21}*k-ma*^{12}*y* *do*^{21}
1s-brother_in_law EXIST
'I have a brother-in-law.' / 'My brother-in-law is around.'

(73)^ *ba*^{1}*k* *do*^{21} *sado*^{21}
tobacco EXIST NOLONGER
'There is no tobacco left.' [tobacco has sold out in shop]

(74) *bɛ*^{3}*m* *do*^{21}*-ɛ*^{31} – *do*^{21} ; *do*^{21} *po*^{21}*w*
plate EXIST-YNQ EXIST EXIST NEG
'Are there plates?' 'Yes [there are].' ; 'No [there aren't any].'

There is also a negative existential marker, *mo*^{3}*n*, which is synonymous with *do*^{21} *po*^{21}*w* 'there is not', as in (75). As seen from (76), however, *mo*^{3}*n* can combine with *po*^{21}*w*, forming a double negation, which may carry extra emphasis with it.

(75) *yi*^{1}*n* *i-mo*^{3}*n-paro*
fish 3s-NEG.EXIST-IMM
'There is no fish left now.'

(76) *kalu*^{3}*f* *mo*^{3}*n* *po*^{21}*w*
mouse NEG.EXIST NEG
'There are no mice.'

The morpheme *do*^{21} plays a key role in non-verbal predicates that express location and possession. Each of these is discussed in turn. First, it is used as a preposition marking location. The prepositional phrase headed by *do*^{21} can appear in clauses with a verbal predicate, as in (77). Alternatively, this prepositional phrase can constitute the predicate by itself, as in (78,79). The use of *do*^{21} in locational predicates can be interpreted as a natural extension of its role as an existential marker. That is, the specification of a location following *do*^{21} restricts the scope of availability.

(77)^ *ya*^{21}*k* *hi*^{21}*p* *mni*^{12}*k* *da*^{1}*ga*^{3}*ŋ* *do-k-wɛ*^{3}*ŋa*
1s rub oil eucalyptus EXIST-1s-leg:1s
'I am rubbing eucalyptus oil on my leg.'

(78)^ *baya*^{1}*y* *i-do* *nu*^{3}*-no*
party 3s-EXIST village-DEM
'There is a party in this village.'

(79) *gata*^{3} *i-do*^{21} *bu*^{3}*k* *sabɛ*^{121}*mta*
bowl 3s-EXIST book chest:3
'The bowl is in front of the book.'[7]

When the argument of prepositional *do*^{21} is queried, we find the construction illustrated in (80). In this example, question status is marked by means of a tonal morpheme – the High Fall /31/. This tone pattern is only found at the right edge of certain questions. As illustrated in (81), question-word questions querying other prepositional phrases are formed in the same way, that is, by realising the stranded preposition with the question-marking tone pattern. There is no ambiguity between the existential and locational interpretations of *do*^{21} in questions: existential predicates are queried as yes/no questions, and these are marked by the clause-final yes/no-question operator *-ɛ*^{31}, as in (74).

(80) *ya*^{21}*w-ma*^{12}*y* *do*^{31} – *ya*^{21}*k-ma*^{12}*y* *do*^{21} *fa*^{3}*n-paye*^{3}
2s-brother_in_law EXIST:Q 1s-brother_in_law EXIST board-topside
'Where is your brother-in-law?' 'My brother-in-law is in Fafanlap.'

(81) *i-bɔ*^{3}*t* *ho*^{3}*y* *pɔ*^{31} – *i-bɔ*^{3}*t* *po*^{21} *fa*^{12}*n-paye*^{3}
3s-reach return FROM:Q 3S-reach FROM log-side
'Where are you returning from?' 'He is returning from Fafanlap [village].'

The existential and locational uses of *do*^{21} are also available in clauses that have a verbal predicate. In such sentences, *do*^{21} can head a prepositional phrase, as in (82), or it can be used without an argument as a marker of assertion or affirmation, as in (83). Other prepositional predicates, marking origin and direction, are constructed in parallel with the stative constructions involving *do*^{21}.

(82) *ya*^{21}*w* *m-amlɛ*^{12} *do*^{31} – *ya*^{21}*k* *k-amlɛ*^{12} *do*^{21} *yi*^{41}*-ba*^{3}*k*
2s 2s-wash_face EXIST:Q 1s 1s-wash_face EXIST water-container
'Where do you wash your face?' 'I wash my face at the waterpoint.'

