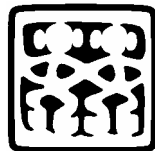# PACIFIC LINGUISTICS

Books in .pdf format

This .pdf (portable document file) document is an image file produced from a scan of an original copy of the book. It may not be reproduced without the author's permission.

Pacific Linguistics

Research School of Pacific and Asian Studies

THE AUSTRALIAN NATIONAL UNIVERSITY

URL: http://pacling.anu.edu.au/index.html

Spices from the East

Papers in languages of eastern Indonesia

Also in Pacific Linguistics

Pacific Linguistics 503

Spices from the East

Papers in languages of eastern Indonesia

edited by

Charles E. Grimes

Pacific Linguistics

Table of contents

Maps, tables and figures

Maps

Tables

Figures

Eastern Indonesia and East Timor
(Central Malayo-Polynesian languages framed)
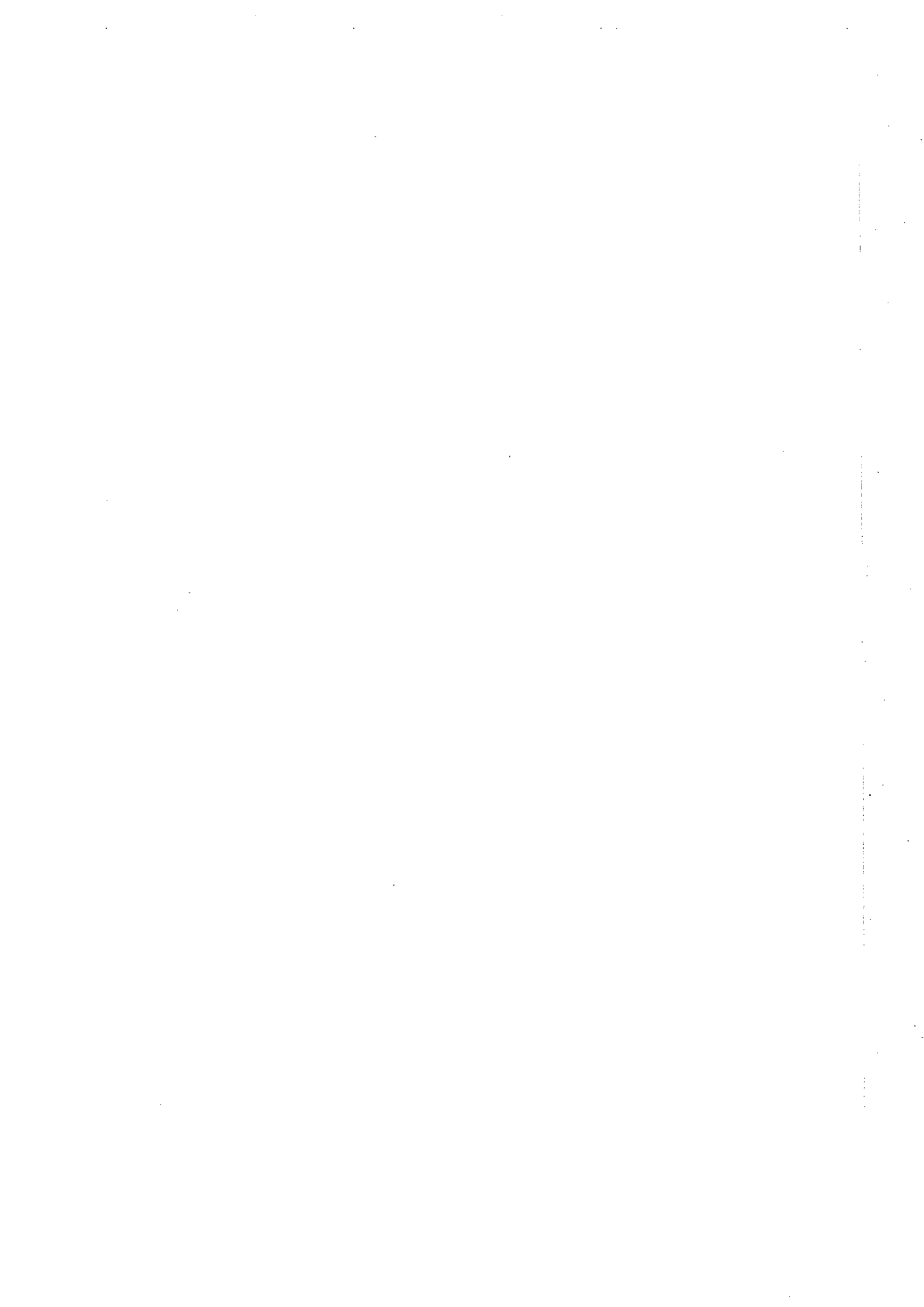
1 Introduction: new information filling old gaps in eastern Indonesia

CHARLES E. GRIMES

The papers in this volume are, in a very real sense, filling gaps in our knowledge of the languages of eastern Indonesia. While most papers can claim to fill a gap of some sort, very little has been published to date concerning the languages covered in this volume. The papers are arranged alphabetically by author. The regions covered by the papers in this volume (moving from west to east), Tukang Besi, Buru, Wetar, Selaru, Fordata, and Aru, are located in the map.

David Coward and Naomi Coward's 'A phonological sketch of the Selaru language', while providing a detailed phonological description of a complex language, also contributes to our understanding of natural classes in phonology by presenting a strong case for the need (in Selaru) to treat a class of glides (G) differently from the class of vowels (V) and consonants (C) on the basis of their unique distributional behaviour. These glides do not interact with the skeletal CV–tier until the morphemes of which they are a part are placed in the context of other morphemes in an utterance. Only then are they interpreted as a C or V.[1]

While for some languages in the Timor–south-western Maluku region one could perhaps claim that certain combinations of subject prefixes with verb roots trigger segmental metathesis, this Coward and Coward paper demonstrates that for Selaru such a phenomenon is more appropriately characterised as a spreading of features in an autosegmental model. This brings reliable data and solid analysis into circulation which will be greatly welcomed by comparative linguists interested in the region who have had access only to Drabbe (1932a). The paper includes an introduction to the various pronominal systems, as well as the genitive and possessive systems. Like some Oceanic languages (see Pawley 1973), Selaru also has a special 'edible possession' construction. Because parts of these systems interact with the glides and morphophonemic processes, these subsystems are also

---

[1]
AN	Austronesian	PANS	Proto Austronesian Stresemann (1927)
C	consonant	PCMP	Proto Central Malayo Polynesian
G	glides	V	vowel
OIN	Original Indonesian	VBG	Verhandelingenvan het Bataviaasch
PAMS	Proto Ambonese Stresemann (1927)		Genootschap
PANS	Proto Austronesian Stresemann (1927)	WMP	Western Malayo Polynesian

introduced in this paper along with a wealth of data on reduplication. This phonology sketch foreshadows a fuller grammar of Selaru (D. Coward 1990).

Mark Donohue's 'Tukang Besi dialectology' explores the speech varieties found on the Tukang Besi island chain located off of South-east Sulawesi, east of Buton. Works as recent as Anceaux (1978) and van den Berg (1991) have noted the almost complete lack of information and lack of linguistic information from this chain of islands. Donohue documents a number of similarities and differences among the speech varieties, looking at lexical similarity, and lexical and phonological innovations from a historical and comparative perspective.

Charles Grimes' paper 'Defining speech communities on Buru Island: a look at both linguistic and non-linguistic factors' takes a multi-disciplinary look at speech varieties on Buru. Collins (1982:84) notes that "no dialect study [of Buru] has ever been undertaken". This paper aims to begin to remedy that situation. In the process it is demonstrated that no single method of looking at the speech varieties is sufficient to unravel the picture, but only by interweaving a variety of approaches do the anomalies begin to make sense. Discrepancies in the historical sound correspondences are explained by looking at the historical record and the contact situation in various parts of the island, particularly the role of a forcible resettling of part of the coastal population of Buru by the Dutch in the mid-1600s. Discrepancies between reported information, lexical similarity, and historical sound correspondences are partially accounted for by looking at a variety of taboo practices. Widespread lexical tabooing is found on the island, but of particular note is the special speech registar known as *Li Garan* which is entirely driven by a motivation of taboo. Other languages are addressed, such as Kayeli and Hukumina, and well as immigrant languages and languages of wider communication. This paper is, in essence, a sociolinguistic case study.

Bryan Hinton's overview of 'The languages of Wetar: recent survey results and word lists, with notes on Tugun grammar' provides valuable data and discussion of this little known corner of the Austronesian world near the island of Timor. There has effectively been a gap in the literature on this island since the travelogue/report by Riedel (1886). So little is known of the languages of Wetar, that the island and its languages go unmentioned in Collins' (1982) survey of the region. In addition to word lists, this paper contains preliminary sketch material of the Tugun language, where Hinton did additional field work following the initial survey trip. Data in the word lists (such as the behaviour of reflexes of *maRi) provide counter-evidence to some of the preliminary claims made by Blust (1990) for subgrouping elsewhere in Maluku. The data in this paper also establish a direct link between the Talur language of Wetar and the Galoli on East Timor, something that has been suggested elsewhere, however, without demonstration.

Jock Hughes' 'The morphology of Dobel, Aru, with special reference to reduplication' provides a phonological sketch and a detailed introduction to the morphology and morphophonemics of Dobel, in the eastern part of the Aru islands. Collins (1982:127) comments that "Perhaps of all the areas in Indonesia, Aru is the least known". The data provide evidence for Dobel to be classified as a split–S language with the Actor proclitics attaching to active intransitive verbs and the Undergoer enclitics attaching to non-active intransitive verbs. The genitive and possessive constructions are explored, as are the functions of a valence decreaser *-r*, and similar derivational affixes. Dobel exhibits a number of forms and functions of reduplication that map functional contrast. Of particular

note is C–reduplication which results in the phonetic gemination of initial C, including initial glottals.

Craig Marshall's 'A phonological description of Fordata' provides a modern phonological description of what is, relatively speaking, one of the better documented languages of Maluku. The language is best known from Drabbe's grammar (1926) and dictionary (1932b). The elaboration of dialect information in Marshall's phonology provides a laboratory for seeing patterns of phonological change at a glance. The detailed phonology argues for certain consonants to be considered ambisyllabic as a clean way to account for otherwise complex shifts in syllable membership. The paper also includes a discussion of the phoneme-driven reduplication found in Fordata.

The classification of the languages described in this volume

With the exception of Tukang Besi, all the languages in this volume fall broadly under the label of the so-called Central Malayo-Polynesian (CMP) languages (Blust 1978, 1990).

Under the framework presented in Blust (1978) Tukang Besi is assumed to group with the Western Malayo-Polynesian languages. However, what languages Tukang Besi links up with is unclear and is still under study (Mark Donohue, pers. comm.). It does not link easily with other nearby language groups of South and South-east Sulawesi, nor with Moluccan languages to the east, nor with languages of the Lesser Sundas to the south.

There is on-going debate about the status of CMP as a subgroup (Blust 1990, Grimes 1991, Pawley and Ross 1993, Ross, 1995). It appears that if CMP is a subgroup of some sort, then it is not of the sort that had a parent language of which all daughter languages share common innovations. We do not yet know what is diagnostic of a CMP language and how this compares and contrasts with other subgroups. Furthermore, if there is such a subgroup, there is little consensus in the literature as to the internal relationships among the languages within the purported scope of CMP. In other words, we do not yet have even a general consensus of how Selaru, Fordata, Dobel, the Wetar languages, and Buru are related to each other, even though there is a common assumption that they are related in some way under the broader umbrella of Malayo-Polynesian.

While the notion of a major grouping of Austronesian (AN) languages in eastern Indonesia seems to have gained widespread and long-standing acceptance in scholarly circles, as recently as March 1990, Robert Blust (1990:3) observed that "to date no published argument in support of the Central Malayo-Polynesian (CMP) hypothesis exists".

Very little has been known about the CMP languages. Clark (1987) said there may be 40–50 languages in the group. Blust (1990) estimates '90-odd' languages. Grimes, Grimes, Ross, Grimes and Tryon (1995) list 150 CMP languages by name with additional information on location, number of speakers, etc.[2]

The languages of eastern Indonesia have long been the subject of debate as to how they fit into the larger Austronesian picture. The debate has tended to be marginal to the discussion of AN as a whole, however, and based on very little data.

---

[2]   This puts CMP numerically on a par with the Philippines which has around 160 languages. The latter is relatively well studied while the former is almost untouched.

The justification for talking about the AN languages of eastern Indonesia as a grouping within AN is based primarily on a series of independent observations and assumptions going back at least as far as Brandes (1884), Jonker (1908, 1913), Friederici (1913), Esser (1938), (many referring to each other) that the languages of Nusa Tenggara (the Lesser Sundas) and Maluku are somehow different from those to the east and west. For the most part, however, the arguments have been by declaration rather than by demonstration.

The 'reverse genitive' was noted by Van Hoëvell (1877) for languages of Ambon-Lease. Brandes (1884) used word order of the genitive construction as a basis for dividing Austronesian languages into west and east (drawing a line between Sulawesi and Maluku, and through Flores known as the 'Brandes line'). Kern (1906) suggested languages of eastern Indonesia might be transitional between east and west (noted in Collins 1983:27). Friederici (1913) supported the classificatory relevance of the reversed genitive. Jonker (1914) rejected Brandes' and Friederici's use of word order as a basis for subgrouping. Cowan (1951–52) and Capell (1976) continued to refer to it. The reverse genitive is likely to be the result of contact-induced change, and thus represents a typological shift that has no bearing on genetic relationships and is thus irrelevant for subgrouping (Grimes 1991b:290– 292).

Stresemann, an ornithologist-linguist who wrote a grammar of Paulohi on south-central Seram (1918) with additional comparative notes on Seram and Buru, returned to Germany and worked side-by-side with Jonker and Otto Dempwolff where he produced a volume (1927) of comparative Austronesian reconstructions arguing for 'Ur-Ambon' as a distinct subgroup within it (i.e. Seram & Buru, excluding Sula, and excluding Geser). Collins (1983:4) notes this was "one of the earliest systematic attempts at subgrouping in the (AN) family". Stresemann's (1927) phonological and morphological claims of innovations for Ur-Ambon are summarised in Collins (1983:12) and evaluated in C. Grimes (1991b:498– 502). Several claims can now be discounted, but they are impressive nevertheless.[3]

Capell (1943–44/13:194) supported the notion of a closer connection between the languages of Timor with the languages of Melanesia to the east than with the languages to the west.

> These [Timor] languages belong to the 'eastern' section of the Indonesian group which is practically Melanesian in structure. It will be remembered that it was from eastern Indonesia that Friederici drew the Melanesians, though his final theory has not justified itself to students of linguistics.

Later on in the study Capell is not so sure.

> The languages of Timor are usually grouped as Indonesian. They depart, however, in many respects from the Indonesian type represented by Battak [sic], Javanese, and the types found in Borneo and Celebes, so much so that some authors have reckoned the beginning of Melanesia from this point. There is much to be said for this, but it will not be assumed here. (1943–44/14:311)

> Some writers accept a more westerly line of demarcation between Indonesian and Melanesian languages. By Brandstetter and Dempwolff the line was placed along the western side of New Guinea; Friederici argued for one farther west, running between

---

[3] For a glimpse of Stresemann's results for lexical reconstruction, see Wurm & Wilson (1975) PAMS and PANS.

Sumba and Timor, across Flores, and between Celebes and the Moluccas. This places Timor, then, in the Melanesian group. This seems the better course for the present writer also, but it does not affect the statement that Timor and certain of its ethnological dependencies still form a regional group or province within that area, be it labelled one or the other. (1944–45/15:20)

Capell was possibly the first to suggest that there might be a distinct *subgroup* whose scope included Timor and surrounding languages to the east and west, and those of central Maluku.

there are 'regional' linguistic types — a fact for which Dempwolff does not seem to the present writer to have made sufficient allowance. It will also be shown that the languages of Timor form one such regional subgroup, distinct not only in its peculiar phonetic laws, but also in certain common grammatical usages. This group includes not only the languages of Timor itself, but also those of Roti, Wetar, Leti and Kisar, and less intimately connected, those of Flores (Sikka) and Solor, and in the conjugation of the verb also showing kinship with Seran. The treatment of OIN [PMP] final consonants is the most outstanding characteristic of the western section of Timor group, including Roti. Curiously enough, Savu is very different. (1944–45/15:19)

Dyen (1965), Haudricourt (1965), and Dahl (1976) seem to include languages of Maluku (eastern Indonesia) with the languages of the west rather than with those of Melanesia.

Chlenov and Sirk (1973) argued for the merger of labials *p and *b for 'Ambonese' languages [a merger not supported by current data for Buru, Sula, Ambelau, or many of the Seram languages, nor for the languages of southern Maluku]. Chlenov (1969, 1976) are lexicostatistical reassessments of the area covered by Stresemann (1927).

Chlenov (1978, 1980) looking at 'cultural' vocabulary divides AN languages of Maluku into 1) South Halmahera, 2) Aru, and 3) Southern Maluku (i.e. Central and South, excluding Aru), but the basis for his divisions is elusive.

Dyen (1978a) looked at data from Kei and Kamarian in a wider discussion of languages in eastern Indonesia. Dyen (1978b) looks at implications of vowels of central Maluku languages for PAn subgrouping.

CMP (with that label) as a subgroup within Austronesian is based on Blust (1974, 1978, 1979, 1981, 1982). It is roughly equivalent to Esser's (1938) Ambon-Timor Group (followed by Salzner 1960). Dyen (1965) also proposed a 'Moluccan Linkage' (Sumba, Sawu, Sikkic (including Endeh), Fordatic, (including Jamden) Letic, Buru, Ambic, Bonfia, Soboyo, Kuiwai) on the basis of his lexicostatistic calculations encompassing approximately the same scope as CMP. Blust clearly groups CMP languages with the east under Central-Eastern Malayo-Polynesian (CEMP).

As late as 1982 Blust (Note 13) says, "The proposed Central Malayo-Polynesian subgroup has thus far been *assumed* without qualitative demonstration. Dyen (1978a) makes a similar assumption, presumably on the basis of his 1965 lexicostatistical results". [Emphasis mine CEG]. The basis of Blust (1982) is that the 'CMP' languages share reflexes of two lexical innovations (i.e. CEMP *kandoRa* 'cuscus' and *mansar* 'bandicoot') that are not found in the WMP languages.

Although Stresemann, Collins, Chlenov and others have looked at subgrouping within areas of Maluku (with varying degrees of credibility) nobody has looked at the over-all relationships of the languages of Nusa Tenggara and Timor with the languages of Maluku in a principled way. Thus, we have had no credible bird's-eye view of first-order divisions within CMP that withstands the scrutiny of additional data.

Blust's 1990 paper 'Central and Central-Eastern Malayo-Polynesian' brings the notion of CMP out of the realm of educated impressions and into the realm of scientific discussion, and is a substantial contribution to our understanding of the AN languages of eastern Indonesia. But in spite of impressions that would lead one to assume that CMP is a 'branch' of AN whose daughter languages all reflect the characteristics of a parent PCMP, a careful reading of Blust's paper supports only the weaker claim that there are innovations found in the region that are not found elsewhere.

Grimes (1991a), Pawley and Ross (1993) and Ross (1995), all conclude that the notion of CMP as a traditional subgroup within a family tree model should not yet be taken at face value. Thus, there remain significant questions yet unanswered regarding the macro- and micro- classification of the languages described in this volume. It is hoped that the data presented here will contribute positively to the discussion, and to a credible refinement of the linguistic picture of eastern Indonesia.

References

Anceaux, J.C., 1978, The linguistic position of south-east Sulawesi: a preliminary outline. In Wurm and Carrington eds, 1978, *Second international conference on Austronesian linguistics: proceedings, fascicle 1 Western Austronesian,* 275–283. Canberra: Pacific Linguistics.

Berg, René van den, 1991, Muna dialects and Munic languages: towards a reconstruction. In Harlow, ed., 1991, *VICAL 2, Western Austronesian and contact languages, Papers from the fifth international conference on Austronesian linguistics,* 21–51. Linguistic Society of New Zealand, University of Auckland: Auckland.

Blust, Robert A., 1974, Eastern Austronesian: a note. *University of Hawaii Working Papers in Linguistics* 6(4):101–107.

1978, Eastern Malayo-Polynesian: a subgrouping argument. In Blust, ed., 1978, *Proceedings of the Second International Conference on Austronesian Linguistics, fascicle 1,* 181–234. Canberra: Pacific Linguistics.

1979, Proto-Western Malayo-Polynesian vocatives. *Bijdragen tot de Taal-, Land- en Volkenkunde* 135:205–251.

1981, The Soboyo reflexes of PAN *S. In Blust, ed., 1981, *Historical linguistics in Indonesia,* Part I. *NUSA* 10:31–45. Jakarta: Badan Penyelenggara Seri *NUSA.*

1982, The linguistic value of the Wallace line. *Bijdragen tot de Taal-, Land- en Volkenkunde* 138(2/3):231–250.

1990, Central and Central-Eastern Malayo-Polynesian. Paper presented at Maluku Research Conference, Honolulu.

Brandes, J.L.A., 1884, *Bijdrage tot de vergelijkende klankeer der westerse afdeeling van de Maleisch-Polynesische taalfamilie.* Utrecht: Proefschrift.

Brandstetter, Renward, 1906, *Ein Prodromus zu einem vergleichenden Worterbuch der malaio-polynesischen Sprachen für Sprachfroscher und Ethnographen.* Luzern.

Capell, Arthur, 1943–44; 1944–45, Peoples and languages of Timor, parts one and two. *Oceania* 14/3:191–219; 14/4:311–337; 15/1:19–48.

1976, General features of Austronesian languages, New Guinea area. In Wurm, ed., 1976, *New Guinea area languages and language study, vol. 2, Austronesian languages*, 5–52. Canberra: Pacific Linguistics.

Chlenov, M.A., 1969, Ocherki po etnicheskoi istorii naradov tsentral'nykh Molukk (Indoneziia). Moscow, dissertation.

1976, *Naselenie Molukkskikh ostrovov*. Moscow: Nauka.

1978, Preliminary classification of the Austronesian languages in the Moluccas. Abstract for the Second International Conference on Austronesian Linguistics, Canberra. (paper not delivered).

1980, Cultural vocabulary as an indicator of interethnic relations: eastern Indonesian evidence. *Bijdragen tot de Taal-, Land- en Volkenkunde* 136/4:426–439. 1969. Ocherki po etnicheskoi istorii naradov tsentral'nykh Molukk (Indoneziia). Moscow, dissertation.

Chlenov, M.A. and Yu Sirk, 1973, Merger of labial phonemes in Ambonese languages. Tartu Riikliku Yulikooli Toimetised. Tyoid Orientalistika Alalt: Acta et Commentationes Universitatis Tartuensis. *Oriental Studies* II/2:58–97.

Clark, Ross, 1987, Austronesian languages. In Comrie, ed., 1987, *The world's major languages*, 901–912. London: Croom Helm.

Collins, James T., 1982, Linguistic research in Maluku: a report of recent field work. *Oceanic Linguistics* 21:73–146.

1983, *The historical relationships of the languages of Central Maluku, Indonesia*. Canberra: Pacific Linguistics.

Collins, James T. and C.L. Voorhoeve, compilers, 1983, Moluccas. In Wurm and Hattori, eds, *Language atlas of the Pacific area* (map No. 45, Part II). Canberra: Australian Academy of the Humanities in collaboration with the Japan Academy.

Cowan, H.K.J., 1951–52, Genetief constructie en Melanesische talen. *Indonesië* 5:307–313.

Coward, David F., 1990, An introduction to the grammar of Selaru. MA thesis, University of Texas at Arlington: Arlington, Texas.

Dahl, Otto Christian, 1976, *Proto-Austronesian,* second revised edition. Scandinavian Institute of Asian Studies Monograph Series 15, Oslo.

Drabbe, P., 1926, Spraakkunst der Fordaatsche taal. *Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen*, Deel 67/1. Batavia: Albrecht and Co.

1932a, Beknopte spraakkunst en korte woordenlijst der Slaroeesche taal. *Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen*, Deel 71/2:1–32.

1932b, Woordenboek der Fordaatsche taal. *Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen*, Deel 71/3:1–118.

Dyen, Isidore, 1965, *A lexicostatistical classification of the Austronesian languages*. IJAL Memoir 19. Baltimore.

1978a, The position of the languages of eastern Indonesia. In Blust, ed., 1978, *Proceedings of the Second International Conference on Austronesian Linguistics, fascicle 1*, 235–254. Canberra: Pacific Linguistics.

1978b, Proto-Ambonic evidence for additional Proto-Austronesian vowels. In Udin, ed., *Essays presented to Sultan Takdir Alisjahbana on his seventieth birthday*, 390–399. Jakarta: Dian Rakyat.

Esser, S.J., 1938, Talen. In *Atlas van Tropisch Nederland,* 9–9b. Amsterdam: Koninklijk Nederlansch Aardrijkskundig Genootschap.

Friederici, G., 1913, *Wissenschaftliche Ergebnisse einer amtlichen Forschungsreise nach dem Bismark-Archipel im Jahre 1908, III: Untersuchungen über eine melanesische Wanderstrasse.* Mitt. aus den deutchen Schutzgebieten. Erganzungsheft no. 7. Berlin: Mittler.

Grimes, Barbara F., Joseph E. Grimes, Malcolm D. Ross, Charles E. Grimes, Darrell T. Tryon, 1995, Listing of Austronesian languages. In Tryon, ed., *Comparative Austronesian dictionary,* 5 Vols., 121–279. Berlin: Mouton de Gruyter.

Grimes, Charles E., 1991a, Central Malayo-Polynesian: an interim evaluation. Paper presented at the Sixth International Conference on Austronesian Linguistics, Honolulu, May 1991.

1991b, The Buru language of eastern Indonesia. PhD dissertation, Australian National University, Canberra.

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

Hoëvell, G.W.W.C. Baron van, 1877, Iets over de vijf voornaamste dialecten der Ambonsche landtaal (bahasa tanah). *Bijdragen tot de Taal-, Land- en Volkenkunde* 24:1–136.

Jonker, J.C.G., 1908, *Rottineesch-Hollandsch woordenboek.* Leiden: E.J. Brill.

1913, Bijdrage tot de kennis der Rottineesche tongvallen. *Bijdragen tot de Taal-, Land- en Volkenkunde* 68:521–622.

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

Kern, H., 1906, *Taalvergelijkende verhandeling over het Aneityumsch met een aanhangsel over het klankstelsel van het Eromanga.* Amsterdam: J. Miller.

Pawley, Andrew K., 1973, Some problems in Proto-Oceanic grammar. *Oceanic Linguistics* 12/1–2:103–188.

Pawley, Andrew and Malcolm D. Ross, 1993, Austronesian historical linguistics and culture history. *Annual Review of Anthropology* 22:425–459.

Riedel, J.G.F., 1886, *De sluik- en kroesharige rassen tusschen Selebes en Papua.* 's-Gravenhage: Martinus Nijhoff.

Ross, Malcolm D., 1995, Some current issues in Austronesian linguistics. In Tryon, ed., *Comparative Austronesian dictionary,* 5 Vols., 45–120. Berlin: Mouton de Gruyter.

Salzner, Richard, 1960, *Sprachenatlas des Indopazifischen Raumes.* Wiesbaden: Otto Harrassowitz.

Stresemann, Erwin, 1918, *Die Paulohisprache: Ein beitrag zur Kenntnis der Amboinischen Sprachengruppe.* The Hague: Martinus Nijhoff.

1927, *Die Lauterscheinungen in den Amboinischen Sprachen.* Zeitschrift für Eingeborenen-Sprachen, Supplement 10. Berlin.

Wurm, Stephen and B. Wilson, 1975, *English finderlist of reconstructions in Austronesian languages (post-Brandstetter).* Canberra: Pacific Linguistics.

2 A phonological sketch of the Selaru language

DAVID COWARD AND NAOMI COWARD

1 Introduction

The Selaru language is spoken on the island of Selaru, the southernmost island in the Tanimbar archipelago in the province of Maluku, Indonesia. Selaru is an Austronesian language located within the area encompassing what are classified as Central Malayo-Polynesian languages (see Blust 1978). The most closely related language is Seluwasan at 56 per cent lexical similarity (Hughes 1987).

The language is spoken by approximately 6,000 people in six of the seven villages on Selaru island (Kandar, Lingat, Namtabung, Eliasa, Werain, and Fursuy-see map), as well as in the village of Latdalam on the island of Yamdena where at least half of the population is comprised of Selaru speakers. There are four other villages of Selaru speakers who have migrated to other islands: Matakus, a small island northeast of Selaru; Lingada, located on the island of Nus Wotar off the west coast of Yamdena in northern Tanimbar; Tenaman and Mitak, on small islands farther north of Nus Wotar. There are also sizeable communities of Selaru speakers in Saumlaki, the district capital, and Ambon, the provincial capital. The total number of Selaru speakers is approximately 8,000.

Dialectal variation within the Selaru language is minimal for those speakers born or raised on the island of Selaru itself. Some variation in intonation patterns exist, with the Selaru spoken in the village of Kandar considered the most 'pleasant sounding' by all those Selaru speakers queried on the subject. There are some minor phonetic differences between [k]-[g] and [f]-[h] between villages.[1] No research has been done with those speakers who have migrated to islands other than Selaru, Yamdena, and Ambon.

---

[1] There may be some sociological significance for the variations encountered; perhaps these are due to a sound change which has not diffused uniformly throughout the community. Drabbe (1932) reports that one of his language helpers viewed the use of /f/ as rough or common language use while /h/ was more refined. In our work we have heard opposing arguments to this by various Selaru speakers, but it seems to us that /f/ is more frequently used by those speakers with less formal education and/or less outside contact. The [f]~[h] alternation is found where there was a historical *b.
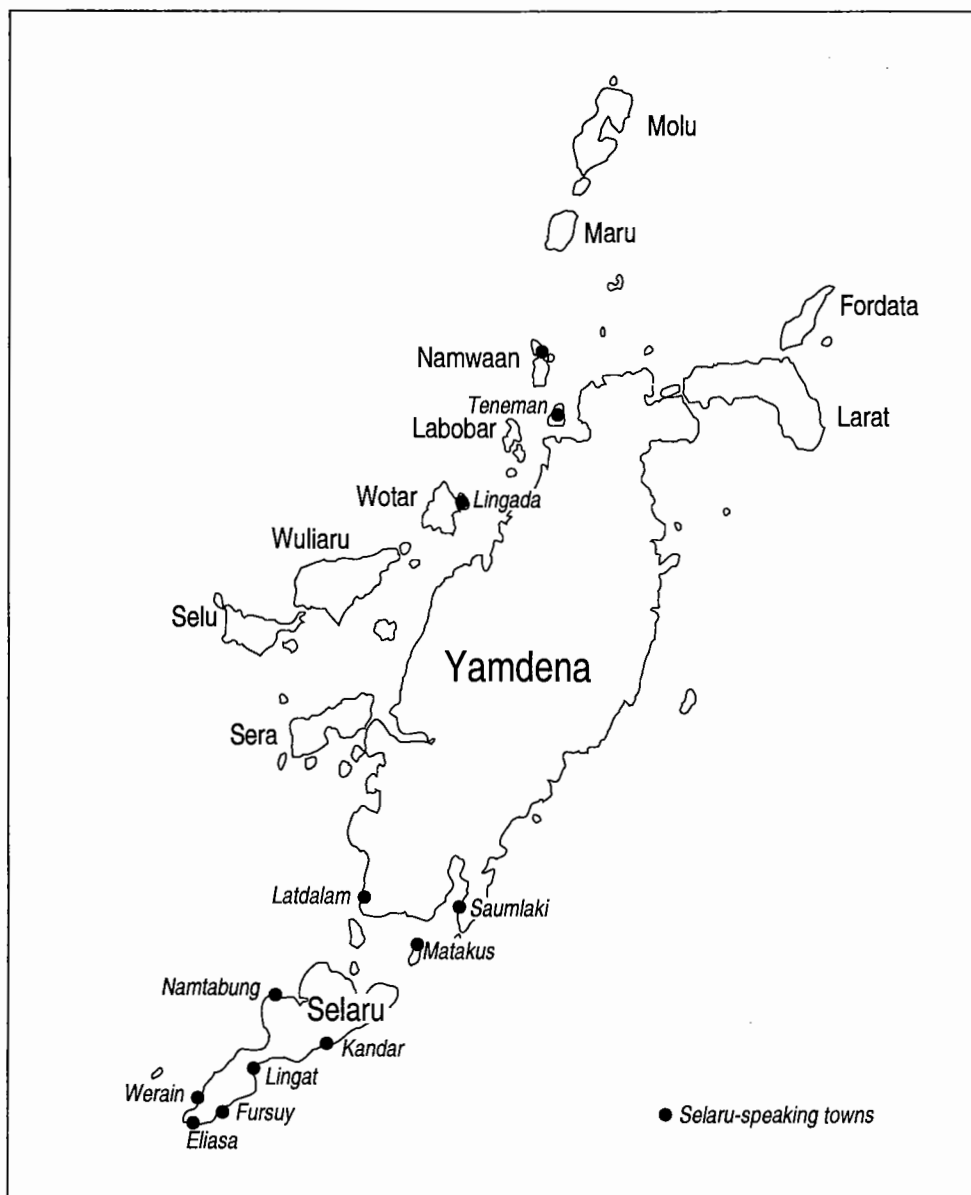
Selaru-speaking towns in the Tanimbar Islands

This research was conducted under a cooperative agreement between the Summer Institute of Linguistics and Pattimura University of Ambon, Maluku. We have spent a total of twelve months resident in Selaru over a period of four years. The majority of our time was in the village of Namtabung, population 1744 (Biro Pusat Statistik, Kantor Statistik Propinsi Maluku 1985). We have visited the other Selaru-speaking villages on the island and also Latdalam. Examples for this paper are taken from a corpus of data including over 60 texts, a beginning lexicon of 2000+ entries and our fieldnotes.[2]

---

[2] Principal language helpers were B. Loblobly (60-year-old male, village head), O. Nureroan (70-year-old male), M. Hidungoran (50-year-old male, *soa* head), B. Sambonu/K (38-year-old female, elementary

Previous linguistic research done on the Selaru language is limited. The Catholic missionary P. Drabbe (1932) includes a grammar sketch and word list, and was the most extensive work (32 pages) published on the Selaru language prior to N. Coward (1989) and D. Coward (1990). Brief mention of Selaru is also included in several comparative studies (Mills and Grima 1980, Mills 1981, 1991, Blust 1980, 1983-84a,b, 1986, Collins 1982, Hughes 1987). We believe all these, but Hughes (1987) and Mills (1991), was based on the Drabbe word list. Selaru is also listed in various language atlases (Salzner 1960, Wurm and Hattori 1981, B. F. Grimes 1992). Simonne Pauwels, a Belgian anthropologist, stayed in the village of Fursuy during 1985 and 1986 and has published several works on Selaru culture (including 1985, 1990a, 1990b).

The present work[3] is a synchronic description of the basic phonological patterns of Selaru. Its aim is to provide a basis for understanding how Selaru works, a foundation for developing the orthography and, perhaps, a basis for comparative studies with other languages in the region. This study is descriptive, the approach eclectic.

2 Inventory of phonemes

2.1 Description of contrastive features

The phonemic system is composed of three classes: consonants, vowels, and glides (this paper demonstrates why considering glides to be a distinct class is warranted). The suprasegmentals of stress, length, labialisation, palatalisation, and nasalisation are important features of the language, though not phonemically distinctive. The Selaru phonemic inventory consists of 12 consonants, five vowels, and two glides. In addition to the phonemes charted in Tables 1 and 2, Selaru speakers also use /p g ŋ ǰ č/ in words borrowed from Malay, but these are not considered part of the inherited Selaru phonemic inventory.

Table 1: Consonantal and glide system

	Labial	Apical	Laminal	Dorsal	Glottal
Stop Voiceless		t		k	ʔ
Voiced	b	d			
Fricative	f	s			h
Nasal	m	n			
Lateral		l			
Trill		r			
Glide	w		y		

Table 2: Vowel system

	Front	Central	Back
Close	i		u
Mid	e		o
Open		a	

Table 1 shows the abundance of apical articulations in Selaru while front and back articulations are fairly sparse. The following display gives the binary features of these phonemes and will be useful later in describing phonological processes. The consonantal and glide systems may be described using ten distinctive features and the vowel system may be analysed in terms of only three features (height, frontness, and roundness).

Table 3: Feature chart

	b	t	d	k	ʔ	f	h	s
Sonorant	−	−	−	−	−	−	−	−
Continuant	−	−	-	−	−	+	+	+
Consonantal	+	+	+	+	+	+	+	+
Coronal	−	+	+	−	−	−	−	+
Voice		−	+					
Back	−			+		−	+	

	m	n	l	r	w	y
Sonorant	+	+	+	+	+	+
Consonantal	+	+	+	+		
Nasal	+	+	−	−		
Lateral			+	−		
Coronal	−	+				
Round					+	−
High					+	+

	i	e	a	o	u
High	+	−	−	−	+
Front	+	+	−	−	−
Round	−	−	−	+	+

3 Description of phonemes

This section will give a brief description of each phoneme in the Selaru inventory, with examples. To prevent the following examples from causing more confusion than insight, the characteristics of the glides /y/ and /w/ are introduced here briefly. These glides appear to be historically /i/ and /u/ that lost their 'vowelness' but have not yet gained full consonant stability; i.e., while phonologically they often act like 'typical' glides, more often they spread to other consonants and even cross word boundaries to do so; they occasionally even regain full vowel status and surface as [i] or [u].

(1) /kwe-ke/ → [gʷege]   'the banana'
     banana-the

     /asw-ke/ → [askʷe]   'the dog'   (/w/ appears to metathesise)
     dog-the

     /asw-Vre/ → [asure]   'the dogs'   (/w/ becomes [u])
     dog-PL[4]

A more complete description of these peculiar characteristics is addressed in §3.3.

3.1 Consonants

3.1.1 Description

3.1.1.1 Stops

Selaru voiced stops /b/ and /d/ devoice syllable-finally. This causes neutralisation of contrast between /d/ and /t/ in this position, but as /p/ is not a phoneme in Selaru, no neutralisation occurs with /b/.

The phoneme /b/ is a bilabial stop most commonly realised as [b]. It is also realised as [p] syllable-finally and sometimes (very rarely) as a voiced bilabial fricative [β] intervocalically:

(2) [bołbboł]   /bolbol/   'tomorrow'
     [ʔɛpno]   /ebno/   'ancestor'
     [kbat]   /kbat/   'lemon'
     [bʸop]   /y-bob/   'I lay down'
     [koban]   /kw-oban/   'I hit'

---

⁴ Abbreviations used in this paper

[ ]	phonetic	hi	high
/ /	phonemic	inam	inanimate
1pi	first person plural inclusive	incl	inclusive
1px	first person plural exclusive	k.o.	kind of
1s, 2s, 3s	first, second, and third person singular	LOC	locative
2p, 3p	second and third person plural	pl	plural
AmSubj	animate subject	POSS	possessive
G	glide in CCGVCG	Qmrkr	question marker
C	consonant	rd	round
GEN	genitive	V	vowel

[ksop]	/kw-sob/	'I pray'
[oβóban]	/ob-oban/	'hitting stick'

The phoneme /t/ is a voiceless apical-dental stop [t]:

(3)
[turi]	/turi/	'machete'
[mɛtmɛt]	/metmet/	'black'
[ʔɛnatɛɫʷ]	/enatelw/	'three'
[kʰʷɛtʰy] ~ [kʷɛtⁱ]	/kw-hety/	'I cut'[5]

The phoneme /d/ is a voiced apical-alveolar stop [d]. As mentioned above, syllable-final devoicing ensures that the phone [d] never occurs word-finally. This causes a loss of contrast between /t/ and /d/ in that position. Generally any phone [t] in this position is posited as /t/ (unless other morphological evidence supports /d/ as the underlying phoneme).

(4)
[dakun]	/dakun/	'also'
[ʔode]	/ode/	'and'
[kduduhan]	/kduduhan/	'smoke'

In conjunction with the glide /y/ in pre-nuclear position, the voiced apical stop can surface as the affricate [ǰ].

(5)
[metdʸet] ~ [metǰet]	/metdyet/	'dark/night'
[desdʸe] ~ [desǰe]	/desy-de/	'that-already'
[katdʸai] ~ [katǰai]	/katdyai/	'totem, heirloom'

The phoneme /k/ is an unaspirated dorsal stop [k]. It often has a voiced realisation [g] intervocalically or in conjunction with a glide in a syllable initial position.

(6)
[kuknam]	/kwknam/	'I eat something. (non-spec.)'
[agʷe]	/aw-ke/	'wood'
[krala]	/krala/	'inside'
[gʸeyer]	/y-keyer/	'he's angry'
[lililik]	/lililik/	'weaving tool'
[gʷeare]	/kwe-Vre/	'bananas'

The phoneme /ʔ/ is a glottal stop [ʔ]. Between two like vowels in normal rapid speech the glottal is lost, resulting in geminate vowels (which sound phonetically long). Where glottal occurs word-medially, it is only at the juncture of two morphemes.

(7)
[ʔenaʔat]	/ena-ʔat/	'four'
[ʔenaːt]	/ena-ʔat/	'four' (rapid speech)
[raʔora]	/r-ʔora/	'they sing'
[kuʔutw]	/kw-ʔutw/	'I spear'
[kuːtw]	/kw-ʔutw/	'I spear' (rapid speech)

The prevocalic glottal on vowel initial words is not phonemic. Also, there is no clear evidence for any words ending in a glottal. Positing /ʔ/ as part of the Selaru phonemic inventory is somewhat tenuous because of its extreme rarity (the examples above constitute

---

[5]   Stops occurring before phrase-final glides may gain a [+aspirated] release (especially if the following glide is /y/) with the additional [+hi][-rd] or [+hi][+rd] features of the glide.

most of the examples found in our data). These forms may be residual evidence for a once more pervasive phoneme.

3.1.1.2 Fricatives

The phoneme /f/ is realised as the labio-dental fricative [f] and the glottal fricative [h]. The phoneme /f/ does not occur intervocalically, but can be realised as [f] or [h] in all the environments in which it does occur. Where /f/ occurs word-finally, [f] is more common than [h]. As mentioned before, there may be social register/dialectal variations associated with this phoneme. In the more isolated Selaru villages, [f] is used more frequently than [h]. We point out here that while [h] is an allophone of /f/, the phoneme /h/ is also found in the inventory (which of course is also realised as [h]). The phoneme /f/ may be realised as a voiceless nasal [m̥] or [n̥] preceding a voiced nasal. Occasionally certain ideolects contain the voiceless bilabial fricative [ɸ] word-finally.

(8)	[ftun], [tʰun]	/ftun/	'star'
	[saf], [sah]	/saf/	'river'
	[fnu], [n̥nu]	/fnu/	'the village'
	[babaf], [babaɸ], [babah]	/babaf/	'short'

The phoneme /h/ is a glottal fricative [h]. It may (rarely) be realised as a velar fricative [x] before velar stops. The phoneme /h/ does not surface as [f] (although /f/ can freely surface as [h]).[6]

(9)	[huruk]	/huruk/	'also'
	[hihi]	/hihi/	'lips'
	[kahas]	/kahas/	'cotton'
	[klaxke]	/klah-ke/	'the mouse'

The phoneme /s/ is an apical-alveolar fricative [s]. It may have a palatalised realisation [ʃ] before /i/ or /y/ in rapid speech:

(10)	[saɬ]	/sal/	'road'
	[sit]	/sit/	'cat'
	[ksʷeak]	/kw-seak/	'I see'
	[ʔeras]	/eras/	'good'
	[desige]	/desy-ke/	'that'
	[ʃinaut]	/sinaut/	'something one looks for'

3.1.1.3 Nasals

As a generalisation, nasals, when clustered with a following stop, optionally assimilate to the point of articulation of that stop. This is not confined to within a morpheme, as this process occurs across morpheme boundaries.

---

[6] If an [h] is ever realised as [f] in any situation, we assume the underlying form to be /f/; however if it is always realised as [h], we posit the underlying form to be /h/.

The phoneme /m/ is a bilabial nasal [m]. It may also assimilate to the point of articulation of a following stop and be realised as [n] or [ŋ].

(11)	[mtubaɫ], [ntubaɫ]	/mtubal/	'dull'
	[manas]	/manas/	'hot'
	[mtu], [ntu]	/mtu/	'old'
	[sima]	/sim-a/	'hand-GEN(3s)'
	[mtasa], [ntasa]	/mtasa/	'cooked'
	[nam]	/nam/	'whatever'
	[ʔaniŋᵂa]	/ani-mw-ka/	'what's your name?'
	[simaŋᵂe]	/sim-a-mw-ke/	'your hand'
	[ŋɛs], [mkes]	/mkes/	'salty'

The phoneme /n/ is an apical-alveolar nasal [n]. It too may assimilate to the point of articulation of the following stop.[7]

(12)	[nisi]	/nisi/	'tooth'
	[kuŋgoɫ]	/kw-nkol/	'I am tired'
	[ʔana]	/ana/	'later'
	[ŋgora]	/nkora/	'hard'
	[lan]	/lan/	'big'
	[oboban]	/ob-oban/	'stick'

There is occasional neutralisation of contrast between /m/ and /n/ before dorsal and apical stops. Although identifying the root may be a problem in future generations, for now one can tell when the root contains an /m/ and when it contains an /n/ in these environments, because, for example, older speakers will still use [m] for /m/ in words such as *mtubal* and *mtu* (in (11) above).

3.1.1.4 *Lateral and trill*

The phoneme /l/ is a lateral [l]. It is realised as a 'dark' or velarised lateral [ɫ] before palatals and in the coda or syllable final position. There may be some optionality to this, but generally it holds.

(13)	[lɛma]	/lema/	'no'
	[tɛla]	/tela/	'neck'
	[moɫmoɫ]	/molmol/	'blue'
	[ɫʸalaᵂ]	/lyalaw/	'young man'
	[kʰʸaɫʸ]	/ky-haly/	'something makes noise'

The phoneme /r/ is an apical trill [r]. It is optionally realised as a flap [ɾ] intervocalically and a voiceless trill [r̥] when in conjunction with a glide phrase-finally.[8] There is also an optional [ə] on-glide word initially when /r/ is in a complex onset (i.e. [#rC...]).

---

[7]   The underlying phoneme for a surface realisation of [ŋ] is assumed to be /m/ if the word ever occurs with [m], or if the morpheme is known and from a closed class (such as /mw/ '2s'). For other occurrences of [ŋ] we posit /n/. We have no clear evidence of /n/ ever going to [m] before a labial.