(83) *ya*^{21}*w* *m-amlɛ*^{12} *do-ɛ*^{31} – *ya*^{21}*k* *k-amlɛ*^{12} *he*^{3}*-do*
2s 2s-wash_face EXIST-YNQ 1s 1s-wash_face COMPL-EXIST
'Have you washed your face?' 'I have washed my face already.'

Second, the existential marker *do*^{21} is also used to communicate possession at the clausal level. As seen from (84) and also from (72), *do*^{21} expresses possession when it follows a possessed noun phrase subject. The construction in (84) can be paraphrased analytically as 'Does your change exist?'. This pattern, whereby possessive clauses are based on an existential construction, is well known from other Austronesian languages (Himmelmann 2005:139). There are two other structures to express clause-level possession, using *ni*^{21} 'to have' (see §3.1.2) and the pronoun *pɔ*^{3}*n* (see §5.3).

---

7 Example (79) also shows the use of deictic postpositions. The nouns *sabɛ*^{12}*1mta* 'chest' and *sabɔ*^{21}*mo* 'back' are used both for spatial and for temporal deixis, conveying 'before, in front of', and 'after, behind', respectively. The postpositional construction can be considered as a possessor-possessed sequence (see also Reesink 1998:624).

(84)^ *ya*^{21}*w* *pi*^{3}*p* *wa*^{3}*y* *do*^{21}*-ɛ*^{31}
2s money small EXIST-YNQ
'Do you have change?'

Moreover, *do*^{21} also contributes the notion of possession in combination with *lo*^{21} 'to want', as in (85). Here *lo*^{21} contributes marking for possession by means of its inalienable pattern of inflection (see §3.1.2). This juncture is structurally similar to other compounds with *lo*^{2}, but now the constituent it combines with is a function morpheme.

(85) *ya*^{21}*k* *lɔ*^{3}*ŋ-do*^{21} *lɔ*^{12}*ŋ*
1s want:1s-EXIST spear
'I want to have a spear.'

Another interesting combination with *do*^{21} involves the noun *mo*^{3}*l* 'debt'. The combination of these two morphemes yields the meaning 'to be owed by'. An example is presented in (86). While this combination can be translated most naturally as a verb, it can best be analysed as an abstract variant of the locational predicate, in which *mo*^{3}*l* heads a possessed noun phrase, and *do*^{21} *ma*^{3}*t-ha* is an abstract expression of location.

(86)^ *peli*^{3}*pu*^{3}*s* *i-mo*^{3}*l* *do*^{21} *ma*^{3}*t-ha*
Pelipus 3s-debt EXIST person-PL
'Several people owe a debt to Pelipus.'

4.2 Predicate nominals

The syntax for nominal predication shows the same pattern reported for verb predicates (§3.1.1) and nominal possession (§2.3). The subject and the predicate are linked by the copular clitic *i-* (87,88), unless the subject is a pronoun (89).

(87) *ak-nɔ*^{3}*ŋ* *i-ma*^{3}*ntri*^{3}
1s-brother:1s 3s-health_visitor (< Malay)
'My brother is a health visitor.'

(88) *Sɛ*^{3}*m* *i-na*^{41}*ŋ*
Sem 3s-tall
'Sem is tall.'

(89) *hafɔ*^{12} *ma*^{3}*tsu-da*^{3}*n*
3p person-garden
'They are farmers.'

5 Discussion and conclusion

In the following subsections, we will consider three more topics in Magey Matbat grammar, all of which have a bearing on the typological characterisation of the language.

5.1 Grammaticalisation

Most genuine function words in Magey Matbat are clause-final markers. In addition, there are many weakly grammaticalised items, which have developed from content word stems. A list of these appears in Table 9. Most prepositions are part of this set.

Table 9: Grammaticalisation in Magey Matbat

Form	Content	Function
fi^{3}	'good'	Modal auxiliary 'like to, tend to'
-u^{21}*n*	'to know'	Modal auxiliary 'be able to'
be^{21}	'to give'	Preposition 'to (dative)'; Conjunction 'in order to'
ba^{41}	'to hit'	Preposition 'with (instrument)'
-u^{21}*t*	'to take'	Grammatical serialisation 'with (instrument)'
na^{21}*w*	'to see'	Grammatical serialisation 'towards (direction)'
dɛ^{12}*l*	'friend'	Preposition 'with (accompaniment)'
du^{3}*m*	'to follow'	Conjunction 'and'
sabɛ^{121}*mta*	'chest'	Postposition 'in front of'
sabɔ^{21}*mo*	'back'	Postposition 'behind'
bɔ^{3}*t*	'to reach'	Grammatical serialisation – perfective aspect

5.2 Lexical classes

Lexical class membership cannot be determined morphologically, for lack of inflections and derivations that combine with any member of a given class. Instead, nouns can be defined as constituents that can head noun phrases, that is, phrases with the constituent sequence represented in A. Verbs can be defined as constituents that can freely combine in the formation of serial verbs.

With respect to adjectives, one option is to postulate a separate lexical class; the other option is to categorise them as (intransitive) verbs (see discussion in Baker 2003). Again, inflection does not provide us with a straightforward heuristic. Only a subset of verbs take agreement-marking prefixes, and both transitive and intransitive verbs are found in each of the four verb classes (see §3.1.1). Instead, a defining characteristic of adjectives is that they be modified by the comparative *nyɛ*^{3}*t*, as in examples (90) and (91).

(90) *Sɛ*^{3}*m* *i-na*^{41}*ŋ* *nyɛ*^{3}*t-aka*
Sem 3s-tall CPV-1s
'Sem is taller than me.'

(91) *iko*^{3} *i-fi*^{3} *nyɛ*^{3}*t* *supermi*^{3}
k.o.mollusc 3s-be_good CPV instant_noodle [Malay loan]
'*Iko*-mollusc is tastier than instant noodle.'

5.3 A parallel between noun phrases and predicates

There is a clear parallel between noun phrases and predicate clauses. The sequence POSSESSOR-NOUN is structurally identical to the structure SUBJECT-PREDICATE. In each case, (a) the specifier precedes the head, and (b), a non-pronominal specifier tends to be linked to the head by the copular clitic *i-*. This parallel is illustrated in examples (92,93).

(92) a. *ha^{3}y i-ho^{41}w*
tree 3s-bloom
'The tree is in bloom.'

b. *ha^{3}y i-ho^{21}w*
tree 3s-flower
'The flowers of the tree.'

(93) a. *ya^{21}w bawo^{21}*
2s run
'You run.'

b. *ya^{21}w na^{3}n*
3s name
'Your name.'

As seen from (94), this parallelly can lead to structural ambiguity. The word *wa^{3}y* in (94a) is homonymous between the noun meaning 'child' and the adjective meaning 'small'. Similarly, a structure like *aw-la^{3}y*, does not by itself contain a conclusive indication as to the word class of the head. Both verbal ('you are wrong') and nominal ('your error') interpretations look plausible. However, there is the compound verb *-a^{21}-la^{3}y* 'to eat in transgression of a taboo', which suggests that *la^{3}y* is a verb.

(94) a. *i-wa^{3}y*
3s-small OR 3s-child
'It is small.' 'His child.'

b. *aw-la^{3}y*
3s-be_wrong
'You are wrong.'

The parallel extends to the fact that both verb predicates and nominal heads can show agreement with their specifier, that is subject (see §3.1.1) and inalienable possessor (see §2.3), respectively. These similarities between inalienably possessed noun phrases and verb phrases may have contributed to the development of inalienable nouns into verb predicates (see §3.1.2). However, it is not only inalienable nouns like *sabɔ^{l2}t* 'sweat:3' in (27) that behave like verb predicates. In example (95), the indefinite pronoun *pɔ^{3}n* is used as a verb. In its basic pronominal use, this pronoun comes in the place of any argument, animate or inanimate, be it a subject, an object (52) or a prepositional modifier (21). In (95), however, it is used as a verb predicate, as is evident from its class III verb inflection. If it were not for this prefix, a noun-phrase interpretation 'our thing' would be conceivable. Just as *ni^{21}* (§3.1.2), this is another function morpheme that has its origin in noun phrase syntax, but that has turned into a verb predicate.

(95) *ya^{21}t batu^{3}p at-t-ɔ^{3}n*
1pi all 1pi-1pi-PRN
'It belongs to all of us.'

In summary, the structure of noun phrases parallels that of verb phrases, and several stems have crossed over from functioning as a noun-phrase head to behaving as a verb-phrase head. Inalienable nouns that undergo this process often retain the pattern of inflection that is characteristic of inalienable nouns.