[8]   The [+hi][-rd] or [+hi][+rd] features of the glide are actually mapped onto the /r/ producing a rounded or spread voiceless [r̥] (written here as [r̥ᵂ] or [r̥ʸ]).

(14)
[rabit]	/rabit/	'clothes'
[ʔur]	/ur/	'rice pot'
[ʔɛras]	/eras/	'good'
[ᵊrbade]	/r-ba-de/	'they've already left'
[mumdirʸ]	/mw-mdiry/	'you stand'
[nur]	/nur/	'coconut'
[mundi̥rʸ]	/mw-mdiry/	'you stand'
[lahlu̥rʷ]	/lahlurw/	'mung bean'

3.1.2 Contrasts

b/f/w

[ʔafri]	/afri/	'cousin'	[besa]	/besa/	'wave'
[ʔibrɛn]	/y-brɛn/	'he plays'	[wela]	/wela/	'eye ridge'
[seraf]	/seraf/	'staple food'	[banban]	/ban-ban/	'lukewarm'
[ʔisrap]	/y-srab/	'he prays'	[wan]	/wan/	'right (side)'
[sifkʸe]	/sify-ke/	'chicken'	[bailola]	/bailola/	'spider'
[yapkʸe]	/yaby-ke/	'handicraft'	[wai]	/wai/	'sibling, same sex'
[sawa]	/sawa/	'wife'	[ʔaʷ]	/aw/	'wood'
[koban]	/kw-oban/	'I hit'	[saf]	/saf/	'river'
[rsaʷ]	/r-saw/	'they stab'			

h/ʔ

[haha]	/haha/	'mouth'	[raʔora]	/r-ʔora/	'they sing'
[ʔɛnaʔat]	/ena-ʔat/	'four'	[rahora]	/r-hora/	'they stir'

ʔ/k

[ʔɛnaʔat]	/ena-ʔat/	'four'
[kakan]	/kakan/	'small'

t/d/r

[tais]	/tais/	'sarong'	[tola]	/tola/	'hanging'
[dai]	/dai/	'earlier'	[dolan]	/dolan/	'marriage wealth'
[kʰatu]	/khatu/	'seed'	[dɛndɛm]	/demdem/	'deep'
[ʔaduk]	/aduk/	'before'	[dɛmdɛm]	/demdem/	'deep'
[tɛnan]	/tenan/	'body'			
[ktura]	/ktura/	'short (in length)'	[tete]	/tete/	'weaving tool'
[kdusa]	/kdusa/	'short (in height)'	[dɛtke]	/det-ke/	'last night'
[sir]	/sir/	'they/them'	[rai]	/rai/	'indigestion'
[sit]	/sit/	'cat'	[dai]	/dai/	'earlier'
[krala]	/krala/	'inside'	[ʔaro]	/aro/	'boat'
[ktahi]	/ktahi/	'leaf'	[ʔaduk]	/aduk/	'before'

[ra]	/ra/	'they eat'	[rean]	/rean/	'high tide'
[ta]	/ta/	'or'	[desʸ]	/desy/	'that'

[rei]	/r-ei/	'they go to'
[tei]	/tei/	'the end'

l/r

[lai]	/lai/	'husband'	[mrin]	/mrin/	'cold'
[rai]	/rai/	'indigestion'	[mlin]	/mlin/	'calm (at sea)'

[kuŋgoł]	/kw-nkol/	'I'm tired'	[nur]	/nur/	'coconut'
[kor]	/kw-or/	'me with'	[nuł]	/nul/	'oil'

[msuł]	/mw-sul/	'you put on head'	[lea]	/lea/	'not yet'
[mbʷesur]	/mw-besur/	'you are full'	[rean]	/rean/	'high tide'

[kala]	/kw-ala/	'I do/make'	[lał]	/lal/	'housefly'
[lara]	/lara/	'blood'	[rał]	/r-al/	'they carry'

m/n

[manas]	/manas/	'hot'	[nam]	/nam/	'goods/things'
[lan]	/lan/	'big'	[naman]	/naman/	'child'

[ʔana]	/ana/	'later'	[sim]	/sim/	'hand'
[ʔama]	/ama/	'father'	[mrin]	/mrin/	'cold'

b/m

[banban]	/ban-ban/	'lukewarm'	[ʔisrap]	/y-srab/	'he prays'
[manas]	/manas/	'hot'	[nam]	/nam/	'whatever'

[koban]	/kw-oban/	'I hit'
[ʔuman]	/uman/	'fingernail'

w/y/h/Ø

[ksʷer]	/kw-ser/	'I turn'	[kutɛn]	/kw-ten/	'I ask'
[ksʷe̞rʸ]	/kw-sery/	'I cry'	[kutɛnʷ]	/kw-tenw/	'I weave'

[rat]	/rat/	'wild cat'	[tɛł]	/tel/	'language'
[ratⁱ]	/raty/	'cemetery'	[tɛlʷ]	/telw/	'three'

[aʷ]	/aw/	'wood'	[ʔo]	/o/	'animal pen'
[aʸ]	/ay/	'fire'	[ʔoʷ]	/ow/	'mango'

[sah]	/sah/	'river	[ʔitⁱ]	/ity/	'we/us (incl)'
[seʷ]	/sew/	'sun'	[ʔitʷ]	/itw/	'seven'
[seʸ]	/sey/	'house'			

[wałʷ]	/walw/	'eight'	[gʷemalaʸ]	/kwe-malay/	'papaya'
[yanʷ]	/y-anw/	'his condition'	[gʸeyer]	/y-keyer/	'he is angry'

3.2 Vowels

3.2.1 Description

All word-initial vowels are spoken with a weak non-phonemic glottal onset inserted at the word boundary. This only occurs in listing and utterance-initial conditions; in streams of speech this glottal onset is not present. As there is no contrast here, these glottals are not phonemic and so are not written phonemically.

The phoneme /i/ is the high front vowel. Its phonetic realisation is [i].

(15)   [ʔika]   /i-ka/   '(AmSubj) is where'
        [kisi]   /kisi/   'meat'
        [ʔir̥ʸ]   /iry/   'person'
        [tasi]   /tasi/   'sea/ocean'

The phoneme /e/ is the low front vowel normally realised as [e]. In closed syllables (and sometimes in open syllables) it is phonetically realised as its open counterpart [ɛ]. There is no contrast between [e] and [ɛ]. The phoneme /e/ is frequently realised at some point between [e] and [ɛ].

(16)   [ʔeras]   /eras/   'good'
        [ʔɛno]   /eno/   'mother (vocative)'
        [lea]   /lea/   'not yet'
        [mɛtmɛt]   /metmet/   'black'
        [gʷeare]   /kwe-Vre/   'bananas'
        [ʔɛte]   /ete/   'don't'

The phoneme /a/ is a low central unrounded vowel [a]. Its phonetic realisation sometimes laxes to become a mid-open vowel [ə] in unstressed syllables.

(17)   [ʔaramʸ]   /aramy/   'we (excl)'
        [ʔinəmɛt]   /y-namet/   'he steals'
        [ʔaduk]   /aduk/   'first'
        [mənasar]   /manasar/   'Makassar'
        [lema]   /lema/   'no'
        [kdusə]   /kdusa/   'short'

The phoneme /o/ is a mid back rounded vowel often realised as [o], in closed syllables (and sometimes in open syllables) it is phonetically realised as the lower variant [ɔ]. As with /e/ it is impossible to predict its exact pronunciation.

(18)   [ʔolas]   /olas/   'thread'
        [bɔkbɔk]   /bokbok/   'white'
        [ʔamo]   /amo/   'father (vocative)'
        [ʔɛnmɔ]   /enmo/   'just'
        [ʔode]   /ode/   'and'
        [tɔɬ]   /tol/   'well.bucket'
        [toto]   /toto/   'son'
        [mɔ]   /mo/   'if'

The phoneme /u/ is the high back rounded vowel. It is phonetically realised as [u].

(19)   [ʔus]   /us/   'rain
        [kusu]   /kw-su/   'I watch'
        [huɬ]   /hul/   'moon'
        [kmʷasu]   /kw-masu/   'I smoke'

One residue item is the variation between [i] and [ɛ] in [hina] and [hɛna], both meaning the possession word for food. The first has been reported by some to be an impolite or 'rough' variation of the second. But further investigation indicates external pressure is causing this 'free-variation.' Other speakers, who are more linguistically aware of their language, tell us that [hina] is in fact the proper form, but since this sounds like a Malay term meaning 'insult/humiliate', many nominal Selaru speakers (those living outside the language center) are shifting its pronunciation to the neutral form [hɛna].

3.2.2 *Contrasts*

<div align="center">i/e</div>

[seri]	/seri/	'side'	[ʔɛsʷ]	/esw/	'wind'
[sir]	/sir/	'they/them'	[ʔisi]	/isi/	'muscle (living)'
[kɛti]	/kw-eti/	'I tie'	[ʔeta]	/eta/	'lower back'
[ʔɛte]	/ete/	'don't'	[ʔita]	/ita/	'whoever'
[muhihi]	/mw-hihi/	'you sleep on your side'			
[muhehe]	/mw-hehe/	'you know'			

<div align="center">u/o</div>

[ʔowan]	/owan/	'cloud'	[soso]	/so-so/	'far'
[ʔuwam]	/uwam/	'fingernail'	[susu]	/susu/	'breast'
[tɔɬ]	/tol/	'water-bucket'	[ʔotʷ]	/otw/	'sea rocks'
[huɬ]	/hul/	'moon'	[ʔutʷ]	/utw/	'louse'
[soɬsʸoɬʸ]	/soly-soly/	'broom'			
[suɬsuɬʷ]	/sulw-sulw/	'hat'			

3.3 Glides

3.3.1 *Ambiguous segments*

The Selaru [-cons,+hi] segments /i u y w/ are closely related but distinct. Determining the underlying phoneme is not always obvious. Generally, if a [-cons,+hi] segment carries stress and/or fills the syllable peak position it is functioning as a vowel (i.e. [+syllabic]) and posited as /i/ or /u/ (unless other evidence forces reinterpretation). In the following, full stops mark syllable breaks.

(20)   [ʔi.ku]   /iku/   'tail'
        [ʔur]   /ur/   'pot'
        [ʔi.ə]   /ia/   'he/she'
        [ʔus]   /us/   'rain'

[ni.si]	/nisi/	'tooth'
[huɫ]	/hul/	'moon'
[ti]	/ti/	'at'
[kʰu.na]	/khuna/	'flower'
[mbi.nan]	/mbinan/	'dish'
[kdu.du.han]	/kduduhan/	'smoke'

In the prenuclear position of the syllable, [-cons,+hi] segments are interpreted as glides and function as consonants:

(21)	[ya.bik.ba.i.nare]	/yabikbai-Vre/	'weaving implements'
	[yaʷ]	/yaw/	'I/ me'
	[wa.i]	/wai/	'sibling of the same sex'
	[wer]	/wer/	'water'
	[ʔa.wa]	/awa/	'sibling of the same sex'
	[sa.wa]	/sawa/	'wife'
	[kkʷe.yer]	/kw-keyer/	'I'm angry'
	[ki.ta.yar]	/ky-tayar/	'something is lost'

In post-nuclear position /y/ and /w/ prove to be the most complex (as mentioned in the introduction to §3). Since the actual phonetic character of these segments in this position is not always clear to the outside analyst, a more detailed explanation of their phonetic production is in order. Briefly, Selaru glides are like /i/'s and /u/'s that have lost all sonorant quality, all syllabic quality, and are not specified as to voice (i.e., they adopt the voicing environment around them) but have retained their lip and tongue shape and [+hi] features. To produce a Selaru glide in isolation, one begins by pronouncing an [i] or an [u], but then stops just short of initiating phonation, i.e., the mouth and tongue are shaped just as for the vowels, but no sound or air is expelled. Producing such a segment in conjunction with a consonant is more complex. For [asʷ] 'dog', as one begins phonation of the [s] segment, one starts moving the lips and tongue into the position of an [u], but does not actually pronounce [u]. The [+hi,+rd] features of /w/ are achieved during the phonation of the [s] segment, basically adding these features to those of /s/ (a spreading process). Note that [asʷ] is a monosyllabic word; there is no extra puffing of air after the [s]. The /y/ is completely analogous to the /w/ in that the lips and tongue form an [i] with no syllabic quality, and its features [+hi,-rd] can spread to adjacent simple (C) consonants.

The [-cons,+hi] phonemes contrast in identical environments:

(22)	[ʔa.su]	/asu/	'cheek'
	[ta.si]	/tasi/	'ocean'
	[ʔasʷ]	/asw/	'dog'
	[tasʸ]	/tasy/	'rope'
	[lu.lu]	/lulu/	'garden village'
	[kɛ.ti]	/kw-eti/	'I tie'
	[luɫʷ]	/lulw/	'front'
	[kɛtⁱ]	/kw-ety/	'I carry'

[ku.tɛn]	/kw-ten/	'I ask'
[rat]	/rat/	'wild cat'
[ku.tɛnʷ]	/kw-tenw/	'I weave'
[ratⁱ]	/raty/	'cemetery'

An outside analyst, when encountering Selaru word-final glides for the first time, does not generally hear them following a consonant word-finally (as in (22) above) and therefore doubts their existence, but for the native speaker there is never any question. Yet it is not always necessary to rely on a 'phonetic ear' to detect their presence in this environment, nor to hear the difference between them and their high vowel counterparts following another vowel (as in /wai/ 'sibling' vs /ay/ 'fire'). An ambiguous segment is always interpreted as a glide if it ever transfers its binary feature bundle to the simple consonant (C) onset of a following root, clitic, or word.

(23)   [ta.si.ke]   [tas.kʸe]
        /tasi-ke/   /tasy-ke/
        'the ocean'   'the rope'

        [kʰa.tu.ke]   [hat.kʷe]
        /khatu-ke/   /hatw-ke/
        'the seed'   'the rock'

        [ru.gu.i.bo]   [rlu.bʸo]
        /r-ukui+bo/   /kw-luy+bo/
        'they only cut (the field)'   'they are only spinning'

        [sa.i.de]   [a.kye]
        /sai+de/   /ay-ke/
        'what?'   'the fire'

        [tas.ke]   [des.dʸe]~[des.ǰe] [9]
        /tas-ke/   /desy+de/
        'the bag'   'that already!'

While these segments should be interpreted as /w/ and /y/ in the post-nuclear, word-final position, they cannot be strictly interpreted as consonants. Their ability to transfer their binary features to other segments is not typical of consonants, and for purposes of penultimate stress placement, /y/ and /w/ count as the ultimate [+syllabic] segment. For example, 'papaya' /kwemalay/ surfaces as [gʷemaláʸ] and not [*gʷemálaʸ]. Yet the final segment is clearly seen to be /y/ (and not /i/) when 'papaya' is combined with the noun phrase article /-ke/:

(24)   /kwemalay-ke/   [gʷemalágʸe]   'the papaya'
        *[gʷemaláige]

---

[9]   /dy/ merges to become [ǰ] in casual speech.

3.3.2 Description

The phoneme /w/ is the bilabial glide [ʷ]. Its phonetic realisation word-finally generally has only the features [+hi,+rd]. If phrase-final following a voiceless stop it may be realised as a voiceless [+hi,+rd] aspirant [ᶣ].

(25)

[wahar]	/wahar/	'ripe'
[ʔasʷ]	/asw/	'dog'
[sawa]	/sawa/	'wife'
[waɬʷ]	/walw/	'eight'
[seʷ]	/sew/	'sun'
[ʔaʷ]	/aw/	'wood'
[ʔatʷ]	/atw/	'100'
[kbʸɛnʷ]	/ky-benw/	'full'
[kɔtᶣ]	/kotw/	'food'
[liakʷ]	/lia-kw/	'my friend'
[wammʷanʸ]	/wamw-many/	'man'

The phoneme /y/ is a laminal glide [ʸ]. Its phonetic realisation word-finally generally has only the features [+hi,-rd]. If phrase-final following a voiceless stop it may be realised as a [+hi,-rd] aspirant [ᶦ]. Also phrase-finally, following a glottal fricative [h] (from either /f/ or /h/), /y/ will combine with the fricative and be realised as [hᶦ] or [i̥].

(26)

[yaʷ]	/yaw/	'I/me'
[hahᶦ]	/hahy/	'pig'
[gʸeyer]	/y-keyer/	'he is mad'
[sihᶦ]	/sify/	'chicken'
[seʸ]	/sey/	'house'
[kʰʷɛtᶦ]	/kw-hety/	'I cut'
[ʔaʸ]	/ay/	'fire'
[makitᶦ]	/makity/	'sour'
[soɬʃʸoɬʸ]	/soly-soly/	'broom'
[moskʸa]	/mw-osy+ka/	'where are you coming from?'

In pre-nuclear position, the glide /y/ in conjunction with an apical stop can surface as the affricate [ǰ].

(27)

[metdʸet]~[metǰet]	/metdyet/	'dark, night'
[desdʸe]~[desǰe]	/desy-de/	'that-already'
[katdʸai]~[katǰai]	/katdyai/	'totem, heirloom'

These descriptions are helpful but do not address the crucial, distinctive characteristics of these glides; i.e. their ability to spread their features to other segments (even across word boundaries), and their acting like consonants (i.e. [-syllabic]) but functioning like vowels for stress placement.[10]

---

[10] The Selaru glides have not always been such. In fact, there is good reason to believe (through historical and comparative evidence) that these w's and y's came about by a historical loss of syllabification to morpheme-final high vowels. Consider the following cognate pairs from Selaru and PAn/PMP (this list is by no means exhaustive):

The CV phonology framework provides some theoretical advantages in addressing Selaru glides.[11] In this framework, the phoneme /u/ is typically depicted as:

(28) V   ←V indicates [-cons,+syllabic]
      |
      u   ←u indicates [+hi,+rd]

A typical (non-Selaru) /w/ could then be depicted as either:

(29) C   C   ←C indicates [+cons,-syllabic]
      | or |
      u   w   ←u/w indicates [+hi,+rd]

Since the characteristics of Selaru glides indicate that it is actually neither fully C nor fully V, we posit that a Selaru /w/ is simply a 'w' (i.e. [+hi,+rd]) in the segmental tier, with no specification (C or V) on the CV skeletal tier:

      Vowel   Glide
(30)   V   ←the CV skeletal tier
        |
        u   w   ←segmental feature bundles [+hi,+rd]

Using this formulation, the underlying form of [as^w] 'dog' is:

(31)   VC
        | |
        a s w
        'dog'

Such a formulation states that in the lexicon the morpheme /asw/ 'dog' has a segment (i.e. a set of features) not associated with anything on the CV skeletal tier (all other segments do). At the time of utterance, phonological rules link or associate this set of features with the CV tier:

(32)   VC
        | |\
        a s w
        'dog'

The formalism expressed in (32) captures the observed phonetic reality that the [+hi,+rd] features of 'w' are merged with /s/ making the surface manifestation of /s/ an [s] with lip rounding.

The /y/ phoneme is completely analogous to the /w/. It is a [+hi,-rd] feature bundle with no associations with any unit on the CV tier.

PAn/PMP	Selaru		
*babuy	*hahy*	'pig'	(*b>h, *uy>y)
*apuy	*ay*	'fire'	(*p>Ø, *uy>y)
*diRi	*-ndiry*	'stand'	(*i>y)
*asu	*asw*	'dog'	(*u>w)

[11] We are using the approach described in Clements and Keyser (1983).

As shown earlier, the high vowels and their corresponding glides are in absolutely contrastive environments. For example, [tas^y] 'rope' and [tasi] 'ocean' would be depicted underlyingly as:

(33)   CVC
        | | | (before phonological linking)
        t a s y
        [tas^y]
        'rope'

(34)   CVCV
        | | | |
        t a s i
        [tasi]
        'ocean'

The monosyllabic nature of /tasy/ 'rope' and the disyllabic nature of /tasi/ 'ocean' is clearly depicted in the CV tier (because even after linking, the CV tier for /tasy/ remains CVC).

3.3.3 The association rules

Selaru glides are underlyingly segments with no C or V specification at all on the CV tier. It is this lack of specification or association with any CV unit that makes these glides a distinct class from either consonants or vowels. All other segments are 'prelinked' to either C's or V's in the lexicon. At the time of utterance the unassociated segments (glides) link to the CV tier in predictable ways. The following rules attempt to delineate this linking process.[12] The rules are basically descriptive, but the underlying or driving force is the need for glides to link up to the CV tier under the constraints of allowable syllable structure (i.e. the CV structure on the CV skeletal tier). In streams of speech, these restrictions in syllable structure apply even across word boundaries.

---

[12] These rules, somewhat altered here, were first suggested by Duane Clouse in May of 1989 while looking over a preliminary draft of this paper. This paper constitutes a working out of his suggestions taking into consideration more data than he had available to him at the time. We are indebted to him for his help. Any inaccuracies though in the final implementation of the CV phonology framework or any of the claims of this analysis are of course our responsibility.

Table 4: Association rules for Selaru

1.	Any CCC string on skeletal tier is restructured as CVCC (CCC violates allowable CV syllable patterns).
2	Unassociated segments link to any adjacent unattached V slot.
(3.	Unassociated-w's delete if between a C and a V at a morphological juncture.)[13]
4.	Unassociated segments associate rightward to the adjacent C, if there is one, and it is a simple (C) onset.
5.	Unassociated segments associate leftward to the adjacent C, if there is one.
6.	Unassociated segments are given a C timing unit on the CV skeletal tier (unless to do so would violate allowable CV syllable patterns) and are then linked to it.
7.	Unassociated segments are given a V timing unit and linked to it.
8.	Unassociated V's surface as [a] (the unmarked Selaru vowel).

All of these rules are crucially ordered as numbered and constitute a sequence of IF THEN-ELSE statements. If an unassociated segment cannot obey Rule 2, it tries Rule 3 and so on. Rule 8 simply states that any V unit on the CV tier that has not been assigned any phonetic value by the end of the derivation will surface as the unmarked phone [a].

(35)   **Rule 5: Associate leftward**

$$\begin{array}{ccccc} \text{VC} & & \text{VC} & & \\ |\ | & \rightarrow & |\ |\ \backslash & \rightarrow & [\text{as}^\text{w}] \text{ 'dog'} \\ \text{a s w} & & \text{a s w} & & \end{array}$$

(36)   **Rule 4: Associate rightward**

$$\begin{array}{ccccc} \text{VC\ \ CV} & & \text{VC\ \ CV} & & \\ |\ |\ \ |\ | & \rightarrow & |\ |\ /\ |\ | & \rightarrow & [\text{ask}^\text{w}\text{e}] \text{ 'the dog'} \\ \text{a s w k e} & & \text{a s w k e} & & \end{array}$$
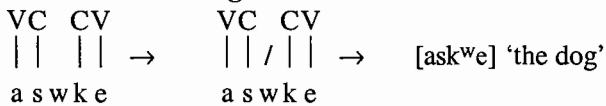
In example (35), the unassociated-w[14] cannot obey Rule 4 (associate rightward), because there is no adjacent C unit on the CV tier in that direction. It, therefore, follows Rule 5 and attaches to the adjacent C on the left. In example (36) there is an adjacent C to its right, and the /w/ attaches rightward (as per Rule 4).

Phonetically, a rightward attachment is written in the reverse order of the underlying segmental sequence (see the phonetic transcription for (36)). This, in a linear framework, would be called 'metathesis,' but CV phonology gives us a more accurate depiction of the phonetic reality. What actually happens in a rightward linking process is that the place of articulation features[15] of the /w/ segment ([+hi,+rd]), in (36), are overlaid on top of the /k/

---

[13] This rule is obligatorily present for verb onset structure (for almost all Selaru speakers), but is either restricted to that domain or more pervasive depending on whether or not this new phonological process is real for the individual speaker. See §5.1 for further discussion.

[14] The terms 'unassociated-w' and 'unassociated-y' are synonymous with /w/ and /y/. In most cases, the shorthand notation /w/ and /y/ is used, but the implication of this notation is that, for Selaru, glides consist of a set of features that are unassociated on the CV tier.

[15] Such a grouping of phonological features has been made possible by Clements (1985). To account for the natural grouping of particular features, Clements recommends a reorganising of the feature bundle into a tree structure, making direct reference to such subgroupings a natural result of the structure.

features yielding a rounded [k]. When this [k] with lip-rounding releases into the following vowel, it has a labialised release giving the impression that there is a [ʷ] segment dividing the [k] from the following vowel.

The following are examples of glides with no adjacent C units.

(37) **Rule 6: C-insertion and linking**

V V   VCV   VCV
| | → | | → | | | →   [awa] 'sister'
a w a   a w a   a w a

(38) **Rule 6: C-insertion and linking**

V   CVC   CVC
| → | → | | | →   [yaʷ] 'I/me'
y a w   y a w   y a w
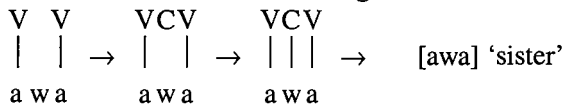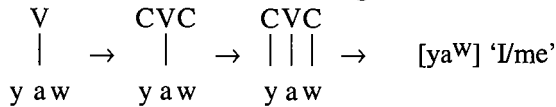
Rule 6 is a two step C-insertion rule which first inserts an empty C unit and then links the unassociated segment (features) to this unit. A C will not be inserted if to do so would violate allowable syllable structure. The resultant V.CV pattern for [awa] 'sister' and the CVC pattern for [yaw] 'I/me' are typical syllable patterns.

When /yaw/ 'I/me' comes in contact with a subsequent word, the final-w spreads its set of features across a word boundary.

(39) [ya.dʷa.kun]
      /yaw+dakun/
      'me+too'

      Underlying   **Surface**
      V +CVCVC   CV CVCVC
      | | | | | | → | | ⁄| | | | |
      y a w+ d a k u n   y a w d a k u n
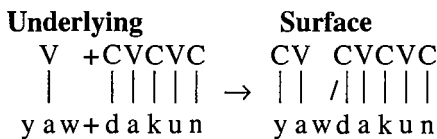
The unassociated-w in /yaw/ does not receive its own C timing unit in this case, but rather attaches to the first adjacent C to its right (hence Rule 4: Associate Rightward before Rule 6: C-insertion). This example not only validates that the final segment in /yaw/ is an unassociated-w (rather than the vowel /u/)[16] but also demonstrates that the association rules operate on a phrase level (and are not confined to only the morpheme level).

Rule 1 relates to CV structure directly, stating that a string of three consonants (CCC) will be restructured to CVCC. This only occurs at morpheme junctures and effectively breaks up the untenable consonant cluster[17] by inserting an empty V timing unit. Rules 2 and Rule 8 relate to empty V timing units on the CV tier. Rule 2 states that unlinked segments link to adjacent empty V units. Rule 8 effectively states that, if by the end of the derivation an empty V unit has not gained any segmental features, then it surfaces as the unmarked phone [a].

---

[16] This distinction is not always clear following a vowel in morpheme-final position.

[17] No unambiguous CCC pattern (i.e. one that does not involve a glide) exists in any Selaru morpheme.

(40) a. [lan.bo]   (CC)   b.   [la.na.ks^ya.lik]   (CVCC)   (V surfaces as [a])
         /lan+bo/   (CC)   /lan+ksyalik/ [18]   (CCC)
         big + just   big+very
         'just big'   'very big'

(41) a. [kun.kun.b^yo] (CC)   b.   [kun.ku.ni.ks^ya.lik]   (CVCC) (/y/ becomes [i])
         /kunkuny+bo/ (CC)   /kunkuny+ksyalik/   (CCC)   (no timing unit for y)
         yellow + just   yellow + very
         'just yellow'   'very yellow'

Example (40a) supports the interpretation that /lan/ 'big' is a simple CVC morpheme and /bo/ 'just' has a CV pattern. In normal speech, these abut and create an utterance with a CVCCV structure. This resultant structure is common, so no restructuring is needed. In example (40b), the modifier /ksyalik/ 'very' has the underlying CV pattern of CCVCVC (where the /y/ has no timing unit). The noun phrase /lan+ksyalik/ causes three C's to abut each other at a word boundary. A Selaru person would never say this phrase as *[lan.ksyalik] (unless there is a significant pause between the words, as in a listing rather than an utterance). There will always be an [a] inserted between these words. According to the association rules, this involves Rule 1 (change CCC to CVCC) and Rule 8 (an empty V surfaces as an [a]). This derivation involves Rule 8 simply because after the CCC string is broken up, there is no segment available to attach to the empty V timing unit.

Example (41a) gives evidence for the morpheme-final /y/ in /kunkuny/ 'yellow'[19] in that the unassociated features [+hi,-rd] from this segment are transferred across a word boundary to the /b/ segment in /bo/ 'just' (involving Rule 4: associated rightward to simple C onsets).

The noun phrase /kunkuny+ksyalik/ 'very yellow', in example (41b), creates a string of three C's (because the /y/ in /kunkuny/ 'yellow' has no CV timing unit). This is restructured to CVCC as per Rule 1.

(42) a. **Rule 1: Restructure CCC to CVCC**

        CVCCV C CC VCVC   CVCCVCV CC VCVC
        | | | | | | + | | | | | | → | | | | | | + | | | | | |
        k u n k u n y k s y a l i k   k u n k u n y   k s y a l i k
        'yellow'   'very'   'yellow'   'very'
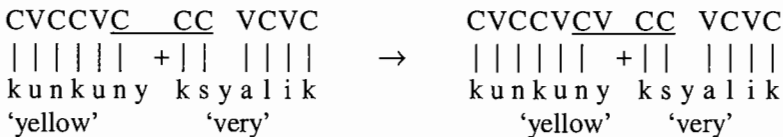
   b. **Rule 2: Link to empty V's**   c. **Rule 5: Associate leftward**

        CVCCVC V CC VCVC   CVCCVCV C C VCVC
        | | | | | | | + | | | | | | → | | | | | | | + | \ | | | |
        k u n k u n y k s y a l i k   k u n k u n y k s y a l i k
        'yellow'   'very'   'yellow'   'very'
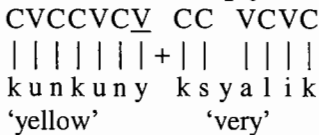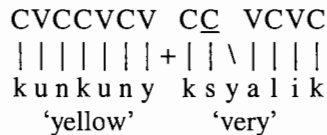
Following the restructuring of (42a), Rule 2 links unassociated segments to empty V timing units, which (as in (42b) causes them to gain the feature [+syllabic], making /w/ or

---

[18] There is some evidence that /ksyalik/ 'very' may actually have derived from /ky-salik/ 'inanimate-opposite side' idiomatically meaning 'it is unexpected', but today it is used as 'very' and neither interpretation changes the CV tier pattern (since /y/ has no timing unit in either case).

[19] The /y/ is nearly imperceptible when /kunkuny/ is said in isolation.

/y/ surface as [u] or [i]. Finally unassociated segments are linked leftward to adjacent C's (in (42c) as per Rule 5).

Rule 3, which deletes unassociated-w's between C and V timing units, and Rule 7, which gives a V timing unit to unassociated segments, will be discussed later under morphophonemics.

3.4 Distribution

3.4.1 Stress

In the phonological word, the stress is evidenced by increased intensity and raised pitch. Stress falls on the penultimate syllable of the root. Stress assignment in Selaru is lexical rather than post-lexical, so that stress is assigned before morphological and syntactic processes occur, i.e. before any affixes attach. After such processes have occurred, any suffix of two syllables in length will carry its own penultimate stress although it will be secondary to the root stress.

At first glance, the following minimal pairs could lead one to assume stress to be contrastive.

(43) a. [tála]   'we make'
    b. [talá]   'we run'

(44) a. [amána]   'octopus'
    b. [ámana]   'his father'

However, these are easily explained when the roots are extracted.

(45) a. [tála]   b. [talá]
        /t-ála/   /t-la´/ [20]
        'we(incl.)-make'   'we(incl.)-run'

(46) a. [ʔamána]   b. [ʔámana]
        /amána/   /áma-na/
        'octopus'   'father-his'

Their underlying form also confirms that primary stress must be a function of root structure and cannot be assigned after affixation.

Morpheme-final glides function as the ultimate syllable in determining penultimate stress placement on the root (even though they have no syllabic beat). Secondary stress can occur on suffixes which are long enough to support a penultimate syllable.

(47) [ʔówan]   'cloud'   [hɛttíłʸ]   'corn'
     [ʔówanke]   'the cloud'   [hɛttíłgʸe]   'the corn'
     [ʔówanàre]   'clouds'   [hɛttílìre]   'corn (pl)'

Generally, roots in Selaru are three syllables or less, but there are words of four syllables (and possibly more) that may or may not be a compound or reduplication. In longer roots,

---

[20] One of the few simple onset verb roots is /la/ 'run' which does not follow the general association rules. The 1pi-verb prefix /t-/ surfaces as [ta-] rather than the more common [t-]. Verb prefix structure is discussed later under morphophonemics.

stress occurs on the first syllable of the word and then again on the penultimate. The penultimate stress is usually the primary one with the secondary stress on the first syllable.

(48)   [kàbarbárak]   'round'
        [yàbikbái-nare]   'weaving implements'

Compound words likewise follow the generalisation of penultimate stress, but with penultimate stress placed on each of the roots. Which root gets the primary stress is mainly determined by intonation of the phrase.

(49) a. [ʔinóhakɔ̀tᵁ]   b.   [kwèmalágʸe]
        /y-nóha   +kótw/   /kwé   +maláy-ke/
        '3s-cooks +food'   'banana +Malay-the'
        'he/she cooks (generic)'   'the papaya'

     c. [sèikatʃáinàre]
        /séy-V[21]   +katdyái-Vre/
        'house-GEN   +totem-plural'
        'family heirlooms'

For nominalised verbs, penultimate stress always holds.

(50)   [kénah]   [ɛnénah]
        /kw-enah/   →   /en-enah/
        'I sleep'   'sleep'

(51)   [kóban]   [obóban]
        /kw-oban/   →   /ob-oban/
        'I hit'   'hitting stick'

Exact reduplication has identical repeated stress.[22]

(52) a. [wérwér]   'wet'   from   /wér/   'water'
     b. [bílabíla]   'weaving tool'   from   /bíla/   (root does not occur in isolation)

Each example in (52) is an exactly reduplicated morpheme, repeating the lexically assigned penultimate stress for each replication. There is no secondary stress in these morphemes (when uttered in isolation); each stressed segment is identical in intensity.

When exactly reduplicated morphemes occur in natural speech, their identical stress patterns are often modified in intensity by phrase or sentence-level intonation, as well as the pragmatics of the language situation (e.g. speaker emotion, etc.). These suprasegmental considerations determine their final phonetic realisation. Hence both [wèrwér] and [wérwèr] can be well-formed utterances:

(53)   [wérwèr yáʷ]   'I am wet'   (frustration)
        [wèrwér ò]   'you are wet'   (pity)

For a discussion of reduplication structure see §5.4.

---

[21] This 'V' indicates an empty-V inserted on the CV tier, and is a function of Selaru genitive possessive structure. As per the association rules, if a glide is adjacent to it, the glide links to it, becoming a full vowel; otherwise, the empty-V surfaces as an [a].

[22] 'Exact reduplication' refers to reduplication where the root morpheme is repeated in its entirety. This does not include reduplication such as /ob-oban/ 'hitting stick' which derives from /-oban/ 'hit'.

3.4.2 Syllables

Using a linear approach, there are 24 syllable patterns represented by the generic shape [CCGVCG]$_\sigma$ (with all C and G segments optionally present).[23]

Table 5: Selaru syllable patterns

CV Pattern	Example		CV Pattern	Example	
V	/o/	'you'	VC	/us/	'rain'
GV	/wa.i/	'sibling'[24]	GVC	/wan/	'person'
CV	/bo/	'garden'	CVC	/hul/	'moon'
CGV	/kwe/	'banana'	CGVC	/kwen/	'liquid'
CCV	/sra/	'above'	CCVC	/mrin/	'cold'
CCGV	/smwe.la/	'waist'	CCGVC	/slyet/	'bean'
VG	/ay/	'fire'	VCG	/atw/	'100'
GVG	/yaw/	'I/me'	GVCG	/walw/	'eight'
CVG	/taw/	'cup'	CVCG	/telw/	'three'
CGVG	/lyoy/	'spinning top'	CGVCG	/kwary/	'bed mat'
CCVG	/kray/	'sword'	CCVCG	/mlosy/	'outside'
CCGVG	/knyaw/	'k.o. wood'	CCGVCG	/klyehy/	'winnow basket'

The only restrictions on the distribution of these syllable patterns within the word are:

1. V(C)G may only occur morpheme- or word-finally

2. CCGV may only occur word-initially (within a morpheme such a pattern is a part of two separate syllables, as in: VC.CGV)

3. A CC(G) syllable onset pattern occurs only in word-initial and morpheme-initial syllables. A C(G) onset pattern may occur in non word-initial or non morpheme-initial syllables.

In the following examples [.] indicates syllable breaks, [-] morpheme breaks, and [+] clitic or word breaks.

(54) a. [yat.lʸo.sik]
/y-atlyosik/
'he cleanses'

   b. [ʔif.teʸ]
/i-ftey/
'she is pregnant'

   c. [ka.kan.bo]
/kakan+bo/
'small only'

   d. [ka.bar.ba.rak]
/kabar-barak/
'round'

   e. [wam.mʷan.i.re]
/wam-mwany-Vre/
sex-male-PL
'men'

   f. [wa.sim.dʷa.e.ra.ke.desʸ]
/wasi-mw+daera-ke+desy/
POSS-2sGEN+area -the+that
'that is your area'

Within morphemes, syllables generally divide: V/V, V/CV, VC/CV, and VC/CGV. But a few words, such as [kubrai] /kw-brai/ 'I do not want to' and [ʔibrɛn] /y-brɛn/ 'he plays',

---

[23] A 'G' denotes a glide in these syllable patterns.

[24] No monosyllabic words have been found to exemplify this pattern or the CCGV pattern, but this does not invalidate these as good examples of these syllable patterns.

divide the syllables as [ku.brai] and [ʔi.brɛn], rather than *[kub.rai] and *[ʔib.rɛn]. Evidence to support this variant syllable division is that the phoneme /b/ never surfaces as the phone [b] syllable-finally.

3.4.2.1 Glides

There are no syllable-final CC consonant clusters found in Selaru that do not involve a glide. To group glides with 'normal' consonants would create a new syllable pattern (CVCC) in a word such as /tasy/ 'rope'. Because of the unique nature of glides, it seems better to at least describe such syllable patterns as CVCG, with G representing the glide.

We have already described glides as actually having no C or V unit at all, but rather surfacing as a V or C, or transferring their features to an adjacent C only after following the association rules. If glides have no C or V of their own then the complete inventory of CV syllabic patterns for Selaru could be reduced to six.

Table 6: Simplified Selaru syllable patterns

CV	V	CVC	VC	CCV	CCVC

However, since this simplified list of syllable patterns does not tell us where the glides can be found, it must be appended: Glides can be found attached either to word-final C's or attached to the C just before a V. Some examples of glides in these positions include:[25]

(55) a. C V C   b. C C V C   c. C C V C
         | | | |
         G   G   G   G
     [tasy] 'rope'   [slyet] 'bean'   [kbyenw] 'full'

     d. V C   e. C V C   f. C V C
         | | | |
         G   G   G   G
     [asw] 'dog'   [kwen] 'place' [kwary] 'bed mat'

In all other cases, Selaru glides surface as full vowels or consonants, following the syllable patterns in Table 6. For example, /yaw/ 'I/me' when said in isolation surfaces as [yaw], which after the association rules has the syllable pattern of CVC, but when joined with /dakun/ 'also', the result is [yadwakun] 'me too', which has the following syllable pattern:

---

[25] The examples in (55) are displayed using a shorthand. The C-G link simply implies that the glide features are added to the underlying features of the original C segment. Compare (55b) to the display used most frequently in this paper:

       C C   V C   C C   V C
       | | | | → | |\ | | →   [slyet] 'bean'
       s l   y e t   s l   y e t

Either way, the syllabic structure is CCVC, and the word is monosyllabic.

(56)   **Underlying**   **Surface**
        V   CVCVC   CV . CV . CVC
        | | | | | | | | | | | | |
        | + | | | | | → | | G | | | |
        y a w   d a k u n | | d a   k u n   'me too'
                               y a   w

Determining whether the underlying phoneme is the vowel /i/ or /u/ or the glide /y/ or /w/ is not always clear. In ambiguous sequences composed of one or more [-cons,+hi] segments the ambivalent segment may be interpreted as the vowel /i/ or /u/ if the segment serves as the peak of a syllable. In all other cases, the [-cons,+hi] segments are interpreted as glides underlyingly. For example: it is not readily clear whether [áur] 'coral' is mono- or disyllablic, so [áur] 'coral' could be interpreted as /aur/ VVC, /aᵘr/ VC, or /awr/ VGC. The [u] segment is not the peak of the phonological word, so the possibility remains that it may be interpreted as something other than a vowel. However, the VC solution is rejected because it posits a new phoneme, the diphthong /aᵘ/. The VGC solution is also rejected because there is no evidence that the [u] segment ever spreads its features to the adjacent /r/ segment (as it would have to if [u] was underlyingly a glide). Therefore, the VVC solution is preferable in that it uses an accepted syllable pattern and does not introduce any new phonemes.[26]

In [ltúat] 'small basket for steaming rice', the [-cons,+hi,+rd] segment must be interpreted as a vowel because it carries the stress in the word. This solution also does not conflict with any syllable pattern or stress rules. If the stress had been different as in *[ltuát], then the word would be interpreted as */ltwat/, which is an acceptable Selaru syllable and stress pattern (though no such word exists).

The word [rahʸeta] 'sister/cousin' could be reinterpreted as either CV.CV.V.CV or CV.CGV.CV since the [-cons,+hi] segment is not the peak of the phonological word. Neither interpretation violates the CV patterns. If the former solution (CV.CV.V.CV) is accepted then /y/ would have to function as the peak of a syllable, but both phonetically and according to native speakers[27] this is not so, therefore we interpret [rahʸeta] as having the underlying form /rahyeta/. But for the same sort of vowel sequence [tlíake] 'his ear' the underlying form must be /tlia-ke/ 'ear-the' (the 3s genitive is null in this case), and not */tlya-ke/, since stress is carried on the [-con,+hi] segment.

3.4.3 Consonants

Limitations in single consonant occurrences: the consonant /d/ occurs only in the pre-nuclear margin of the syllable, i.e., in onsets; the phoneme /ʔ/ occurs morpheme-initially; while /f/ does not occur intervocalically.

---

[26] The word /aur/ 'coral' is also confirmed by native speakers as having two syllables: speakers were given a list of words and asked to tell the number of syllables in each. Some words contained ambiguous segments while many did not. The results were insightful but of course can not be held to as dogma.

[27] That is, speakers naturally syllabify /rahyeta/ with three syllables when asked to pronounce slowly.

There are more than 60 consonant cluster combinations evidenced in our data with more than half of these being ambiguous sequences (i.e. involving glides). Since there are enough unambiguous consonant clusters to establish a CCV syllable pattern, all consonant clusters can be interpreted as sequences with each segment filling one slot.

The possibility of interpreting the ambivalent sequences /mb/, /nd/, and /nt/ as prenasalised stops is rejected, since there are other unambiguous nasal plus stop sequences such as /md/ and /mk/, and it is more economical not to posit any new phonemes.[28]

The cluster [th] is also interpreted as a sequence mainly because it is in free variation with the unambiguous sequence /ft/. Underlying /f/ may be realised as [f] or [h] and when clustered word-initially with /t/ the [h] is realised as a release: /ftun/ → [ftun] ~ [tʰun] 'star'.

The combination [hn] is also interpreted as a sequence. It is often realised as a voiceless nasal plus [n] cluster, i.e. [n̥n]. Because there are no occurrences of [n̥n] or [hn] which could not be said as [fn], and /h/ cannot be realised as [f], then the underlying form for [n̥n] and [hn] must be the sequence /fn/.

All the consonant clusters involving /y/ or /w/ are also interpreted as sequences, with each phoneme filling one slot underlyingly. There is no evidence to interpret these clusters as underlyingly single-unit palatalised or labialised consonant phonemes. (Though, as we have seen, on surfacing, these segments do combine features to create blended phones.)

Distribution of consonant clusters is limited in several ways. The sequence limitations within the syllable (which may involve more than one morpheme) are as follows:

---

[28] This interpretation does not agree with Drabbe (1932) who posits /mb/ and /nd/ as separate letters (i.e. phonemes).

Table 7: Sequence limitations for consonants

> STOP + STOP: sequences include both geminate and non-geminate stops, but no voiced stops have been encountered.
>
> STOP + FRICATIVE: voiced stops do not occur, /ts/ only between morphemes.
>
> STOP + NASAL: voiced stops do not occur.
>
> STOP + LATERAL: /d/ does not occur.
>
> STOP + TRILL: /d/ does not occur.
>
> NASAL + STOP: no /nb/.[29]
>
> NASAL + FRICATIVE: /ms/ only, with /mf/ between morphemes.
>
> NASAL + NASAL: /mn/ only.
>
> NASAL + LATERAL: /n/ does not occur.
>
> NASAL + TRILL: /n/ does not occur.
>
> FRICATIVE + STOP: limited to /ft/, /sb/ and /sky/ only (*sk* by itself has not been found yet).
>
> FRICATIVE + NASAL: no /hm/ or /fm/.
>
> TRILL + CG: /r/ never heads a CCG sequence (but can occur within one, i.e. CrG, as in [kr^We.kan] /kw-rekan/ 'I think').
>
> LATERAL + STOP: no /ld/ sequence within the syllable (/lb/, /lt/, and /lk/ do occur).
>
> LATERAL + NASAL: no restrictions.

3.4.4 Vowels

Any vowel can occur in any syllable pattern in any position,[30] though there are restrictions on vowel sequences. (These are between syllables only since no VV occurs within the syllable.) No sequences of identical vowels can occur except in rapid speech where /ʔ/ is deleted. A high vowel may only be followed by the other high vowel or /a/. The phoneme /e/ has a limited distribution, such that it never occurs filling the second slot of a vowel sequence, and occurs in the first slot only with /i/ or /a/ filling the second.

Table 8: Sequence limitations for vowels

	_i	_e	_a	_o	_u
i_	—	—	/ia/	—	/iu/
e_	/ei/	—	/ea/	—	—
a_	/ai/	—	—	/ao/	/au/
o_	/oi/	—	/oa/	—	/ou/
u_	/ui/	—	/ua/	—	—

---

[29] Possibly /nb/ is absent since nasals tend to assimilate to the point of articulation of the following stop and we have found no clear evidence to reinterpret an /mb/ sequence as /nb/. But on considering the NASAL + LATERAL and NASAL + TRILL limitation of no /n/ in the sequence, one could infer that the nasal sequences are all /m/ initial and this /m/ needs further investigation.

[30] Drabbe (1932) states that /e/ and /o/ never occur word-finally. He claims that if a word appears to end in /e/ or /o/, another vowel actually follows it, though it is not heard. This may or may not be a reference to word-final glides, but either way, this interpretation is not tenable.

Co-occurrence restrictions with the glides also exist such that the high vowels seldom occur with their corresponding glide. There are examples that result from affixation: [msʷuɬ] /mw-sul/ 'you put something on your head' and [mbʸih] /my-bih/ 'you pl.-inspect', but they are rare. There is evidence that the glide of the verb subject-prefix is being dropped by younger speakers when it occurs adjacent to its full vowel segment.

All consonants can co-occur with any vowel, with the exception of /f/ which occurs word-medially and word-finally with all vowels except /e/ and /o/.

4 Suprasegmentals

4.1 Duration and length

Duration or lengthening of consonants or vowels is not phonemic in Selaru, but can occur for various non-phonemic reasons.

Geminate consonants occur in Selaru and are manifested by phonetic lengthening.

(57) a. [hettiɬgʸe]   /hettily-ke/   'the corn'
      b. [nukke]   /nuk-ke/   'the wing'
      c. [wammʷanʸ]   /wam-mwany/   'man/male'[31]

In rapid speech there is no phonetic (durational) difference between geminate and single consonants, but in slow, deliberate speech, the voiceless stops in (57a) and (57b) have a delayed release, and the /mm/ sequence in (57c) is slightly lengthened.[32]

Elision of /ʔ/ between like vowels results in geminate vowels. For example: /enaʔat/ can surface as either [ʔɛnaʔat] or [ʔɛnaːt]. The elision is a result of rapid speech, and the resulting geminate /a-a/ sounds like a lengthened [aː].

Vowels can be lengthened for pragmatic reasons (such as oratorical effect). Lengthening the stressed syllable of a word adds intensity to its basic meaning. (This is usually accompanied by a heavier stress on the vowel and an amplification of the basic phrase-level intonational pattern).

(58) [kuŋgóːɬ]   /kw-nkol/   'I am <u>tired</u>'
      [mdéːdan]   /mdedan/   '<u>heavy</u>'
      [máːɬ]   /mal/   '<u>thick</u>'

Duration is also a significant discourse feature and is used while 'talking *adat*.'[33] Examples include: 1) a new speaker interrupting a discussion to make a comment during

---

[31] The exact glosses for the morphemes /wam/ and /mwany/ in this compound are not clear, but probably mean 'sex' and 'man' respectively. /wam-mfwet/ means 'woman or female'. The words, /seri-mfwet/ lit. 'woman's side' and /seri-mwany/ lit. 'man's side', are marriage relationship terms.

[32] The fact that there are double m's in /wam-mwany/ was confirmed in a language game similar to pig Latin in which the syllables are broken up and a [da] is inserted. For /wam-mwany/ the game produced [wadam mʷadanʸ] with a clear break between the /m/'s. The noun /nuk/ 'wing' is pluralised as [nukare] rather than *[nuare] or *[nunare], verifying that the root ends in a /k/.

[33] '*Adat*' is an Arabic loan covering a wide range of traditional customs and behaviour--the morés and folkways of each society. Therefore, to 'talk *adat*' means discussing accepted behavior during a traditional get-together, for example, discussing brideprice during a marriage ceremony, or discussing the possible reasons for a child's death at his funeral.

traditional get-togethers; 2) 'formal' language registers such as reciting legends or traditional prayer; 3) frequently as a pause marker, much like the American English 'uh...'; and 4) extensively used by the town crier for public announcements. It most frequently occurs with the vowel /a/ (especially in 'adat' situations), but also occurs with the other vowels. Duration in these cases refers to a highly protracted phonation of the vowels. One example would be the town crier calling attention to the village: [a---------mt^{y}omolu---------haratke------] 'hey, you all listen to the pronouncement' (the lines denoting continuous phonation of the preceding vowel). Such use of duration continues throughout the announcement, not just at the beginning.

4.2 Nasalisation of vowels

Nasalisation of vowels, though not contrastive, is very prominent in Selaru. Phonetically, Selaru is more nasal sounding than languages in other areas of Maluku. The nasalisation of vowels seems to be viewed as the more traditional way to speak.[34] This comes into play most clearly in *adat* situations and when relating legends. *Adat* discussions are carried out in almost continuous, very pronounced nasalisation of all vowels. In legends, when one of the actors (an ancestor of theirs) is quoted, his speech is nasalised to the extreme. Traditional *adat* songs are also entirely nasalised.

4.3 Intonation and cadence

It is beyond the scope of this paper to include a thorough analysis of intonation, but a few features will be mentioned. As a general rule, falling pitch marks the end of the phonological sentence and rising pitch the end of the phonological phrase. Pauses do not necessarily come between syntactic sentences, but do often come between syntactic phrases.

Discourse-level connectors and time markers are marked by a rising-then-falling pitch. The peak of the pitch is on the stressed syllable, which can be lengthened. These connectors include [lemáde] 'so then' and [desi-gé-o] 'and then' in narratives, and in procedural discourse [maktéːi-o] or the longer [maktéi-bon^y-oː] (both mean 'after that's done then...'). These may be said to mark the beginning of the phonological paragraph.

Throughout a discourse there are also marked changes in speed and cadence. At the beginning, e.g. when introducing characters in a story, the narration is very slow and methodical, while near the climax of the discourse, the stream of speech becomes very rapid; where literally no phonological breaks are heard for several syntactic sentences. Discourse also has marked changes in intonational range. During long stretches of plain narration (no quotation or peak features) the intonation becomes a flat mid tone whereas in more active scenes the voice range becomes very extended, often rising and falling many levels in rapid succession.

---

[34] We observed an educated and fluent speaker of both Selaru and Indonesian switch back and forth between the two languages at a meeting. He started out speaking Indonesian in a clear unnasalised voice and then switched to Selaru in mid-sentence. His voice instantly switched to a more fronted and much more nasalised voice. He switched back and forth several times with the same effects.

5 Morphophonemics

5.1 Subject-prefixes on the verb

There are three sets of surface forms for the person-number subject-prefixes on verbs. These surface forms are conditioned by the prenuclear syllable structure of the verb to which they attach.

	1. Vowel Initial	2. Complex (CC) Onset	3. Simple (C) Onset
(59) a.	[koban]	[kuknam]	[kbʷa]
	/kw-oban/	/kw-knam/	/kw-ba/
	'I hit'	'I eat'	'I go'
b.	[ʸoban]	[iknam]	[bʸa]
	/y-oban/	/y-knam/	/y-ba/
	'he hit'	'he eats'	'he goes'
c.	[roban]	[raknam]	[rba]
	/r-oban/	/r-knam/	/r-ba/
	'they hit'	'they eat'	'they go'

Though there are exceptions,[35] the general rule is vowel-initial verbs with no prenuclear margin take Set 1 (where the labial glide /w/ is lost, while /y/ is unaffected), verbs with complex (CC) onsets take Set 2 (where the glides /w/ and /y/ become vocalic [+syllabic]), and verbs with simple consonant onsets take Set 3 (where the glides transfer their features to the simple C onset).

Table 9: Verb subject-prefixes

Verb root begins with	1s	2s	3s	inam	1pi	1px	2p	3p
SET 1 -V	k-	m-	y-	ky-	t-	aramy my-	my-	r-
SET 2 -CC	ku-	mu-	i-	ki-	ta-	aramy mi-	mi-	ra-
SET 3 -C	kCw-	mCw-	Cy-	kCy-	t-	aramy mCy-	mCy-	r-
Underlying structure								
CV tier	C	C		C	C	VCVC C	C	C
Segmental tier	kw	mw	y	ky	t	a r a my my	my	r
Shorthand form	kw-	mw-	y-	ky-	t-	aramy my-	my-	r-
Possible Historical source	*(a)ku	*(ka)mu	*i(a)		*(ki)ta	*(ka)mi	*(ki)mi *mi(u)	*(si)Da

---

[35] In our lexicon there are at least 20 counterexamples to Set 3 with about half monosyllabic and half disyllabic simple onset roots. These follow Set 2 instead of Set 3. There are at least 20 examples of monosyllabic simple onset roots which do follow Set 3. Other exceptions are few: two examples of Set 1 roots following the pattern of Set 2, one example of Set 3 roots following the Set 1 pattern, and one example of a Set 1 root following the Set 3 pattern.

Table 9 presents the person-number verbal subject prefixes 1) according to their three surface form sets, 2) their theoretical underlying form, 3) their 'shorthand' form, used extensively in this paper, and 4) their possible historical forms.

Concerning their underlying forms, note that each prefix (except 1pi and 3p) has an unassociated final segment /y/ or /w/, and that the third singular prefix is posited as merely a glide, i.e. an unassociated /y/ segment, by itself, with no CV tier specifications at all. As stated before, depending on the onset of the verb root, the prefix will act in one of the three ways. The motivation behind these three sets of surface forms lies in the unassociated segments of these prefixes, and not merely because a verb root begins with some particular CV pattern. As explained in §3.3.3, association rules predict how these glides will interact with other segments.

Verbs with simple onsets can be exemplified using the verb root /-tomolu/ 'to hear/listen' with the 1s, 3s, and 1pi subject prefixes.

(60) **Rule 4: Associate rightward**

```
C    CVCVCV          C  CVCVCV
|    ||||||    →     | / ||||||     →    [ktʷomolu]  'I hear'
kw - t omo l u        kw t omo l u
```

(61) **Rule 4: Associate rightward**

```
     CVCVCV              CVCVCV
     ||||||    →     / ||||||     →    [tʸomolu]  'he hears'
y -  t omo l u          y t omo l u
```

(62) **No rules apply**

```
C   CVCVCV
|   ||||||                        →    [tːomolu]  'we hear'
t - t omo l u
```
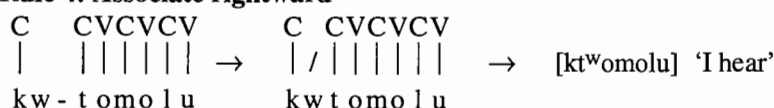
If a verb root has a complex onset, then the unassociated segment of the subject prefix becomes a vowel. This is explained by Rule 1: Restructure CCC to CVCC, and Rule 2: Link to adjacent empty V. In the following example the root for 'eat' is /-knam/.

(63)a. **Rule 1: Restructure CCC and Rule 2: Link to empty V**

```
C    CCVC       CVCCVC       CVCCVC
|    ||||   →   | ||||   →   ||||||     →    [kuknam]  'I eat'
kw - k n am     kwk n am      kwk n am
```

b. **Rule 1: Restructure CCC and Rule 2: Link to empty V**

```
C    CCVC       CVCCVC       CVCCVC
|    ||||   →   | ||||   →   ||||||     →    [miknam]  'you all eat'
my - k n am     myk n am      myk n am
```
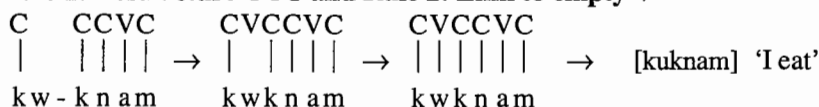
When /kw-/ '1s' in (63a) is added to the root /knam/ 'eat', the skeletal tier is filled with a sequence of three C's, i.e., the C unit, associated with the /k/ of /kw-/, 'bumps' up against the CCVC pattern of the verb root. Because a sequence of three C's is not allowed in Selaru, the CV pattern is restructure to CVCCVC (as per Rule 1). This inserted V unit is empty, i.e. is not associated with any phonemic segment. The unassociated-w in /kw-/ then links to this empty-V gaining full vowel quality (as per Rule 2). The process for [miknam] 'you all eat' in (63b) is completely analogous.

When this same verb root is conjugated with a subject prefix that does not have an unassociated segment morpheme-finally (e.g., 1pi) Rule 8: Empty V Surfaces as [a], also comes into play.

(64)   **Rule 1: Restructure CCC and Rule 8: Empty V Surfaces as [a]**

C CCVC   CVCCVC   CVCCVC
| | | | | → | | | | | → | | | | | | →   [taknam] 'we eat'
t - k n a m   t k n a m   t a k n a m

In this example, the CCC concatenation is restructured to CVCC, but the empty V has no unassociated segment to attach to and therefore surfaces as the unmarked phone [a] at the end of the derivation. This process of restructuring CCC units and the eventual surfacing of the empty V as the unmarked phone [a] is not restricted to verbal prefixes, but commonly occurs in general discourse (where two words abut, forming a CCC sequence, as shown in §3.3.3).[36]

When a complex CC onset verb root is conjugated with the 3s prefix, it involves a completely different process than discussed thus far. Since the 3s prefix is simply a /y/, and this phoneme is unspecified on the CV tier, the 3s prefix is completely by itself. It cannot cause any CCC restructuring (as in preceding examples). The first rule that the 3s prefix could possibly obey is Rule 6 (unassociated segments are given a C unit, unless to do so would violate allowable CV syllable patterns). But in this case, adding a C to the CV tier for /y-/ would create a CCC string, and this is not an acceptable pattern. So then Rule 7 must give the still unassociated /y-/ prefix a V timing unit and link the /y/ to it:

(65)   **Underlying   Rule 6: get C not allowed   Rule 7: get a V unit**

CCVC   *CCCVC   VCCVC
| | | | → | | | | → | | | | → [iknam] 'he/she eats'
y - k n a m   y k n a m   y k n a m

Vowel-initial verb roots present the first inconsistency in the rules, in that the unassociated-w's in the 1s and 2s prefixes drop out before vowel-initial verb roots, while the unassociated-y's in other prefixes do not.

(66) a. C   VCVC   C VCVC
      | | | | | → | | | | | →   [kasar] 'I grill'
      kw - a s a r   k a s a r

b. C   VCVC   C VCVC
   | | | | | → |\ | | | | →   [mʸasar] 'you all grill'
   my - a s a r   m y a s a r

It is not clear why the /w/ in [kasar], (66a), would drop out while the /y/ in [mʸasar] in (66b) does not, although there is some evidence that an unassociated-y is 'stronger' than an unassociated-w. A word-final /y/ seems to have more force or puff of air after voiceless

---

[36] It might appear more straight forward to posit the 1pi and the 3p prefixes as being /ta-/ and /ra-/ respectively (and from a historical perspective, more satisfying), but this creates the more serious problem of stating when and where the /a/ segments are deleted (as in [toban], not *[taoban], from /t-oban/ 'we hit'; and [tnauk], not *[tanauk], from /t-nauk/ 'we get down'), and misses the generalisation that the restructuring process occurs in several other grammatical processes not involving these subject prefixes.

stops than a /w/ (which is totally airless), and after other morphological processes, like reduplication, the reduplicated final-y is retained where the final-w is not: [nelnʸelʸ] 'clean' (from /nely-nely/), but [harharʷ] 'new' (from /harw-harw/).

This inconsistency in the verb paradigm set requires the /w/ deletion rule (association Rule 3), which basically states:

(67) **Rule 3:** Unassociated-w's delete in the following environment:

    C   V
    | |
    x w - x   where 'x' refers to any segment and the '-' is a morpheme break. [37]

This deletion rule must be ordered before the first of the consonantal association rules (Rule 4: Associate rightward). There appears to be some evidence that this rule is fairly recent and has a wider application than just the verb prefix. For example:

(68) /imbatw-aw   -ke/ →   [imbatagʷe] ~ [imbatʷagʷe]
     tuber-wood   -the   'the cassava'

The first phonetic realisation, [imbatagʷe], is used by all speakers middle-aged and younger, whereas [imbatʷagʷe] is used only by the older speakers (and not all of them). Also, a few older speakers have corrected our pronunciation of /kw-oban/ 'I hit' from [koban] (the most common way) to [kʷoban].

The association rules (outlined in §3.3.3) correctly generate the attested forms for all of the other person-number prefixes on vowel-initial verbs: the 1pi /t-/ and the 3p /r-/ prefixes do nothing, and the 3s /y-/ prefix has no adjacent C to associate with, and therefore is given a C timing unit as per Rule 6 (unassociated segments are given a C timing unit and linked to it, unless to do so would violate allowable CV syllable patterns).

The complex prefix form for 1px /aramy my-/ involves the 1px pronoun /aramy/ 'we (exclusive)'. This person-number subject shares the same prefix form (/my-/) with second person plural. It is mandatory in an utterance to include /aramy/ when referencing 1px subjects, for the prefix /my-/ alone will always be interpreted as referring to 2p.

5.2 Plural enclitic

There are four surface forms for the plural enclitic. The plural enclitic attaches to the last word in a noun phrase, and its form is a function of the immediate environment.

(69) a. words ending in a glide: hahy + PL →   [hahi̱re̱]   'pigs'
                                     asw   + PL →   [asu̱re̱]   'dogs'
     b. words ending in C: sorib + PL →   [soriba̱re̱] 'daggers'
        or monosyllabic roots: kwe   + PL →   [kwea̱re̱]   'bananas'
     c. other words ending in V: turi + PL →   [turi̱na̱re̱] 'machetes'

---

[37] If a morpheme break were not specified, this deletion rule would require morphemes like [kʷe] 'banana' to be posited as /wke/ rather than the more expected /kwe/ form. There is no evidence to suggest /wke/ constitutes a more accurate depiction.

Our current understanding of the plural enclitic leads us to posit the somewhat abstract form:

Plural Enclitic

(70)   V C V
         | | the shorthand form being /-Vre/
         - r e

The underlying form for the plural enclitic, displayed in (70), implies the plural morpheme is disyllabic, and this form, with the association rules (from §3.3.3), account for all forms but the [-nare] form in (69c).

It is not clear why some nouns ending in a V take the [-nare] form. Generally, most nouns which take this form are at least disyllabic. And many are possessed using the genitive construction (rather than using a possession word, see §5.3); many refer to things from nature, such as animals, plants, earth and ocean. No solid explanation yet accounts for the /n/ in this form, but it could be related to the 3s genitive enclitic. For now, the following 'patch' rule will have to suffice.

(71)   Non-monosyllabic words ending in V take the plural form: [-nare][38]

The plural form which surfaces is not dependent on its head noun but on its immediate environment. As mentioned before, the plural enclitic attaches to the last word in the noun phrase. If this word is an adjective, then the adjective's structure is what determines the surface manifestation of the plural marker.

(72) a. [as<u>ure</u>   rba   de]   [asl^wan<u>are</u>   rba   de]
        /asw-Vre   r-ba   de/   /asw-lan-Vre   r-ba   de/
        dog-PL   3p-go   already   dog-big-PL   3p-go   already
        'the dogs left already'   'the big dogs left already'

     b. [turi<u>nare</u>   mtu   de]   [turi   harhar<u>ure</u>   mtu   de]
        /turi-Vre   mtu   de/   /turi   harharw-Vre mtu   de/
        machete-PL   dull already   machete   new-PL   dull already
        'the machetes are already dull'   'the new machetes are already dull'

5.3 Genitive possessive enclitics

The following table gives the shorthand form for the Selaru genitive possessive enclitics. As before, the /w/ and /y/ segments are unassociated on the CV tier (see §3.3).

Table 10: The genitive possessive enclitics

1s	2s	3s	inam	1pi	1px	2p	3p
-kw	-mw	Ø/-n	Ø	ity- -t	ara- -my	-my	-t

The first person plural forms /ity- -t/ 'ours (inclusive)' and /ara- -my/ 'ours (exclusive)' are split morphemes attaching to each end of the possessed nominal or the possessing word

---

[38] There are a few exceptions, e.g. [aroare] /aro-Vre/ 'boats' and [abuare] /abu-Vre/ 'black-eyed peas'.

(which are /wasi/ 'have (generic)'[39] and /hina/ 'have (food)'); all others come in the postposition. The third-singular genitive has two forms, the second is always used for plural head nouns, but is also used with some singular head nouns which do not take a possessing word.[40]

(73) a. [wasikw   aroge]
         /wasi-kw   aro-ke/
         have-1sGEN   boat-the
         'my boat'

   b. [hinamhʷahkʸe   desʸ]
       /hina-mw   hahy-ke   desy/
       have-2sGEN   pig   the
       'that is your pork (food)'

   c. [arawasimsʸegʸe   desʸ]
       /ara-wasi-my   sey-ke   desy/
       1pxGEN-have-1pxGEN   house-the   that
       'that is our (exclusive) house'

   d. [itʸamatke   mʸatǰe]
       /ity-ama-t-ke   y-maty   de/
       1piGEN-father-1piGEN-the   3s-dead   already/
       'our (inclusive) father is already dead'

   e. [matamgʷe   koskosa   e]
       /mata-mw-ke   koskosa   e/
       eye-2sGEN-the   sleepy   Qmrkr
       'Are you sleepy?'(lit. 'your eye is sleepy?')

   f. [matake   koskosa]
       /mata-Ø-ke   koskosa/
       eye-3sGEN-the   sleepy
       'he is sleepy' (lit. 'his eye is sleepy')

The example genitive noun phrases in (73a) and (73c) use the generic possessing word /wasi/ 'have'. In the second case, /wasi/ is affixed by the 1px genitive in both pre- and postpositions. (73b) uses the possessing word for food /hina/, which forces /hahy/ 'pig' to be reinterpreted as 'pork' (/wasi-mw hahy-ke/ means 'your pig'). The head noun /ama/ 'father', in (73d), takes no possessing word; rather the genitive enclitics attach directly to it (in this case, in both pre- and postpositions). The head noun /mata/ 'eye', in (73e) and (73f), likewise needs no possessing word; the genitive enclitics attach to it directly. In the latter case, the 3s genitive enclitic is 'Ø' or empty (which it is for *all* possessing word constructions). With some other head nouns that do not take possessing words, the 3s genitive is /-n/:

---

[39] The generic possessing word for 1pi and 3p is the collapsed form /wai/ which conjugates as /ity-wai-t/ 'our (inclusive-generic)' and /wai-t/ 'their (generic)' respectively.

[40] These include kinship terms, body parts, etc., and is commonly referred to as 'inalienable possession' but which actually has a broader application than nominals which can not be disassociated from their possessor.

(74) a. [anamgʷe   bʸa   de]   b. [ananke   bya   de]
          /ana -mw-ke   y-ba   de/   /ana-n-ke   y-ba   de/
          child-2sGEN-the 3s-go   already   child-3sGEN-the 3s-go   already
          'your child left already'   'his child left already'

It is not clear yet why the 3s genitive enclitic is 'Ø' in some cases and /-n/ in others. If no article is used to close the noun phrase, and the head noun is the subject of the clause, the 3s genitive can surface as [-na]:[41]

(75)   [amana   iso   e]
          /ama-n(a)   i-so   e/
          father-3sGEN   AmSubj[42]-there Qmrkr
          'Is his father over there?'

The genitive enclitic for 3p is /-t/ (rather than the expected */-r/). This form is neutralised with 1pi, meaning there is no distinction between them. This requires the inclusion of /ity-/ 'we (inclusive)' (the full pronoun for 1pi) preposed to the head noun or possessing word in order to differentiate 1pi from 3p genitive possession.

When the plural enclitic is added to the genitive enclitics the surface forms follow the association rules.

(76)   mata-kw   +   -Vre   →   matakure   'my eyes'
        mata-mw   +   -Vre   →   matamure   'your (sing.) eyes'
        mata-n   +   -Vre   →   matanare   'his eyes'
        ara-mata-my +   -Vre   →   aramatamire   'our (excl.) eyes'
        ity-mata-t   +   -Vre   →   itmʸatatare   'our (incl.) eyes'
        mata-my   +   -Vre   →   matamire   'your (pl.) eyes'
        mata-t   +   -Vre   →   matatare   'their eyes'

5.4 Reduplication

Reduplication in Selaru can produce nouns from verbs, and can give rise to adjectives. Nominalisation of verbs takes the following pattern:

(77) | **Verb** | **Meaning** | **Redup.** | **Noun** | **Meaning** |
      | --- | --- | --- | --- | --- |
      | /-kur/ | 'to wipe' | → | /kurkur/ | 'towel' |
      | /-enaf/ | 'to sleep' | → | /enenaf/ | 'sleep (n.)' |
      | /-oban/ | 'to hit' | → | /oboban/ | 'a switch/stick' |
      | /-lakut/ | 'to walk' | → | /laklakut/ | 'gait' |

The noun forms in (77) were not produced by simple syllabic reduplication, for that would have produced unattested forms like *[ooban], *[lalakut]. Rather, Selaru uses a strategy described by Marantz (1982), which specifies that a reduplication template is what restricts allowable reduplication patterns. Selaru uses a 'CVC' reduplication template

---

[41] Phrasal phonotactics appear to account for the 3s genitive surfacing as [-na] rather than [-n]. This also appears to account for the [a] that is added to the 2s and 3s pronouns (/o/ and /i/ respectively) in subject noun phrases, yielding [oa] and [ia].

[42] The gloss 'AmSubj' is for 'animate subject' and marks the following demonstrative as referring to a living subject.

pattern, and also follows Marantz's Conditions (1982:447), which specify that association lines are attached **from left to right** (for prefixes) and are **phoneme driven** (e.g., association lines attach from the segmental tier to the CV tier and not the other way around).

How this strategy works out for the noun /laklakut/ 'gait' is as follows (the CVC reduplication template is underlined):

(78) a. **Root**

C V C V C
| | | | |
l a k u t   →
'to walk'

b. **Copying the Root**

C V C V C
| | | | |
l a k u t- l a k u t

c. **Add Redup. Template**

<u>C V C</u>   C V C V C
         | | | | |
l a k u t- l a k u t →

d. **Link up to Template**

<u>C V C</u>   C V C V C
\ \ \ | | | | |
l a k u t- l a k u t →
[laklakut] 'gait'

First, the segments from the root morpheme are copied over, (78b); then the CVC reduplication template is added to the CV tier, (78c). Then association lines are drawn from the segments up to the template, going from the left segment to right (consonants to C's and vowels to V's) (78d). Once the final C is linked up, the remainder of the duplicated segments (/ut/) are deleted out (following Marantz's Condition B (1982:446)).

The noun /enenaf/ 'sleep', when linking up, skips the first C in the template.

(79) a. **Root**

V C V C
| | | |
e n a f   →
'to sleep'

b. **Copying the Root**

V C V C
| | | |
e n a f − e n a f

c. **Add Redup. Template**

<u>C V C</u>   V C V C
        | | | |
e n a f - e n a f   →

d. **Link up to Template**

<u>C V C</u>   V C V C
 \ \ | | | |
e n a f - e n a f →
[enenaf] 'sleep (n.)'

The first segment in the reduplicated root morpheme is a vowel. Since linking is phoneme-driven, the /e/ skips the first C in the CVC template and attaches to the V; then the /n/ links to the adjacent C (association lines may not cross). As before, the leftover segments /af/ delete out; also, in this example, leftover CV unit in the reduplication template that are not linked to any segments are thrown out (e.g. the first C in the template for /enenaf/ 'sleep').

The noun /oboban/ 'switch' is formed in exactly the same way as /enenaf/ 'sleep'.

(80) **Verb**   **Noun**

VCVC   <u>CVC</u>   VCVC
| | | | \ \ | | | |
o b a n 'to hit' →   o b a n - o b a n   → [oboban] 'switch/stick'

The segmental tier must initiate association and not the CVC template. If the associations were driven by the CVC template, the reduplication process would produce ill-formed nominals:

(81) **Verb**   **Noun**
     V C V C   C V C - V C V C
     | | | | \ \ \ | | | |
     o b a n   →   o b a n - o b a n   →   *[banoban]

The results for this example would also be ungrammatical if the association lines were attached right to left (whether phoneme driven or skeletal driven); this agrees with the left to right generalisation for prefixes.

The verb root /-ba/ 'to go' becomes /ba-ba/ 'journey' when reduplicated.

(82) **Verb**   **Noun**
     C V   C V C C V
     | | \ \ | |
     b a   'to go'   →   b a - b a   →   [baba]   'journey'

Here again, the unused C in the reduplication template is deleted.

Adjectives in Selaru nearly all come in an inherently reduplicated form, for example, 'white' is [bokbok]; 'sleepy' is [koskosa], etc. These surface forms are already reduplicated, but in many cases the unreduplicated root never occurs, e.g. *[bok], *[kosa]. More intensive forms are reduplicated again as in [bokbokbok] 'pure white' and [koskoskosa] 'very sleepy'. Some adjectives result from a productive process: [werwer] 'wet' comes from the noun /wer/ 'water'; [butbut] 'narrow' comes from the unattested root /but/ (which also produces the verb /-abutak/ 'to make narrower'); [helhela] 'wide' probably comes from /-hela/ 'open up/spread out'.

This inherent reduplication is not limited to adjectives, though they are certainly the most common, as seen in /bolbol/ 'tomorrow' (e.g., /bol/ cannot be extracted from /bolbol/ 'tomorrow' and said by itself).

All such inherently reduplicated morphemes have already undergone a reduplication process before surfacing. We posit this even for those morphemes where the underlying root is never evidenced as a free-standing morpheme (e.g. /bok/, /kosa/, /bol/, etc.). Inherent reduplication uses the same CVC template as nominalisation. The underlying root is copied over and then linked to the CVC reduplication template.

It is clear that a CVC reduplication template will correctly produce [bokbok] 'white', [werwer] 'wet', and [koskosa] 'sleepy' from their underlying root forms (/bok/ (unattested), /wer/ 'water', and /kosa/ (unattested)).

Intensification requires that these same root forms be copied over again for each level of intensity. And for each level of intensity, a new CVC reduplication template must be added.

(83) **Root**   **Adjective**   **Intensive Adj**
     C V C V   C V C   C V C V   C V C   C V C   C V C V
     | | | | \ \ \ | | | | \ \ \   \ \ \ | | | |
     k o s a → k o s a - k o s a   → k o s a - k o s a - k o s a
                     [koskosa]   → [koskoskosa]
     (unattested)   'sleepy'   'very sleepy'

As before, any segments or CV units remaining unattached after the linking process are deleted.

This strategy also works for functors like /nini/ 'toward' which can be intensified to /ninini/ 'continue on and on'. Again, it is the basic root form (in this case the unattested /ni/) which is copied over for each level of intensity (analogous to /ba/ 'go' and /ba-ba/ 'journey' above).

(84) **Root**   **Adjective**   **Intensive Adj**.

    CV   CVC CV   CVC CVC CV

    | | \ \ | | \ \   \ \ | |

    n i → n i - n i → n i - n i - n i

               [nini]   [ninini]

    (unattested) 'toward'   'continue on and on'

Allowing /ni/ to be a valid root form is supported by other morphemes, such as the demonstrative /so/ 'there (far)', which can be reduplicated in a similar manner, e.g. producing /soso/ 'far away' and /sosososo/ which basically means 'way over there so far away you can't even imagine.'[43]

Using this same reduplication strategy, but allowing for the unique behaviour of Selaru glides, will produce /sols^{y}ol^y/ 'broom' from the verb /-sol^y/ 'to sweep'.

(85) **Root**   **Reduplication**   **Associations**

    CVC   CVC CVC   CVC CVC

    | | | → \ \ \ | | | → | | | /| | |\

    s o l y   s o l y s o l y   s o l y s o l y

    /soly/   [sols^{y}ol^y]

    'sweep'   'broom'

Because the unassociated-y can obey the association rules, it is able to link up to C units on the skeletal tier and not be deleted as other segments would be (as in /laklakut/ 'journey', where the [ut] segments are not successfully linked up and are deleted out of the reduplicated part). It appears that only the unassociated-y is allowed to link up in the reduplication process. This restriction follows the discussion of /y/'s apparent 'stronger' nature over /w/. An example of this is found in the adjective /harhar^w/ 'new'. If the unassociated-w in the reduplicated segments were able to link, the results would be: */harh^{w}ar^w/.

See Appendix A for more examples of Selaru reduplication.

---

[43] It is possible to reduplicate /nini/, /soso/ and others basically without limit. As many as eight reduplication levels have been encountered in natural text. The more reduplicated the root, the more intensive the meaning.

6 Miscellaneous residue

6.1 Demonstrative *desy*

One puzzling anomaly in the behaviour of Selaru glides involves the demonstrative /desy/ 'that'. In all cases but one, the final glide in /desy/ acts as predicted by the association rules, but when /desy/ is followed by the determiner /-ke/ 'the', the surface form is /desike/ and not */deskye/:

(86) a. [as<u>d</u>ʷesike   itayar]   b. [sai   des<u>d</u>ʸe]
        /asw   desy-ke   y-tayar]   /sai   desy   de/
        dog   that-the   3s-lost   what   that   Qmrkr
        'that dog is lost'   'what is that?'

This may indicate the functor /desi-ke/ has become frozen, even though the two morphemes are still used extensively, and separately, in other areas of the language.

6.2 Verbs

There are a few verbs which have only simple (C) onsets and yet never allow the glide in the subject prefix to associate to that C. One such example is /-tenw/ 'weave' which conjugates as: [kutenʷ] 'I weave', [itenʷ] 'she weaves', [ratenʷ] 'they weave';[44] another is /-ʔora/ 'sing', producing [kuʔora] 'I sing', [iʔora] 'he sings', [raʔora] 'they sing', etc. We need to either say such verbs are marked (in that they are irregular and do not follow the general association rules) or posit a more abstract underlying form (such as an extra C or V morpheme-initially):

(87) a. CCVCV   b. VCVCV
        | | | | or | | | |
        ʔ o r a   ʔ o r a

The complex CC onset, in (87a), would account for the irregular prefix behavior for these roots, i.e. acting as if they have a complex two C onset, but look like they have only one. The empty V, in (87b), would also produce the attested conjugate forms without complicating any rules, because the unassociated segments would immediately associate with the empty-V thereby gaining full vowel status.[45] Positing such abstract forms is nearly like marking these verbs as simply 'irregular', except that it attempts to give some explanation as to why they do not appear to follow the general pattern (i.e. if they do have an extra C or V in their onset margins, then they actually do follow the general rules). Needless to say, this is speculation, unless some other evidence arises to support such abstraction.

---

[44] These never follow the predicted forms: *[ktʷenʷ], *[tʸenʷ], *[rtenʷ], etc.

[45] Bob Mugele (pers. comm.) was first to suggest the possibility of an empty-V verb-initially. This deserves consideration. If complex onset verb roots historically were vowel initial (e.g. /-knam/ 'eat' was *Vknam), then the current association rules could be reduced and simplified. As it is now, we are assuming that it is the allowable syllable patterns which motivate a V insertion, but if an empty-V is already present, no resegmentation would be necessary.

6.3 Conclusion

Certainly the Selaru glides are interesting phonemes. We think the arguments herein support classifying them in a distinct class from either consonants or vowels. The association rules needed to describe their peculiar behaviour can probably be improved on through further investigation, but as they are, they simplify much of the 'irregular' structure of pluralisation, genitive possession, reduplication, and a few other grammatical processes beyond the scope of this paper. We feel fairly confident that the association rules are accurate for verbal structure, for predicting structure across morpheme breaks in compounded morphemes, and across word boundaries. However, as with any generalisation, there are exceptions. Yet the purpose of a generalisation is to account for the *majority* of the data. Creating rules to account for *all* of the data would require positing abstractions and/or altering the generalisations and thereby lose their inherent generality.

Appendix A: Reduplication data

verb (root)	noun	meaning: verb/noun
-ba	*baba*	to go/journey
-enah	*enenah*	to sleep/sleep
-het	*hetheta*	to beat/beater
-holat	*holholat*	to shut the door/door
-karan	*karkaran*	to grate/grater
-kety	*ketkyety*	to saw/saw
-kuly	*kulkuly*	to husk/husker
-kur	*kurkur*	to wipe/towel
-lakut	*laklakut*	to walk/gait
-lar	*larlar*	to wind/winder (for weaving)
-lilik	*lililik*	to make a shed/shedder (for weaving)
	(the final *k* may be a causitive morpheme)	
-morif	*mormorif*	to resurrect/life
-oban	*oboban*	to hit/a switch
-rasy-k	*rasryasy*	to approach/near
-skyer	*skyerker*	to make sour dough/the bread
	(this is irregular)	
-soly	*solsyoly*	to sweep/broom
-sulw	*sulsulw*	to put something on one's head/hat
-teta	*tetteta*	to chisel/chisel

adjective	intensive	meaning: basic/intensive
bokbok	*bokbokbok*	white/very
kunkuny	*kunkunkuny*	yellow/very
	(for some reason the /y/ does not spread to the /k/ but remains on the /n/ and is not written)	
mermer	*mermermer*	red/very
molmol	*molmolmol*	blue/very
bakbak	(unattested)	dry/??

betbetar	(unattested)	wet/??
bolbol	*bolbolbol*	very early tomorrow/morning
butbut	*butbutbuta*	narrow/very
harharw	*harharharw*	new/very
helhela	*helhelhela*	wide/very
koskosa	*koskoskosa*	sleepy/very
minmina	*minminmina*	sweet/bland (note: opposite meanings)
narnarw	*narnarnarw*	long/very
nelnyely	*nelnyelnyely*	clean/very
nini	*nininini*	continued on/and on and on
werwer	*werwerwer*	wet/very
kduhan	*kduhduhan*	smoking/billowing smoke
	(function of the word-initial k- is unknown; the root should be /duhan/ but is unattested)	

Appendix B: Sample text

The following is an interlinearised text of a legend told by U. Subitmele, an 85–year–old male, which includes the (1) phonetic transcription of the recorded text, (2) orthographic representation, (3) morphemic analysis, (4) English gloss, and (5) the English free translation. In the phonetic line the slash marks indicate breath segments. There are some differences between the first two lines, mostly due to changes in the text made after the transcribed story was checked with several Selaru readers (showing the difference between spoken and written language).

An interlinearised example text

1. **ʔírn^{y}egánia   tánlur^{y}áma   ʔisórnúske/   ʔode**
   Irnyeke   ania   Tanluryama   Isornuske,   ode
   iry -ne -ke   ani -a   Tanluryama   Isornus-ke   ode
   man -this-the   name-Ø   Tanlurayama Isornus-the and

   b^{y}á   ti   siná^{u}t//   b^{y}á   tíka   tási//
   bya   ti   sinaut,   bya   ti ika   tasi.
   y -ba   ti   sinaut   y -ba   ti y -ka   tasi
   3s-go   to   look for   3s-go   to 3s-work   ocean
   'There once was a man named Tanluryama Isornus and he went to work the ocean.'

2. **b^{y}á   nínia   r^{y}áma   lákt^{w}in^{w}e/   láktin^{w}eke**
   Bya   nini   ryama   laktinwe,   laktinweke
   y -ba   nini   y -rama   laktiw-ne   laktiw-ne -ke
   3s-go   kept.on 3s-spear fish-this   fish -this-the

```
?ilá        ka      bólkʷáitke      bá//
ila         ka      bolkwaitke      ba.
y-la        ka      bolkwait-ke     ba
3s-run      take    spear   -the    go
```
'He continued working and speared a fish, but it ran away with his spear.'

3.
```
lɛma    wásia   bólkʷáⁱtke/     de    ?írsyálik       wáⁱt//    lemáde
Lema    wasia   bolkwaitke      de    irsyalik        wait,     lemade
lema    wasi-a  bolkwait-ke     de    iry    -salik   wai-t     lemade
not     own -Ø  spear   -the    but   person-another  own-3p    so.then
```

```
yúlak               ?iba    ti    yála            ?írnʸe          húruk/
yulak               iba     ti    yala            irnye           huruk.
y -ulak             y -ba   ti    y -al    -a     iry    -ne      huruk
3s-turn.around      3s-go   to    3s-fetch-Ø      person-this     again
```
'He didn't own the spear, but another man owned it, so he turned around to go get this other man.'

4.
```
rahʸoyama waⁱseraneke            wási    ?úrke     ode    sakmʷágʸe//
Rahyoyama Waiseran    neke,      wasi    urke      ode    sakmwakye.
Rahyoyama Waiseran    ne -ke     wasi    ur-ke     ode    sakmway      -ke
Rahyoyama Waiseran    this-the   own     pot-the   and    rice.paddle-the
```
'Rahyoyama Waiseran is the one who owned the rice pot and rice spoon [as well as the spear].'

5.
```
yétbʸa              tia     yál     ti    wási    ?ároke//
Yety        bya     tia     yal     ti    wasi    aroke,
y -ety      y -ba   ti-a    y -al   ti    wasi    aro -ke
3s-bring    3s-go   to-Ø    3s-take in    own     boat-the
```

```
?ode    bʸá     ti    yóbak       bólkʷáⁱt     ne//
ode     bya     ti    yobak       bolkwait     ne.
ode     y -ba   ti    y -obak     bolkwait     ne
and     3s-go   to    3s-seek     spear        this
```
'He took [the rice pot and spoon] and went and put [them] in his boat and went to look for the spear.'

6.
```
bʸá     níni    ?imlár/     lemáde    yál     ?úr    desmʸodíti
Bya     nini    imlar,      lemade    yal     ur     desy ma     yoditi,
y -ba   nini    y -mlar     lemade    y -al   ur     desy ma     y-oditi
3s-go   kept.on 3s-hungry   so.then   3s-take pot    that  until 3s-seat
```

```
?ode    yayák       sákmʷagʸe          tia    dái      ?úrke/
ode     yayak       sakmwakye          tia    dai      urke
ode     y -ayak     sakmway      -ke   ti-a   dai      ur -ke
and     3s-stir     rice.paddle-the    in-Ø   earlier  pot-the
```

désigo	sérkʰek	kbʸénʷ	ʔúr	desʸ//
desikeo	serahkye	kbyenw	ur	desy.
desi-ke-o	serahy-ke	ky-benw	ur	desy
then	rice -the	it-full	pot	that

'He went on until he was hungry then he took the rice pot and set it down and stirred the rice paddle in the pot and then the pot was full of rice!'

7.
lemáde/	kʸásʷ	seráhdʸesike	ma	nʸá
Lemade	kyasw	serahy desike	ma	nya
lemade	y -kasw	serahy desy-ke	ma	nya
so.then	3s-spoon.out	rice that-the	until	he eats

maktéⁱ/	ʔode	bʸá	húruk//
maktei,	ode	bya	huruk.
ma -ktei	ode	y -ba	huruk
until-finished	and	3s-go	again

'Then he spooned out that rice and ate it until it was all gone and then went on his way again.'

8.
bʸá	nínia	kʸóman	ma	ʔimlárde/	lemáde
Bya	ninia	kyoman	ma	imlarde	lemade
y -ba	nini-a	y -koman	ma	y -mlar -de	lemade
3s-go	kept.on-∅	3s-feel	until	3s-hungry-already	so.then

yodíti	ʔúr	desʸ/	ʔode	yál	sákmʷagʸe	ma
yoditi	ur	desy,	ode	yal	sakmwakye	ma
y -oditi	ur	desy	ode	y -al	sakmway -ke	ma
3s-put.down	pot	that	and	3s-take	rice.paddle-the	to

iswáⁱl	desigeo	serahy	ʔalʸagʷe	kbʸénʷ	ʔúrke//
iswail,	desikeo	serahy	alyakwe	kbyenwa	urke,
y -swail	desi-ke-o	serahy	alyaw -ke	ky-benw-a	ur -ke
3s-stir	then	rice	k.o.rice-the	it-full-∅	pot-the

serahdʸesike	kbʸénʷ	ʔúrke	ma	ʔína	maktéⁱ/
serahy desike	kbyenw	urke	ma	ina	maktei,
serahy desy-ke	ky-benw	ur -ke	ma	y -na	ma -ktei
rice that-the	it-full	pot-the	until	3s-eat	until-finished

riyékan	éliko//	ʔána	kubáⁱ	lúlgʷe	negána	
ryekan	eliko	ana	kubai	lulkwe	neke,	ana
y -rekan	eliko	ana	kw-ba-i	lulw -ke	ne -ke	ana
3s-think	very	later	1s-go-LOC	front-the	this-the	later

ʔírire	rbóha/	kumgéskʷa/
irire	rbohe	kumkeskwaw.
iry -Vre	r -bohe	ku-mkeskwaw
person-s	3p-say	1s-witch

'He continued on until he felt hungry again and got out the rice pot and took the spoon and stirred it. Then the pot was full of rice and he ate it all up. Then he thought hard, "When I go onward, people will say I am a witch."'

9.   **lemáde akt^wólara/ ʔúrko   sákmagé   tia wén ne//**

```
    lemáde   aktʷólara/  ʔúrko        sákmagé          tia   wén   ne//
    Lemade   a-ktwolar   urke    o    sakmwakye        tia   wen   ne,
    lemade   a-kw-tolar  ur -ke  o    sakmway    -ke   ti-a  wen   ne
    so.then  ∅-1s-pitch  pot-the and  rice.paddle-the  at-∅  place this


    lemáde/  ʔúrgode       sákmʷaɤ   désike     tɤólar   tia
    lemade   urke    ode   sakmway   desike     tyolar   tia
    lemade   ur -ke  ode   sakmway   desy-ke    y -tolar ti-a
    so then  pot-the and   rice.paddle that-the  3s-pitch in-∅


    sáh    sitólar  só/
    sah    Sitolar  so.
    sah    Sitolar  so
    river  Sitolar-over.there
```
'So I should pitch the pot and the rice paddle at this spot". So then the rice pot and spoon, he pitched into the Sitolar river over there.'

10. **máne sáh   désike   rakíti   sitólar// ʔáli**

```
    mane  sah    desike    rakiti   " Sitolar" ali
    mane  sah    desy-ke   r-kiti   sitolar    ali
    so    river  that-the  3p-name  " Sitolar" because


    rtólara  ʔúrke    ʔo  sákmʷagé         ti//
    rtolar   urke    o   sakmwakye         ti.
    r -tolar ur -ke  o   sakmway    -ke    ti
    3p-pitch pot-the and rice.paddle-the   at
```
'So the river is called "Sitolar" because they pitched the rice pot and spoon there. '

11. **tɤólar   ti makteió   ʔíso   ʔiba tia**

```
    tyolar    ti  makteio        iso       iba    tia
    y -tolar  ti  ma    -ktei-o  y -so     y -ba  ti-a
    3s-pitch  to  until-done-∅   he-there  3s-go  to-∅


    ʔimelóa                     lamdésar  tímurke  so//
    imeloa                      Lamdesar  timurke  so.
    y -melo            -a       Lamdesar  timur-ke so
    3s-raised.a.family-∅        Yamdena   east -the there
```
'After he got rid [of them], he went and raised a family in yonder East Yamdena.'

References

Blust, Robert, 1978, Eastern Malayo-Polynesian: a subgrouping argument. *Proceedings of the Second International Conference on Austronesian Linguistics, fascicle 1*, 181–234. Canberra: Pacific Linguistics.