5.4 Conclusion

In this conclusion, I will situate Matbat in the context of typological frameworks that have been developed in the study of languages of Eastern Indonesia. Throughout this study, we have repeatedly encountered characteristics that identify Magey Matbat as a preposed-possessor language, in terms of the typological framework for non-Oceanic Austronesian languages presented in Himmelmann (2005). Type-affirming characteristics include SVO word order, preposed-possessor sequence in possessive noun phrases, the alienable/inalienable distinction, person-marking prefixes on verbs, clause-final negators, and the fact that quantifiers follow the head. These correspondences are to be expected, given the geographical location of Magey Matbat: 'Preposed-possessor languages in this sense are the non-Oceanic Austronesian languages of Timor, the Moluccas and West Papua, as well the Pidgin-derived Malay varieties.' (Himmelmann 2005:113)

Conversely, however, many Austronesian languages of Central and Eastern lack possessor-possessed word order (Klamer 2002), and similarly clause-final negation is not ubiquitous in the region (Florey, this volume). In an alternative, more geographically restricted typology, Klamer presents an overview of the characteristics that distinguish the Austronesian languages of Central and Eastern Indonesia from those to the west. Magey Matbat shares several features with other Austronesian languages of Central/Eastern Indonesia. In the morphology, there is the expression of emotion predicates using complexes that relate to body-part nouns; subject indexation on verbs through affixes and clitics; and the already mentioned alienable/inalienable distinction among nouns. In the syntax, Magey Matbat patterns along with Austronesian languages of Central and Eastern Indonesia in terms of its verb-medial constituent order and the absence of a passive construction.

Reesink (1998) presents a survey of shared grammatical features in the Papuan languages of the Bird's Head of New Guinea. Several of the above-mentioned characteristics are equally common there, including SVO word order, verb serialisation, sentence-final negation, and the originally Austronesian contrast between in- and exclusive first person plural. In addition, there are some interesting parallels and differences between Matbat and the Papuan languages of the Bird's Head. First, in addition to sentence-final negation, '[all of the languages surveyed] have a sentence-final aspectual marker, translatable in most cases as 'already': (…)' (Reesink 1998:617). Second, most of the Papuan languages surveyed have spatial nouns, which express location relative to the preceding noun. Both of these phenomena appear in Magey Matbat – see §3.3.4 and footnote 7, respectively. By contrast, a phenomenon that sets Magey Matbat apart from Papuan languages, is its system of classifiers. Subsets of nouns are distinguished in two ways. Firstly, there are the distinct patterns of marking for alienably versus inalienably possessed nouns – body-parts and kinship terms versus the rest. Secondly, there is the numeral classifier system, based on a number of semantic properties. In contrast, the type of classifier system considered characteristic for Papuan languages, both those on Bird's Head (Reesink 1998:621) and elsewhere (Foley 1986:81), involves a sex-based distinction in the pronominal system, usually referred to as gender. Clearly, the Magey Matbat classifiers are not sex-based, nor do they affect the pronominal system.

In summary, there are clear similarities between Magey Matbat and Austronesian languages in general, and with those of Central and Eastern Indonesia in particular. These shared features support an Austronesian classification, as hypothesised by Blust (1978). However, when similarities between Magey Matbat and the Papuan of the mainland of New Guinea are taken into consideration, and between the latter and the Austronesian

languages of Central and Eastern Indonesia as a whole, it becomes clear that Matbat is part of a transitional area of shared Austronesian and Papuan characteristics.

Acknowledgements

I thank the people of Magey for their warm welcome, for teaching me about their language, and for sharing community life with me. Their hospitality encouraged me to go back and continue the study of Matbat after the first fieldwork trip. In particular, I thank my hosts, the Hamui family, for the hospitality I enjoyed in their home.

I thank the organisers of the Fourth Eastern Nusantara Linguistics Conference for giving me the opportunity to present this study, and I am indebted to the editors and to the reviewer for valuable feedback on earlier versions. The first fieldwork trip to Magey (8 days on-site in March 2001) and the analysis of the resulting data were funded by the Royal Netherlands Academy of Arts and Sciences (KNAW). The second fieldwork trip (45 days on-site in May and June 2003) and the analysis of the resulting data was funded by a postdoctoral grant of the Netherlands Organization for Scientific Research (NWO) to Vincent van Heuven (grant no. 355-70-014). The institutional support of the Max Planck Institute for Evolutionary Anthropology (Jakarta field station) and Atma Jaya University were crucial to make the second fieldwork trip possible. I am very grateful to Uri Tadmor of the Jakarta field station for support and hospitality in Jakarta. Transport to Misol was provided by Pt. Ramoi and Pt. Yellu Mutiara. In particular I thank the manager of Pt. Yellu Mutiara on Misol, Taufik Wijaya. Finally, I wrote up this chapter while on a postdoc grant funded by the UK Arts and Humanity Research Council (Research Grant no. MRG-AN11781/APN19394). I thank these institutions for their support.