1980, Austronesian etymologies. *Oceanic Linguistics* 19:1–181. Honolulu: University of Hawaii.

1983–1984a, Austronesian etymologies--II. *Oceanic Linguistics* 22/23:1–2:29–149. Honolulu: University of Hawaii.

1983–1984b, More on the position of the languages of eastern Indonesia. *Oceanic Linguistics* 22/23:1–2:1–28. Honolulu: University of Hawaii.

1986, Austronesian etymologies--III. *Oceanic Linguistics* 25:1–2:1–123. Honolulu: University of Hawaii.

Clements, George N., 1985, The geometry of phonological features. *Phonology Yearbook* 2:225–52.

Clements, George N. and Samuel J. Keyser, 1983, *CV phonology: a generative theory of the syllable*. Cambridge: MIT Press.

Collins, James T., 1982, Linguistic research in Maluku: a report of recent fieldwork. *Oceanic Linguistics* 21:1–2:73–150. Honolulu: University of Hawaii.

Coward, David F., 1990, An Introduction to the grammar of Selaru. MA Thesis. Arlington: University of Texas at Arlington.

Coward, Naomi E., 1989, A phonological sketch of the Selaru language. *Workpapers in Indonesian Languages and Cultures* 7:1–42. Ambon: Pattimura University and the Summer Institute of Linguistics.

Drabbe, P., 1932, *Beknopte spraakkunst en korte woordenlijst der Slaroeëesche taal*. Bandung, Indonesia: A.C. Nix.

Grimes, Barbara F., ed., 1992, *Ethnologue: languages of the world*, 12th edition. Dallas: Summer Institute of Linguistics.

Hughes, Jock, 1987, The languages of Kei, Tanimbar and Aru: a lexicostatistic classification. *NUSA* 27:71–111. Jakarta: Universitas Atma Jaya.

Kantor Statistik Propinsi Maluku, 1985, *Maluku dalam angka*.

Marantz, Alec, 1982, Re Reduplication, *Linguistic Inquiry* 13, 435–82.

Mills, Roger F., 1981, Additional addenda. *NUSA* 10:59–82. Jakarta: Universitas Atma Jaya.

1991, Tanimbar-Kei: an Eastern Indonesian subgroup. In Robert Blust, ed., *Currents in Pacific Linguistics: papers in Austronesian languages and ethnolinguistics in honour of George W. Grace*, 241–263. Canberra: Pacific Linguistics.

Mills, Roger F. and John Grima, 1980, Historical developments in Lettinese. *Austronesian Studies: Papers from the Second Eastern Conference on Austronesian Languages*. Paz Buenaventura Naylor, ed., Michigan Papers on South and Southeast Asia No. 15, 273–284. Ann Arbor: University of Michigan.

Pauwels, Simonne, 1985, Some notes on the importance of relationships in a Tanimbarese society. Seminar Penelitian Indonesia Bagian Timur: Manado, 23–29 Juli 1985. Lembaga Ekonomi dan Kemasyrakatan Nasional--LIPI dan UNSRAT.

1990a, From Hursu Ribun's 'three hearth stones' to Metenleru's 'sailing boat': a ritual after the harvest. In C. Barraud and J. Platenkamp, eds, *Rituals and socio-cosmic order in eastern Indonesian societies, Part II, Maluku*. Bijdragen tot de Taal-, Land- en Volkenkunde 146:21–34.

1990b, The *awaire* relationship in a Tanimbarese society. Paper presented at Conference on Halmahera Research and its consequences for the study of Eastern Indonesia, KITLV, Leiden, October 1990.

Salzner, Richard, 1960, *Sprachenatlas des Indopazifischen raumes*. Wiesbaden: Otto Harrassowitz.

Wurm, S.A. and S. Hattori, eds, 1981, *Language atlas of the Pacific area*. Canberra: The Australian Academy of the Humanities in collaboration with the Japan Academy.

3 *Tukang Besi dialectology*

MARK DONOHUE

1 Previous studies

Very little linguistic attention has previously been paid to the Tukang Besi islands.[1] Anceaux (1978), in a short but generally accurate survey of south-east Sulawesi, concluded that there was only one language spoken on the Tukang Besi islands. In studies looking further afield, Esser (1938) and Salzner (1960) linked Bonerate, found on some of the islands south of Selayar in the Sea of Flores, with Tukang Besi in the Tukang Besi-Bonerate sub-group, a part of the Muna-Buton group. [The Bonerate islands are indicated on the map in the introduction to this volume, while the remaining islands are indicated on the map in this paper.] The exact relationship shared by Tukang Besi and Bonerate, which actually extends over the islands of Bonerate, Karompa and Madu (and not Kalao Toa, as claimed by Esser) (Broch 1981), has not previously been investigated. A short and idiosyncratic synchronic description of the Bonerate dialect, dealing mainly with some morphological features of the language, has been completed by Manyambeang et al. (1985).

The internal relationship of the isolects spoken on the four islands that make up the Tukang Besi archipelago in south-east Sulawesi is essentially unknown. Sneddon (1987) listed Wanji (=Wanci) and Binonko (=Binongko) as separate dialects, but did not provide details of the other islands or of how separate the dialects are. In Bhurhanuddin's (1979) and Kaseng's (1987) surveys of the languages of South-east Sulawesi, Tukang Besi is listed, but without internal variation (under the name *Wakatobi*).[2]

The precise nature of the language spoken in the many Tukang Besi communities in eastern Indonesia, usually known locally as *orang Buton* or *orang Binongko*, remains to be investigated; apart from their abundance, nothing is known about these communities.

---

[1] The Tukang Besi islands are located between the Sea of Flores and the Banda Sea, in Kabupaten (Regency) Buton, Sulawesi Tenggara province.

[2] This name is derived from the first two letters of the name of each of the islands. It is one of the names in current use by the Tukang Besi people.
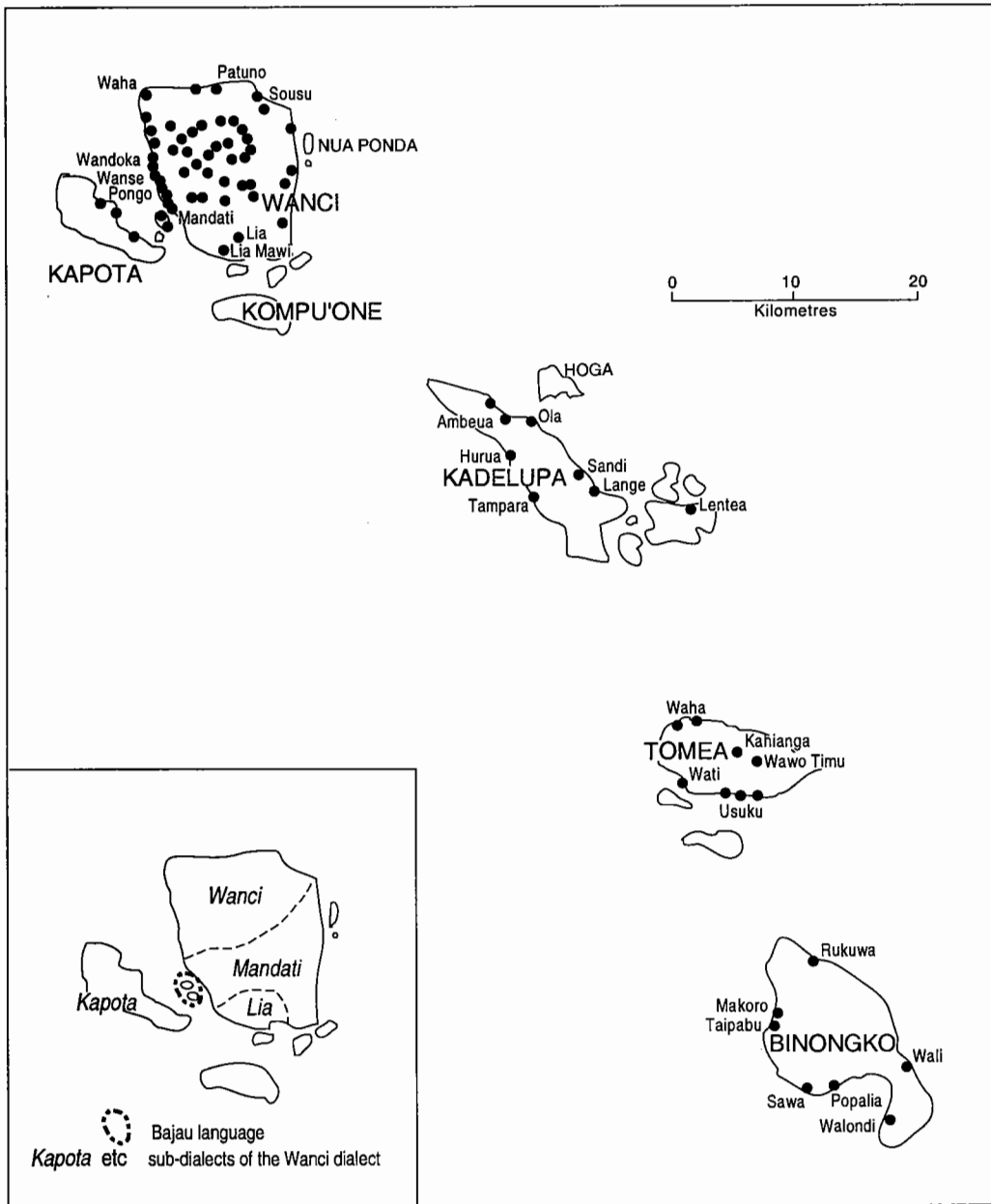
The Tukang Besi language

2 Lexicostatistics

Cognate counts were computed from the comparison of 215-item word lists[3]; the lexical content is that of the word lists at the end of this paper, and was adapted from the SIL 226-item survey word list, with thirteen deletions and two additions: 'bark', 'chicken louse', 'feather' and 'warm' were deleted as being poly-morphemic compounds of morphemes occurring elsewhere in the corpus; 'grandfather', 'elder sister', 'husband', 'wife', 'uncle' (twice, FB and MB[4]) and 'aunt' (twice, FZ and MZ) were deleted as occurring elsewhere in the list, and 'lake' was rejected as culturally irrelevant (there are no lakes on the dry islands); 'roof' and 'paddle' were added as partial compensation. Pairs of words were examined for lexical similarity; no attempt was made to eliminate possible borrowings. Similarity was determined by inspection of the phonetic character of the words, with allowance being made for sound changes that regularly occurred between two isolects. The following figures for cognate percentages were calculated:

Wanci				
85	Kaledupa			
82	83	Binongko		
79	80	87	Tomea	
79	81	85	81	Bonerate

Figure 1: Lexicostatistic results

The figures for cognate percentages point to a primary division between Wanci-Kaledupa on the one hand (85% cognate) and a chain composed of Tomea-Binongko-Bonerate on the other. Within the second grouping, Tomea and Binongko are clearly closer in the chain than Binongko-Bonerate. Note that all dialects show a higher cognate percentage with Binongko, due to the considerable numbers of Binongko living away from their home island, mainly in Maluku, where their role in the production of cloves lends importance to their dialect. The central position of Binongko is reflected in the position of Binongko in Figure 1, where it appears out of its geographic position. The figures obtained when the other dialects were compared with Binongko are inflated due to the regular contact that the Binongko have with other Tukang Besi speakers.

The lexicostatistic figures establish that the communalects in question are closely related to each other. The question of whether we are dealing with one language that chains across the islands, or a language chain situation where there is actually more than one separate language without clear borders, cannot be answered from the lexicostatistic data alone. The

---

[3] Word list sources were Friberg's (n.d.) Bonerate word list; C. Grimes and B. D. Grimes' (1987) language survey book; and the author's own field notes (n.d.)

[4] Abbreviations used in this paper:

[-/]	rising intonation	MB	mother's brother
[-_]	falling intonation	MZ	mother's sister
*^{N}C	prenasalised consonant	PAn	Proto Austronesian
<Ind.	derived from Indonesian	pl.	plural
<MB	derived from Muna Buton	PTB	Proto Tukang Besi
ex.	exclusive	sg.	singular
FB	father's brother	s.o.	someone
FZ	father's sister	v.	verb
in.	inclusive		

lowest cognate percentages are on the borderline (80%) of the generally agreed-upon limit that is indicative of the existence of a separate language, rather than dialects of the one language (although it has been shown that thresholds do vary from case to case – see J. Grimes (1988)). It is helpful to look at some sound correspondences that can aid the division of the isolects.

3 Sound correspondences

There are several sound changes evident between the northern and southern isolects of the Tukang Besi language, both systematic and semi-systematic. These are both in the nature of sound changes that have occurred in a stage ancestral to the dialect groups, and differences in the major allophonic realisation of a phoneme, or phonemic interpretation of a particular phonetic segment. Both of these are discussed below with examples.[5]

*PAn/ *PTB	*p	*q	*S	*g
Wanci	P [6]	ʔ	ʔ	g
Kaledupa	p	ʔ	ʔ, Ø	g
Tomea	h	Ø (ʔ)	Ø	ŋŋ
Binongko	h	Ø (ʔ)	Ø	ŋŋ
Bonerate	h	Ø (ʔ)	Ø	ŋ

Figure 2: Sound changes

	C/V__σ#	l/a,o,u__	h/__u	*NC	Stress
Wanci	C	l, r	/h/ (~ [ɸ])	NC	[¯]
Kaledupa	CC	l, r	/h/ (~ [ɸ])	NC	[\ ]
Tomea	CC	ɭ	/w/ (~ [ɸ])	NN	[\ ]
Binongko	CC	ɭ	/w/ (~ [ɸ])	NN	[\ ]
Bonerate	C	l	h	NC	[\ ]

Figure 3: Allophonic and phonemic differences

There is a strong division into two groups, with the first group comprising the northern islands of Wanci-Kaledupa, which show many instances of unlenited /p/s, glottal stops corresponding to PAn *q, /g/ in *baga* 'cheek' (= *baŋŋa* in Tomea, Binongko), and irregular *1 > r (through **ɭ). The southern isolects display /h/ for *p, Ø (and sometimes /ʔ/) for *q, a geminate nasal in place of the /g/, and show retroflexion as an allophone of /l/ before back vowels. Illustrations of the sound correspondences are:

---

5   Apart from the word lists, Tukang Besi is transcribed into an orthography for typographical convenience. /ng/ = [ŋ], /w/ = [β] or [ɸ], /'/ = [ʔ], and /u/ = [ʊ]. All other letters have their expected phonetic values. Sequences of N + C are prenasalised unit phonemes.

6   As is the case with many of these sound changes, they are not necessarily absolutes; there are cases of *p > /h/ in Wanci dialect, and of *q > /ʔ/ in Binongko as well. They do, however, represent a useful isogloss in many words.

English	'what'	'tree'	'man'	'wood'	'you'
PAn	(PTB: *paira*)	*puqun*	*ma-Ruqanay*	*kaSiw*	*kaSu*
Wanci	*paira*	*hu'u*	*mo'ane*	*ka'u*	*iko'o*
Kaledupa	*paira*	*hu'u*	*mo'ane*	*kau*	*iko'o*
Southern	*haira*	*hu'u*	*moane*	*kau*	*ikoo*

English	'cheek'	'root'	'inside'	'body hair'	'leg'
PAn	(PTB: *baga*)	*akaR*	*Dalem*	*buluq*	*qaqay*
Wanci	*baga*	*aka*	*laro*	*wulu*	*ae* [⁻_]
Kaledupa	*baga*	*akka*	*laro*	*wulu*	*ae* [_]
Southern	*baŋŋa*	*akka*	*laɭo*	*ɸuɭu*	*ae* [_]

Figure 4: Examples of the sound correspondences

Whilst the line is fuzzy, because not all of the isoglosses meeting to form one strong division, there is a clear difference between the southern isolects and Wanci and Kaledupa. The realisation of stressed syllables is interesting in the Wanci dialect. Rather than representing a prominence in the word manifested by pitch and loudness, stressed syllables seem to have acquired a consistently high, level pitch as their primary cue, rather than having a pitch change that is led by the intonation pattern of the sentence and co-occurs with increased loudness, as is the case in English and the other dialects.

The southern isolects have also reinterpreted /hu/ sequences as /wu/. This can be explained by reference to the usual allophonic realisations of /w/ and /h/. In Wanci and Kaledupa isolects, /w/ has the allophones [ʋ], [β], and [ɸ], in decreasing order of frequency, and /h/ is always realised as [h] except before a /u/ ([ɯ]), in which environment it is often realised as a [ɸ]. In Tomea and Binongko, however, the most frequent allophone of /w/ is [ɸ], leading to reinterpretation of the [ɸ] before a /u/ as being an underlying /w/, now showing allophonic variation with [ʋ] and [β]. For example:

'give'	Northern	Southern
Phonemic:	/huʔu/	/wuʔu/
Phonetic:	[ɸɯʔɯ] ~ [hɯʔɯ]	[ɸɯʔɯ] ~ [βɯʔɯ] ~ [ʋɯʔɯ]

The many prenasalised stop phonemes that occur in Tukang Besi find themselves often realised as geminated nasals in the speech of Tomea and Binongko people. This occurs also in the northern islands, although less frequently; on Binongko, and especially Tomea, the most frequent allophone is a geminated nasal:

'frog'	Northern	Southern
Phonemic:	/tumbekeke/	/tumbekeke/
Phonetic:	[tɯmbɛ'kɛkɛ]	[tɯmːɛ'kɛkɛ]

The examples above reveal dialectal divergences of development; with *hu'u* the phonetic realisations are the same but the phonological interpretations have diverged, however with *tumbekeke* the phonological representations have remained invariant but the phonetic realisations have changed. Positing a phonological representation of the form /tum:ekeke/ for the southern dialects would complicate the phonotactic statement with the addition of phonemically geminate sequences, which occur only where a prenasalised stop is the expected reflex.

4 Sociolinguistics

I have not had the opportunity to extensively examine language attitudes across all the communities mentioned here, nor to conduct intelligibility tests, but several impressionistic details can be presented.

Wanci people claim to be able to understand the speech of all the other islands, and indeed it seems that a set of the population can do this. Most adults who live on the coast or who have travelled (not necessarily to other Tukang Besi islands, but usually on trading trips to other communities of Binongko origin in eastern Indonesia) have an extensive passive knowledge of many of the words used in the dialects of the other islands. On Wanci, when mispronouncing *tee* 'tea' ([tɛ:]) with some hesitation ([tɛʔɛ]), I have been told that I had asked for 'water'; the word for water on Wanci is *uwe*, but on neighbouring Kaledupa the word is *te'e*. Thus, whilst a word is not in active use, there is passive familiarity with it.

Children, however, and people from the mountain who have less experience with outsiders, seem unable to understand Tomea or Binongko speakers, and do not have this ready command of non-native words. This implies that knowledge of the lexicon not native to one's own island is learned through contact with outsiders, and not simply picked up in everyday life. Binongko speakers are unable, unless they have had previous experience with Wanci speakers, to understand the Wanci dialect, due to (I am told) the lack of gemination present in Wanci dialect, and the large number of different lexical items, which are not in common knowledge on Binongko. Indeed, in the markets on Wanci, people from Binongko or Kaledupa who have come to sell produce or handicrafts use Wanci dialect words, rather than using their own dialect forms, when speaking to the locals; they are also faster to point out the source of a word as being 'Wanci' or 'Kaledupa'. The acquisition of Wanci dialect forms by people from other islands is an active and conscious process, and not so much the casual approach evidenced in the case of the knowledge of non-Wanci words amongst the Wanci people.

5 Lexical innovations

Variance in vocabulary is one of the major features of Tukang Besi dialectology, grammatical differences being minimal and phonological differences mainly regular (as seen in §3). I list 215 diagnostic lexical items in the word lists in the appendix. The number of lexical isoglosses is as follows, arranged by island or group of islands that share them; the number beside the name of the island or islands shows how many words are peculiar to that group, as against all the other isolects, taken from 488-item word lists (the SIL Sulawesi umbrella word list). There are, for example, 22 lexical items unique to Wanci speech amongst the Tukang Besi isolects, and three that are found in Kaledupa or Bonerate, but not in other islands.

Unique:		Shared:	
Wanci	22	Wanci-Kaledupa	17
Kaledupa	5	Kaledupa-Tomea	1
Tomea	2	Tomea-Binongko	1
Binongko	0		
Bonerate	2		
Wanci-Bonerate	1	Kaledupa-Bonerate	3
Tomea-Bonerate	1	Binongko-Bonerate	1

It can be clearly seen that the greatest amount of lexical diversity is in the north, the southern isolects showing remarkably little in the way of unique lexical items. The diagnostic lexical items that are unique to each island or group of islands are as listed in the following tables (numbers refer to the word lists in the Appendix), listing the forms from the five isolects for each word.

5.1 Lexical items that are unique to Wanci

βeŋka 'betel' 76;
likeʔe 'wake up' 205;
uβe 'water' 96;
iri/ kaβea 'wind' 90;
ika 'fish' 59;
pod'a 'knife' 108;
metaŋku 'near' 119;

morondo 'night' 169;
lagu 'sing' 173;
saʔasa 'one' 147;
βande 'rain' 89;
ulaŋiʔe 'repeat' 172;
henenasi 'vomit' 195.

Word 96 *uβe* < MB languages; 90 (*kawea*) < MB languages; ika 59 < Malay *ikan*; *morondo* 169 shows cognates with several MB languages, but could equally be an independent development; *lagu* 173 < Malay *lagu*; rain 89 *wande* means 'wind' in the other island dialects, and in the Lia sub-dialect spoken in the south of Wanci.

The following items also occur in Wanci but not in other dialects.

	'butterfly'	'day'	'finger'	'house'	'medicine'
Wanci	*mberombero*	*'oloo*	*wunga*	*wunua*	*ro'o*
Kaledupa	*loma*	*moina*	*ranga*	*sapo*	*kombi*
Tomea	*loma*	*moina*	*ranga*	*sapo*	*kombi*
Binongko	*loma*	*moina*	*ranga*	*sapo*	*kombi*
Bonerate	*loma*	*moina*	*ranga*	*sapo*	*kombi*

The word *'oloo* has the meaning of 'sun' in all dialects; *ro'o* has the meaning of 'leaves' in all dialects.

	'raincloud'	'remember'	'shy'	'thunder'
Wanci	*komo*	*rodo*	*ambanga*	*paha*
Kaledupa	*ruto*	*hetira'a*	*mo'ini*	*tondu*
Tomea	*ruto*	*hetira'a*	*mo'ini*	*tondu*
Binongko	*ruto*	*hetira'a*	*mo'ini*	*tondu*
Bonerate	*ruto*	*hetira'a*	*mo'ini*	*tondu*

5.2 Words that are unique to Kaledupa

d'i mburi 'behind' 164   *motind'ou* 'know' 170
kend'esa 'cough' 193   *ŋgolo* 'lie down' 200
tanta 'dry' (v.) 186

Tanta 186 means 'dry (with towel)' in Kaledupa, not 'dry (in the sun)', and thus probably represents an elicitation error rather than a real lexical difference; *motind'ou* 170 < Ciacia, Kumbewaha *motindou*; *b'o'o* nose 5 < Ciacia *bo'o*

5.3 Words that are unique to Tomea

ke'e 'cry' 174;   *herihɖu* 'wash' 187

5.4 Words that are unique to Bonerate

kudara 'green' 140 < Kalao *kudarang*
paupau 'tired' 201

5.5 Lexical items shared by Wanci and Kaledupa

to'oge 'big' 114;   *rod'oŋo* 'hear' 176
talinga 'ear' 10 < Ind. *telinga*   *ko'o* 'smoke' 102
ahu 'fire' 101   *leama* 'good' 116

gana and *ganna* 'four' 150 < Ind. *genap*
mo'aro 'hungry' 131 cognate with many MB languages
maβi 'sea' 91 cognate with Kaimbulawa (MB) *mawi*

The following items are also shared exclusively by Wanci and Kaledupa.

	'correct'	'firefly'	'learn'	'lightning'	'plant (v.)'
Wanci	*kobe*	*kalipopo*	*henehenai*	*kinda*	*hembula*
Kaledupa	*kobe*	*kalipopo*	*henehenai*	*kind'a*	*hembula*
Tomea	*mantotu*	*olo'onda*	*hisihisinga*	*lalo*	*amo*
Binongko	*mantotu*	*olo'onda*	*hisihisinga*	*lalo*	*amo*
Bonerate	*mantotu*	*olo'onda*	*hisihisinga*	*lalo*	*amo*

Note that 'firefly' < Buton languages.

	'play'	'sole': _ *nu ae*	'sweet'
Wanci	*'aka'aka*	*randa*	*momeke*
Kaledupa	*'aka'aka*	*rand'a*	*momeke*
Tomea	*kolia*	*lalo*	*moni*
Binongko	*kolia*	*lalo*	*moni*
Bonerate	*kolia*	*lalo*	*moni*

5.6 Lexical items shared by Kaledupa and Tomea

kotu'a 'forest' 99

5.7 Lexical items shared by Tomea and Binongko

Both *b'ahɸuɭi* in Tomea and *b'ahoɭi* in Binongko mean 'small' 115

5.8 Lexical items shared by Wanci and Bonerate

βungka 'mountain' 98

5.9 Lexical items shared by Kaledupa and Bonerate

Kaledupa *ŋusu* 'lip' 7; and Bonerate *βiβi ŋusu* 'lip' 7, but all other dialects have *ŋusu* meaning 'mouth' 5; thus lip < mouth
Kaledupa *d'i liku* and Bonerate *i liku* mean 'outside' 166

5.10 Lexical items shared by Tomea and Bonerate

keaβa 'swim' 189

5.11 Lexical items shared by Binongko-Bonerate

Binongko *ha ia* and *ɲia* in Bonerate both mean 'this' 142
Many lexical forms that are divergent in Wanci have cognates in the languages of Buton, which is geographically closer to Wanci than to the other Tukang Besi islands. The Bonerate forms often have cognates in the Kalao language, spoken nearby on several of the islands in the Sea of Flores.

6 Conclusions and discussion

The cumulative effect of the phonological, lexical and social differences displayed by the different speech varieties is that there are two closely related but separate languages on the Tukang Besi islands. Inherent intelligibility between these two varieties is marginal, and inadequate for effective communication, requiring some degree of bilingualism in or prolonged contact with the target variety.

The fact that these two languages are often identified as one by the native speakers is probably a reflection of the much greater similarities that the two languages, North Tukang Besi and South Tukang Besi, share with each other when compared with the languages of the nearby Muna-Buton region. The languages of Buton form a separate subgroup, and the

only close similarity that Tukang Besi shares with this subgroup is with Lasalimu, and that only through extensive family and trade ties with the town from Wanci.

The strong cultural identity shared by the Tukang Besi people, and frequent contact between peoples of the different islands, are also powerful factors in the native speaker assessment that the languages are the same. Factors such as mono-directional intelligibility, non-intelligibility amongst speakers who have not learned the lexical items from other speech forms, and the need to learn a 'standard' for communication, whilst retaining awareness of the differences in one's own speech variety, tip the scales in favour of classifying the speech forms as two languages.

The language situation on each of the islands is much more complicated than this summary would indicate; in the Wanci area alone there are four commonly recognised speech varieties, that of the north, west and hill lands (*pogau Wanse*), that of the south-west, southern hill lands and east coast (*pogau Mandati*), that of the south (*pogau Lia*), and the divergent speech of Kapota island, which has been influenced by immigrants from Binongko (*pogau Kapota*). The differences are mainly lexical, but there are regular allophonic changes as well: north coast dialects, for instance, usually drop the final vowel in words, producing an impressionistically very different speech form. The intonation of Kapota island is markedly different from the rest, often rising on the final syllable, and the use of the phrase-final particle *sa* is a hallmark of Kapota speech. Doubtless such variation exists on the other islands as well, but the gross differences are, as may be expected, marked by the greater water boundaries.

A further complication is present in the sociolinguistic diversity of Wanci and, to a lesser extent, the other islands. In addition to the local Tukang Besi dialect, there is a great deal of fluency in Moluccan Malay (Collins 1983; B.D.Grimes 1991), and to a lesser extent Standard Indonesian; additionally, Straits Malay is sometimes used by those who have worked in Singapore. The Wolio language is known by many of the nobility, from the days when it was the lingua franca of the Sultanate (Anceaux 1952), and proximity and family bonds make a limited ability in Cia-Cia or Lasalimu commonplace amongst people on the west coast. Finally, both Wanci and Kaledupa have large (4,000 and 1,500 people) Bajau communities, who remain a distinct linguistic unit in both of the island communities. All of these factors lead to a greater sense of unity amongst the speakers of Tukang Besi isolects than simple intelligibility might indicate.

Appendix: Word lists from the Tukang Besi dialects

	1. head	2. hair	3. face	4. eye	5. nose
Wanci	*kapala*	*hotu*	*aropa / liau*	*mata*	*ŋoʔo*
Kaledupa	*kapala*	*hotu*	*rou*	*mata*	*bʼoʔo*
Tomea	*kapaɭa*	*hotu*	*muka*	*mata*	*ŋoʔo*
Binongko	*kapaɭa*	*hotu*	*liau*	*mata*	*ŋoʔo*
Bonerate	*kapala*	*hotu*	*rou*	*mata*	*ŋoʔo*

	6. mouth	7. lip	8. tongue	9. tooth	10. ear
Wanci	ŋusu	βiβi	ela	koni	taliŋa
Kaledupa	ŋaro	ŋusu	ella	koni	taliŋa
Tomea	ŋusu	βiβi	ella	koni	tuʄi
Binongko	ŋusu	βiβi	ella	koni	tuʄi
Bonerate	ŋaro	βiβi ŋusu	ela	koni	tuli

	11. neck	12. hand	13. nail	14. breast	15. stomach
Wanci	kaβuʔu	lima	kuku	titi	kompo
Kaledupa	kaβuʔu	lima	kuku	titi	kompo
Tomea	ɸuʔu	lima	kuku	titi	kompo
Binongko	ɸuʔu	lima	kuku	titi	kompo
Bonerate	βuʔu	lima	kuku	titi	kompo

	16. leg	17. knee	18. body hair	19. skin	20. flesh
Wanci	ae	tuʔu	βulu	kuli	ramo
Kaledupa	ae	tuʔu	βulu	kuli	ramo
Tomea	ae	tuʔu	ɸulu	kuʄi	ramo
Binongko	ae	tuʔu	ɸulu	kuʄi	ramo
Bonerate	ae	tuʔu	βulu	kuli	ramo

	21. fat	22. bone	23. heart	24. blood	25. liver
Wanci	moʔoβu	b'uku	b'ake	raha	ate
Kaledupa	moʔoβu	b'uku	b'ake	raha	ate
Tomea	monuɸu	b'uku	b'ake	raha	ate
Binongko	mouβu	b'uku	b'ake	raha	ate
Bonerate	taba	buku	tiŋkula	raha	ate

	26. urine	27. faeces	28. person	29. man	30. woman
Wanci	leleʔe	taʔi	mia	moʔane	βoβine
Kaledupa	leleʔe	taʔi	mia	moʔane	βoβine
Tomea	meʔe	taʔi	mia	moane	βoβine
Binongko	meʔe	taʔi	mia	moane	ɸoɸine
Bonerate	meʔe	taʔi	mia	moane	βoβine

	31. father	32. mother	33. child	34. first born	35. last born
Wanci	ama	ina	ana	tumpe	kaipu
Kaledupa	ama	ina	ana	tumpe	-
Tomea	ama	ina	ana	tumpe	-
Binongko	ama	ina	ana	tumpe	bonso
Bonerate	ama	ina	ana	ana ikaka	ana pusi

	36. grand-child	37. grand-parent	38. ancestor	39. older sibling	40. younger sibling
Wanci	*ompu*	*ompu*	*ompu rambi*	*ikaka*	*iai*
Kaledupa	*ompu*	*ompu*	*ompu rambi*	*ikaka*	*iai*
Tomea	*ompu*	*ompu*	*ompu rambi*	*ikaka*	*iai*
Binongko	*ompu*	*ompu*	*ompu rambi*	*ikaka*	*iai*
Bonerate	*ompu*	*ompu*	*ompu rambi*	*ikaka*	*iai*

	41. slave	42. guest	43. friend	44. I	45. you (sg.)
Wanci	*tud'ua*	*tamu*	*kene*	*iaku*	*ikoʔo*
Kaledupa	*tud'ua*	*tamu*	*kene*	*iaku*	*ikoʔo*
Tomea	*tud'ua*	*tamu*	*pobaβa*	*iaku*	*ikoo*
Binongko	*tud'ua*	*tamu*	*kene*	*iaku*	*ikoo*
Bonerate	*tudua*	*tamu*	*kene*	*iaku*	*ikoʔo*

	46. he/she	47. we (excl.)	48. we (incl.)	49. you (pl.)	50. they
Wanci	*ia*	*ikami*	*ikita*	*ikomiu*	*amai*
Kaledupa	*ia*	*ikami*	*ikita*	*ikomiu*	*ammai*
Tomea	*ia*	*ikami*	*ikita*	*ikomiu*	*ammai*
Binongko	*ia*	*ikami*	*ikita*	*ikomiu*	*ammai*
Bonerate	*ia*	*ikami*	*ikita*	*ikomiu*	*ammai*

	51. horn	52. tail	53. bird	54. egg	55. louse
Wanci	*tandu*	*iku*	*kad'ad'i*	*goraʔu*	*kutu*
Kaledupa	*tandʳu*	*iku*	*kad'ad'i*	*goraʔu*	*kutu*
Tomea	*tandu*	*iku*	*kad'ad'i*	*goraʔu*	*kutu*
Binongko	*tandu*	*iku*	*kad'ad'i*	*goraʔu*	*kutu*
Bonerate	*tandu*	*iku*	*kado-kadola*	*gorau*	*kutu*

	56. bat	57. mosquito	58. snake	59. fish	60. rat
Wanci	*honiki*	*βai*	*saʔa*	*ika*	*βoleke*
Kaledupa	*honiki*	*βai*	*saʔa*	*kenta*	*βoleke*
Tomea	*honiki*	*ɸai*	*saʔa*	*kenta*	*ɸuleke*
Binongko	*honiki*	*ɸai*	*saʔa*	*kenta*	*ɸuleke*
Bonerate	*honiki*	*βai*	*saʔa*	*kenta*	*βoleke*

	61. dog	62. tree	63. leaf	64. root	65. wood
Wanci	*ʔob'u*	*huʔu*	*roʔo*	*aka*	*kau ~ kaʔu*
Kaledupa	*ʔob'u*	*huʔu*	*roʔo*	*akka*	*kau*
Tomea	*ʔob'u*	*hɸuʔu*	*roʔo*	*akka*	*kau*
Binongko	*ʔob'u*	*hɸuʔu*	*roʔo*	*akka*	*kau*
Bonerate	*ʔobu*	*huʔu*	*roʔo*	*aka*	*kau*

	66. fruit	67. flower	68. thorn	69. banana	70. coconut (ripe)
Wanci	*b'aʔe*	*kamba*	*ruhi*	*loka*	*kaluku*
Kaledupa	*b'aʔe*	*kamba*	*ruhi*	*loka*	*kaluku*
Tomea	*b'aʔe*	*kamba*	*rihi*	*loka*	*kuʃou*
Binongko	*b'aʔe*	*kamba*	*ruhi*	*loka*	*kaʃuku*
Bonerate	*baʔe*	*kamba*	*rihu*	*loka*	*kaluku motika*

	71. coconut (unripe)	72. coconut shell	73. bamboo	74. sago palm	75. rattan
Wanci	*osimpu*	*b'enu*	*βemba*	*rumbia*	*ʔue*
Kaledupa	*simpu*	*b'enu*	*βemba*	*rumbia*	*ʔue*
Tomea	*niu*	*b'enu*	*ɸemba*	*rumbia*	*ue*
Binongko	*simpu*	*b'enu*	*ɸemba*	*rumbia*	*ue*
Bonerate	*kaluku moruɲaŋ*	*keʔua*	*βemba*	*rumbia*	*ue*

	76. betel	77 alang-alang	78. pandanus	79. seed	80. seedling
Wanci	*βeŋka*	*pad'a*	*pad'a*	*b'eni*	*βine*
Kaledupa	*βeŋka*	*pad'a*	*herommo*	*b'eni*	*pembulo*
Tomea	*βeŋka*	*heʔe*	*pad'a*	*b'eni*	-
Binongko	*ɸeŋka*	*pad'a*	-	*b'eni*	-
Bonerate	*βeŋka*	*pada*	*ponda*	*b'eni*	*amola*

	81. field rice	82. rice	83. cooked rice	84. sun	85. moon
Wanci	*b'ae*	*b'ae*	*b'ae*	*ʔoloo*	*komba*
Kaledupa	*b'ae*	*b'ae*	*b'ae*	*ʔoloo*	*komba*
Tomea	*b'ae*	*b'ae*	*b'ae*	*ʔoloo*	*komba*
Binongko	*b'ae*	*b'ae*	*b'ae*	*ʔoloo*	*komba*
Bonerate	*bae*	*bae*	*maɲa*	*ʔoloo*	*komba*

	86. star	87. sky	88. cloud	89. rain	90. wind
Wanci	*βetuʔo*	*laɲi*	*lono*	*βande*	*iri/ kaβea*
Kaledupa	*βituʔo*	*laɲi*	*lono*	*mond'a*	*βand'e*
Tomea	*ɸituʔo*	*laɲi*	*lono*	*monda*	*βande*
Binongko	*ɸituʔo*	*laɲi*	*lono*	*monda*	*βande*
Bonerate	*βituʔo*	*laɲi*	*lono*	*monda*	*βande*

	91. sea	92. sand	93. earth	94. salt	95. sugar
Wanci	*maβi*	*ʔone*	*βuta*	*gara*	*goʃa*
Kaledupa	*maβi*	*ʔone*	*βuta*	*gara*	*goʃa*
Tomea	*tai*	*ʔone*	*ɸuta*	*gara*	*goʃa*
Binongko	*tai*	*ʔone*	*ɸuta*	*gara*	*goʃa*
Bonerate	*tai*	*ʔone*	*βuta*	*gara*	*gola*

	96. water	97. spring	98. mountain	99. forest	100. river
Wanci	*uβe*	*mata nu uβe*	*βuŋka*	*motokaʔu*	*umala*
Kaledupa	*teʔe*	*mata nu teʔe*	*gunu*	*kotuʔa*	*umala*
Tomea	*teʔe*	*mata nu teʔe*	*gunu*	*kotuʔa*	*umaʟa*
Binongko	*teʔe*	*mata nu teʔe*	*gunu*	*laʟa aʟa*	*umaʟa*
Bonerate	*teʔe*	*mata nu teʔe*	*βuŋka*	*laro hobo*	*umala*

	101. fire	102. smoke	103. ashes	104. stone	105. canoe
Wanci	*ahu*	*koʔo*	*aβu*	*βatu*	*b'aŋka*
Kaledupa	*ahu*	*koʔo*	*koʔo*	*βatu*	*b'aŋka*
Tomea	*lukapi*	*gopo*	*gopo*	*ɸatu*	*b'oti*
Binongko	*lukapi*	*gopo*	*daʟika*	*ɸatu*	*b'aŋka*
Bonerate	*lukapi*	*gopo*	*βau*	*βatu*	*baŋka*

	106. mortar	107. pestle	108. knife	109. machete	110. rope
Wanci	*tumbuʔa*	*ana nu tumb.*	*pod'a*	*kab'ali*	*kalara*
Kaledupa	*tumbuʔa*	*ana nu tumb.*	*soka*	*kab'ali*	*kalara*
Tomea	*tumbuʔa*	*ana nu tumb.*	*soka*	*kab'aʟi*	*kaʟara*
Binongko	*ina nu tumb.*	*ana nu tumb.*	*soka*	*kab'aʟi*	*kaʟara*
Bonerate	*tumbua*	*ana nu tumb.*	*soka*	*kab'ali*	*hau*

	111. paddle	112. roof	113. road	114. big	115. small
Wanci	*b'ose*	*ato*	*sala*	*toʔoge*	*kiʔkiʔ*
Kaledupa	*b'ose*	*ato*	*sala*	*toʔoge*	*b'ahiti*
Tomea	*b'ose*	*ato*	*inteŋa*	*toʔoha*	*b'ahᶲuʟi*
Binongko	*b'ose*	*ato*	*inteŋa*	*toʔoha*	*b'ahoʟi*
Bonerate	*bose*	*helombo*	*inteŋa*	*toʔoha*	*paipai*

	116. good	117. dry	118. far	119. near	120. new
Wanci	*leama*	*motiti*	*melai*	*metaŋku*	*βoʔou*
Kaledupa	*leama*	*motiti*	*mellai*	*med'ani*	*βoʔou*
Tomea	*ndeu*	*karii*	*mellai*	*med'ani*	*ɸoʔou*
Binongko	*ndeu*	*motiti*	*mellai*	*med'ani*	*ɸoʔou*
Bonerate	*ndeu*	*motiti*	*melai*	*medani*	*βoʔou*

	121. old	122. aged	123. fat	124. hot	125. cold
Wanci	*moleŋo*	*mansuana*	*moʔobu*	*mena*	*keŋku*
Kaledupa	*moleŋo*	*mansuana*	*moʔoβu*	*mena*	*keŋku*
Tomea	*moʟeŋo*	*mansuana*	*monuβu*	*morara*	*keŋku*
Binongko	*moʟeŋo*	*mansuana*	*moʔobu*	*mena*	*keŋku*
Bonerate	*moleŋo*	*mansuana*	*moʔobu*	*mena*	*keŋku*

	126. short	127. long	128. blind	129. deaf	130. thirsty
Wanci	*med'umpu*	*melaŋka*	*kandala*	*kab'oŋo*	*motindoʔu*
Kaledupa	*med'umpu*	*melaŋka*	*kandʳala*	*kab'oŋo*	*motindʳoʔu*
Tomea	*med'umpu*	*melaŋka*	*kandaɭa*	*kab'oŋo*	*motindoʔu*
Binongko	*med'umpu*	*melaŋka*	*kandaɭa*	*kab'oŋo*	*motindoʔu*
Bonerate	*medumpu*	*melaŋka*	*kandala*	*kab'oŋo*	*motindoʔu*

	131. hungry	132. all	133. many	134. round	135. full
Wanci	*moʔaro*	*sab'aʔane*	*koruo*	*b'uloli*	*b'uke*
Kaledupa	*moʔaro*	*sab'aʔanne*	*koruo*	*b'uloli*	*b'uke*
Tomea	*moʔommuru*	*sab'aʔanne*	*koruo*	*b'uɭoɭi*	*b'uke*
Binongko	*moʔommuru*	*sab'aʔanne*	*koruo*	*b'uɭoɭi*	*b'uke*
Bonerate	*moʔomuru*	*b'aʔanu*	*koruo*	*bulili*	*buke*

	136. white	137. black	138. yellow	139. red	140. green
Wanci	*mohute*	*b'iru*	*mokuri*	*meha*	*ijo*
Kaledupa	*mohute*	*b'iru*	*mokuri*	*meha*	*ijo*
Tomea	*mohᶲute*	*b'iru*	*mokuri*	*meha*	*iijo*
Binongko	*mohᶲute*	*b'iru*	*mokuri*	*meha*	*iijo*
Bonerate	*mohute*	*biru*	*mokuri*	*meha*	*kudara*

	141. not	142. this	143. that	144. here	145. there
Wanci	*mbeaʔe*	*ana*	*aso*	*i ana*	*i atu*
Kaledupa	*umbea*	*ana*	*iso*	*ka ana*	*ka iso*
Tomea	*umea*	*ana*	*iso*	*anaʔe*	*isoʔe*
Binongko	*ʔumbed'a*	*ha ia*	*iso*	*ha iaʔe*	*haʔiso*
Bonerate	*mbeʔa*	*ʔia*	*atu*	*ma ʔia*	*ma ʔatu*

	146. yonder	147. one	148. two	149. three	150. four
Wanci	*i aso*	*saʔasa*	*d'ua*	*tolu*	*gana*
Kaledupa	*ka isoo*	*assa*	*d'ua*	*tolu*	*ganna*
Tomea	*ia ʔe*	*assa*	*d'ua*	*toɭu*	*paʔa*
Binongko	*ha ʔisoo*	*assa*	*d'ua*	*toɭu*	*paʔa*
Bonerate	*ma iso*	*asa*	*kedua*	*ketolu*	*kehaʔa*

	151. five	152. six	153. seven	154. eight	155. nine
Wanci	*lima*	*noʔo*	*pitu*	*alu*	*sia*
Kaledupa	*lima*	*noʔo*	*popitu*	*alu*	*sosia*
Tomea	*lima*	*kenoʔo*	*kehitu*	*akkeaɭu*	*akkesia*
Binongko	*lima*	*noʔo*	*pitu*	*aɭu*	*sia*
Bonerate	*kelima*	*kenoʔo*	*kehitu*	*kealu*	*kesia*

	156. ten	157. twenty	158. hundred	159. thousand	160. west
Wanci	*ompulu*	*d'uahulu*	*sahatu*	*sariβu*	*βaha*
Kaledupa	*ompulu*	*d'uahulu*	*sahatu*	*sariβu*	*βaha*
Tomea	*ompuɭu*	*ruaɸuɭu*	*asahatu*	*asariβu*	*βaha*
Binongko	*ompuɭu*	*d'uaɸuɭu*	*assahatu*	*assariβu*	*ɸaha*
Bonerate	*ompulu*	*duahulu*	*sahatu*	*saribu*	*βaha*

	161. east	162. under	163. above	164. behind	165. front
Wanci	*timu*	*d'i βoru*	*d'i βaβo*	*d'i taliku*	*d'i aropa*
Kaledupa	*timu*	*d'i βoru*	*d'i βaβo*	*c*	*d'i aro*
Tomea	*timu*	*d'i ɸoru*	*d'i ɸaɸo*	*d'i taɭiku*	*d'i muɭa*
Binongko	*timu*	*d'i ɸoru*	*d'i ɸaɸo*	*d'i taɭi*	*d'i aro*
Bonerate	*timu/ umbosa*	*i βoru*	*i βaβo*	*i taliku*	*i rope*

	166. outside	167. inside	168. edge	169. night	170. know
Wanci	*d'i luara*	*d'i laro*	*puria*	*morondo*	*d'ahani*
Kaledupa	*d'i liku*	*d'i laro*	*puria*	*uutu*	*motind'ou*
Tomea	*d'i luara*	*d'i laɭo*	*uju*	*uutu*	*d'ahani*
Binongko	*d'i luara*	*d'i laɭo*	*umbu*	*hϕuutu*	*d'ahani*
Bonerate	*i liku*	*i laro*	*sisi*	*utu*	*tahani*

	171. say	172. repeat	173. sing	174. cry	175. laugh
Wanci	*pogau*	*ulaɲiʔe*	*lagu*	*d'oito*	*koni*
Kaledupa	*bisara*	*heppuleʔe*	*hekad'aŋki*	*d'oito*	*koni*
Tomea	*bisara*	*heppuleʔe*	*hekad'aŋki*	*keʔe*	*koni*
Binongko	*bisara*	*heppuleʔe*	*hekad'aŋki*	*d'oito*	*koni*
Bonerate	*bisara*	*hempuleʔe*	*kad'aŋki*	*doito*	*koni*

	176. hear	177. see	178. eat	179. drink	180. bite
Wanci	*rod'oŋo*	*ʔita*	*maŋa*	*moroʔu*	*kaha*
Kaledupa	*rod'oŋo*	*ʔita*	*maŋa*	*moroʔu*	*kaha*
Tomea	*hikid'oŋo*	*ʔita*	*maŋa*	*moroʔu*	*kaha*
Binongko	*hikid'oŋo*	*ʔita*	*maŋa*	*moroʔu*	*kaha*
Bonerate	*hiŋkidoŋo*	*ʔita*	*maŋa*	*moroʔu*	*kaha*

	181. fall	182. drop	183. burn	184. pound	185. die
Wanci	*b'uti*	*nab'u*	*henunu*	*tumbu*	*mate*
Kaledupa	*b'uti*	*pab'utiʔe*	*rampu*	*tumbu*	*mate*
Tomea	*monab'u*	*nab'uʔe*	*tunu*	*tumbu*	*mate*
Binongko	*b'uti*	*b'utiʔe*	*tunu*	*tumbu*	*mate*
Bonerate	*morunu*	*nabuʔe*	*tunu*	*tumbu*	*mate*

	186. dry in the sun	187. bath	188. bathe s.o.	189. swim	190. fly
Wanci	*puaiʔe*	*hesoβui*	*hesoβuiʔe*	*naŋu*	*lola*
Kaledupa	*tanta*	*soβui*	*hesoβuiʔe*	*naŋu*	*lolla*
Tomea	*hommaine*	*herihᵠu*	*herihᵠuʔe*	*keaβa*	*loɭa*
Binongko	*hommaine*	*hesoɸui*	*hesohᵠuiʔe*	*naŋu*	*loɭa*
Bonerate	*homaine*	*susui*	*hesuβuiʔe*	*keaβa*	*lola*

	191. kill	192. give	193. cough	194. spit	195. vomit
Wanci	*hokomate*	*hᵠuʔu*	*more*	*hepaluti*	*henenasi*
Kaledupa	*hokomate*	*huʔu*	*kendʳesa*	*hepaluti*	*tokkumbo*
Tomea	*hokomate*	*ɸuʔu*	*more*	*hepaɭuti*	*tokkumbo*
Binongko	*mateakone*	*ɸuʔu*	*more*	*hepaɭuti*	*tokkumbo*
Bonerate	*hokomate*	*houʔu*	*more*	*kapera*	*tokkumbo*

	196. itch	197. walk	198. stand	199. sit	200. lie down
Wanci	*mohaʔato*	*βila*	*tad'e*	*ked'e*	*mbale*
Kaledupa	*mohato*	*βila*	*tad'e*	*ked'e*	*ŋgolo*
Tomea	*mohato*	*ɟila*	*tad'e*	*ked'e*	*mbaɭe*
Binongko	*mohato*	*ɟila*	*tad'e*	*ked'e*	*mbaɭe*
Bonerate	*mohato*	*βila*	*tade*	*kede*	*mbale*

	201. tired	202. sleep	203. dream	204. wake up	205. awaken
Wanci	*motuturu*	*moturu*	*homonihi*	*like*	*likeʔe*
Kaledupa	*motutturu*	*moturu*	*homonihi*	*b'aŋu*	*b'aŋune*
Tomea	*motutturu*	*moturu*	*homonihi*	*b'aŋu*	*b'aŋune*
Binongko	*motutturu*	*moturu*	*homonihi*	*b'aŋu*	*b'aŋune*
Bonerate	*paupau*	*moturu*	*homonihi*	*dahaniako*	*baŋune*

	206. come	207. pregnant	208. name	209. bride price	210. what
Wanci	*mai*	*hotokompo*	*ŋaa*	*kamondo*	*paira*
Kaledupa	*mai*	*hotokompo*	*ŋaa*	*kamondʳo*	*paira*
Tomea	*mai*	*hotokompo*	*ŋaa*	*kaɭasara*	*haira*
Binongko	*mai*	*hotokompo*	*ŋaa*	*masakaβi*	*haira*
Bonerate	*-*	*hotokompo*	*ŋaʔa*	*popolo*	*haira*

	211. who	212. where	213. how many	214. how	215. why
Wanci	*emai/ ieʔe*	*d'i ʔumpa*	*popia*	*βana ʔumpa*	*nohaʔa*
Kaledupa	*ieʔe*	*d'i ʔumpa*	*popia*	*βana ʔumpa*	*nohaʔa*
Tomea	*ieʔe*	*ima ʔumpa*	*akkehia*	*ɸana ʔumpa*	*nohaʔa*
Binongko	*ieʔe*	*ha ʔumpa*	*akkehia*	*ɸana ʔumpa*	*nohaʔa*
Bonerate	*iemai*	*ima ʔumpa*	*popia*	*βana ʔumpa*	*nohaʔa*

References

Anceaux, J.C., 1952, *The Wolio language. Koninklijk Instituut voor Taal-, Land en Volkenkunde* No. 11. Leiden: KITLV Press.

1978, The linguistic position of South-East Sulawesi: a preliminary outline. In S.A. Würm and L. Carrington, eds, *Papers from the Second International Conference on Austronesian Linguistic*: 275–282. Canberra: Pacific Linguistics.

Bhurhanuddin, B. H., 1979, Bahasa-bahasa daerah di Sulawesi Tenggara. Kendari, ms.

Broch, Harald B., 1981, Cultural variation on the islands in the Sea of Flores. In *Archipel* 22, 43–53.

Collins, J., 1983, Syntactic change in Ambonese Malay: the possessive construction. In James T. Collins, ed., *Studies in Malay dialects part 2: Malay dialects in Indonesia. NUSA* 17, 28–41.

Donohue, M., n.d., Field notes in the languages of Southeast Sulawesi.

Esser, S.J., 1938, Talen. In *Atlas van tropisch Nederland*. Amsterdam: Koninklijk Nederlandsch Aardrijkskundig Genootschap.

Friberg, T., n.d., Bonerate word list. Ujung Pandang: Summer Institute of Linguistics and Universitas Hasanuddin.

Grimes, Barbara Dix, 1991, The development and use of Ambonese Malay. In H. Steinhauer, ed., *Papers in Austronesian linguistics* No. 1, 83–123. Canberra: Pacific Linguistics.

Grimes, Charles E. and Barbara D. Grimes, 1987, *Languages of South Sulawesi*. Canberra: Pacific Linguistics.

Grimes, Joseph E., 1988, Correlations between vocabulary similarity and intelligibility. In *Notes on Linguistics* No. 41: 19–32.

Kaseng, Syahruddin, et al., 1987, *Pemetaan bahasa-bahasa di Sulawesi Tenggara*. Ujung Pandang: Departemen Pendidikan dan Kebudayaan.

Manyambeang, A. Kadir, et al., 1985, *Struktur bahasa Binongko*. Ujung Pandang: Proyek Penelitian Bahasa dan Sastera Indonesia.

Noorduyn, J., 1991, *A critical survey of studies on the languages of Sulawesi*. Koninklijk Instituut voor Taal-, Land en Volkenkunde Bibliographical Series 18. Leiden: KITLV Press.

Salzner, Richard, 1960, *Sprachenatlas des indopazifischen raumes*. Wiesbaden: Otto Harrassowitz.

Sneddon, J., 1987, 'Sulawesi'. In S.A.Wurm and Shiro Hattori, eds, *Language atlas of the Pacific area*, sheet 44. Canberra: Pacific Linguistics.

4 Defining speech communities on Buru Island: a look at both linguistic and non-linguistic factors

CHARLES E. GRIMES

1 Introduction

Buru is one of the largest islands in the eastern Indonesian province of Maluku.[1] The size of the island, all 9,800 square kilometers of it, is overwhelming when the primary means of transportation on the island is by foot. But size, distance, and rugged terrain, rising to 9,000 feet in elevation, are only some of the factors that make pinning down the language picture a challenge. On the one hand, both people indigenous to the island and the *Language atlas of the Pacific area* (Wurm & Hattori 1981–83) claim there is only one language on the island (in addition to Ambelau and Kayeli). On the other hand, the picture is complicated by language taboos, overlapping dialect names, old and recent migrations, lack of correlation between political entities and speech varieties, ignorance of other areas by most of the inhabitants, complex marriage alliance networks, hard-to-obtain origin myths, power tied to place, the politics of modernisation, the influx of immigrants from other parts of Maluku and elsewhere, and historical geo-political forces. All these factors work together to make the language picture, in fact, extremely elusive. It is furthermore intriguing as to why people all over the island of Buru insist that two speech varieties which show only 61%–64% lexical similarity are intelligible with each other and are the same language.

It should come as no surprise to those who are familiar with the realities of complex language situations that it has taken months of mobile residence spread over a period of years, and a degree of proficiency in several speech varieties along with an intensive study

of the grammar, lexicon and culture to begin to know what questions are significant and to be able to appropriately interpret the information received.[2]

A summary of current conclusions on the language picture is presented below with detailed discussion following. In Figure 1, capitals represent language names, lower case letters represent dialect names, and italics represent sub-dialects. Alternate names and alternate spellings are in parentheses. Li Garan is a special taboo register spoken by the Rana subdialect and will be discussed separately.

AMBELAU	-dialects unknown
BURU	
Masarete	
Wae Sama	
Rana	
Rana	
[Li Garan]	
Wae Geren	
Wae Kabo	
Wae Tina	
Lisela (Li Enyorot)	
Lisela (Licella)	
Tagalisa	
Wae Geren	
Leliali (Liliali)	
Kayeli (Wae Apo, Unit-unit, Mako)	
Fogi (Li Emteban)	
Fogi (Vogi, Bobo)	
Tomahu	-extinct
HUKUMINA (Bambaa)	
Bara	-extinct
Hukumina	-1 speaker in 1989
Palumata (Palamata, Balamata, Pala Mada)	-extinct
KAYELI (Cajeli, Gaeli)	
Leliali (Liliali, Marulat)	-extinct (as of March 1989)
Kayeli	-4 speakers in 1989
Moksela (Maksela, Opselan)	-extinct (1974)
Ilat	-extinct
Lumaete (Lumaiti, Lumaite, Lumara)	-extinct

Figure 1: Status of speech communities of Buru Island

---

2 It should further come as no surprise that this will not be the last update; the language picture on Buru is complex enough that it continues to warrant further investigation and testing.

2 Influences on language from history

When one strips away the superficial veneer of a hierarchical political structure that has a *raja* with intermediate-level heads ruling over ten clans in seventy villages giving the illusion of a state, or incipient state, one is brought up short with the reality that Buru society traditionally focused around local kin groups living in scattered houses or small clusters of houses. Survival is dependant on continuous hunting and foraging in the jungle (sometimes for 3-5 months at a time) combined with shifting agriculture focused around the cultivation of tuber crops and sago. Marriage is exogamous and is the basis for symmetric alliances between local kin groups (B.D. Grimes 1990). Traditional leadership is informal, focusing around the authority of elders within the kin group. Persons of equal rank in the kin structure have no authority over each other either within the kin group or across kin groups. Leadership is dependent upon both lineage and charisma — one without the other doesn't work. Headhunting and revenge killing (of people in other kin groups) were tied with territory, lack of alliances, and kin.[3]

Both oral history and the historical record combine to present a fascinating picture. Each kin group has an ancestral stream and many have an ancestral mountain. Each kin group is charged with guarding the spiritual power of their territory against all comers — whether from Buru or elsewhere. The *epkitan* 'fighting champion' of each kin group who were charged with guarding areas along the coast from raiding parties frequently inhabited caves in high cliffs overlooking the sea, from which place they would descend and attack any boats that landed. Stories of this era abound about *epkitan* attacking the *geba lano* who are variously interpreted in different parts of the Buru to be Papuan, from Seram, or from Tobelo and Galela (Halmahera, north Maluku).[4]

This traditional picture is obscured in the more accessible and acculturated parts of the island by external political hierarchies imposed first by Ternate, then the Dutch, and now the Indonesian government systems, and by the Islam and Christian religions.

2.1 Ternate and Islam

Against first the Portuguese in the 16th century and later the Dutch, Ternate used Islam as the unifying focus for resistance in the spice trade (van Fraasen 1983). By the middle 1600s most places that afforded favourable anchorages along the Buru coastline had small settlements of immigrants adhering to Islam. What seems to have escaped the notice of Ternate, and subsequently the Portuguese and the Dutch, is that with minor exceptions these coastal populations were settled by Muslim immigrants from other parts of the archipelago, most notably from Buton, Sula and Makassar and not by people indigenous to Buru.

It was the leadership of these non-Buru Muslim communities with whom Ternate, the Portuguese, and the Dutch interacted, punished and transported elsewhere, and tried to use to coerce the cooperation of the indigenous mountain population to little avail.

---

[3] All of the above activities (except headhunting) continue into the present.

[4] *Epkitan* is a loan word via Malay kapitan from Portuguese *capitão* or Spanish capitan 'ship captain, military officer'.

In the 1600s, the Sultan of Ternate, through his 'governors' on Buru had set up the position of four *Mat.gugul*[5] who were responsible to extract 'tribute' (*enati* lit.'that which is set down') from the populace on behalf of the Sultan. The notion of power, authority and decision-making being concentrated in a single individual rather than in a group of elders was a concept foreign to Buru culture.[6]

Thus the 'mountain people' (*geb.fuka*) or interior population of Buru were left almost untouched by Ternate, while the 'coastal people' (*geb.masi*) who were mostly not indigenous to Buru became involved on the periphery of Ternate's struggle with European powers over the spice trade.
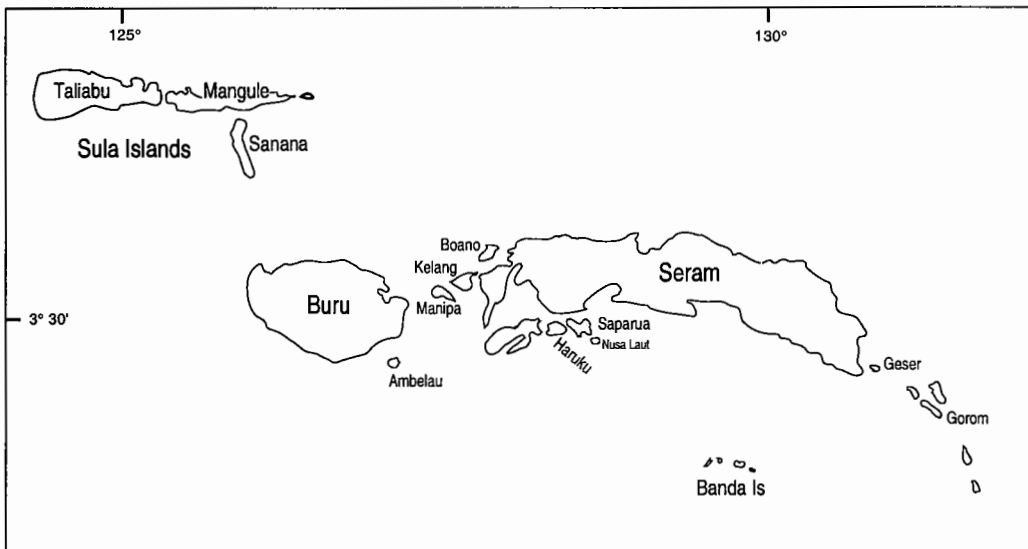
Map 1: Central Maluku, eastern Indonesia

Sometime between 1558 and 1650 the Sultan of Ternate set up at least two Ternatan governors on Buru, bringing Islam to the north and east coast as the organising force in the resistance to the Portuguese and later the Dutch (van Fraassen 1983:7). The Ternatan governors set up a hierarchical government structure to try and control both the coastal and

---

5   *Gugul* is a Ternate loan associated with the position of a ruler. The convention of a word-internal full stop indicates phonological elision in compounding. Thus *geb.haa* is reduced from /geba haa/, *Mat.gugul* reduced from /mate gugul/, etc. See C. Grimes 1991:69ff.).

6   Similarly, most positions of authority focused in an individual on Buru are traceable to Ternate, the Dutch, or the Indonesian political structures. Thus, not only Matgugul, but also Porwisi, Portelo, Raja, Kawasan, Emrimo, Kepala Kampung, are all outside terms and concepts.

interior populations and to exact tribute from them. The governorships organised clusters of kin groups under a single leader — a concept foreign to the traditional political structure.[7]

2.2 The consequences of Dutch economic policy

The demise of the Ternatan power structure and the escalation of Dutch power is described by van Fraassen (1983:17):

> The war ignited by the 1651 rebellion continued until 1656. The rebels received Macassarese support, and the war was in no way restricted to Hoamoal [west Seram]. Arnold de Vlamingh van Oudtshoorn, the Dutch commander-in-chief in this war, was also launching attacks on the VOC's opponents elsewhere, among other places in Buru and east Seram...The Ambonese region was formally removed from Ternatan control and the institution of the Ternatan governorship was abolished. Hoamoal...was completely depopulated...The chiefs of Hoamoal had a place of residence assigned to them in Batu Merah, in the vicinity of the VOC's chief fortress in Ambon. The population of the islands of Boano, Kelang and Ambelau was transferred to the island of Manipa, where the Company had a small fort. *All the Muslim chiefs of the coastal areas of Buru were obliged to settle in the neighbourhood of the Company's fort at Kayeli...The evacuated areas were systematically destroyed and rendered unfit for reoccupation.* Contracts were concluded with the chiefs of Buru and North and East Seram in which the latter conceded themselves to be subordinate to the VOC; promised to entertain no relations with other nations or rulers, to keep out all Macassarese, Malays, and Javanese...[emphasis mine, CEG].

The significance of the above event on the language picture is that the Dutch gathered 12 (according to oral history) hostage-puppets around them at the fort at Kayeli on the southern shore of Namlea bay. They are still referred to in Kayeli as the "12 Raja Patti" or the "12 Latu Patti". Each *raja* or *latu* set up his own village, his own mosque, wells, etc, and built up a community around him of people from his own area speaking whatever variety of speech was distinctive to their area.[8] Six of these mosque-village complexes were clustered west of the Kayeli river, including Kayeli, Lisela, Tagalisa, and Fogi. Six others were clustered on the east side of the Lumaiti river nearby, including Masarete, Hukumina, Lumaiti, and Palumata. The Dutch fort was in the middle, between the two rivers along with a 'Chinese Village' (*Kampung Cina*) and a 'Christian Kayeli' (*Kayeli Kristen*). Willer (1858:138) lists the additional villages in 1847 as Wae Sama, Marulat, Leliali and Tomahu. In the early 20th century the Dutch government, the Chinese, and the Christian village moved out of the swamp at Kayeli to a dry area across the bay which became the present-

---

[7] Abbreviations used in this paper:

6ICAL	Sixth International Conference on Austronesian Linguistics	PAn	Proto Austronesian
		s.t.	something
KITLV	Royal Netherlands Institute of Language and Culture	tr	transitive
		VCV	vowel, consonant, vowel
k.o.	kind of	VICAL	Fifth International Conference on Austronesian Linguistics
LEKNAS	National Co-ordinating Body for Economic Development (Indonesia)		
		VOC	Dutch

[8] The written record uses Ternate and Malay terms for the "chiefs" clustered around the fort at Kayeli. Willer (1858:138) lists a sengaji for Lumaete; raja for Kayeli, Lisela, Tagalisa and Leliali; patti for Marulat, Hukumina and Fogi; orang kaya for Wae Sama, Palamata and Masarete; and an orang tua for Tomahu. Rumphius (1910) says 13 leaders were removed to Kayeli, and includes Hukumina and Ilat, but not Fogi.

day town of Namlea. By the time of a detailed Dutch map in 1915, the two groupings of six villages had assimilated into just two villages – Masarete and Kayeli.

At the time of writing, the village of Kayeli is using the Lisela mosque (the Kayeli mosque being in ruins) and the locations of the former villages of Tagalisa and Fogi are still known. The village of Masarete stands seaward from the site of the former village of Hukumina whose foundations are still to be found hidden under water in a nearby sago swamp. Remnants of the inhabitants of the extinct villages are still identifiable by their kin group affiliations. And some of the older people still remember bits and pieces of the different speech varieties as either first language or second language speakers. Details of this are presented later.

3 Territorial divisions of Buru

A sometimes frustrating exercise is to try and make sense of the virtual plethora of names of different types of territorial areas, political entities such as "kingships" or clan alliance networks, village names, river names, river valley systems, distinct speech varieties, and so forth. For example, the term "Masarete" may variously refer to (a) a village in east Buru, (b) a "kingship" in south Buru, (c) an alliance of ten kin groups, (d) a territory defined by two river systems, (e) a distinct speech variety, and several variations on each of these. The picture is further clouded in that 1. the range defined by the political boundaries has changed many times over the course of several different administrations, and 2. there may be only vague correlation between the boundaries of the political, territorial, and linguistic uses of the same term, each being different in scope, and in some cases, location.

Various versions of origin myths divide the island into either two or three 'divisions' (*petak*, a Malay loan). The dual division variously names Masarete and Lisela, or Masarete and Kayeli, but basically identifies a north-south division from the north-west to the south-east. The three-way division variously names Masarete, Lisela and Wae Sama, or Masarete, Lisela and Kayeli.

3.1 Regentschap

The Dutch divided the island into 'regencies' (*regentschap*) for control of the population. The *raja/latu* was charged with being the ruling power's liaison with the populace, communicating edicts and collecting tribute (*enati*). At various times there have been anywhere from 4-15 *regentschap* (Malay *petuanan*). There is no indigenous term for this division. There are currently eight recognised regentschaps as shown in Map 2 below.

To understand the locations of speech varieties discussed later in the paper, the figure below presents the *Regentschaps* as they were in 1840. Maps obtained after that year no longer record the locations of the *Regentschaps* of Hukumina, Palamata, Tomahu, Marulat, Ilat, or Lumaete.
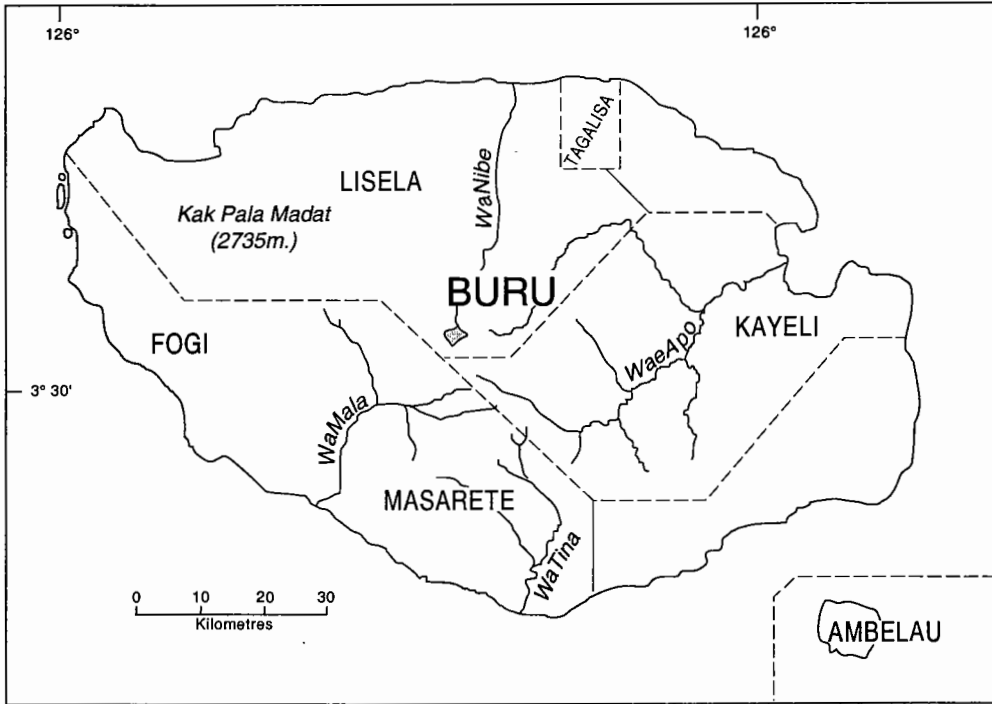
Map 2: *Regentschaps* still recognised on Buru Island in 1991
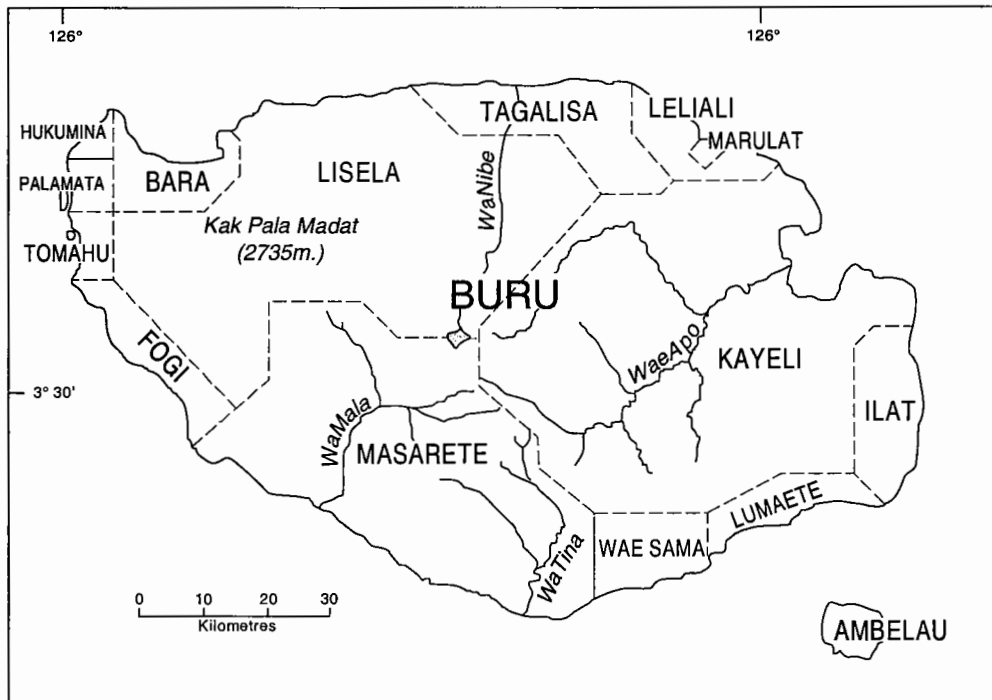
Map 3: *Regentschaps* on Buru Island in 1840

3.2 Kecamatan

The island is currently divided into three administrative districts (Indonesian *kecamatan*) as shown in Map 4 below. The division of the north into two separate administrative units took place in 1977-78 due to the logistical difficulty in administering the entire area from Namlea.

3.3 Traditional territorial divisions

General territories are referred to by people all over Buru in discussing the hunt, marriage alliances, cash crops, clan origins, and other topics. Such areas are territorially defined and are not politically or linguistically motivated. They are vague in their definition, not having finite boundaries, as illustrated in Map 5 below. Smaller divisions such as more localised names of areas of the jungle or names of sections of river systems are not shown.

4 Migrations and marriage alliances

To complicate the above picture in which there is only a vague correlation between the various names for territories and the scope to which the terms refer, one must also begin to grapple with migrations and marriage alliances on the island.

4.1 Migrations

Shifting agriculture depletes the soil after a period of years, forcing those working the soil to look for new fields (see Bellwood 1985 for a more detailed discussion). On Buru, the need to move on is not consciously associated with the depletion of soils and its effect on the tuber crops, but rather on the resulting rise in disease and subsequent in-fighting and death that cause a place to be declared a 'bad place' (*neten boho*) and be abandoned in favour of a new location where the spirits and the ancestors can again show favour.

Migrations may also be sparked by external pressures. For example, before the turn of the century large segments of several Masarete kin groups fled to the north-west part of the island to escape the Dutch-imposed forced labour when building the stone dock at Tifu. Part of the Nalbessy kin group migrated from Rana in the centre of the island around 1880–90 to the north coast to evade Dutch missionaries bringing the gospel from the south. Other migrations may be sparked by extreme violence or mass poisoning.

Migrations on Buru are commonplace, with people moving further into the mountains, moving to the coast, shifting a few kilometers upstream or downstream, or even moving from the north coast to the south coast or vice versa. A migration may be carried out as an entire village unit, but may also be a fragmentation or dispersion of a local group. In some cases distant migrations can be traced back 30 or more generations, but with kin group, affiliation, knowledge of ancestry and origins, ancestral river and linguistic identity still intact.
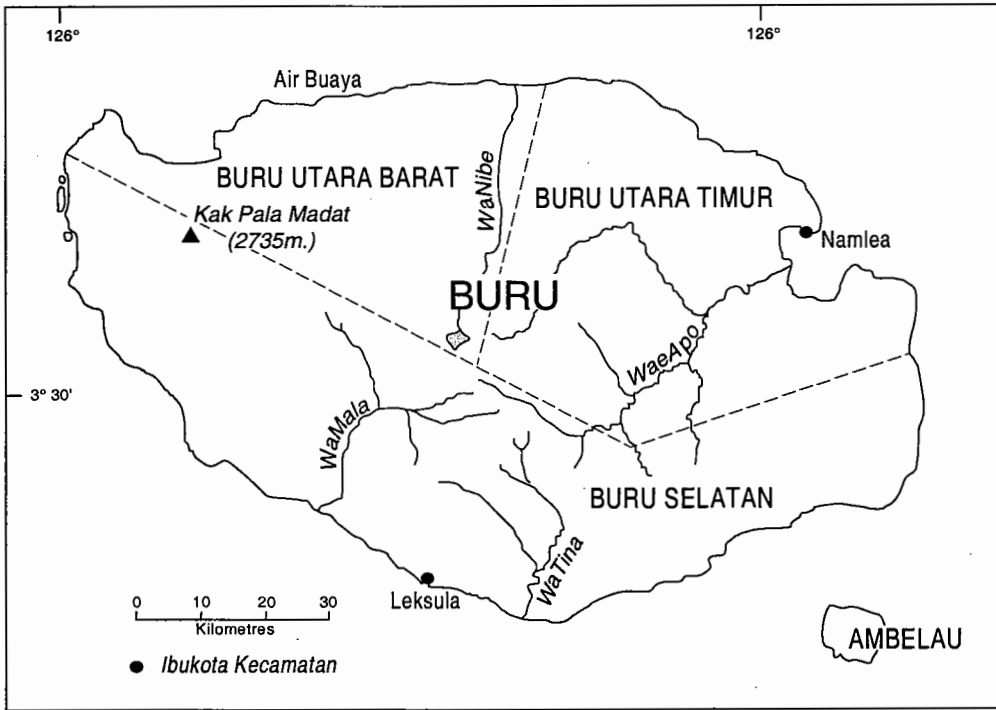
Map 4: *Kecamatans* on Buru Island
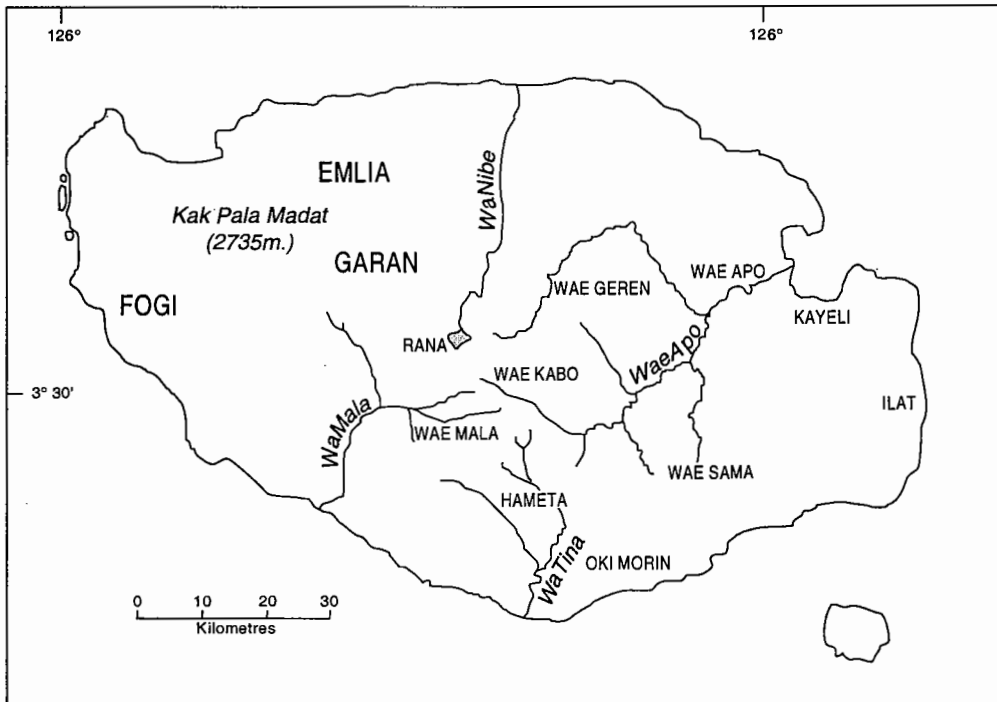
Map 5: Traditional territorial divisions on Buru

Thus, when trying to find out what the language picture is in any part of the island, it is essential to know what kin groups one is dealing with and where they "belong". In a short stretch along the north coast one can find a 100-year old migration from Rana, a 90-year old migration from the south coast, a 10-year old and continuing migration from Rana, a concentration of teachers and government officials who originate from the south coast, and a village of Sula immigrants with several hundred Buru people who belong on the north coast and can trace 26 generations there. Only the latter can give the speech variety that represents the area accurately. Furthermore, because of religion and other factors there is very little interaction between those who "belong" and those who have come more recently.

4.2 Marriage alliances

Marriage on Buru is exogamous. Symmetric marriage alliances produce a preponderance toward pairing of kin groups in any given locale (B.D. Grimes 1990). In any given clustering of houses (*hum.lolin*), there are found both the ideal pattern of marrying one's mother's brother's daughter (*emdaa*), and wives taken from many different kin groups, both near and far. Kin groups from within the same territorial/alliance group are preferred over those from other areas, but kin groups residing near the boundary with another area (see Map 5) marry freely both directions. Some marriages are arranged over distances of four to eight day's walk, from the centre of the island to the coast, or from the north coast to the south coast.

Until the bridewealth is fully paid, residence tends to be uxorilocal (with wife's parent's); after that residence is patrilocal. This means that one constantly finds adult men or women out of "their" area, and this must be taken into account when going to new areas to check out speech varieties. One may find eager informants who turn out to be from other areas.

5 Lexical similarity and how to interpret it

Figure 2 below presents the percentages of lexical similarity found between several speech communities on Buru. The word list used was a 205-item modified Swadesh 200 word list. Both the word lists and the tabulations represent an extended period of field study and are quite different from what would be found after a quick survey. The word lists were supplemented by a 4000-item Masarete lexicon, two 1300-item word lists from the eastern and western stretches of Li Enyorot, and a 700-item word list from Kayeli all compiled by the author. Only the items from this supplementary material that were related to the original 205-item word list were used for comparison, however. The Ambelau word list was taken by Edgar Travis in October 1986. The Wae Sama, Masarete and Rana lists have been adjusted to account for synonyms (see discussion below) because of the familiarity of the author in documenting these speech varieties. The others have not been adjusted, accounting for the big drop in percentage of lexical similarity between Erdapa (Rana) and Silewa (Li Enyorot).

Wae Neven (Wae Sama)													
90	Pelat Puun (Wae Sama)												
93	93	Wae Nama Olon (Wae Sama/Masarete)											
89	91	96	Wae Loo (Masarete)										
89	91	97	99	Fakal (Masarete)									
89	91	97	98	100	Wae Katin (Masarete)								
81	81	87	89	89	89	Kaktuan (Rana)							
79	79	85	88	87	88	94	Erdapa (Rana)						
64	65	70	71	69	70	77	79	Silewa (Lisela)					
63	63	68	69	69	69	77	79	89	Wamlana (Lisela)				
61	61	65	66	66	66	73	74	88	92	Wae Poti (Lisela)			
46	46	48	48	48	49	51	49	52	53	54	Jiku Merasa (KAYELI)		
41	41	44	44	45	45	47	46	50	51	50	77	Kayeli (KAYELI)	
40	42	43	43	44	44	46	45	41	44	43	52	50	Ulima-AMBELAU

Figure 2: Percentages of lexical similarity on Buru

Wae Neven and Pelat Puun are both Wae Sama dialect. Wae Nama Olon interacts vigorously with both the Wae Sama dialect and the Masarete dialect, and this is also reflected in the percentage figures. Wae Loo, Fakal, Wae Katin represent the Masarete dialect. Kaktuan and Erdapa are from the Rana subdialect of the Central dialect. Silewa, Wamlana and Wae Poti represent the Lisela dialect. Jiku Merasa is the Leliali dialect of the Kayeli language.
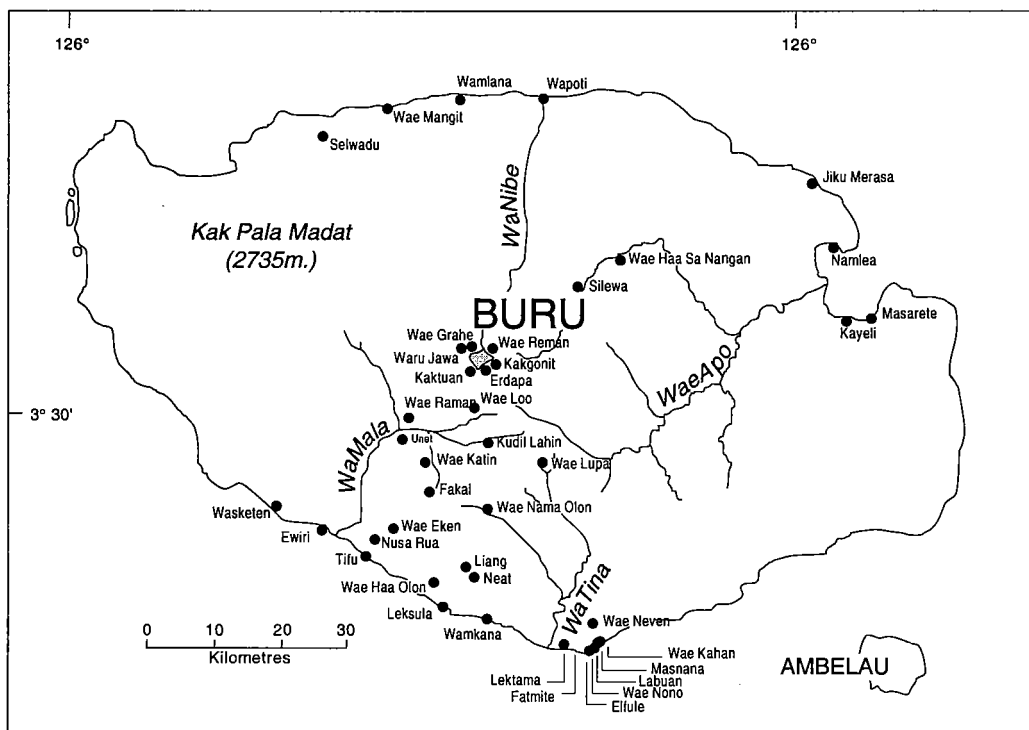
Map 6: Locations referred to in this paper

The percentages in the above figure of lexical similarity have been averaged in the figure below to highlight the statistically significant groupings. Note the pattern of dialect chaining (cf. Simons 1977) between Wae Sama, Masarete, Rana, and Lisela.[9]

Wae Sama					
90	Masarete				
80	88	Rana			
63	68	76	Lisela		
44	46	48	52	Kayeli	
41	44	46	43	51	Ambelau

Figure 3: Reduced matrix of major groupings

While it is generally recognised that thresholds between language and dialect, group and subgroup, need to be calibrated anew for each area and study, a significant point of this paper is that it argues for classifying Lisela as belonging to the same language as Rana, Masarete and Wae Sama even though the percentages of lexical similarity go down to 63%.

While it is generally accepted that the threshold between language and dialect (by whatever definition) falls somewhere in the neighbourhood of 70%–81% cognate, it has also been documented (J. Grimes 1988) that some cases over 90% probable cognate must be considered separate languages (when considering such factors as functors, intelligibility and socio-political identity), while other cases as low as 60% must be considered the same language. J. Grimes (1988), considering built-in methodological problems with lexicostatistics and looking at empirical data, suggests anything falling into the range of 60%–90% is in a grey area that needs further checking with other methods.[10]

Lexicostatistics as a method is only precise enough to tell us that two speech varieties are 1. clearly the same, 2. clearly different, or 3. unclear, such that other factors and other methods of testing must be taken into consideration to interpret the numbers. From the two figures of lexical similarity above, the only real unclear case is how to classify Lisela, as a dialect of Buru or as a separate language.

5.1 The character of the lexical differences

Differences between the dialects of what I am calling "Buru" are primarily lexical. However, the lexical differences are not restricted to obscure low-frequency items, but are also found on what are normally considered basic vocabulary, everyday cultural items, and even on a few functors. In the figure below, blanks indicate the form is uncertain for that dialect. Where alternate forms are indicated, both are in common usage. The list is not exhaustive, but gives an indication of the range of parts of speech and vocabulary that is *different*. The majority of forms not listed are identical or similar for all dialects.

---

[9] Grimes and Grimes (1994) note that Buru relates to the languages of Sula in the range of 27-33% lexical similarity. The statistically significant gap between Sula and Buru versus Buru and Kayeli and Ambelau suggests quite a different picture of over-all language relationships from that presented by Collins (1981) and Wurm & Hattori (1981), in which Buru and Sula are grouped together, while Kayeli and Ambelau are said to relate more closely to the languages of west Seram. However, the different conclusions were arrived at by different methods.

[10] J. Grimes also suggests (pers. comm.) that where language taboos are involved, 50% may be a more realistic lower threshold.

GLOSS	MASARETE	RANA	LISELA	WAE SAMA
sand	*ena*	*ena/sea*	*sea*	*ena*
mountain	*kaku/fuka*	*kaku*	*fude*	*fuka*
island	*fuka*	*fuka*	*bual*	*pulo*
root	*lahin*	*lahin*	*laden*	*lahin*
ocean	*masi*	*olat*	*olat*	*masi*
fire	*bana*	*bana*	*bana*	*gemat*
dust	*kahit*	*kahit*	*kahit*	*lafo*
wave	*emhein*	*ahut*	*wae fatan*	
rock	*fatu*	*fatu/miat*	*miat*	*fatu*
hole	*fefan*	*fefan*	*lefat*	
sun	*lea*	*lea*	*hangat*	*lea*
star	*tolot*	*tolot*	*gae*	*tolot*
today	*lea naa*	*lea naa*	*pahi naa*	
big	*haat*	*haat*	*bagut*	*haat*
straight	*dofo*	*dofo*	*tolon*	
all	*hansiak*	*hansiak*	*haluuk*	*hansiak*
many	*edemen*	*edemen*	*waro*	*edemen*
quickly	*rabo*	*rabo*	*spati*	
fast	*friken*	*friken*	*spati*	
hard	*giwe*	*giwe*	*meit*	
hot	*lepoton*	*lepoton*	*himdunun*	*erarat*
black	*mitet*	*mitet/medet*	*medet*	*medet*
white	*botit*	*botit/gawat*	*gawat*	*botit*
husband	*gebhaa*	*gebhaa/namorit*	*mori*	*gebhaa*
wife	*finhaa*	*finhaa/gefina*	*bilea/gefina*	*finhaa*
companion	*roko*	*roko/tawe*	*tawe*	*pemnaken*
bird	*manut*	*manut*	*pani*	*manut*
cat	*mau*	*mau*	*sika*	
marsupial	*tonal*	*tonal/blafen*	*blafen*	*lemet*
civet cat	*kamyoho*	*yoho*	*ngewaet*	
louse	*koto*	*koto*	*yemin*	*koto*
mosquito	*inhadat*	*senget*	*mimaun*	*senget*
firefly	*bana lafun*	*bana lafun*	*brama*	*ebrama*
crocodile	*emhalat*	*emhalat*	*ubaa*	
skin	*okon*	*kolin*	*usan*	*okon*
mouth	*fifin*	*fifin*	*muen*	*fifin*
lips	*fifin okon*	*fifin kolin*	*biban*	*fifin okon*
shoulder	*malan*	*malan*	*fesan*	
abdomen	*fukan*	*fukan*	*tian*	*fukan*
knee	*enolon*	*enolon*	*tonen*	*enolon*
breast	*soson*	*soson*	*hono*	*soson*
wing	*panin*	*panin*	*ahin*	
fat	*minan*	*minan*	*niwae*	*emnobo*
yawn	*duba*	*duba*	*psusun*	*duba*
sleep	*bage*	*bage/ine*	*ine*	*bage*
awake	*hosak*	*hosak/fango*	*fango*	*hosak*
urinate	*stefo*	*stefo*	*thei*	*thefo*

GLOSS	MASARETE	RANA	LISELA	WAE SAMA
defecate	*stei*	*thei*	*buu*	
kill	*epmata*	*epmata*	*mdau*	*epmata*
see	*kita*	*kita*	*tine*	*kita*
swim	*uka*	*nango*	*nango*	*uka*
make, do	*puna*	*puna/loa*	*loa*	
slip	*spisa*	*prisa*	*hori*	
follow	*hai*	*hai*	*taga*	
enter	*rogo/oso*	*rogo/oso*	*sii*	*oso*
give	*tuke/duwe*	*tuke/huke*	*huke*	*tuke/duwe*
request	*laha*	*laha*	*sea*	
forbid	*erei*	*erei*	*heke*	
play	*midi*	*midi*	*eptomi*	
speak	*prepa*	*prepa*	*bina*	*prepa*
hear	*caan*	*caan/prenge*	*prenge*	*caan*
remember	*nanbeta*	*nanbeta*	*halaik*	
hunger	*glada*	*emlapa*	*emlapa*	*emlapa*
burn(tr)	*pefa*	*pefa*	*sigi*	*pefa*
knife	*katanan*	*irit*	*irit*	*katanan*
machete	*katuen*	*todo*	*todo*	*todo/fae*
cloth	*wagun*	*wagun/ate*	*ate*	
door	*subu*	*sufen*	*sufen/karen*	
thatch	*atet*	*atet/abat*	*abat*	
citrus	*puhat*	*puhat*	*hosi*	
cassava	*mangkau*	[*kasbiit*]	[*kasbiit*]	
yam	*mangat*	[*obiit*]	[*obiit*]	
nettle	*pelat*	*pelat*	*saha*	
outside	*kako*	*kako/gidan*	*gawan*	*gidan*
Preposition	*gam/fi*	*fa/fi*	*fa*	*gam/fi*
up	*fi saka*	*fa saka*	*fa rete*	
distal	*fi dii*	*fa dii*	*fnii*	*fi dii*

Figure 4: Differences in vocabulary among Buru dialects

5.2 Historical-comparative evidence

Sound correspondences from Proto Austronesian (PAn) are uniform throughout the Buru dialects, with two exceptions noted below. Other non-lexical, non-phonological differences are that the Rana dialect uses a pre-posed, rather than a post-posed genitive marker in the vocative (contact-induced influence from Sula), and the Masarete dialect shows a collapse of the entire genitive system to the third person singular (C. Grimes 1991).