References

Aikhenvald, A.Y., 2000. *Classifiers; A typology of noun categorization devices*. Oxford: Oxford University Press.

Baker, M.C., 2003. *Lexical categories; Verbs, nouns and adjectives*. Cambridge Studies in Linguistics 102. Cambridge: Cambridge University Press.

Blust, Robert, 1978. Eastern Malayo-Polynesian; A Subgrouping Argument. In S.A. Wurm and L. Carrington, eds, *Second International Conference on Austronesian Linguistics: Proceedings*, 181–234. Canberra: Pacific Linguistics Series.

Bowden, John, 2001. *Taba: Description of a South Halmahera Austronesian language*. Volume 521. Canberra: Pacific Linguistics.

Foley, William A. and Mike Olson, 1985. Clausehood and verb serialisation. In Johanna Nichols and Anthony C. Woodbury, eds, *Grammar inside and outside the clause; Some approaches to theory from the field*, 17–60. Cambridge: Cambridge University Press.

— 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.

Florey, Margaret. Negation in Moluccan languages. (This volume)

Givón, Talmy, 1991. Serial verbs and the mental reality of 'event': Grammatical vs. cognitive packaging. In E.C. Traugott and B. Heine, eds, *Approaches to grammaticalisation; Volume I: Focus on theoretical and methodological issues*, 81–128. Amsterdam: John Benjamins

Himmelmann, Nikolaus P., 2005. The Austronesian languages of Asia and Madgascar: typological characteristics. In K.A. Adelaar and N.P. Himmelmann, eds, *The Austronesian languages of Asia and Madagascar*, 110–81. London: Routledge.

Klamer, Marian, 2000. Phrasal emotion predicates in three languages of Eastern Indonesia. In G. Booij and J. van Marle, eds, *Yearbook of Morphology 2000*, 97–122. Dordrecht: Kluwer Academic Publishers.

— 2002. 'Typical features of Austronesian Languages in Central/Eastern Indonesia, *Oceanic Linguistics* 41:363–83.

Remijsen, Bert, 2001. *Word-prosodic systems of Raja Ampat languages*. LOT dissertation series vol.49. Utrecht.

— 2002. Lexically contrastive stress accent and lexical tone in Ma'ya. In C. Gussenhoven and N. Warner, eds, *Laboratory Phonology 7*, 585–614. Berlin: Mouton de Gruyter.

— 2007. Lexical tone in Magey Matbat. In V.J. van Heuven and E. van Zanten, eds, *Prosody in Indonesian languages*, 9–34. Utrecht: LOT Occasional Series vol.9.

Reesink, Ger, 1998. The Bird's Head as Sprachbund. In J. Miedema, C. Odé and R.A.C. Dam, eds, *Perspectives on the Bird's Head of Irian Jaya, Indonesia – Proceedings of the Conference*, *Leiden 13–17 October 1997*, 603–42. Amsterdam: Rodopi.

Smits, L. and Voorhoeve, C.L., 1992. *The J.C. Anceaux collection of wordlists of Irian Jaya Languages. A: Austronesian languages*. Leiden: Leiden University, Department of Languages and Cultures of South-East Asia and Oceania (Irian Jaya Studies Project).

Staden, Miriam van and Ger Reesink, 2008. Serial verb constructions in a linguistic area. In G. Senft, ed., *Serial verb constructions in Austronesian and Papuan languages* 17–54. Canberra: Pacific Linguistics.

van der Leeden, A. C., no date. Ma'ya morphology and syntax. (Unpublished ms, Radboud University, Nijmegen.)

van der Leeden, A. C., 1993. Ma'ya: a language study: A. Phonology. Jakarta: Lembaga Ilmu Pengetahuan and Leiden University.

Wallace, A.R., 1869. *The Malay archipelago*. London: Macmillan.