The first exception to uniform sound correspondences throughout the dialects of Buru is with the back vowels /u/ and /o/. Historically both /u/ and /o/ derive from PAn *u,* thus reflecting a split.

PAn	BURU	GLOSS
*inum	*ino*	drink
*ma-takut	*em-tako*	afraid
*uliq	*oli*	return home, go back
*suqan	*sua-n*	dibble stick
*qaniCu	*nitu*	*ghost, spirit of dead//dead person
*tuak	*tua-t*	*palmwine/tree yielding palmwine

Figure 5: Split of PAn *u* to Buru /u/ and /o/

However, when both vowels of a disyllabic root are back vowels and there is a consonant between them (VCV), a process of vowel harmony comes into play such that in Wae Sama, Masarete, and the dialects both vowels become /o/, whereas in Lisela both vowels appear as /u/. [NOTE: Buru (-L) means all the other dialects of Buru minus Lisela].

PAn	*tuRun* 'descend'	*qulu* 'head'	*pusuq* 'heart'	'hot (water)'
Buru (-L)	*toho*	*olo-n*	*poso-n fua-n*	*(wae) poto-t*
Lisela	*tuhu*	*ulu-n*	*pusu-n fua-n*	*(wae) putu-t*

Figure 6: Vowel harmony of back vowels

The second exception to uniform sound correspondences relates to PAn *p*.

PAn	*p*	*b*	*R*	*t*
Buru (-L.)	p	f	h	t
Lisela	p/h	f	h	t
Kayeli	h	b	l	t
Ambelau	f	b	h	r

Figure 7: Reflexes of PAn *p* and other selected correspondences

Li Enyorot does not show any identifiable patterns as to which items reflect /p/ and which reflect /h/. Instead of positing a historical split, those items reflecting /h/ may be easily accounted for as borrowings from Kayeli. And in fact, the items which reflect /h/ are identical in the Kayeli language.

PAn/PMP	*pusuq* 'heart'	*S-in-ipi* 'dream'	*pulut* 'birdlime'	
Buru (-L)	*poso-n*	*em-nipi*	*polo-n*	
Lisela	*pusu-n*	*em-nipi*	*pulu-n*	
Kayeli	*hoso-ni*	*em-nihi*	*??*	
PAn/PMP	*paRi* 'stingray'	*peñu* 'turtle'	*qapuR* 'lime, chalk'	*pajey* 'rice'
Buru (-L)	*pahi*	*peno*	*apu*	*pala*
Lisela	*hali*	*heno*	*ahul*	*hala*
Kayeli	*hali*	*heno*	*ahul*	*hala*

Figure 8: Items reflecting PAn *p* in Buru dialects

The Kayeli language was spoken along the north-eastern coastal strip of Buru in an area that is subsumed within the Lisela-speaking area. Lisela (in this broader context usually referred to as Li Enyorot) is best characterised as a *lingua franca* that developed along the

northern and eastern coastal strips for communication between speakers of Kayeli and other native inhabitants of Buru, and between Sula, Ternate, Buton and other migrants who settled along the north coast of Buru with native Buru inhabitants who had also settled along the coast. Thus, the presence of borrowings from Kayeli (or Sula, Ternate and Buton, for that matter) is not surprising.

In summary, the inherent sound correspondences of Lisela match the sound correspondences of the other dialects of Buru, but not those of Kayeli or Ambelau.

5.3 Figures deflated from taboos

It is well known that the impression of linguistic convergence can occur as the result of culture contact. Among the Sama-Bajau languages in the southern Philippines, Pallesen (1985) found that to obtain an appropriate indicator of linguistic similarity, scores of lexical similarity needed to be adjusted downward to weed out the great degree of convergence that resulted from culture contact — an adjustment which could only be done on an item-by-item basis.

A similar thing seems to be happening on Buru, but in the opposite way. Lexical taboos results in scores of lexical similarity that are lower than appropriate. As Pallesen could demonstrate borrowing (convergence) on a word by word basis, it can also be demonstrated on a word by word basis that taboos artificially deflate scores.

Simons (1982) presents a survey of tabooing practices found in the Austronesian world and their implications for comparative study. He notes (1982:162):

> Word tabooing requires the speakers of a language to have at their disposal at least two ways of saying certain things. When all the speakers of a language control synonymous forms for the same wordlist item and the choice between them is culturally rather than semantically determined, then a lexicostatistic method which records and compares only one response for each wordlist item can grossly miscalculate the true lexical relationship between languages. *The error will be in the direction of yielding cognate percentages that are lower than the actual reality....this has the effect of overestimating linguistic divergence.* [emphasis mine, CEG]

Simons goes on to present a study of Malaitan (Solomons) speech communities giving three different figures, 1. a standard lexicostatistic score, 2. the score adjusted to account for synonyms, and 3. the score further adjusted to include cognates that have a shift in meaning. The overall effect of these adjustments to the percentage figures was to increase them by an average of 9.3% when accounting for synonyms, and an additional 8.3% when accounting for cognates with meaning change. Some individual cases were adjusted upward by as much as 27%.

On Buru there are many types of taboos. Some speakers are consciously aware that taboos (*koit*) are associated with language variation.

Affinal taboos are both behavioural and linguistic, preventing one from uttering the name of parents-in-law, children-in-law, and siblings-in-law. One may refer to the person by kin term, but not by name (such a phenomenon is actually the norm for all adult relationships, but is further associated with a lowering of the eyes and voice, among other ways of avoidance, for in-laws). In the more traditional areas where personal names continue to be the names of plants, animals, places or seasons, the taboo extends to avoiding mentioning that particular item.

There is no taboo associated with the names of dead ancestors, as is found in other areas. On the contrary, ancestors (even recent ones) are frequently addressed directly by name in various circumstances and locations.

The most common kind of taboos on Buru are territorially restricted. In a given locale while hunting, travelling or just living, there are certain words that may not be uttered. In the Wae Lupa area of Wae Tina, one may not say *menjangan* 'deer' (a borrowed substitute itself assimilated by all dialects), but one must hunt *wadun* 'deer' (normally means 'back of neck'). In many of the streams along the slopes facing the southern coast one hunts *uran* 'crayfish' (PAn *uDang*), but in most interior areas of the island one must hunt *sehe* 'crayfish' (normally means 'to reverse, retreat, back up'). Examples of this sort are numerous, but a thorough listing is beyond the scope of this paper. In some places a topic may not be addressed at all in any language. In others cases, word substitutes or circumlocutions are listed for the newcomer before entering the taboo area (*net.koit*). Most of these territiorially restricted taboos have an associated myth or legend explaining why those particular items may not be used in that particular area.

Some taboos are not so much associated with a territory as much as an activity. For example, during particularly heavy east monsoons when groups of men will go hunting and foraging in the jungle for 3–5 months, there are special behavioural and linguistic taboos that must be strictly adhered to.

The taboos are not seen so much as a deception of the spirit world, but rather as a cooperative effort with the ancestors to assure themselves and their family and descendents long life and good fortune by avoiding the harm, disease and ill fate associated with breaking taboos. Failure to observe the taboos may result in a sudden and violent deterioration in the weather resulting in branches blowing down, or roofs blowing off, hurting or killing someone. Crops may be destroyed, or ill-health, miscarriage or still-birth may occur. A temporary failure to 'get s.t., obtain s.t., succeed, receive blessing' (*dufa*) on the hunt, in the fields, or while foraging can also be the result of failure to observe taboos.

Word taboos on Buru are formed by substituting a new item for the one to be avoided, rather than by altering the original form.

One type of substitution is by semantic shift. For example, while other dialects of Buru say *manut* 'bird' [PAn *manuk* 'bird'] and *pani-n* 'wing' [PAn *panij* 'wing'], Lisela says *pani* 'bird'. When talking about species of birds in different areas, the genera are different but the differentia are the same – *man.samal/pan.samal* 'seagull', *man.kumul/pani kumul* 'k.o. large dove'.

Similarly, a term may be retained with a shift in meaning while the original meaning is taken by a completely different item. Lisela has *nango* 'swim' [PAn *nanguy* 'swim']. In other dialects of Buru 'swim' is *uka* while *nango* is retained meaning 'wade'.

Another common strategy on Buru is to use circumlocutions or abstractions as a substitute for a lexical item. For example, *innewet* 'snake (generic)' literally means 'the living thing'; *isaleu* 'python' literally means 'a thing that goes ahead'. Lisela *senget* 'mosquito' [PAn *senget* 'sting'] is replaced by Masarete *inhadat* 'mosquito' (lit. 'the thing that bites'). The lexicon is full of examples of this sort.

5.4 Li Garan

The extreme case of taboo is found in an uninhabited section of the jungle in the north-west quadrant of the island called Garan. Garan takes two days of walking to traverse. The

taboo is that the vernacular language of Buru may not be spoken there. Malay (or any other language) is acceptable, but Buru is not. Li Garan 'the language of Garan' has developed entirely around a strategy of taboo. It is only known by those Rana people who have reason to travel to the north-west coast. They teach it to their children, and different levels of proficiency are recognised. A secondary use of the language is as a secret language in the presence of people from other parts of Buru.

On the basis of some texts and around 400 lexical items collected by the author, a few generalisations can be made. The syntax of Li Garan is the same as Buru. Nouns, verbs and 'adjectives' are different, but functors such as pronouns, prepositions, conjunctions, aspect markers and 'adverbs' are the same. Vocabulary is adequate for most domains, and there is even vulgar vocabulary. The nature and vocabulary of Li Garan seems to be fairly stable when compared with a general description by Schut (1919) and a word list from Jansen (1933). A fuller description of Li Garan is found in C. Grimes and Maryott (1994) and in C. Grimes (1991).

5.5 Figures deflated due to peculiarities of locale

Another phenomenon which gives the illusion of divergence in the figures of lexical similarity is that different areas of the island reflect features that are specific to the locale in which the word list is elicited.

For example, the geological forces which have shaped Buru have caused a great southern uplift with a northern tilt (see Bellwood 1985:3-8). The result of this process is that mountains along the southern coast in a band from 15-40 km inland are sharp, shattered coral protrusions that have been thrust up from the ocean bottom to heights of 1600+ metres with very little soil. In the northern part of the island, however, the mountains are great piles of soil, complete with landslides and silty streams (rare in the south). Thus, when one asks in the north what is the word for "mountain", one is given *fude*. In the south one is told *fuka*. It is only after extensive experience on the trails over a wide part of the island that it becomes clear that *fude* means 'the kind of mountain that is made out of soil' and *fuka* means 'island, with an extension to mean mountain'. *Fude* is also used in the south in the rare place where one can find a hill or mountain made of soil rather than coral. Incidently, both north and south use *kaku* to mean 'ridge, mountain'.

Similarly, when one asks for the word for "grass", one sometimes gets *mehet* and sometimes *rei*. Further study reveals that *mehet* is the saw-toothed variety and *rei* is the smooth-edged variety, given according to which variety is dominant in the immediate area.

There are additional examples of this nature giving the illusion that elicited word lists are in fact more divergent than they really are.

5.6 Figures deflated due to ambiguous generic-specific relations

Word list items tend to be generic. The problem for tabulation comes in that the responses given may be either generic or specific.

For example, when asking for "snake" on Buru one is normally given *isaleu* 'reticulating python', because that is the snake most salient, i.e. the one most talked about and feared. Every once in a while, in a word list situation, one may be given *innewet* 'snake (generic)', or *karapapa* 'k.o. small snake'. Similar problems occur for other nouns, such as generic and

specifics for types of bamboo. But it isn't until the researcher gains proficiency and broad experience in the language(s) in question that problems of this sort can be weeded out.

A more difficult problem to sort out is for some areas of action that may be lexically rich in the target language. For example, I have catalogued over thirty legitimate lexical responses in Buru when asked for "cut", and seventeen for "carry", depending on manner, instrument, goal, etc. (listed in C. Grimes 1991).

6 Oral reports and communication

The above discussion suggests that the figures of lexical similarity between Lisela and the other dialects of Buru are artificially low. But is there corroborating evidence?

One finds on Buru both a great majority who are completely ignorant of other parts of the island, and a minority who have, through marriage alliances, migrations, pursuit of cash crops, chasing down a runaway wife, or whatever reason, become familiar enough with some or several other parts of the island enough to be conversant in the names of streams and taboo areas, and know what kin groups "belong" in those areas. Both the ignorant majority and the informed minority vigorously insist that Buru has one language and only one.

Everyone on Buru is aware of some commonly known lexical differences, but the emphasis is on the similarity. On the one hand, Wae Sama, Masarete, Rana, Lisela, and Fogi can be identified by various people (the informed minority) as having speech varieties different enough to be identifiable, while insisting that they can communicate well. Some are aware enough to be able to say that Rana, Masarete and Wae Sama are closer to each other than they are to Lisela/Li Enyorot (a claim supported by the figures of lexical similarity). On the other hand, those from the Masarete and Wae Tina area who have actually tried to communicate with speakers of Lisela also admit that they must sometimes switch to Malay so as to avoid miscommunication. On the basis of intelligibility, therefore, this indicates that Lisela is marginal as to whether it should be considered a dialect, or a separate language (also corroborated by the figures of lexical similarity).

Women who have been purchased (their term) from other areas report a period of a few days to a few weeks to adjust their speech when they arrive in their husband's area.

Being aware of lexical similarity figures around 60% – 65%, and being aware of some of the great differences in basic vocabulary (see Figure 4), initially I remained sceptical at the insistence by speakers of Buru that there is just one language on the island.

However, after studying the Masarete and Rana dialects and gaining a degree of proficiency in them, I found opportunity to then return to the Lisela area. I spent two full days in Wae Mangit in north-west Buru talking with a man from the Waernangan kin group (that belongs there on the coast and is a native speaker of the Lisela/Li Enyorot speech variety). He could trace a direct line of his ancestors who had been in the immediate area through 26 generations. We talked in the vernacular about history, marriage, alliance, brideprice, travel, food, knowledge, research, and the outside world. We had to stop every once in a while to ask each other what some words meant, but communication at a significant level was both possible and enjoyable. I had a similar experience in Jiku Merasa in north-east Buru with a man from the Toraha kin group (which belongs there and also is a native speaker of Lisela/Li Enyorot) whose ancestors had been in that area through 14 generations. In both cases, both I and the Lisela speakers adjusted our speech toward what we knew of the other's speech variety.

Simons' (1982) discussion about taboo is built around the assumption that all speakers know what all the synonyms are. Such a uniform knowledge is clearly not the case in Buru.

This leaves us with the knowledge that Lisela is marginal, and the uncertainty as to whether the ability to communicate is more a reflection of inherent intelligibility or of bilingualism. Clearly, more rigorous testing a la Casad (1974) is still required with an adequate cross-section of Buru society.

7 Language and demographic information

In the discussion below, one confusion that must be sorted through is the repeated appearance of the same names (e.g. Lisela, Leliali and Kayeli) under different languages. This is a necessary confusion for several reasons, 1. to keep the names tied to the names used in the literature, 2. to keep the names tied in with the names used by the people themselves, and 3. because in these areas there were two different ethnolinguistic groups interacting with each other and residing in the same area.

7.1 The non-indigenous populations of Buru

The non-indigenous populations of Buru may be summarised as follows:

a. The northern coastal strip of Buru from Namlea to Bara is predominantly Muslim and is settled primarily by immigrants from Sula (Sanana). Many of these have been on Buru ten generations or more. They are the majority in most villages in this area and have the ethnic numbers to maintain their own language and not learn the vernacular language of Buru. (see Grimes & Grimes 1994 for a summary of the language picture in the Sula Islands).

b. Interspersed among the Sula villages are small numbers of individuals who trace their origins to Ternate, Galela (both north Maluku), Bugis (generic for several groups from South Sulawesi), Buton (a generic term for the islands off of South-east Sulawesi), and merchants generally calling themselves "Arab". These immigrants have banded together with the ethnic Buru people (who are also a minority group in the Sula communities) and many of them have learned to speak Lisela/Li Enyorot.

c. The west coast is inhabited almost exclusively by *geb.Binongko*, the Buru term for those from the Muna-Buton area off of south-east Sulawesi, most particularly from the Tukang Besi islands (See Donohue, this volume). This area provides ideal shelter for the Butonese sailing vessels plying their trade between Java and Irian Jaya. As many as 200 of these boats have been known to cluster in the area and then convoy to Java together. Many of these communities have been there for over 12 generations. With the recent exception of those who have migrated from other parts of Buru, such as Rana, and those who have made a complete shift to Malay, the west coast is devoid of indigineous inhabitants.

d. The south coast consists of Buru villages interspersed with several Buton villages and a few villages and village segments from "Kisar" from the south-western part of Maluku near Timor. Where these immigrants are minorities within a larger Buru community they have learned the Buru language.

e. The lower Wae Apo river valley in north-east Buru is an ethnolinguistic checkerboard. There is a large Bugis village at the mouth of the Wae Apo river entering the bay. Interspersed between the various Javanese transmigration units in the area are villages of indigenous Buru inhabitants whose kin groups "belong" in the area and indigenous Buru inhabitants who have migrated from the Masarete area since World War II to seek their fortune distilling rubbing oil (*gelan*) from the *Melaleuca kajuputi* plant. There is beginning to be some marriage between Buru people and Javanese, usually Buru men taking Javanese wives.

7.2 Languages of wider communication

Indonesian, the national language, is the language of government, education, and the media. Ambonese Malay (Collins 1980; C. Grimes 1985; B.D. Grimes 1991) is the language of commerce and inter-ethnic communication along the coastal areas of Buru. The Malay of much of north Buru is shifting from North Moluccan Malay (Taylor 1983; Voorhoeve 1983) to Ambonese Malay with more and more children sent to Ambon for schooling, and with more and more teachers and government officials working in the area coming from Ambonese Malay speaking areas.

7.3 Population estimates

Estimates of number of people belonging to each ethno-linguistic grouping are based on 1987 government statistics (Kantor Statistik Kabupaten Maluku Tengah 1987a, 1987b, 1987c) combined with the author's and informants' knowledge of ethnic composition of coastal villages and field notes on numbers of houses in many hamlets in the interior of the island not included in the government figures. The breakdown according to religion highlights the distinctive character of several of the demographic groupings. The numbers in the figure below represent best estimates on the number of members of kin groups belonging to that speech variety, rather than the number of active speakers (except for the last seven speech varieties, whose numbers represent active speakers). Observations on language use follow below. The numbers for each speech variety take into account migrations. For example, the high number of Christians in the Rana dialect is due to the large communities that have migrated closer to the coast and are found in such villages as Wae Poti, Wae Nibe, Slealale, Neat, and Liang. The range of error in the figures could be as much as 10%.

Speech Variety	TOTAL	Moslem	Christian	Traditional
AMBELAU	5,700	5,700	-----	-----
BURU (total)	44,902	17,022	21,695	6,185
Wae Sama	6,622	4,350	972	1,300
Masarete	9,600	2,550	6,650	400
Buru (Ambon)	2,000	500	1,500	-----
Rana	14,258	150	9,623	4,485
Lisela	11,922	8,472	3,450	-----
Fogi	500	500	-----	-----
HUKUMINA	1	1	-----	-----
KAYELI [11]	4	4	-----	-----
Kayeli	4	4	-----	-----
Leliali	0	-----	-----	-----

Figure 9: Population estimates for speech varieties of Buru

In addition to those indigenous to Buru, there are several major immigrant groups clustered in different places. There are 20,000 Javanese, mainly in the Wae Apo transmigration area, with another Javanese community around the Waenibe Wood Industry plywood factory at the mouth of the Wae Nibe river. There are 12,600 Butonese (a generic term for people from South-east Sulawesi) concentrated mostly along the west and south coasts. 12,500 people from the Sula Islands (to the north of Buru) live along the north and north-east coasts. There are an additional 10,000 people from various other parts of Maluku and Sulawesi scattered around the coastal areas of Buru, the largest subgroups of which are Bugis, Ternate, from southern Maluku (Kei, Tanimbar, Luang, Kisar — mostly Catholic), Ambon-Lease (Muslim & Christian), Seram (Christian), Chinese (mostly Buddhist and Christian), and 'Arab'. These immigrant groups adhere to Islam except where noted otherwise.

7.4 Language information

Some generalisations can be made regarding language use on Buru when looking at the map above. The communities that adhere to Islam tend to be multi-ethnic, with the indigenous Buru population usually a minority. In these communities, the vernacular language is not normally the primary means of communication, even in the home. The vernacular is used with varying degrees of proficiency and success in limited domains, and in attempts to exclude outsiders during such things as negotiations. These communities may be characterised as well along in a language shift to Malay, particularly in the north, east and west coastal areas of the island. Language use in Christian communities is vigorous, except for those in the towns of Ambon, Namlea, some in Leksula, and some in the multi-ethnic checkerboard of the Wae Apo valley. For those who adhere to traditional beliefs, the vernacular is their primary, and sometimes exclusive, means of communication.

---

[11] The entire ethnic group is possibly 800 people.
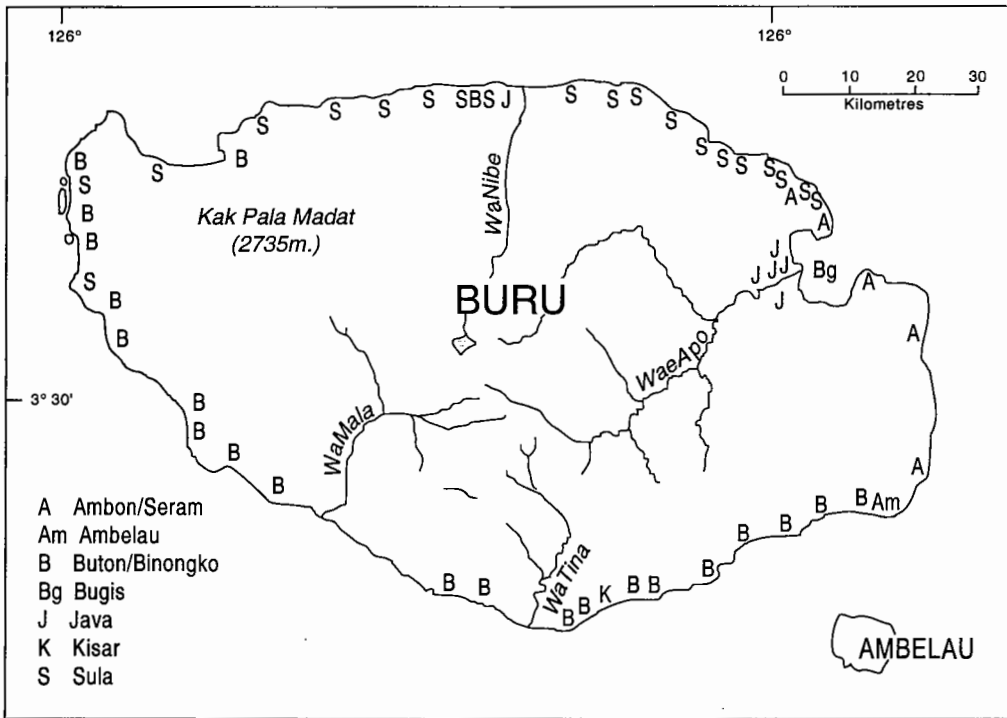
Map 7: Immigrant communities on Buru

7.4.1 Ambelau

The language of Ambelau is spoken in the villages of Elara, Salasi, Siwar, Kampung Baru, Ulima, Masawoy and Lumoy on the island of Ambelau, and in the village of Wae Tawa across the straight on the south-east tip of Buru. There are 5,700 speakers of Ambelau, all of whom adhere to Islam. The rocky terrain and abundance of destructive wild pigs make the tilling of gardens for food crops on Ambelau a futile effort. The village of Wae Tawa on the coast of Buru is an Ambelau colony established for the purpose of supplying vegetables and tubers. Cloves and copra are the main cash crops. There is insufficient data at this point to make intelligent statements about any dialect variations on Ambelau.

The language of Ambelau is not intelligible with Buru. Origin myths on Ambelau trace a connection with Nusa Laut in the Ambon-Lease islands.

7.4.2 Buru

The language of Buru is spoken throughout the inhabited areas of the island. Those adhering to Islam are found scattered along the north and east coasts of Buru and in most coastal villages in the Wae Sama area to Namrole bay on the southern tip of the island. Traditional beliefs are held to in the inland Wae Sama area, Wae Kabo, Wae Tina, Rana, and Wae Geren, although churches are beginning to appear in some places. The Masarete area (coastal and mountain) is predominantly Christian. Migrations from one geographical

region to another are common (see previous discussion) but where migrations have included large families, the speech variety of their place of origin is maintained.

There are 45,000 speakers of Buru, of whom around 30,000 use it vigorously. The five major dialects of Buru are indicated on Map 8 below.
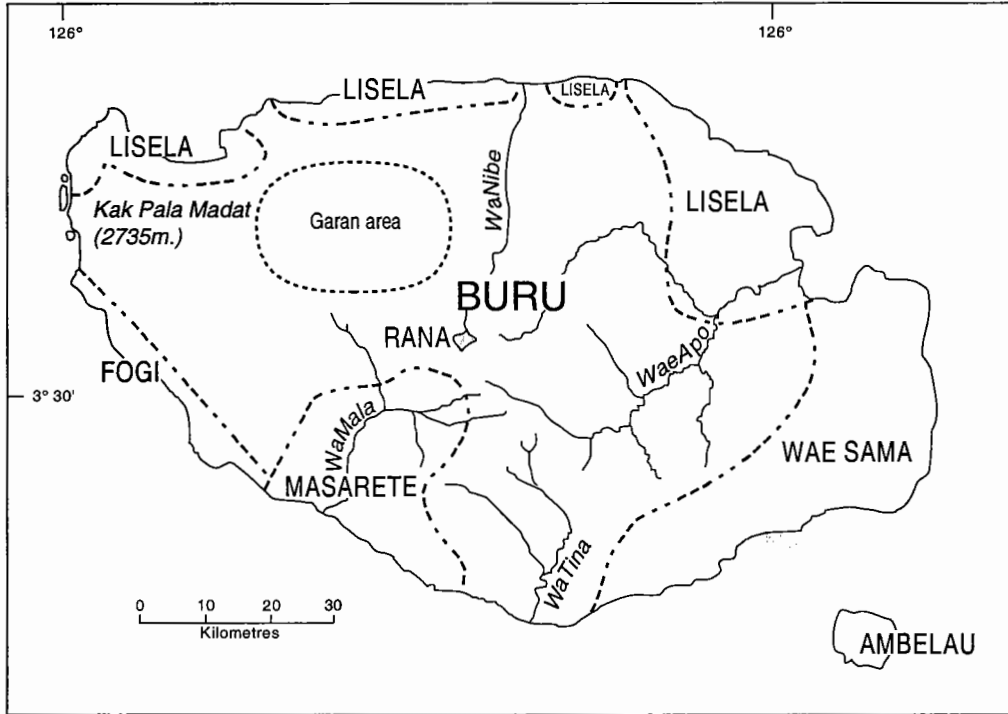
Map 8: Major dialects of the Buru language

7.4.2.1 *Wae Sama dialect*

Those in the interior of the Wae Sama area adhering to traditional beliefs are commonly seen by others on the island as being the most isolated, backward, and the most violent. While these stereotypes are not without foundation, they are a reflection of an us-them distinction that has a basis in speech differences and cultural variation. Some of the better known differences of the Wae Sama dialect are intonation, saying *fiet* rather than *fuat* for 'banana', and saying *fae* 'machete', *lemet* 'cuscus', and having the prothetic /e/ tend toward /i/.

7.4.2.2 *Masarete dialect*

Those in the Masarete area have had the longest exposure to Europeans, education and Christianity, of all the areas in the southern, central, and western parts of the island. Thus it is not surprising that most of the ethnic Buru people in Ambon, Jakarta, and the Netherlands are from Masarete.

Hendriks (1897) gives notes on the grammar, a short lexicon, and several texts for Masarete. Two Masarete Holle lists are provided by Stokhof (1982) dated 1895 and 1904, one filled out by an Ambonese, and the other by either a Dutchman or an Ambonese. Both lists must be evaluated on an item-by-item basis, as they mix Malay, give descriptive phrases rather than the appropriate lexical item, and have significant problems with transcription. Wallace's (1869, reprinted 1962) word list for 'Massaratty' is indeed a Masarete word list.

Masarete stands apart from other dialects by using *nunu* rather than *nini* for the third person plural possessive, saying *katuen* instead of *todo* 'machete', *tongi* 'also, as well' and *tirin* 'very' where other dialects use *pee* for both senses. The genitive enclitic system in Masarete has collapsed to the third person singular form for all person and number combinations (C. Grimes 1991:282ff.).

7.4.2.3 Rana dialect

There are several varieties (subdialects) of the Rana dialect spoken both in their areas of origin and in the areas to which some speakers have migrated. Even in the migrations to the coast that occurred around 100 years ago language use remains vigorous by all age groups.

7.4.2.3.1 Rana subdialect (Rana dialect)

Rana covers all the inhabitants of the greater Rana (Wae Kolo) valley surrounding the large lake in the centre of the island. Large groups have migrated over the last 100 years and set up communities on both the north and south coasts, as well as inland about one day's walk from the coast.

Some speakers from Rana also speak Li Garan (see discussion above). I estimate 3,000–5,000 people who would be knowledgeable in Li Garan to one degree of proficiency or another.

7.4.2.3.2 Wae Geren subdialect (Rana dialect)

The upper Wae Geren area has periodic traffic and intermarriage with Rana, and the two areas have overlapping hunting grounds.[12]

7.4.2.3.3 Wae Kabo subdialect (Rana dialect)

Those from the Wae Kabo area have also migrated into the Wae Tina and upper Wae Mala areas and so the speech variety in these areas is fairly uniform.

---

[12] The lower Wae Geren area is part of Wae Apo and speaks Li Enyorot.

7.4.2.4 *Lisela (Li Enyorot) dialect*

As mentioned above, Li Enyorot is the variety of Buru used along the north and east coasts in the areas inhabited primarily by outsiders and in the area formerly identified with the Kayeli language. It seems to have functioned as a *lingua franca* along the flat coastal strip to relate to people from other language groups.

The term 'Li Enyorot' is what speakers of the Rana dialect use to describe this divergent but often intelligible speech variety spoken along the northern and eastern coastal strips, and in the lower Wae Geren and Wae Apo valleys.[13]

Speakers of Li Enyorot describe their own speech variety as 'coarse', 'crude', 'mixed (adulterated)', 'not pure', 'not sweet'. They describe the Rana and Masarete dialects as 'pure' and 'pleasant sounding'.[14]

As a generalisation, use of Li Enyorot is not vigorous and a language shift to Malay is under way. Many people under 30 years of age know only bits and pieces of the language or can function only in a very few domains. As they have tended to fully participate in the greater Malay Islamic coastal culture, the value of their own culture and language has not been upheld. This situation is best characterised as 'language shift in process' rather than 'stable bilingualism'.

Li Enyorot has been variously referred to in the literature as Lisela (Licella) and *liet enjorot* (Schut 1919). Besides being spoken along the five kilometer wide northern coastal strip of Regentschap Lisela, it is also spoken along the coastal strip of Regentschaps Tagalisa, Leliali, and Kayeli. In relation to the literature it must be further noted that in Leliali and Kayeli, two different languages were spoken in the same area, both referred to as Leliali or Kayeli. The two languages were Li Enyorot and two dialects of Kayeli. The Kayeli region is also referred to as Wae Apo, Mako, 'transmigrasi', and 'unit-unit', the latter three focusing on the area where Javanese transmigrants are being settled in an attempt to turn the lower Wae Apo river valley into a wet-rice growing area.[15]

Wallace's (1869) 'Wayapo' word list is Li Enyorot. Holle list 55c (Stokhof 1982) dated 1896, that is identified as 'Kayeli', is in fact Li Enyorot.

7.4.2.5 *Fogi (Li Emteban)*

Li Emteban is a term noted by Schut (1919) to refer to a variety of speech used along the south-west coast in the Fogi and Tomahu areas. (The term 'Li Emteban' means 'the language that is cut short at the end', referring to perceptions by those in other parts of the island toward the variety of Buru spoken around Fogi). The area is generally referred to as Fogi, following the name of the Regentschap. The Tomahu speech variety is extinct, and the Tomahu kin group itself is also nearly extinct. The Fogi area is completely dominated by Butonese with a few people from Sula. Reports from both the north and south coasts as well

---

[13] Li is a normal cliticisation of *liet* 'language'. *Oro* means 'sag, droop' as when a bird alights on a small branch. Thus, the term '*Li Enyorot*' gives the picture by metaphor of 'a language that droops'.

[14] The negative attitudes expressed toward their own dialect and the positive attitudes expressed toward the Rana dialect may be one indication that vernacular literature in the the Rana dialect might be acceptable to speakers of Li Enyorot. Further study is needed.

[15] The process was begun in the late 1960s using political prisoners but now is just another one of the government's nationwide transmigration projects.

as personal contact with ethnic Buru people from the Fogi area all indicate that the language shift to Malay is complete in this area among the 500 or so people in the area who are indigenous to Buru. I have talked with both adults and school children from the area who do not even have a partial knowledge of the vernacular.

7.4.3 Hukumina (Bambaa)

Hukumina was referred to by Hendriks (1897) as a dialect of Buru spoken in the districts of Hukumina, Palumata and Tomahu (these were districts in the north-west corner of Buru). He gives no other details other than his impression that one could shift easily between Hukumina and other dialects of Buru. A Hukumina word list is published in the Holle lists (Stokhof 1982) with very little information other than the date 1896.

All reports indicate the language is extinct in the Hukumina area (north-west Buru), but I was able to find one old woman from the former village of Hukumina that used to be located behind the present village of Masarete near the fort at Kayeli (north-east Buru). My data are even more divergent than the Holle list (the latter includes several items that are clear borrowings from Buru). The language of Hukumina is very distinct from Buru in intonation, lexicon and sound correspondences. For example, PAn *k is Buru /k/ and Hukumina /c/ intervocalically.

The language is functionally extinct, meaning that in 1989 there was only one 80-year-old toothless woman whose mind wandered that still remembered phrases and sentences, but hadn't had anyone to speak it with for two decades. The language is referred to in the villages of Kayeli and Masarete as the 'Bambaa' language, which, in Hukumina means 'there isn't any'. The informant herself calls the language Hukumina.

7.4.3.1 Palumata (Palamata, Balamata)

This speech variety is also extinct, having been spoken in north-west Buru, an area that is now completely inhabited by Butonese. Although oral history in the Kayeli area remembers Palumata as distinct from Hukumina, the former Regentschaps of Palumata and Hukumina were adjacent to each other. Hendriks (1897) identifies the 'dialect' of Hukumina as being spoken in the districts of Hukumina, Palumata, and Tomahu. My assumption at this point is that Palumata and Hukumina were varieties of the same language.

7.4.4 Kayeli (Cajeli, Gaeli) language

The Kayeli language was spoken in former days around Namlea Bay in north-east Buru and along the coast as far as the Samalagi river. Except for the four remaining speakers of the Kayeli language, indigenous Buru people in this area have shifted to Malay and the Li Enyorot dialect of Buru.

The king's family of Kayeli and the king's family of Leliali continue to marry each other in cross-cousin marriage as they have for several generations. This, and several other factors indicate that the notion of a Kayeli ethnic group remains intact despite language shift and language death.
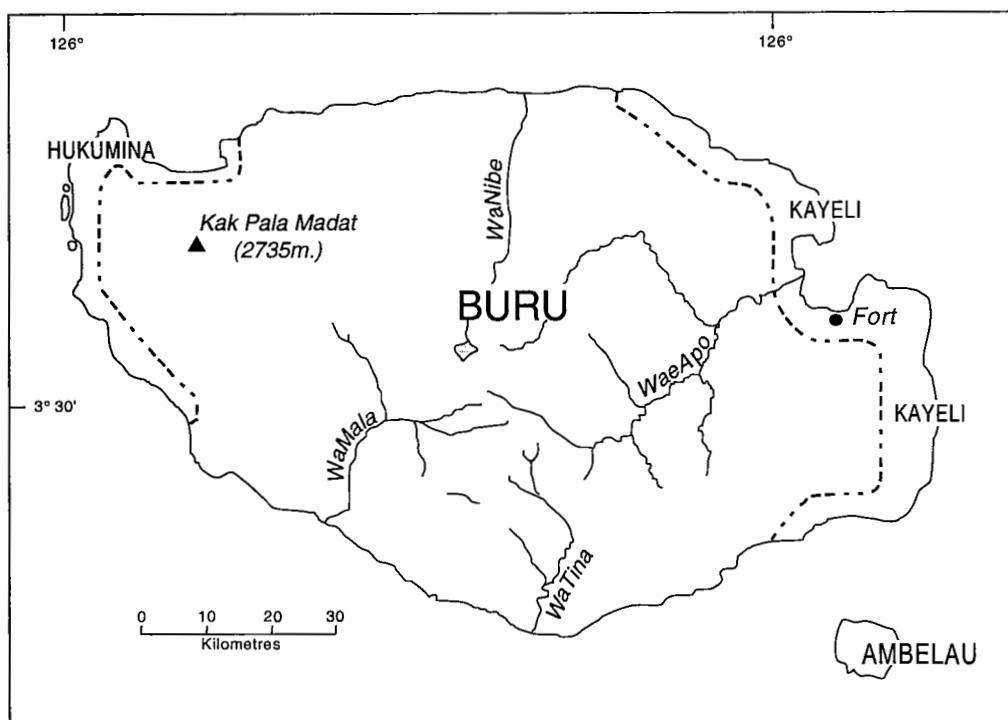
Map 9: Probable language picture on Buru around 1600

7.4.4.1 Kayeli dialect

The Kayeli dialect of Kayeli was spoken around the southern region of Namlea bay in the villages of Kayeli and Masarete, and in Namlea and Namete on the north coast of Namlea bay where villages of Kampung Cina and Kayeli Kristen moved after the turn of the century.

The four remaining speakers of this dialect live in the villages of Kayeli and Masarete and are all over 60 years old. Several people over 35 years of age know isolated words and phrases of Kayeli, but none of them control the language in any single domain. As the four remaining speakers of the language do not use it among themselves, the language is best characterised as nearly extinct, or functionally extinct.[16]

Two "Kayeli" word lists are given in the Holle lists (Stokhof 1982) both dated 1896. The data in list 55c is actually the Li Enyorot dialect of Buru and it is described as "the language spoken by the Alfuru from Kayeli, Licella, Tagalisa and Liliali", the areas in which Li Enyorot is spoken. List 55a is a Kayeli word list. Wallace's (1869) 'Cajeli' word list is indeed Kayeli.

---

[16] When eliciting data from the four remaining speakers of Kayeli, one could visibly see the unfolding horror and agony as it dawned on them that even basic vocabulary items that none of them could remember were, as they said, "gone forever and irrecoverable."

7.4.4.2 Leliali (Liliali) dialect

I was able to get a 205-item Leliali word list from an 81 year old speaker on a field trip in 1983. He died two months later. On a return trip in May 1989 I was told that the last speaker, a woman who was actually born in Kayeli, had died in March 1989, two months prior to my visit.

My word list from 1983 is sufficient to confirm that Leliali was a dialect of the Kayeli language, and not a separate language.

7.4.4.3 Lumaete (Lumaiti, Lumaite, Lumara)

This recently extinct speech variety had speakers clustered around Kayeli in the past and there is still the memory of people who spoke something called 'Lumaiti'. There is a Lumaiti river near Kayeli, but there was also, in 1840, a Regentschap Lumaete in the south-east corner of Buru. Hendriks (1897) identifies the Kayeli 'dialect' of Buru as being spoken in the districts of Kayeli, Ilat, and Lumara, the latter being the name of a river in south-east Buru in the area of the former Regentschap Lumaete.

Although the linguistic classification of Lumaete is impossible in an absence of data, there are a couple of clues which help. One is that this variety of speech was spoken somewhere in east Buru. Another clue is from the name itself — *luma* 'house' follows the Kayeli reflex of /l/ for PAn *R, rather than the Buru reflex, which is /h/. My assumption at this point is that Lumaete was a dialect of Kayeli.

7.4.5 Moksela (Maksela, Opselan)

This speech variety has been extinct since 1974, when the last speaker died. No clues other than the name of a stream east of Kayeli called Moksela, give any indication as to where it was spoken or what it was like. If it was spoken from the stream by that name eastward, then chances are likely that it was also a variety of the Kayeli language. People in the Kayeli area remember nothing more than the name of the language, who in the community spoke it before they died, and that it was somehow different enough to have its own identity.

8 Summary

While one might expect wide linguistic variation on a mountainous island the size of Buru, the indigenous inhabitants insist there is only one language. A perusal of word lists sampling speech from around the island indicates wide diversity, with some varieties sharing as little as 61% similarity. This paper has shown that the complex language picture on Buru can be credibly unravelled only by taking into account synonym sets, lexical taboos, marriage patterns, internal migrations, local and regional history, sound correspondences, and patterns of language use in multi-ethnic societies. The status of Lisela (Li Enyorot) in relation to the other main Buru dialects was shown on a number of criteria to be marginal between a divergent dialect of Buru, and a separate language. A number of

speech varieties on Buru have become recently extinct, while others are well along a shift to Malay in a process that has been on-going for at least since 1658.[17]

References

Bellwood, Peter, 1985, *Prehistory of the Indo-Malaysian Archipelago*. New York: Academic Press.

Casad, Eugene, 1974, *Dialect intelligibility testing*. Dallas: Summer Institute of Linguistics.

Collins, James T., 1980, *Ambonese Malay and creolization theory*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

  1981, Preliminary notes on Buru, Sula and Taliabo and Ambelau. In Robert A. Blust, ed., *Historical linguistics in Indonesia*, Part I. *NUSA* 10:31–45.

Collins, James T., ed., 1983, *Studies in Malay dialects*, Part II. *NUSA* 17.

Donohue, Mark, 2000, Tukang Besi dialectology. In this volume, Charles E. Grimes, ed., *Spices from the East: papers in languages from eastern Indonesia*, 59-76. Canberra: Pacific Linguistics.

Fraassen, Ch.F. van, 1983, Historical introduction. In K. Polman, *The Central Moluccas: an annotated bibliography*. KITLV:1–59.

Grimes, Barbara Dix, 1990, The return of the bride: affiliation and alliance in Buru. MA thesis, Australian National University.

  1991, The development and use of Ambonese Malay. In H. Steinhauer, ed., *Papers in Austronesian linguistics*, No. 1, 83–123. Canberra: Pacific Linguistics.

  1994, Cloves, nutmeg, traders and wars: language contact in the Spice Islands. In Tom Dutton and Darrell Tryon, eds, *Language contact and change in the Austronesian world*, 251–274. Berlin: Mouton de Gruyter.

Grimes, Charles E., 1985, Ambonese Malay: a brief orientation. *Bits & Pieces*, April 1985: 14–28. Abepura, Irian Jaya: Summer Institute of Linguistics.

  1991, The Buru language of eastern Indonesia. PhD dissertation, The Australian National University.

Grimes, Charles E. and Barbara D. Grimes, 1994, Languages of the North Moluccas: a preliminary lexicostatistic classification. In the proceedings of the Seminar Penelitian Maluku dan Irian Jaya, LEKNAS-LIPI: Jakarta, 35–63.

Grimes, Charles E. and Kenneth Maryott, 1994, Named speech registers in Austronesian languages. In Tom Dutton and Darrell Tryon, eds, *Language contact and change in the Austronesian world*, 275–319. Berlin: Mouton de Gruyter.

Grimes, Joseph E., 1988, Correlations between vocabulary similarity and intelligibility. *Notes on Linguistics* 41:19–33. Dallas: Summer Institute of Linguistics.

Hendriks, H., 1897, *Het Burusch van Masarete*. s'Gravenhage: Koninklijk Intituut voor Taal-, Land- en Volkenkunde.

Jansen, H.J., 1933, Gegevens over Boeroe (1928). *Adatrechtbundels* 36:463–489.

---

[17] See B.D. Grimes (1994) for a more thorough discussion of the dynamics involved in this language shift.

Kantor Statistik Kabupaten Maluku Tengah, 1987a, *Kecamatan Buru Selatan dalam angka.* Leksula.

1987b, *Kecamatan Buru Utara Barat dalam angka.* Air Buaya.

1987c, *Kecamatan Buru Utara Timur dalam angka.* Namlea.

Pallesen, A. Kemp, 1985, *Culture contact and language convergence.* Manila: Linguistic Society of the Philippines Monograph 24.

Polman, Katrien, 1983, *The Central Moluccas: an annotated bibliography.* Dordrecht: Foris Publications.

Rumphius, G. E., 1910, De Ambonsche historie, behelsende een kort verhaal der gedenkwaardigste geschiedenissen zo in vreede als oorlog voorgevallen sedet dat de Nederlandsche Oost Indische Compagnie het besit in Amboina gehadt heeft. *Bidragen tot de Taal-, Land- en Volkenkunde van Nederlandsche-Indië* 64.

Schut, J.A.F., 1919, Een Tijl Uilenspiegel verhaal in de Boeroetaal. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indie.* 75:303–362.

Simons, Gary F., 1977, Recognizing patterns of convergence and divergence in a matrix of lexico-statistic relations. In Loving and Simons, eds, *Language variation and survey techniques.* Workpapers in Papua New Guinea Languages 21:107–134. Ukarumpa: Summer Institute of Linguistics.

1982, Word taboo and comparative Austronesian linguistics. In A. Halim, L. Carrington, & S.A.Wurm eds, *Papers from the third international conference on Austronesian linguistics, vol. 3: accent on variety,* 157–226. Canberra: Pacific Linguistics.

Stokhof, W.A.L., ed., 1982, *Holle lists: vocabularies in languages of Indonesia Vol. 3/4. Central Moluccas: Ambon (II), Buru, Nusa Laut, Saparua.* Canberra: Pacific Linguistics.

Taylor, Paul Michael, 1983, North Moluccan Malay: notes on a 'substandard' dialect of Indonesian. In James T. Collins, ed., *Studies in Malay dialects,* Part II. *NUSA* 17:14–27.

Voorhoeve, C.L., 1983, Some observations on North-Moluccan Malay. In James T. Collins, ed., *Studies in Malay dialects,* Part II. *NUSA* 17:1–13.

Wallace, Alfred Russel, 1869, *The Malay archipelago.* Macmillan and Co. (Reprinted in 1962 by Dover Publications, New York).

Willer, T.J., 1858, *Het eiland Boeroe: zijne exploitatie en Halfoersche instellingen.* Amsterdam: Frederik Muller.

Wurm, S.A. and Shiro Hattori, eds, 1981–83, *Language atlas of the Pacific area.* Part 1. Canberra: The Australian Academy of the Humanities/The Japan Academy. Canberra: Pacific Linguistics.

5 *The languages of Wetar: recent survey results and word lists, with notes on Tugun grammar*

BRYAN HINTON

1 Introduction

To date, very little information has been available on the languages of Wetar. This study presents a current overview of the language situation on this eastern Indonesian island, using a lexicostatistical study to calibrate other information. The perspective offered here is quite different from what was previously known about the island. Also included is a brief sketch of Tugun phonology and grammar, as well as representative word lists from each of the five language groups, in order to provide data that otherwise are unavailable on these Austronesian languages.

The majority of the data for this study were collected on the island of Wetar, with additional data collected on nearby Kisar island, for two weeks in April of 1989 by the author, and colleague Lee Steven. Subsequent visits of three weeks each in June of 1989 and January 1990 by the author were spent living in a village in the Tugun language group collecting texts and data on morphology and syntax, as well as checking previous survey data. [These islands are identified on the map in the introduction to this volume.]

Wetar is located 60 km north of East Timor, in the south-west corner of the province of Maluku. It is a mountainous island roughly 95 km long by 30 km wide, with peaks rising to 1,430 meters. The total population of the island is just under 4,000, according to official government statistics (Kantor Statistic Propinsi Maluku 1987).

An attempt was made to visit every village that has a Wetar-speaking native population, and to elicit a 210-word list from the local dialects. Two of the twenty-three villages on Wetar are comprised entirely of recent migrants from Kisar Island that still use the Kisar language, and therefore were not included in this study.[1] The survey was carried out under a

---

[1] The Kisar-speaking villages of Neumatang and Arnau were not included in this study. Lurang village was completely deserted when we arrived, except one old woman who was either too sick or too afraid to come out of her house. The attempt to take a word list through the window failed. However, on a subsequent

Charles E. Grimes, ed., *Spices from the East: papers in languages of eastern Indonesia*, 105-129. Canberra: Pacific Linguistics, 2000.
© Bryan Hinton

co-operative agreement between the Summer Institute of Linguistics and Pattimura University in Ambon, Maluku.

2 Previous studies

Previous works that have attempted to delineate the speech groups of Wetar or contribute to knowledge about them are discussed below. Wetar has been considered an Austronesian language with several dialects, classified into a larger 'Ambon-Timor Group' of languages (Esser 1938, Salzner 1960, Dyen 1965, Voegelin and Voegelin 1977). More recently, Blust (1977) places the languages of southern Maluku (including Wetar) within the Central Malayo-Polynesian subdivision of Austronesian.

2.1 Riedel

Riedel (1886) identifies five speech communities on Wetar: Iliwaki, Tutunohan, Hahutau, Limera, and Welemur. Included throughout his chapter on Wetar are numerous examples from the Iliwaki language. The following map shows the location of Wetar dialects according to Riedel:
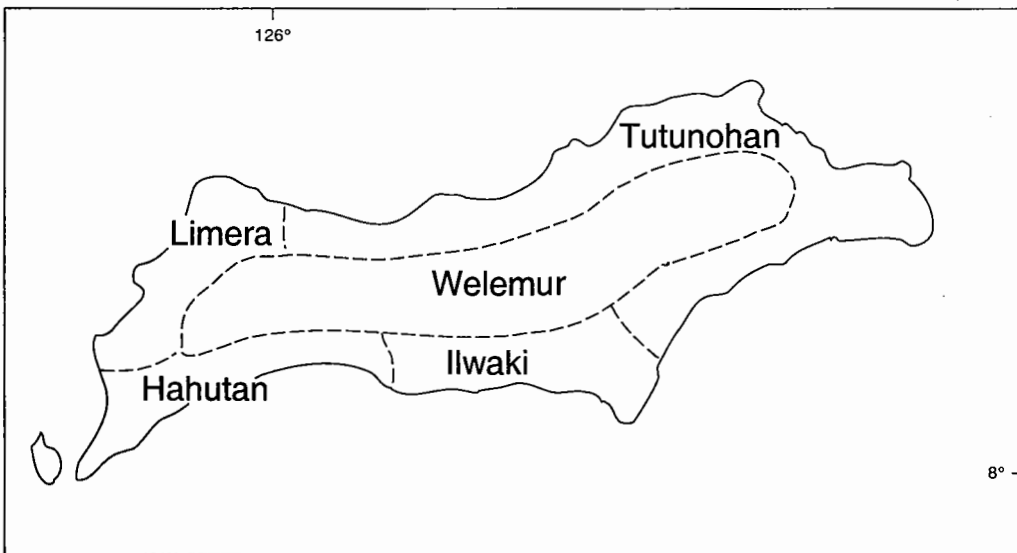
Map 1: The dialects of Wetar Island, from Riedel (1886)

On the map above, a dialect called Welemur is shown to be located in the interior region of the island. During the recent survey, however, it became apparent that this group no longer exists as a separate dialect or language. Local officials and the local inhabitants report that no-one lives in the interior anymore. Finally, the head of Lurang village reports

---

visit in Jan. 1990, a word list was successfully taken, and it was explained that the woman was not sick as had been previously assumed, but had already passed away.

that the Welemur people have all been absorbed into the villages of Lurang, Ilwaki, and Ilputih, and no longer speak a distinct dialect.

2.2 Holle lists

"The Holle lists", published in 1980 (W.A.L. Stokhof, ed.) included a 1,400-word list from Wetar, plus some brief notes on the grammar. The list was taken in 1895 by W. Heldt, from the "Lir Talo dialect". Based on a comparison with the current survey results, this list came from what is presently called the Talur language group.

(1)	Gloss	Holle list	Talur	All other Wetar languages
	'house'	*uma*	*umah*	*raha*
	'we' (incl)	*gita*	*gita*	*ita*
	'yellow'	*mara*	*mara*	*pola*

2.3 J.P.B. de Josselin de Jong

The most significant work on a language of Wetar to date is de Josselin de Jong's *Studies in Indonesian culture II: the community of Erai (Wetar)* (1947), based on texts gathered during an eighteen day stay in a village on the west coast of Wetar. The book includes ethnographic notes, a brief sketch on some aspects of the grammar of the language found there, a 1,200 word dictionary, and transcripts of the texts with English translation.

2.4 Salzner

In R. Salzner's 1960 *Sprachenatlas des Indopazifischen Raumes* (Language Atlas of the Indo-Pacific Region), Wetar is considered to be one language, and is divided into six dialects: Ilwaki, Hahutan, Limera, Tutunohan, and Welemur on Wetar island, and Galoli on E. Timor. He includes Wetar in the Ambon-Timor group. The significance of Salzner's map is the inclusion of a dialect of Wetar located in E. Timor, not found in any other work (see discussion under §5.1 below).

2.5 Voegelin and Voegelin

Voegelin and Voegelin's 1977 *Classification and index of the world's languages* includes Wetar in the Ambon-Timor group, following Esser (1938) and others. This work differs from Salzner (1960) in that Erai is included as one of the speech groups of Wetar, while Galoli is not.

During the recent survey, it became apparent that Erai refers to a village on Wetar in the Ili'uun language group (see Map 2), rather than a separate language or dialect.
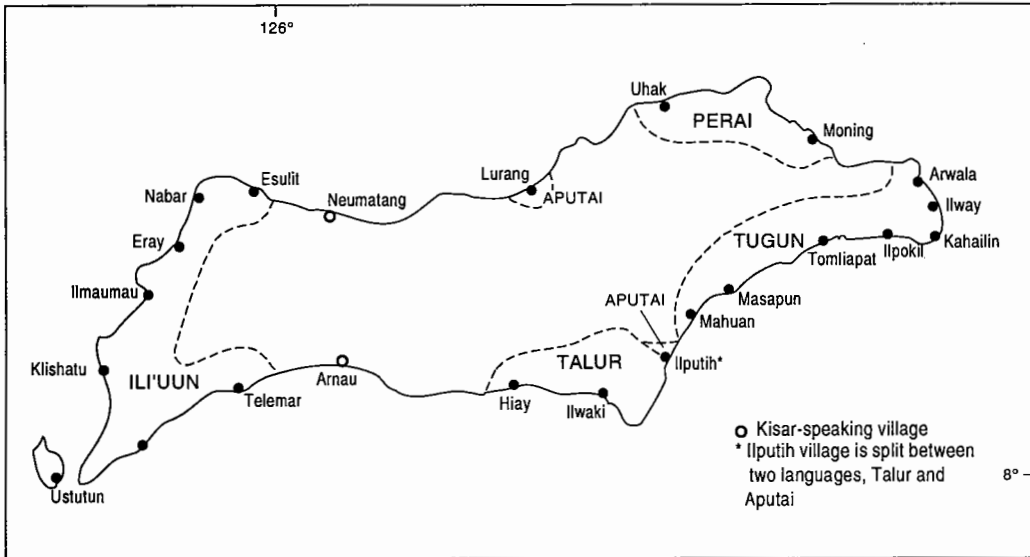
Map 2: The languages of Wetar Island, 1989 survey

2.6 Wurm and Hattori (eds)

The Wurm and Hattori, eds (1981) map on the region that includes Wetar presents all the linguistic groups on the island as dialects of the same language, although they note "The differences between the dialects of Wetar are considerable, and there are numerous subdialects".

Wurm and Hattori's classification of these languages is as follows:

Timor Area Group
Timor and Islands Subgroup
Wetar (dialects)
Ilwaki
Tutunohan
Limera
Hahutan
Welemur

Figure 1: Classification of Wetar dialects, from Wurm and Hattori, eds

The dialect names above are the same as those used by Riedel (1886), except for a slight spelling difference (Hahutan vs Hahutau). The location of the dialects on both maps are also practically identical.

2.7 Taber

Taber (1990), in an overview survey of south-western Maluku languages, has tabulated the percentage of lexical similarity between word lists from the entire area, including one from each of the languages on Wetar. Taber places the languages of Wetar in what he calls

the South-West Islands Stock, which includes the Wetar Family, the Kisar-Roma Family, the Luang Family, the South Damar Family, and the Teun-Nila-Serua Family.[2]

The data for Taber's study originally included a word list from only four of the five language groups on Wetar. The data therefore was not sufficient to obtain a clear perspective on where the present day speech community boundaries on Wetar are, or the range of variation between the dialects. In addition, the existence of a speech community living in the interior of the island was not confirmed (see discussion under Riedel above).

The present survey was proposed to obtain detailed information on the number, location, and range of variation of present day speech communities on Wetar, which was beyond the scope of Taber's study. The procedures and results of the present study are discussed in the following sections.

3 Procedures

3.1 Word list

A 210-word list was used in this survey, which includes most of the Swadesh 100 and 200-word lists. The list was translated into Indonesian and elicited in Indonesian/Ambonese Malay. Some of the items were excluded from the analysis after the survey was completed. Due to the confusion and hesitation encountered during elicitation, six words were considered unreliable for analysis.[3] A seventh word 'bark' was excluded since it always appears as a polymorphemic lexical item consisting of the words for 'tree' and 'skin', both of which were elicited separately. The remaining 203 items were used for comparison.

3.2 Data collection

Care was taken to be sure all the word lists were elicited from native speakers of the languages who were also offspring of native speakers. Where possible, all the lists were taken in the villages on Wetar, although three were elicited nearby at the government district center.[4] For most of the villages two word lists were taken as a check on the reliability of the data. Of the 38 word lists collected, 21 (one from each village) were chosen for comparison.

The opinions of the native population of Wetar concerning the language situation were taken into account, with a high degree of correlation between their reported information and the results of this study.

---

[2] In July of 1987, colleagues Mark Taber and David Coward of the Summer Institute of Linguistics conducted a language survey of the south-west part of Maluku, under a cooperative agreement with Pattimura University. Included among the word lists they elicited was one from Wetar (Ilwaki village). In addition, John Christensen, working on nearby Kisar island, elicited three of the Wetar word lists that were included in Taber's study. The final word list, as well as information on the location of the Wetar languages, was contributed by the author.

[3] The words bangunkan 'wake up', jatuhkan 'drop', bulat 'round', bagaimana 'how', berbaring 'lie down', and mandikan 'bathe s.o.' were excluded. Their morphological and semantic complexity regularly caused confusion.

[4] Word lists for the villages of Hiay, Mahuan, and Masapun were taken on Kisar.

3.3 Data analysis

In the analysis of the word list, words considered to be phonetically the same or similar were grouped together into the same cognate set,[5] using the inspection method (Gudschinsky 1956), providing they also had the same or similar meaning. The computer program WORDSURV[6] was used to compile the percentages of lexical similarity. The computer tabulated the choices made by the linguist to arrive at the percentage of lexical similarity for each pair of word lists.

4 A lexicostatistical perspective

Figure 2 below presents the percentages of lexical similarity between the word lists from Wetar. Two word lists are included from Ilputih village since there are two language groups living together there. Names in lower case on the chart are villages which the word lists represent.

4.1 Discussion of chart

In using the term 'language' I am referring to the distinct groupings in the chart. I have used the range of 75 to 80 per cent to determine the division between languages and dialects of the same language. Consequently, word lists grouping above the 80 per cent level were considered dialects of the same language.

The language division choices were also influenced by the opinions and reports of the native speakers of these languages, who consistently reported little or no intelligibility between the languages. Intelligibility testing (Casad 1974) needs to be done to verify the degree of intelligibility between groups that fall close to the language/dialect threshold, such as Perai and Aputai.

The figures in the chart in Figure 2 are all relative to each other. A researcher using a shorter or a longer word list would most likely come up with different percentages. What the chart demonstrates is the internal cohesion of what I call the Wetar Cluster (Perai, Aputai, Ili'uun and Tugun) as opposed to the Talur language. The figures from Taber (1993) also support these divisions. The split between the Talur language and the languages of the Wetar Cluster is graphically illustrated in the figure below, which is based on the percentages of lexical similarity:

---

[5] 'Cognate' is not an entirely accurate term for this study, since it implies a rigorous study of sound correspondences to determine genetic relationships and weed out borrowings.

[6] John Wimbish's WORDSURV program (1989) greatly reduced the time and difficulty of comparing 21 lists. After the lists were entered and words grouped by the linguist into sets, it took under 60 seconds to finish the tabulations.

```
Ilputih (a)
99  Ilwaki      TALUR
94  94  Hiay
58  57  56 | Ilputih (b)      APUTAI
55  54  55 | 93  Lurang
52  51  52 | 79  75 | Moning
51  50  50 | 80  77 | 93  Uhak       PERAI
52  51  52 | 75  75  76 | Mahuan
53  52  52 | 74  76  76 | 97  Masapun
54  53  52 | 74  76  76 | 94  97  Tomliapat
52  51  52 | 72  76  75 | 93  96  97  Ilpokil      TUGUN
53  52  52 | 73  75  75 | 92  92  94  95  Kahailin
52  52  51 | 74  76  76 | 93  95  96  97  96  Ilway
53  52  52 | 73  76  76 | 93  95  96  97  97  99  Arwala
53  54  51 | 69  67  67  72  73  73  71  71  72  73 | Telemar
51  52  50 | 68  66  67  72  73  74  71  72  72  73 | 96  Karbubu        ILI'UUN
52  53  51 | 69  67  67  72  73  73  71  71  72  73 | 96  98  Ustutun
51  52  51 | 70  68  67  73  74  74  72  72  73  74 | 96  96  96  Klishatu
53  54  51 | 70  67  67  73  74  75  73  72  74  74 | 97  96  97  97  Ilmaumau
52  53  51 | 69  66  67  72  73  73  71  71  71  72 | 97  97  97  97  97  Eray
51  51  49 | 68  63  64  71  72  72  70  70  71  71 | 94  94  93  94  95  96  Nabar
51  52  50 | 70  66  67  72  73  73  71  71  72  73 | 95  96  96  96  96  97  97  Esulit
```

WETAR CLUSTER

Figure 2: Lexical similarity of Wetar word lists
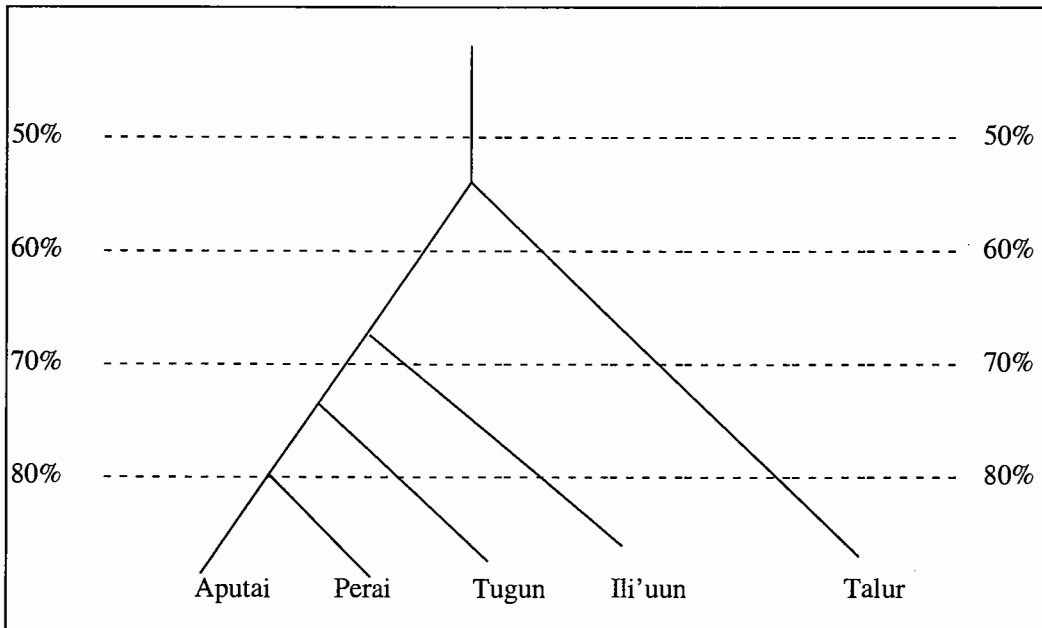
Figure 3: Tree diagram of Wetar languages

4.2 Further evidence for the Wetar cluster

Further evidence of the split between the Wetar Cluster and the Talur language comes from sound correspondences between the two groups, and in the possessive construction. In the Talur language, intervocalic Proto Austronesian *R is lost, while the other languages of Wetar reflect it as [r] (trilled or flapped):

| (2) | *DuRi* | 'thorn' | [ruin] | Talur |
| | | | [rurin] | others |

| (3) | *ma-iRaq* | 'red' | [mea] | Talur |
| | | | [mera] | others |

| (4) | *daRəq* | 'earth' | [rea] | Talur |
| | | | [rare] | Tugun, Ili'uun, Aputai |

(5)	*baqəRu*	'new'	[heu]	Talur
			[feru]	Perai, Tugun, Aputai
			[heruŋ]	Ili'uun
			[herun]	

(6)	*daRaq*	'blood'	[raan]	Talur
			[rara]	others
			[raraŋ]	

The possessive construction for the two groups is also very different. Talur, like Galoli in East Timor, has a full set of possessive suffixes for certain body parts and terms of relationships. The other four languages of Wetar do not use possessive suffixes.[7] Talur possessive suffixes are given below with the forms *mata-* 'eye', *ama-* 'father', and *ee-* 'leg':

		'eye'		'father'		'leg'
1s	*ga*	*mata-k*	*ga*	*ama-k*	*ga*	*ee-k*
2s	*go*	*mata-k*	*go*	*ama-k*	*go*	*ee-k*
3s	*ni*	*mata-n*	*ni*	*ama-n*	*ni*	*ee-n*
1pi	*gita*	*mata-r*	*gita*	*ama-r*	*gita*	*ee-r*
1pe	*gami*	*mata-r*	*gami*	*ama-r*	*gami*	*ee-r*
2p	*mi*	*mata-r*	*mi*	*ama-r*	*mi*	*ee-r*
3p	*sia*	*mata-r*	*sia*	*ama-r*	*sia*	*ee-r*

Figure 4: Talur possessive construction

5 The languages of Wetar

The following is a description of the languages of Wetar, including alternate names, location, and population figures. I have also compared the results of this survey with the

---

7   In Wetar languages, the final *n* of certain part-whole relationships like *ai-funan* 'flower' (lit. tree-flower), from PAn *buŋa*, may be a frozen 3s possessive suffix.

previous works mentioned above. The language names used reflect the current usage based on responses from the majority of the inhabitants I spoke with. As discussed above, the languages of Wetar fall into two major divisions, the Wetar Cluster (Perai, Ili'uun, Aputai, and Tugun languages), and the Talur language group.

All of the villages of Wetar are located on the coast, with garden plots scattered over miles of rugged mountainous terrain. The people practice swidden agriculture, with maize being the main crop. All of the villages on the island have accepted the Christian religion. In addition to the local languages, most of the people are also bilingual to some extent in the regional variety of Malay.

5.1 The Talur language

Talur is spoken in three villages on the south-central coast of Wetar (Hiay, Ilwaki, and Ilputih). Several alternate names are used for the language: Galoleng, Ilwaki, Lir Talo, and Ilmedu (Ilputih village). The total population is 675 (on Wetar, see discussion below). Hiay village is reported to be comprised mainly of immigrants to the area from Kisar who have learned the local language, and is therefore included in this population figure.

While in the Talur language area, I was told that the language was the same as the language spoken in East Timor in the region around Manatuto. This roughly corresponds to the Galoli language area shown in the Wurm and Hattori (1981) map of Timor. Salzner (1960) even includes Galoli as a dialect of Wetar located in E. Timor. An alternate name for the Talur language is Galoleng, giving further indication of a connection. Capell (1944) notes "...the language of Wetar is almost identical with that of the Galoli country in Timor..." (he assumed data from Ilwaki was representative of the whole island). Comparing data from Capell (1944) with data from this study, the two languages are 86% lexically similar. This Timor connection for the Talur language could explain why it is divergent from the language groups of the Wetar Cluster. Further data from East Timor is needed to clarify the relationship.

5.2 The Wetar cluster languages

5.2.1 Perai

Perai is spoken on the north-east coast of Wetar in the villages of Uhak and Moning. The total population is 278. Kisar speakers have intermarried with the people of Moning to the point where they comprise nearly 50% of the population. It is claimed that outsiders who marry into the group learn the local language. The people living today claim to be descendants of people who originally lived in the interior.

Perai does not appear in the maps of Riedel (1886), Salzner (1960), or Wurm and Hattori (1981). However, during the recent survey people all around the island identified Perai as a separate language. Elbert (1912) also was familiar with a group called Perai, living in the same area that Perai speakers occupy today.

Perai is most closely related to the Aputai language (79–80%), right at the threshold between language and dialect. Since the speakers consider them separate languages, I have classified them separately pending further research on intelligibility.

5.2.2 Aputai

Aputai is spoken in Ilputih village on the south-central coast of Wetar, and in the village of Lurang, on the north-central coast of the island. An alternate name given in Ilputih was Opotai. In the previous works, Aputai appears only in Elbert (1912). The population for this group is approximately 150, split between the two villages.

A trail connects the two villages, though it is more than a two day hike. Some of the residents of Ilputih were born in Lurang. In the last few years a gold mining operation has begun near Lurang, which may potentially cause significant changes in the area.

Originally the speakers of this language lived in two villages in the interior of the island. Prior to World War II they were moved to the coast, one group to the north, and one to the south. The people that moved to the south coast were placed with a village of Talur speakers. Consequently, many (if not all) Aputai speakers living in Ilputih village also speak Talur. This contact between the languages may explain why the Aputai dialect from Ilputih has a higher percentage of lexical similarity with Talur (average 57%) than the other languages of the Wetar Cluster (average 52%). The pattern in the figures of lexical similarity for Talur and Aputai suggests what Simons (1977) calls a 'pattern of convergence'. Talur being the dominant language of the area may explain why Ilputih Aputai shows a pattern of convergence with Talur in the matrix.

5.2.3 Ili'uun

Ili'uun is spoken in the villages of Telemar, Karbubu, Klishatu, Ilmaumau, Eray, Nabar, and Esulit on the western end of Wetar, and in Ustutun village on Lirang, a small island off the south-west tip of Wetar. The total population is 1400+. Ustutun is the most influential village in the area and the most progressive. Because of the extreme tidal range, the beach is used for boat repair by people from around the region.

In Riedel (1886), Salzner (1960) and Wurm and Hattori (1981), the present day Ili'uun language area was divided into the Hahutan and Limera dialects. On the lexicostatistic data alone, there is no basis for dividing the Ili'uun language into dialects today, though there may be other linguistic or social factors that would. Also, no-one I talked to was familiar with the names Hahutan and Limera. The village head of Telemar reports that there used to be three dialects in this area, Jeh, Ili'uun and Juru, but now all use the Ili'uun dialect. Ustutun is in the old Ili'uun dialect area, and being the most influential village on the west end of the island, it seems that the other dialects have converged with it or died out.

5.2.4 Tugun

The Tugun language is spoken in the villages of Mahuan, Masapun, Tomliapat, Ilpokil, Kahailin, Ilway, and Arwala on the south-east end of Wetar. The people report little or no variation between the villages, except for Arwala, which is claimed to have different intonation. Hunters use a special speech register, consisting of a limited set of unique vocabulary items which are only used in the interior regions of the island while hunting game. The total population for this group is 1200+.

The only alternate name, Tutunohan, comes from the earlier works of Riedel (1886), Salzner (1960) and Wurm and Hattori (1981). The Tutunohan dialect area included the present day Tugun language area, covering the region between Ilputih on the south coast,

the villages on the east end of the island, and the villages on the north coast to Neumatang. The results of this survey, however, show a completely different picture. Now there are three distinct language groups living in what was previously called the Tutunohan area. As for the name Tutunohan, I could find no-one who recognised it.

6 Notes on Tugun phonology and grammar

In June of 1989 and January of 1990, I was able to return to the Tugun area, and collected data each time for three weeks in several villages. The following is a brief summary of some aspects of Tugun phonology and grammar.

6.1 Phonology

Vowels: Tugun has five vowels: /i e a o u/. The vowel /e/ has allophones [e] and [ɛ]. The details of the distribution of [ɛ] are still unclear, though it always occurs before /r/ in a closed syllable: *ɛɛr* 'water', *tɛɛr* 'ask'. The vowel u has allophones [u] and [ʊ]. Following /m/, [ʊ] is the allophone of /u/: *mʊʊ* 'banana', *mʊmʊ* 'machete' though only in a few examples have been observed at this point in the analysis.

Consonants: The consonant phonemes are given below. (See Appendix B for phonetic values)

		labial	apical	laminal	dorsal	glottal
Stops	vl	p	t	c	k	ʔ
	vd			j	g	
Fricative	vl	f	s			h
	vd	v				
Nasal		m	n		ŋ	
Trill			r			
Lateral			l			

Figure 5: Tugun consonants

The consonant phoneme /v/ has two allophones [v] and [w] which alternate with each other. Examples: *vatin ~ watin* 'place'; *vina ~ wina* 'my mother'; *vai ~ wai* 'big'. During elicitation, my main informant, a 45-year old male, gave the v–form first, and if questioned on the pronunciation, gave the w-form. A 15-year old informant gave only the w–form.

The phoneme /r/ tends to be trilled if it occurs word-final or in slow speech, and elsewhere is flapped. Glottal /ʔ/ is of questionable phonemic status as it occurs: medially in only a few words, e.g. *namʔii* 'good' (but *namii* 'he dreams'), and *-iʔai* 'cook'; in verb morphology (1pe *amʔ*, see §6.3 below); and before word-initial vowels, where it is non-contrastive.

6.2 Stress placement

In general stress falls on the penultimate syllable of words. In the following examples the stressed syllable is underlined: *nalun* 'food', *juun* 'find', *kaliran* 'spouse'. There are a few

exceptions, where often the aspectual clitics *-e* and *-me* shift the stress to the ultimate syllable of the root, e.g. *mu-falik* 'I will-return' but *mu-falik-me* 'I will-return-ASP'.

6.3 Personal pronouns

The personal pronouns are classified by number (singular or plural), by person (first, second, or third), and by inclusive versus exclusive reference in the first person plural. The free pronouns do not change form for different syntactic functions. [8]

Tugun has four sets of subject agreement affixes on verbs, determined by phonotactics and mood. The relevant phonotactics are whether the verb stem is consonant-initial or vowel-initial. The mood distinction is between irrealis and realis. The various sets are given below:

	Free pronouns	C-Initial Irrealis	Realis	V-Initial Irrealis	Realis
1s	*au*	*mu-*	*u-*	*mu-*	*v-*
2s	*oo*	*om-*	*o-*	*om-*	*m-*
3s	*nii*	*ma-*	*ø-*	*man-*	*n-*
1pe	*ami*	*am-*	*am-*	*am ?-*	*am ?-*
1pi	*ita*	*ka-*	*it-*	*kat-*	*it-*
2p	*mii*	*mar-*	*mi-*	*mar-*	*mir-*
3p	*hira*	*mar-*	*ra-*	*mar-*	*r-*

Figure 6: Tugun pronouns and subject agreement forms

Duals with the free pronouns are formed by combining them with the numeral root for two *rua*: *ita rua* 'we (incl) two'; *mi-rua* 'you two'; *hira rua*, *hi-rua* 'the two of them'.

Duals in the subject agreement system are derived from the root for 'two' *rua*: *ka-ru-* '1du', *ru-* '3du'.

(7)   *ka-ru-falik*   'Let's us two return'
      ru-kefe   'two-carried'

6.4 Possessives

Possession is expressed by placing the personal pronouns (full or cliticised form) before the possessed object.

1du	first person dual	3s	third person singular
1pe	first person exclusive	ASP	aspectual clitic
1pi	first person inclusive	incl	inclusive
1s	first person singular	lgs	languages
2p	second person plural	vd	voiced
2s	second person singular	vl	voiceless
3p	third person plural		

(8) *au mumu*   'my machete'
     o-raha   'your house'
     ni-ama   'his father'
     ita ana   'our (incl) child'
     ami raha   'our (excl) house'
     hira vatin   'their place'

The first-person singular bound pronoun has two forms, /u-/ and /v-/. For nouns beginning with a consonant, /u-/ is used. For vowel–initial nouns, /v-/ is used.

(9) *u-mumu*   'my machete'
     u-raha   'my house'

     v-ama   'my father'
     v-ina   'my mother'

Possession is also expressed by using the verbal *ene* 'to own'. This follows the patterns described above for verbal inflection.

(10) *ni-ene mumu ge*   'that's his machete'
      au v-ene   '[it's] mine'

7 Summary

Wetar has been a relatively little studied island in the past, and though the present study brings to light some significant advances on what was previously understood, much research remains to be done. The significance of this study is that it not only provides data that can not be found elsewhere, but it also provides a comprehensive survey of Wetar, covering every community on the island. Past literature concerning Wetar has given a much different picture on the languages of this island. Some of the differences may be explained in terms of migration or convergence, but most probably resulted from a lack of data originally. Five distinct groups are apparent from the survey data, and have been tentatively considered separate languages. Dialect intelligibility testing, as well as sociolinguistic research, would greatly enhance the understanding of the languages of this island.

Appendix A: Information on word lists

Word list number	Village name and (language)	Name of primary informant(s) (age)
1.	Moning (Perai)	Sakius (54)
2.	Uhak (Perai)	Iace Makesso (25); Y.S. Makesso (47)
3.	Ilputih(b) (Aputai)	Joseph Manuka (37)
4.	Ilputih(a) (Talur)	village head (60+)
5.	Hiay (Talur)	Otis Maunari (22)
6.	Ilwaki (Talur)	Octovianus Naharuik (62)
7.	Mahuan (Tugun)	Solemar Lutpuru (46)
8.	Masapun (Tugun)	Ismael Malali (46)
9.	Tomliapat (Tugun)	Soleman Lainjais (57); Karel Panrati (14)
10.	Ilpokil (Tugun)	Gerson Pakolay (33)
11.	Kahailin (Tugun)	Andanas Maumaja (22)
12.	Ilway (Tugun)	Samuel Samat (30); Thomas Unukoly (45)
13.	Arwala (Tugun)	Zekarias Sindjadje (30); Oktupianus Maukaut (38)
14.	Telemar (Ili'uun)	Pietar Maray (45); Jehudas Maehen (41)
15.	Karbubu (Ili'uun)	Moses Maia (46); Ingalina Maia (29)
16.	Ustutun (Ili'uun)	Lazarus Malau (40); Edward Maika (47)
17.	Klishatu (Ili'uun)	Abraham Mamonu (34); Set Makima (47)
18.	Ilmaumau (Ili'uun)	Domingus Malikang (18)
19.	Eray (Ili'uun)	Agustinus Debanse (34); Yesayaw Heruk (37)
20.	Nabar (Ili'uun)	Yan Mara (27); Benjamin Mamaga (47)
21.	Esulit (Ili'uun)	Samuel Magareng (15); Paulos Magumi (46)

Appendix B: Comparative word lists

Word lists elicited during this survey are reproduced below, with a description of the symbols used. Representative word lists were chosen from each language group as follows: Uhak (Perai); Ilputih(b) (Aputai); Ilwaki (Talur); Ilway (Tugun); Ilmaumau (Ili'uun).

[Editor's note: Since writing this paper the author has spent ten months living in the Tugun speaking area doing research on that language. He writes, "I know that there are no monosyllabic words in Tugun, and I suspect the same is true for the others. When I took the word lists, I was not in tune to vowel length, and didn't record it. Only after spending more time in the Tugun language did I analyse it as a sequence of like vowels." Since this is also true for languages in the wider region, monosyllabic lexical words (i.e. non-functors) in the following lists should be reinterpreted as having sequences of like vowels. E.g. /un/ should be assumed to be /uun/, /ron/ to be /roon/, and so forth. Word stress remains penultimate. --CEG]

Symbols

VV	Long vowels are marked by a double symbol.
ṽ	Vocoid with tilda [~] diacritic is nasalised.
'	Stress in Wetar languages is generally penultimate. It is marked with an apostrophe preceding the stressed syllable where different.

Contoids

p	Voiceless bilabial stop	*j*	Voiced alveopalatal grooved affricate
f	Voiceless labiodental fricative	*k*	Voiceless velar stop
v	Voiced labiodental fricative	*g*	Voiced velar stop
t	Voiceless alveolar stop	*ʔ*	glottal stop
d	Voiced alveolar stop	*h*	Voiceless glottal fricative
r	Alveolar trill or flap	*m*	Voiced bilabial nasal
l	Lateral	*n*	Voiced alveolar nasal
s	Voiceless alveolar grooved fricative	*ŋ*	Voiced velar nasal
c	Voiceless alveopalatal grooved affricate	*w*	Non-syllabic /u/

Vocoids

i	High tense front unrounded vocoid	*a*	Low central unrounded vocoid
e	Mid tense front unrounded vocoid	*u*	High tense back rounded vocoid
ɛ	Mid lax front unrounded vocoid	*o*	Mid tense back rounded vocoid

	1.	2.	3.	4.	5.
English	root	tree	leaf	thorn	coconut shell
Indonesian	akar	pohon	daun	duri	tempurung
PAn[9]	*(w)akaR	*kaSiw/*puqun	*DaSun	*DuRi	
Perai	*akar*	*ai*	*ron*	*ruriŋ*	*kanuruŋ*
Aputai	*ramuk*	*un*	*ron*	*rurin*	*suaŋ*
Tugun	*akar*	*ahan*	*roon*	*rurin*	*tau*
Ili'uun	*aka*	*un*	*ron*	*rurin*	*taknuru*
Talur	*amut*	*un*	*ron*	*ruin*	*kanurak*

	6.	7.	8.	9.	10.
English	coconut	banana	betelnut	rattan	bamboo
Indonesian	kelapa	pisang	sirih	rotan	bambu
PAn	*niuR			*quay	
Perai	*nor*	*mũ*	*malu*	*ua*	*ulak*
Aputai	*nor*	*mũ*	*malu*	*ua*	*ulak*
Tugun	*noor*	*mũũ*	*malu*	*ua*	*ulak*
Ili'uun	*no*	*mũ*	*malu*	*ua*	*o*
Talur	*nõ*	*mu ʔu*	*huru ron*	*ua*	*petun*

	11.	12.	13.	14.	15.
English	flower	fruit	kind of grass	sago	hulled rice
Indonesian	bunga	buah	kusu-kusu	sagu	beras
PAn	*buŋa	*buaq			*bəRas/*pajəy
Perai	*funaŋ*	*fuaŋ*	*leteŋ*	*kamariu*	*reha*
Aputai	*funan*	*fuaŋ*	*nau*	*kamariu*	*rehar*
Tugun	*funan*	*fuaŋ*	*nau*	*kambriu*	*paci*
Ili'uun	*hunã*	*huã*	*nau*	*kamariu*	*hala*
Talur	*hunan*	*huan*	*kli*	*kamdiu*	*papreas*

	16.	17.	18.	19.	20.
English	I	you	we(excl)	we(incl)	sun
Indonesian	aku	engkau	kami	kita	matahari
PAn	*i-aku		*i-kami	*i-kita	*qaləjaw
Perai	*au*	*o*	*ami*	*ita*	*leo*
Aputai	*au*	*o*	*ami*	*ita*	*leo*
Tugun	*au*	*oo*	*ami*	*ita*	*leo*
Ili'uun	*au*	*o*	*ami*	*ita*	*leo*
Talur	*ga ʔu*	*go*	*gami*	*gita*	*lelo*

---

[9] [EDITOR'S NOTE: The comparative and historical notes found in this appendix are editiorial additions. The reconstructions are variously at the level of PAn, PMP, or PCEMP, and are found in the general Austronesian literature.]

	21.	22.	23	24.	25.
English	moon	star	water	rain	stone
Indonesian	bulan	bintang	air	hujan	batu
PAn	*bulan	*bituqən	*waSir	*quZaN	*batu
Perai	*fulaŋ*	*tu*	*ɛr*	*usa*	*fatu*
Aputai	*fulan*	*asa*	*ɛr*	*usa*	*fatu*
Tugun	*fulan*	*tu*	*ɛɛr*	*usa*	*fatu*
Ili'uun	*hulã*	*tu*	*e*	*husã*	*hatu*
Talur	*hulan*	*ada*	*we*	*usan*	*hatu*

	26.	27.	28.	29.	30.
English	sand	island	land	cloud	wind
Indonesian	pasir	pulau	tanah	awan	angin
PAn	*qənay	*nusa	*DaRəq		*aŋin
Perai	*moeŋ*	*nusa*	*rere*	*kakaŋ*	*aniŋ*
Aputai	*moeŋ*	*nusa*	*rare*	*natas*	*anin*
Tugun	*moen*	*nusa*	*rare*	*natas*	*anin*
Ili'uun	*moẽ*	*nusa*	*rare*	*nata*	*ani*
Talur	*dadean*	*nusa*	*rea*	*natas*	*anin*

	31.	32.	33.	34.	35.
English	sea	forest	mountain	fire	smoke
Indonesian	laut	hutan	gunung	api	asap
PAn	*laSud	*alas		*apuy	*abu 'dust'
Perai	*lor*	*alas*	*uluŋ*	*aimoriŋ*	*mahun*
Aputai	*lor*	*ala*	*huhun*	*aimoriŋ*	*mahun*
Tugun	*loor*	*ala*	*utur*	*aimoriŋ*	*mahun*
Ili'uun	*lo*	*letẽ*	*otu*	*aimorin*	*mahun*
Talur	*lor*	*alas*	*huhun*	*aimorin*	*masun*

	36.	37.	38.	39.	40.
English	ashes	hot	burn/roast	cold	night
Indonesian	abu	panas	bakar	dingin	malam
PAn	*abu	*ma-panas	*tunu	*ma-diŋdiŋ	*bəŋi
Perai	*karono*	*pana*	*tunu*	*mohoŋ*	*sema*
Aputai	*alaman*	*panas*	*tunu*	*mohon*	*sema*
Tugun	*afun*	*pana*	*tunu*	*mohon*	*cema*
Ili'uun	*krahu*	*pana*	*tunu*	*ciri*	*heni*
Talur	*ahun*	*panas*	*tunu*	*nadin*	*henin*

	41.	42.	43.	44.	45.
English	fish	bird	egg	dog	flying fox
Indonesian	ikan	burung	telur	anjing	kalong
PAn	*Sika?ən	*manuk	*qatəluR	*asu	*paniki
Perai	*meti*	*manu*	*telur*	*ahu*	*faniki*
Aputai	*meti*	*manu*	*telun*	*ahu*	*rego*
Tugun	*ika*	*manu*	*telur*	*ahu*	*faniki*
Ili'uun	*meti*	*manu*	*telu*	*asu*	*hiki*
Talur	*i?an*	*manu*	*telun*	*asu*	*kni?ĩ*

	46.	47.	48.	49.	50.
English	louse	mosquito	rat	snake	tail
Indonesian	kutu	nyamuk	tikus	ular	ekor
PAn	*kutu	*ñamuk	*ka-labaw	*sawa	*ikuR
Perai	utu	karhoik	palaho	so	ikur
Aputai	utu	utuŋ	palaho	sao	ikur
Tugun	utu	namu	laso	sao	ikur
Ili'uun	utu	sisi	laso	sao	iku
Talur	utu	namu	plaho	sau	iʔun

	51.	52.	53.	54.	55.
English	black	white	red	yellow	green
Indonesian	hitam	putih	merah	kuning	hijau
PAn	*ma-qitəm	*ma-putiq	*ma-iRaq		
Perai	metaŋ	puti	mera	pola	mosoŋ[10]
Aputai	metaŋ	puti	mera	pola	moso
Tugun	metan	puti	mera	pola	moso
Ili'uun	metã	puti	mera	pola	mosõ
Talur	metan	puti	mea	mara	moso

	56.	57.	58.	59.	60.
English	one	two	three	four	five
Indonesian	satu	dua	tiga	empat	lima
PAn	*asa / *əsa	*DuSa	*təlu	*Səpat	*lima
Perai	hasa	rua	telu	aak	lima
Aputai	hasa	rua	telu	aak	lima
Tugun	hasa	farua	fatelu	faʔaat	falima
Ili'uun	eha	harua	hatelu	haʔaa	halima
Talur	nehe	irua	itelu	inhat	ilima

	61.	62.	63.	64.	65.
English	six	seven	eight	nine	ten
Indonesian	enam	tujuh	delapan	sembilan	sepuluh
PAn	*ənəm	*pitu		*siwa	*puluq
Perai	neŋ	itu	kau	sia	sanulu
Aputai	neŋ	itu	kau	sia	sanuluŋ
Tugun	faneŋ	faʔitu	fakau	fasia	sanulu
Ili'uun	hane	haʔitu	hakau	hasia	sanulu
Talur	inen	ihitu	ihau	isia	sanulu

---

[10] Compare Makassar *moncoŋ* 'green'.

	66.	67.	68.	69.	70.
English	twenty	one hundred	one thousand	all	many
Indonesian	dua puluh	seratus	seribu	semua	banyak
PAn		*Ratus	*Ribu		
Perai	*felerua*	*para*	*riuŋ*	*tema*	*hoe*
Aputai	*felerua*	*atus*	*rihun*	*tatema*	*aru*
Tugun	*felrua*	*ratu*	*rihun*	*tatema*	*aru*
Ili'uun	*helerua*	*atu*	*rihuŋ*	*puhi*	*aru*
Talur	*ruanulu*	*atus*	*rihun*	*tatema*	*waʔi*

	71.	72.	73.	74.	75.
English	big	small	long	short	near
Indonesian	besar	kecil	panjang	pendek	dekat
PAn					*qadaŋi
Perai	*wai*	*tawai*	*poha*	*kapoir*	*katesi*
Aputai	*wai*	*kamais*	*poha*	*kapoir*	*karinik*
Tugun	*vai*	*anan*	*poha*	*poir*	*rani*
Ili'uun	*ai*	*ana*	*poha*	*poi*	*rani*
Talur	*waki*	*kenok*	*naru*	*padak*	*ranik*

	76.	77.	78.	79.	80.
English	far	full	new	good	dry
Indonesian	jauh	penuh	baru	baik	kering
PAn	*Zauq	*ma-pənuq	*baqəRu		*kəRaŋ
Perai	*so*	*penu*	*feru*	*eaŋ*	*geraŋ*
Aputai	*so*	*penu*	*feru*	*eaŋ*	*geraŋ*
Tugun	*soo*	*penu*	*faferu*	*namʔii*	*geran*
Ili'uun	*so*	*penu*	*heruŋ*	*raja*	*gera*
Talur	*so*	*penu*	*haheu*	*lalan*	*namgetun*

	81.	82.	83.	84.	85.
English	to dry	no	this	that	here
Indonesian	menjemur	tidak	ini	itu	di sini
Perai	*rinaŋ*	*tora*	*ge*	*ko*	*ge*
Aputai	*rinaŋ*	*tora*	*ge*	*ko*	*ge*
Tugun	*rinan*	*toa*	*ge*	*ko*	*gee*
Ili'uun	*rese*	*tora*	*ge*	*go*	*ge*
Talur	*reʔar*	*taik*	*ileʔe*	*ilaa*	*ileʔe*

	86.	87.	88.	89.	90.
English	there	inside	above	outside	underneath
Indonesian	di situ	di dalam	di atas	di luar	di bawah
PAn		*Daləm			
Perai	*ko*	*ralaŋ*	*otur*	*milaŋ*	*leŋ*
Aputai	*go*	*ralaŋ*	*sasiŋ*	*liur*	*pair*
Tugun	*koo*	*ralan*	*legar*	*milan*	*loin*
Ili'uun	*go*	*ralan*	*jela*	*kakeo*	*orã*
Talur	*ilaa*	*ralan*	*upus*	*tuʔan*	*rea*

English	**91.** in front	**92.** behind	**93.** edge	**94.** skin	**95.** meat/flesh
Indonesian	di depan	di belakang	pinggir	kulit	daging
PAn				*kulit*	*isi*
Perai	*oeŋ*	*tinaŋ*	*seriŋ*	*ɛlik*	*ihi*
Aputai	*oiŋ*	*tina*	*seriŋ*	*ɛlik*	*musan*
Tugun	*oe*	*tuan*	*seriŋ*	*ulik*	*musan*
Ili'uun	*õẽ*	*liu*	*seriŋ*	*uli*	*ihi*
Talur	*ain*	*tuʔan*	*serin*	*ulit*	*isin*

English	**96.** blood	**97.** bone	**98.** fat	**99.** hair	**100.** body hair
Indonesian	darah	tulang	gemuk	rambut	bulu
PAn	*DaRaq*	*duRi*			*buluq*
Perai	*raraŋ*	*ruriŋ*	*mair*	*kuru*	*fuluŋ*
Aputai	*raraŋ*	*ruriŋ*	*mair*	*kuru*	*fulun*
Tugun	*raran*	*rurin*	*mair*	*kuroon*	*fulun*
Ili'uun	*rara*	*rurin*	*mai*	*kuru'ron*	*huluŋ*
Talur	*ran*	*ruin*	*poʔur*	*gare'ron*	*hulur*

English	**101.** head	**102.** ear	**103.** eye	**104.** nose	**105.** mouth
Indonesian	kepala	telinga	mata	hidung	mulut
PAn			*mata*		*babaq*
Perai	*kuruʔatu*	*kiniŋ*	*mata*	*nunur*	*hafa*
Aputai	*kaʔatun*	*kiniŋ*	*mata*	*nunur*	*hafa*
Tugun	*kuratun*	*kinin*	*mata*	*kunai*	*kanar*
Ili'uun	*karatuŋ*	*kini*	*mataŋ*	*kunai*	*nũnũ*
Talur	*garenhatun*	*kina*	*matar*	*inur*	*hahar*

English	**106.** lip	**107.** tooth	**108.** tongue	**109.** nail	**110.** foot
Indonesian	bibir	gigi	lidah	kuku	kaki
PAn		*ŋisi*			
Perai	*ipur*	*ni*	*na*	*kalusuŋ*	*luku*
Aputai	*ipur*	*ni*	*nã*	*kalusuŋ*	*luku*
Tugun	*ipur*	*niin*	*naan*	*kilusu*	*lukuŋ*
Ili'uun	*ipu*	*ni*	*na*	*kalusuŋ*	*luku*
Talur	*ipur*	*nis*	*nanar*	*lumatan*	*en*

English	**111.** knee	**112.** hand	**113.** stomach	**114.** neck	**115.** breast
Indonesian	lutut	tangan	perut	leher	susu
PAn	*luluj*	*lima*	*tian/*kə(m)puŋ*		*susu*
Perai	*tarulu*	*lima*	*teŋ*	*genu*	*susu*
Aputai	*aturulu*	*lima*	*teŋ*	*genun*	*susu*
Tugun	*tiluru*	*liman*	*kanen*	*genuŋ*	*susu*
Ili'uun	*klulu*	*lima*	*kapuŋ*	*genuŋ*	*susu*
Talur	*laloun*	*liman*	*apun*	*nuan*	*susu*

English	116. liver	117. face	118. faeces	119. urine	120. deaf
Indonesian	hati	muka	tahi	air kencing	tuli
PAn	*qatay		*taqi	*təDiq	
Perai	rumaŋ	oe	te	tarik	pu
Aputai	rumaŋ	oe	te	tarik	puŋ
Tugun	fua	oe	tee	tarik	upu
Ili'uun	ateŋ	õẽ	te	tari	puŋ
Talur	aten	ain	tek	sumer	par

English	121. blind	122. person	123. man	124. woman	125. husband
Indonesian	buta	orang	laki-laki	perempuan	suami
PAn	*buta		*maRuqanay	*bata	*qasawa
Perai	peke	aitopu	amene	afata	riu
Aputai	puta	aitopu	amaŋ	inaŋ	riu
Tugun	purur	aitopu	mamane	fafata	kaliran
Ili'uun	puta	atou	mamane	hahata	riu
Talur	puta	itopu	mamane	papata	soun

English	126. wife	127. father	128. mother	129. name	130. grandfather
Indonesian	istri	bapak	ibu	nama	tete/kakek
PAn	*qasawa	*ama	*ina	*ŋajan	*u(m)pu/*laki 'male'
Perai	riu	ama	ina	neaŋ	apulaik
Aputai	riu	ama	ina	neaŋ	apulaik
Tugun	kaliran	ama	ina	nean	apulaik
Ili'uun	riu	ama	ina	nẽã	pauŋ
Talur	soun	aman	inan	naran	tete

English	131. grandmother	132. elder sibling	133. younger sibling	134. grandchild	135. uncle
Indonesian	nenek	kakak	adik	cucu	om
PAn	*bəi 'aunt'		*Sua(n)ji	*u(m)pu	
Perai	apufeik	po	ali	kapu	pani
Aputai	apufeik	pou	ali	kapu	opu
Tugun	apufeik	pɔu	ali	kapu	pani
Ili'uun	pauŋ	pouŋ	ali	peŋ	lai
Talur	nene	poun	ali	pein	aki

English	136. aunt	137. friend	138. guest	139. rope	140. road
Indonesian	tante	kawan	tamu	tali	jalan
PAn		*kawan		*tali	*zalan
Perai	foi	kawair	panege	asar	sala
Aputai	foi	kawair	panege	asar	sala
Tugun	foi	talin	tamu	asar	sala
Ili'uun	hai	aliŋ	panage	tali	salã
Talur	pani	huwalan	panege	tali	salan

	141.	142.	143.	144.	145.
English	boat	sail	machete	knife	rice mortar
Indonesian	perahu	layar	parang	pisau	lesung
PAn		*layaR*			*ləsuŋ*
Perai	*loi*	*lar*	*perik*	*perik*	*kanehuŋ*
Aputai	*loi*	*lar*	*mūmū*	*sihik*	*kanehuŋ*
Tugun	*loi*	*laar*	*mūmū*	*perik*	*kanehuŋ*
Ili'uun	*loi ʔina*	*la*	*api'suŋ*	*mūmū*	*knehuŋ*
Talur	*ro*	*lan*	*itur*	*sihit*	*nesun*

	146.	147.	148.	149.	150.
English	rice stamper	salt	sugar	what	who
Indonesian	alu	garam	gula	apa	siapa
PAn	*qaSəlu*	*ma-qasin 'salty'*		*sapa*	*i-sai*
Perai	*alu*	*masiŋ*	*gula*	*hala*	*hari*
Aputai	*alu*	*masiŋ*	*gula*	*hala*	*ari*
Tugun	*alu*	*mahin*	*gula*	*hala*	*hari*
Ili'uun	*alu*	*masiŋ*	*gula*	*hala*	*hari*
Talur	*alu*	*masin*	*gula*	*la*	*se*

	151.	152.	153.	154.	155.
English	where	why	how many	thirsty	bite
Indonesian	di mana	kenapa	berapa	haus	menggigit
PAn			*pija*		
Perai	*'oinume*	*oinala*	*fa ʔira*	*mete ɛr*	*popar*
Aputai	*lamei*	*alame*	*fa ʔira*	*paro*	*la ʔi*
Tugun	*oinumee*	*onihaala*	*fa ʔira*	*matroo*	*gaga*
Ili'uun	*aci'me*	*nohila*	*ha ʔira*	*mate e*	*kosa*
Talur	*gamai*	*labe*	*ihila*	*apro*	*hisi*

	156.	157.	158.	159.	160.
English	drink	hungry	eat	see	hear
Indonesian	minum	lapar	makan	melihat	mendengar
PAn	*ma-inum*	*lapaR*	*ka ʔən*		
Perai	*minu*	*hahi*	*mua*	*gua*	*liuŋ*
Aputai	*minu*	*palar*	*moa*	*nauŋ*	*liuŋ*
Tugun	*minu*	*matlaar*	*mua*	*coco*	*liu*
Ili'uun	*rinu*	*kla*	*ra*	*oco*	*liu*
Talur	*renu*	*apla*	*a ʔan*	*roga*	*dahar*

	161.	162.	163.	164.	165.
English	sleep	know	wake up	dead	swim
Indonesian	tidur	tahu	bangun	mati	berenang
PAn	*qinəp*	*taqu*	*[ma]-baŋun*	*maCey*	*naŋuy*
Perai	*pu*	*sasak*	*pono*	*mete*	*nani*
Aputai	*lole*	*nauŋ*	*pano*	*mate*	*nani*
Tugun	*nipuur*	*sɛrek*	*pono*	*mate*	*nani*
Ili'uun	*rapu*	*nanau*	*panu*	*mate*	*nani*
Talur	*rena*	*nau*	*mesu*	*mate*	*nani*

English	**166.** walk	**167.** sing	**168.** vomit	**169.** laugh	**170.** itch
Indonesian	berjalan	bernyanyi	muntah	tertawa	gatal
PAn			*ma-qutaq	*malip	*gatəl
Perai	*akak*	*loli*	*muta*	*mali*	*gataŋ*
Aputai	*ahak*	*loli*	*muta*	*mali*	*gataŋ*
Tugun	*akat*	*sani*	*muta*	*mali*	*gatar*
Ili'uun	*raloa*	*sani*	*muta*	*mali*	*gatã*
Talur	*sole*	*loli*	*pura*	*mali*	*gatar*

English	**171.** sleepy	**172.** child	**173.** sky	**174.** to fly	**175.** come
Indonesian	mengantuk	anak	langit	terbang	datang
PAn		*anak	*atas 'above'		*maRi/*mai
Perai	*fɛrak*	*ana*	*mota*	*sopaŋ*	*ma*
Aputai	*pulir*	*anuran*	*mota*	*palia*	*ma*
Tugun	*fɛrak*	*ana*	*natas*	*sopar*	*maa*
Ili'uun	*hera*	*ana*	*nata*	*palia*	*ma*
Talur	*dukur*	*ana*	*natas*	*nemo*	*mai*

English	**176.** stand	**177.** cough	**178.** spit	**179.** cry	**180.** dream
Indonesian	berdiri	batuk	meludah	menangis	bermimpi
PAn	*DiRi			*taŋis	*S-um-ipi
Perai	*piri*	*koser*	*pita*	*nage*	*nami*
Aputai	*piri*	*koser*	*pik'pre*	*rage*	*rami*
Tugun	*piri*	*koho*	*pita aprei*	*kua*	*nami*
Ili'uun	*piri*	*koho*	*pita apre*	*kua*	*rami*
Talur	*apri*	*mahu*	*puituper*	*tanis*	*nami*

English	**181.** pregnant	**182.** old	**183.** river	**184.** kill	**185.** sit
Indonesian	mengandung	tua	sungai	membunuh	duduk
PAn	*tian 'belly'				
Perai	*koru*	*laik*	*oraŋ*	*eja*	*tur* [11]
Aputai	*kamutia*	*laik*	*oraŋ*	*ega*	*tur*
Tugun	*kapkorun*	*agin*	*pair*	*eja*	*tuur*
Ili'uun	*magutia*	*lai*	*pai*	*eja*	*pasu*
Talur	*sau*	*laʔit*	*meta*	*ega*	*adei*

---

[11] Compare Tetun *tuur* 'sit'.

	186.	187.	188.	189.	190.
English	bathe	fall	give	speak	leech
Indonesian	mandi	jatuh	memberi	bicara	lintah
PAn	*diRus				
Perai	*topur*	*mosu*	*ne*	*salaŋ*	*heti*
Aputai	*topur*	*fonuk*	*ne*	*oruŋ*	*makak*
Tugun	*topur*	*pole*	*nee*	*nihcalan*	*makat*
Ili'uun	*topu*	*pole*	*ne*	*coca*	*maka*
Talur	*rius*	*ledir*	*ne*	*teten*	*makat*

	191.	192.	193.	194.	195.
English	thick	thin	sharp	rotten	shoulder carry
Indonesian	tebal	tipis	tajam	busuk	pikul
PAn	*kapal	*ma-nipis	*taZim	*buRuk	
Perai	*magas*	*mis*	*taiŋ*	*pouŋ*	*teruŋ*
Aputai	*tun*	*mis*	*nagan*	*fouŋ*	*fara*
Tugun	*kavar*	*miis*	*tain*	*poon*	*fara*
Ili'uun	*kabaŋ*	*mi*	*taiŋ*	*hõ*	*hara*
Talur	*tuʔun*	*mis*	*preʔat*	*hõn*	*san*

	196.	197.	198.	199.	200.
English	run	hit	wound	house	village
Indonesian	lari	pukul	luka	rumah	kampung
PAn	*laRiw	*pukpuk/*palu		*RumaH	
Perai	*koe*	*fu*	*mani*	*raha*	*ili*
Aputai	*koe*	*kalu*	*maniŋ*	*raha*	*ili*
Tugun	*huuk*	*fuu*	*maniŋ*	*raha*	*ili*
Ili'uun	*ragarusu*	*kalu*	*mani*	*raha*	*ili*
Talur	*lai*	*hehas*	*mani*	*umah*	*ili*

	201.	202.	203.
English	brain	sweat	coral
Indonesian	otak	berkeringat	karang
PAn	*qutək		
Perai	*gutaŋ*	*kamorah*	*kura*
Aputai	*gutaŋ*	*kamemaŋ*	*kura*
Tugun	*gutaŋ*	*kaneus*	*aur*
Ili'uun	*gutaŋ*	*klema*	*krosa*
Talur	*utak*	*kmeman*	*nele*

References

Blust, Robert, 1977, The Proto-Austronesian pronouns and Austronesian sub-grouping: a preliminary report. *Working Papers in Linguistics*, University of Hawaii 9/2:1–15.

Capell, A., 1943–44, Peoples and languages of Timor. *Oceania*, 14/3:191–219, 14/4:311–337, 15/1:19–48.

Casad, Eugene, 1974, *Dialect intelligibility testing*. Summer Institute of Linguistics publications in linguistics and related fields 38, Norman: The Summer Institute of Linguistics of the University of Oklahoma.

Dyen, Isadore, 1965, *A lexicostatistical classification of the Austronesian languages*. International Journal of American Linguistics, Memoir 19. IJAL: Baltimore.

Elbert, Johannes, 1912, *Die Sunda-Expedition des Vereins für Geographic und Statistik*. 2 vols. Frankfurt: Hermamm Minjon.

Esser, S.J., 1938, Talen. In *Atlas van tropisch Nederland*. Amsterdam: Koninklijk Nederlandsch Aardrijkskundig Genootschap.

Gudschinsky, Sarah, 1956, The ABC's of lexicostatistics (glotto-chronology). *Word* 12:175–210. Reprinted in Dell Hymes, ed., 1964, *Language in culture and society*. New York: Harper and Row.

Josselin de Jong, J.P.B. de, 1947, *Studies in Indonesian culture, vol II: the community of Erai (Wetar)*. Verhandelingen Der Koninklijke Nederlandsche Akademie Van Wetenschappen, AFD. Letterkunde, Amsterdam.

Kantor Statistic Propinsi Maluku, 1987, *Maluku dalam angka*. Ambon: Kantor Statistic Maluku.

Riedel, J.G.F., 1886, *De sluik- en kroesharige rassen tusschen Selebes en Papua*. 's-Gravenhage: Nijhoff.

Salzner, Richard, 1960, *Sprachenatlas des Indopazifischen Raumes*. Wiesbaden: Otto Harrassowitz.

Simons, Gary, 1977, Recognizing patterns of convergence and divergence in a matrix of lexico-statistic relations. In Loving and Simons, eds, *Language variation and survey techniques*. Workpapers in Papua New Guinea Languages, vol 21, 107–134. Ukarumpa: Summer Institute of Linguistics.

Stokhof, W.A.L. ed., 1981, *Holle lists: vocabularies in the languages of Indonesia vol 3/1 Southern Maluku; Central Maluku:Seram (1)*. Canberra: Pacific Linguistics.

Taber, Mark, 1990, A Lexicostatistic Survey of South-Western Maluku. Paper presented at the Maluku Research Conference, University of Hawaii at Manoa, March 16–18, 1990.

1993, Toward a better understanding of the indigenous languages of southwestern Maluku. *Oceanic Linguistics* vol 2, no 32, 389–441. University of Hawaii Press, Honolulu.

Voegelin, C.F. and F.M. Voegelin, 1977, *Classification and index of the worlds languages*. New York: Elsevier.

Wimbish, John S., 1989, *WORDSURV: a program for analyzing language survey word lists*. Dallas: Summer Institute of Linguistics.

Wurm, Stephen A. and Shiro Hattori, eds, 1981–83, *Language atlas of the Pacific area*. 2 volumes. Canberra: The Australian Academy of the Humanities and The Japan Academy. Pacific Linguistics.

6 *The morphology of Dobel, Aru, with special reference to reduplication*

JOCK HUGHES

1 Introduction

The Dobel[1] language is spoken in the Aru islands, which are in the south-east of the Province of Maluku, Indonesia. Dobel is one of the larger languages of the Aru Islands, spoken by about 6,500 people in the language area itself. There are probably at least another 1,000 speakers outside the area, primarily in Dobo and Ambon. Dobel is spoken mainly on the eastern half of the large central island of Kobror and in some villages on neighbouring islands just across the narrow straits of Barakai to the south, and Manombai to the north.

There are three principal dialects of Dobel, Northern, Straits, and South-eastern [see map]. The largest is the Northern dialect (2700), which is spoken in four villages in the north-east of Kobror Island, and in Karwai village, which is just across the mouth of the Manombai strait on the south-eastern tip of Wokam Island. The second is the Straits dialect (1800), spoken in the villages that border the eastern half of the Barakai Strait and the rivers that lead into it, also in one inland village, Jirlay, on the banks of a river that leads northwards to the Manombai strait. The smallest is the South-Eastern dialect (1400), spoken on the east coast of Kobror Island from Ponom, at the mouth of the Barakai Strait, northwards to Warjukur. Within these dialects there is minor variation from village to village. The South-Eastern dialect has several phonological differences from the other two, but this presents little difficulty for intelligibility.

Bordering the Dobel area to the south-east are the three villages of Koba Dangar on Baun Island, and Koba Seltimur and Koba Selfara on Fukarel Island (combined population: 600), where Koba, a language closely related to Dobel, is spoken. Although Koba has some significant phonological and lexical differences from Dobel, it is close enough to Dobel so

---

[1] The name of the language, Dobel [do'bel] is the name used by speakers of the South-Eastern and Straits dialects, and also by outsiders when referring to the language. In the Northern dialect it is known as Doibel [doy'bel].

that, in certain speech domains, speakers of the two languages can sometimes communicate with each other successfully without recourse to Malay.
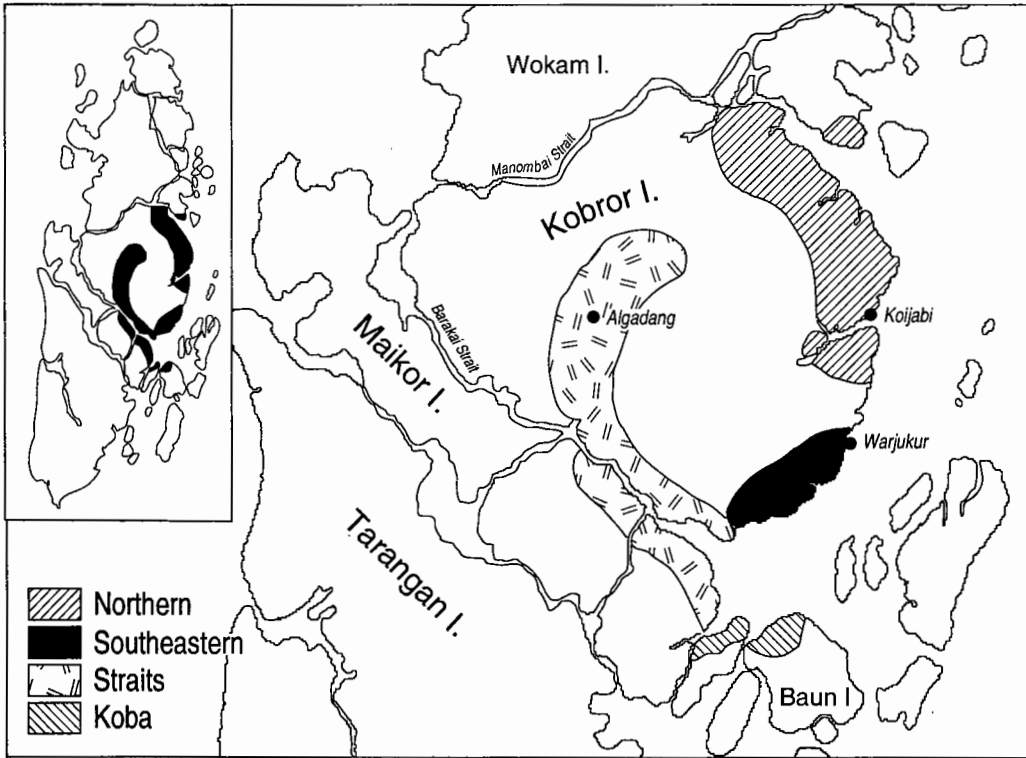
The central Aru Islands, showing the dialects of Dobel
Dobel and the closely related Koba language

References to Malay refer not to Standard Malay, but to Aru Malay, a dialect of Ambonese Malay, which is the lingua franca of the Aru Islands. The following description of Dobel refers to the Northern dialect, specifically as spoken in Koijabi, unless otherwise noted.[2]

---

[2]   The data for this paper were collected while the author was working under the auspices of a cooperative agreement between Pattimura University in Ambon, and SIL.

2 Summary of the phonology and morphophonemics

2.1 Inventory of consonants

Dobel has fourteen consonant phonemes:

		Labial	Apical	Laminal	Dorsal
Plosives	vl.		t		ʔ
Labialised	vl.				kʷ
	vd.	b	d		
Fricatives		ɸ		s	
Nasals		m	n		ŋ
Lateral			l		
Trill			r		
Semivowels		w		y[3]	

Figure 1: Dobel consonants

The /t/ tends to be dental, whereas the other Apical consonants are alveolar.[4]

The phoneme /kʷ/ has variants [kʷ] preceding non-back vowels and [k] preceding back vowels, thus:

/'kʷakʷa/	['kʷakʷə]	'small, child'
/kʷel/	[kʷel]	'younger sibling'
/'rakʷin/	['rakʷin]	'leaf'

/'kʷoyar/	['koyər]	'dog'
/'kʷotu/	['kotu]	'if'
/ʔu-'yakʷur/	[ʔu'yakur]	'I pound'

The phoneme /ɸ/[5] has variants [p] following the phoneme /m/ in word-initial position, and [ɸ] elsewhere, thus:

| /m-'ɸaʔi/ | [m'paʔi] | 'you (sg.) use' |
| /m-'ɸo/ | [m'po] | 'you carry' |

| /'ɸaɸa/ | ['ɸaɸə] | 'ground' |
| /ɸo/ | [ɸo] | 'with' |

The phonetic sequence [ka] occurs rarely. In this environment [k] is best treated as a variant of /ʔ/. The sequence [ka] occurs as a root-initial stressed syllable in some verbs

---

[3]   The laminal semivowel is written as /y/, as in the Dobel orthography, rather than the IPA /j/.

[4]   See Hughes (1989) for a full description of each of the phonemes in Dobel. A word list of Dobel may be found in Hughes (1995b).

[5]   The voiceless bilabial fricative phoneme /ɸ/ will hereafter be written using the symbol /f/, following the Dobel orthography.

where it is in free variation with [ʔa] in the 3s form only[6] There are, however, a few such roots where [ka] may not be replaced by [ʔa].[7]

2.2 Inventory of vowels

Dobel has five vowel phonemes:

	Front	Central	Back
Close	i		u
Mid	e		o
Open		a	

Figure 2: Dobel vowels

The open central vowel phoneme /a/ is realised in stressed syllables as [a] and in unstressed syllables is raised to a schwa-quality allophone (symbolised [ə]). In syllables that carry secondary stress (two syllables before the primary stressed syllable) it is realised as [a].

2.3 The syllable and phoneme distribution

The syllable in Dobel is defined as a rhyme with an obligatory onset, in which any consonant phoneme may occur. A geminate pair of consonants may also occur in the onset.[8] The rhyme contains an obligatory nucleus with an optional coda. The nucleus may consist of any vowel and any member of the consonant sub-class /l r m n w y/ may occur as the coda.[9] The tree in Figure 3 summarises the general features:

---

[6]   Out of eleven verbs with root initial stressed /ʔa/, seven were found where /ʔa/ could freely vary with [ka]. In the Straits dialect, as spoken in Algadang, there are several other verbs where the [k] appears in free variation with [ʔ], and not always in the 3s form nor always before [a].

[7]   I have identified four such morphemes, e.g. *-kaliba* 'coil (rope)'.

[8]   See §4.1 The phonological forms of reduplication.

[9]   The phoneme /ŋ/ may occur in the coda only following morphophonemic changes to the morpheme *-ŋu* [see §2.5]
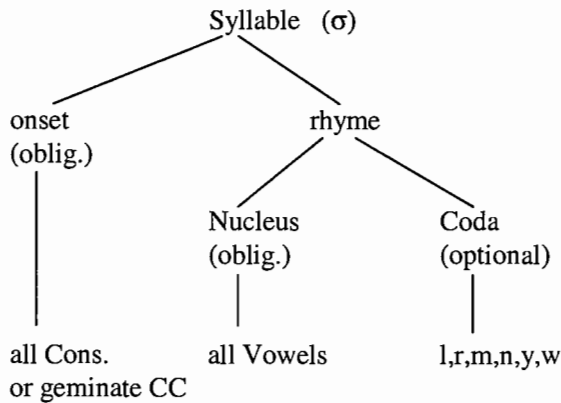
Figure 3: Syllable structure

The fundamental syllable patterns in Dobel are thus CVC and CV, with C_iC_iVC and C_iC_iV patterns occurring only as the stressed syllable of a phonological word, when that word carries the reduplication morpheme [see §4 on reduplication]. There are no vowel clusters in Dobel. The above patterns undergo some restructuring effects under certain morphophonemic processes to produce the otherwise anomalous sequences:

$$Ca^u\eta \qquad\qquad CV^un \qquad\qquad CV^in$$

involving in the first instance an unusual final -η and a preceding unstressed vocoid (^{u}), where the morpheme -ηu metathesises following a root final *a*, and in the second and third instances a root-final *y* or *w* which is reinterpreted as an intrusive vocoid between syllable nucleus and coda [see §2.5 for details].

Other than the geminate consonants which occur initially as a result of reduplication, initial consonant clusters occur in two circumstances. The phoneme /m/ may occur word initially before a syllable-initial consonant, when the 2s Actor proclitic occurs (as it always does) before a root initial consonant, e.g. /m-bana/ 'you go'. The phoneme /n/ also occurs word initially before a syllable-initial consonant in /n'da/, the shortened form of /na?u'da/ 'no, not'. In these two cases the nasal sounds syllabic. Since these exceptions are restricted to the examples given above, they do not necessitate the addition of a new syllable type.

2.4 The phonological word

In Dobel the phonological word is defined as a stress group. Each phonological word is the domain of one primary stress which occurs on one syllable of the phonological word. The following is a list of phonological word types that have been found in Dobel, where the symbol 'S denotes a syllable carrying primary stress and S denotes a syllable not carrying primary stress. This list consists only of monomorphemic words:

'S	/'tay/	'sea water'
'SS	/'suwan/	'naughtiness'
S'S	/ma'del/	'wave'
'SSS	/'lofa?u/	'leaking'
S'SS	/kʷa'lisan/	'sibling (same sex)'
SS'S	/taŋa'lay/	'like, similar to'

Stress always occurs on the grammatical root, and may occur on any of the last three syllables of the root. In the majority of roots it occurs on the penultimate syllable, but there is a significant minority of words where stress occurs on the final syllable.[10] This is a contrastive feature of the language and there are several pairs of words that contrast for stress, such as:

/da-'tabay/	'they carry (on shoulder)'
/da-ta'bay/	'they hit'
/'ʔala-y/	'its/his skin'
/ʔa'lay/	'kind of lemon'
/'tamin/	'house materials'
/ta'nin/	'genuine'

However, there are some cases where final stress is to some extent predictable. Nearly all roots that have the mid vowels /e/ and /o/ in the final syllable have final stress.[11] The following are some examples:

/ʔara'ʔe/	'small sp. of bat'
/ʔa'ye/	'some'
/ʔata'ler/	'vegetable'
/na-fu'rer/	'he coughs'
/ʔa'ro/	'eight'
/fa'no/	'village'
/ŋi'ro/	'mouse'
/ta'ron/	'room'

Stress on the antepenultimate syllable is also predictable, this occurs when the final syllable is a voiceless plosive (/t/ or /ʔ/) followed by /u/.[12] The following are some examples:

/'minaʔu/	'firearm'
/'netaʔu/	'chisel'
/ʔa-'kʷataʔu/	'he feeds'
/'malatu/	'small adze'
/na-ʔa'ʔikatu/	'he pinches'
/'lofaʔu/	'leaking'

This sequence may then be called an extrametrical syllable, as it is not counted in the positioning of the stress. Stress is not affected by the addition of suffixes or enclitics.

---

[10] In a list of 1,920 Dobel polysyllabic roots, there were 388 which had final stress, that is 20%. In this paper stress is only marked when it occurs on the final syllable, or where it is relevant to the discussion.

[11] There are a very few exceptions, such as /da-'bele/ 'praise' and /'sere/ 'characteristic'.

[12] Interestingly the Dobel words ending in voiceless plosive followed by /u/ which have cognates in the West Tarangan language of southern Aru end in a voiceless plosive in West Tarangan, e.g. *minak* 'firearm', (not possible in Dobel syllable structure), and have penultimate stress. (Richard Nivens, pers. comm.). The one exception to this rule in Dobel, /ta'matu/ 'person', is cognate with *ta'mata* in West Tarangan and other Aru languages; in this case the /tu/ is not an extrametrical syllable.

It may be noted at this point that the phonological word does not in every case coincide with the grammatical word. In the word /ʔa-'fan+ti/ '3sA-fall-PERF (he has fallen)', the morpheme "ti" is a clitic which attaches itself to the last word of the clause. 'He has fallen from the steps' is then /ʔa-'fan 'fay 'lola+ti / '3sA-fall from steps-PERF'. It can be seen that, by the principle of moveability, *ti* is grammatically a separate word, although not a separate phonological word since it does not carry stress. There are other clitics which behave similarly.[13]

2.5 Morphophonemic processes

All but one of the morphophonemic processes which occur in Dobel concern affixes or enclitics.[14] These effects do not occur throughout the Dobel area, in fact the first, second, third and fifth, listed below, are found only in the Northern dialect.[15] The majority of these processes are efforts to conform to a penultimate stress structure.

1) The morpheme *-ŋu* '1sU' or '1sG' metathesises to *-^{u}ŋ* when it follows a root final /a/,[16] thus:

'sama + -ŋu	→ /'sama^{u}ŋ/	'my father [father-1sG]'
ʔa-'yok^{w}a + -ŋu	→ /ʔa'yok^{w}a^{u}ŋ/	'he sees me [3sA-see-1sU]'

The same process also occurs after root-final /u/ and (in one case) /e/, but then the resultant /^{u}/ elides,[17] thus:

---

13 The following abbreviations are used throughout this paper:

1s	First person singular (also, 2s, 3s).		Adj.	Adjective
1p	First person plural (also, 2p, 3p).		AN	Animate noun numeral prefix
1pi	First person plural inclusive.		CPr	Contrastive Pronoun
1pe	First person plural exclusive.		DEM	Demonstrative Pronoun
3sa	Third person singular animate.		DUP	Reduplication morpheme
3sn	Third person singular inanimate.		EM	Existential Marker
1sA	First person singular Actor proclitic (also 2sA, 1peA, 3pA, etc.).		IMM	Imminent clitic
1sU	First person singular Undergoer enclitic (also 3saU, 3snU, 1peU, 2pU, etc.).		INTENS	Intensifier
			NF	Non-finite verb prefix
1sG	First person singular genitive suffix (also 3sG, 2pG, etc.).		LocPr	Locative Pronoun
			PERF	Perfective clitic
1sPoss	First person singular general possession word (also 3sPoss, etc.)		Qmk	Question marker
			REL	Relative Pronoun
1sPr	First person singular free pronoun (also 3pPr, etc.)		TagQ	Tag question
			VR	Valency Reducing prefix

14 Throughout this paper suffixes and enclitics are written morphophonemically with the exception of *-m*, which is not written when it elides following a root-final consonant.

15 In the Straits dialect there is a further morphophonemic process, not found in the Northern dialect, which is that the 2s Actor proclitic, *m-*, undergoes homorganic nasal substitution before the initial consonants of all verb roots: *m-bana* 'you go/leave', *n-ten* 'you cry', *ŋ-k^{w}oy* 'you die' etc.

16 Morphophonemic changes to this morpheme cause two phenomena rare in the language, these are the occurrence of the sequence a^u, and the only occurrence (in the Northern dialect) of word final /ŋ/.

17 There is no change to *-ŋu* after root-final /i/, e.g. /k^{w}aliŋu/, and there have been no examples of *-ŋu* following final /o/.

'muru + *-ŋu*   →   /'muruŋ/   'my back'
ʔe + *-ŋu*   →   /'ʔeŋ/   'my desire'

2) The suffix *-ni* '3saU' metathesises to *-in* following the verb *nal*[18] 'take'.

     nal + *-ni*   →   /'nalin/   'he takes it (animate)'

3) The suffix *-ni* '3saU' often drops its final vowel and becomes *-n* following a root final vowel, /y/ or /w/. There is thus resyllabification, causing the semivowels /y/ and /w/ to be reinterpreted as unstressed vocoids within the syllable [see §2.3 above]. This process is optional, but it usually occurs in the Northern dialect.

     ʔu-'yokʷa + *-ni*   →   /ʔu'yokʷan/   'I see him'
     bu'lay + *-ni*   →   /bu'laⁱn/   'he is hungry'
     saw + *-ni*   →   /saᵘn/   'onto him'

4) The suffix *-ye* '3pU' when it follows root final /a/, merges with the root resulting in the root final vowel becoming /e/; some speakers retain the /a/, simply dropping the /y/. This rule is optional. Thus:

     'sina + *-ye*   →   /'sine/, /'sinae/[19] or /'sinaye/   'they are big'

5) The suffix *-m* '2sG' is deleted after root final /l r m n/.[20] This is true for all three dialects of Dobel. Thus:

     taʔar + *-m*   →   *taʔar*   'your ears'

6) When the phoneme /y/ follows a syllable-final consonant, the consonant is doubled and the /y/ elides. This rule applies across morpheme boundaries and across word boundaries. Thus:

     ʔa-'luŋun + *-ye*   →   /ʔa'luŋunne/   'he hides them'
     'kʷalar + *'yabil*   →   /'kʷalar 'rabil/   'inside the house'

     Compare: /'mola 'yabil/   'inside the hut'

7) The prefix, *ʔin-*, which derives nouns from active verbs [see §3.2.1.3], has been found only with verbs whose roots begin with /t/, and in this case the /t/ elides, as in:

     ʔin- + *talar* (sit)   →   *ʔi'nalar*   'the act of sitting'
     ʔin- + *toran* (agree) →   *ʔi'noran*   'affirmation'
     ʔin- + *ten* (cry)   →   *ʔi'nen*   'crying' (Adj.)

---

[18] The process does not occur with other verbs ending in /l/, e.g. *dasolni* 'they ask him'. (For the paradigm of *nal* see §3.2.1.4)

[19] This process can result in vowel clusters across morpheme boundaries, the only way vowel clusters can occur in Dobel.

[20] The word /larm/ 'your voice' is an exception to this rule.

3 Morphology

3.1 Nouns and pronominals

Nouns and pronominals[21], in contrast to verbs, are not morphologically complex. Here we discuss them as the heads of noun phrases.

3.1.1 *Noun classes*

All nouns in Dobel belong to one of two noun classes, animate and inanimate, distinguished by the following:

a)   different Undergoer enclitics that substitute for them,
b)   numeral agreement, and
c)   the use of different sets of demonstratives.

In each of the above mentioned phenomena, with the exception of numeral agreement, it is, however, only in the third person *singular* forms that the animate and inanimate distinction is made. The *plural* Undergoer enclitics and demonstratives are the same for both noun classes.

The noun classes are given the names *animate* and *inanimate* because all nouns that refer to living things, both (whole) plants and (whole) animals, are in the animate class (body parts and plant parts are usually inanimate). Besides these nouns, there are many nouns that refer to semantically inanimate things, but which are nevertheless classified in the Dobel system as animate. The animate class is further subdivided into *human* and *non-human* on the basis of different numeral affixation.

Full descriptions of the agreement systems for undergoer enclitics, numerals and deictics are given under their respective sections. However, at this point, the following examples are sufficient to illustrate the agreement patterns for animate and inanimate nouns:

Undergoer enclitics:

(1)   *Tamatu   ne   ssoba-ni.*
        person   DEM   good-3saU
        'That person is good.'

(2)   *K"alar   ne   ssoba-ø.*
        house   DEM   good-3snU
        'That house is good.'

Numerals:

(3)   *k"alar   ʔawa*   (4)   *tamatu   ʔawa-ye*   (5)   *k"oyar   ʔay-ʔawa*
        house   four   person   four-3pU   dog   AN-four
        'four houses'   'four people'   'four dogs'

The deictic system is complicated and is beyond the scope of this present paper. However it is worth noting, while discussing noun classes, that when demonstratives occur following a certain specialised marker (the existential marker) the class of the noun referred

---

[21] This author presented some notes on Dobel morphology in Hughes (1995a). This present work represents a fuller understanding of the subject.

to by the pronoun determines the choice of demonstrative. If the noun is *animate* the demonstrative is chosen from the following set:

nay	(near speaker)
ne	(mid distance or near addressee)
nno	(far away)
ni?a	(in focus but not visible)

If the noun is *inanimate* then the demonstrative is chosen from the set:

wa	(near speaker)
re	(mid distance or near addressee)
nno	(far away)
ri?a	(in focus but not visible)

3.1.2 Pronominals

Dobel has several pronominal systems, which are discussed in turn below. The following table summarises them:

	Free Pronouns	Actor Proclitics[22]		Undergoer Enclitics	Genitive Suffixes	General Possession	Numeral agree (human)
1s	*sa?u*	*?u*	*?o-*	*-ŋu*	*-ŋu*	*?ana*	-
2s	*?a*	*m-*	*mo-*	*-?a*	*-m / ø*[23]	*?amu*	-
3sa	DEM	*?a-*	*na-*	*-ni*	*-y/ V# → i*	*?ani*	*-ni*
3sn	DEM	*?a-*	*na-*	*-ø / V# → i*	*-y/V#→i*[24]	*?ani*	-
1pe	*?ama*	*ma-*	*ma-*	*-?ama*	*-ma*	*?ama*	*-?ama*
1pi	*?ita*	*ta-*	*ta-*	*-da*	*-da*	*?ita ?ada*	*-da*
2p	*?emi*	*mi-*	*mina-*	*-?ami*	*-mi*	*?ami*	*-?ami*
3p	*?iri*	*da-*	*da-*	*-ye / -di*[25]	*-di*	*?ada*	*-ye*

Figure 4: Dobel pronominal systems

3.1.2.1 Free pronouns

There is a set of free pronouns which may occur as subjects or objects. In both positions their presence brings the subject or object into higher prominence, the unmarked forms being (a) the absence of a free pronoun as subject – person and number are indicated by the Actor proclitics – and (b) the use of Undergoer enclitics.

---

[22] The first set occurs with roots that take stress on the initial syllable, and on which the valency reducing prefix *r-* does not occur. The second set occurs with roots that have non-initial stress or before the prefix *r-*.

[23] On roots that end in /l r m n/.

[24] Also some root changes, such as *mata-* 'eye' → *may* 'eye-3sG', *yakʷa-* 'spouse' → *yasi* 'spouse-3sG'.

[25] The Undergoer of non-active verbs is never encoded by *-di* in the Northern dialect.

1s	2s	3s	1pe	1pi	2p	3p
sa?u	*?a*	-	*?ama*	*?ita*	*?emi*	*?iri*

Figure 5: Free pronouns

It will be noted that there is no 3s free pronoun. This is because demonstratives are used in lieu of 3s pronouns. The plural demonstratives may be used optionally instead of the 3p free pronoun *?iri*. Note that the forms *wa* and *re* are used only for non-count (mass) nouns, whereas *nay* and *ne* are used for count nouns.

	3s (count nouns)	3s (mass nouns)	3p
Near speaker	*nay*	*wa*	*wadi*
Mid-distance or near addressee	*ne*	*re*	*de*
Far away	*nno*	*nno*	*do*
In focus but not visible	*ni?a*	*ri?a*	*di?a*

Figure 6: Demonstratives (used in lieu or 3s and 3p pronouns)

3.1.2.2 Actor proclitics

The syntactic subject of active verbs is marked by an Actor proclitic on the verb, which occurs obligatorily on finite verb forms of all active verbs except those which follow another verb in serial constructions.[26] The Actor proclitic marks the person and number of the syntactic subject. It has the following forms which occur when the verb root has initial stress and the valency reducing prefix *r-* [see §3.2.1.1] does not occur:

Person:	First	Second	Third
Singular:	*?u-*	*m-*	*?a-*
Plural:	*ma-* (excl.) *ta-* (incl.)	*mi-*	*Da-*

Figure 7: Actor proclitics, first allomorph set

There is an allomorph set as follows whose members occur with verb roots with non-initial stress, or before the prefix *r-*:

Person:	First	Second	Third
Singular:	*?o-*	*mo-*	*na-*
Plural	*ma-* (excl.) *ta-* (incl.)	*mina-*	*da-*

Figure 8: Actor proclitics, second allomorph set

The following are examples of different allomorphs:

---

[26] The form of these pronominals is reported here, their function is dealt with in the introduction of §3.2, Verbs.

(6)	*ʔa-'bana*	(7)	*na-ba'ʔarum*	(8)	*na-r-'boy*
	3sA-leave		3sA-sink		3sA-VR-be.tired
	'he leaves'		'he sinks'		'he is tired'

3.1.2.3 Undergoer enclitics

The undergoer of an action is encoded in the surface structure in Dobel. The undergoer is marked by an *undergoer enclitic* which attaches to the verb phrase. Figure 9 shows the paradigm for Undergoer enclitics. The third singular animate (3sa) and inanimate (3sn) forms are shown separately:

1s	2s	3sa	3sn	1pe	1pi	2p	3p
-ŋu	*-ʔa*	*-ni*	*-ø* (O/S$_O$) or last vowel $\rightarrow$ i (S$_O$)	*-ʔama*	*-da*	*-ʔami*	*-ye* (O/S$_O$) *-di* (O)

Figure 9: Undergoer enclitics

Either 3p form may be used to encode the object (O)[27] of active verbs, although the undergoer enclitic *-ye* is preferred in the Northern dialect, whereas the subject (S$_O$) of non-active verbs may only be encoded by *-ye*.

The undergoer may be the syntactic object of an active verb, as in (9), or the syntactic subject of a non-active verb, as in (10) [See §3.2 below for a full discussion]. The Undergoer enclitic attaches to the verb phrase. Such a verb phrase may function as the predicate in a main or subordinate clause.

(9)	*ʔAni*	*kʷoyar*	*nay*	*ne,*	*m-yok ʷa-ni?*
	3sPoss	dog	EM	DEM	2sA-see-3saU
	'That's his dog, do you see it?'

(10)	*ʔAni*	*kʷoyar*	*nay*	*ne,*	*lo ʔar-ni,*	*ni?*
	3sPoss	dog	EM	DEM	beautiful-3saU	TagQ
	'That's his dog, it's beautiful, isn't it?'

The following evidence is presented for the Undergoer enclitic's status as a clitic.

(a)   When the Undergoer enclitic is the object of a transitive verb it cannot co-occur with a full object noun phrase or a free pronoun as object. This is shown in the following examples:

(11)	*Ta-wata*	*lli*	*nay.*
	1piA-fold	leaf.blanket	DEM
	'Let's fold this leaf blanket.'

(12)	*Ta-wata-ni.*
	1piA-fold-3saU
	'Let's fold it.'

Note that in (11) there is no enclitic, because there is a full NP$_O$, whereas in (12) the object is referred to only by the Undergoer enclitic.

---

[27] The designations O and S$_O$ are discussed in §3.2

(b) A verb may be intensified with the intensifier *yu?u*, which follows the verb. When *yu?u* is present the undergoer enclitic cliticises to it, rather than to the verb itself. The following examples illustrate this:

(13) *Da-dayar-ni.*
3pA-hit-3saU
'They are hitting him.'

(14) *Da-dayar   yu?u-ni.*
3pA-hit   INTENS-3saU
'They are hitting him hard.'

(15) *Tamatu   ne   soba-ni.*
person   DEM   good-3saU
'That person is good.'

(16) *Tamatu   ne   soba   yu?u-ni.*
person   DEM   good   INTENS-3saU
'That person is very good.'

For an active verb which is reflexive or reciprocal, the Undergoer enclitic is coreferential with the Actor proclitic [see §3.2.1.1.2].

If the verb is non-active, the verb phrase may function as a qualifier of the noun head within the noun phrase, (and the verb root is then reduplicated [see §3.2.1.1.2]). In this case the undergoer enclitic is obligatorily present, as in the following example.

(17) *Tamatu   ssoba-ni   ne   ?a-k"oy   ti.*
person   DUP-good-3saU   DEM   3sA-die   PERF
'That good person has died.'

The Undergoer may also be governed by a preposition, in which case the enclitic attaches to the preposition, as in (18).

(18) *Mol   beda   ne   ya?a-ṇu   da.*
(2sA)give bush.knife DEM to-1sU   IMM
'Please give that bush knife to me.'

To refer to 3s animate nouns, the form *-ni* (3saU) is used as in examples (9) and (10) above. The following examples serve to illustrate the variations in how the undergoer is marked for third person singular *inanimate* nouns (3snU) (see Figure 9, above). For the undergoer of active verbs and that of many non-active verbs there is no enclitic (marked -ø), as in (19) and (20) respectively. However, for some non-active verbs, the final vowel of the root (usually *a*) mutates to *i*, as in (21) [see (10) above].

(19) *?Ani   k"alar   nay   re,   m-yok"a-ø?*
3sPoss house   EM   DEM   2sA-see-3snU
'That's his house, do you see it?'

(20) *?Ani   k"alar   nay   re,   k"aṇay-ø,   ni?*
3sPoss house   EM   DEM   bad-3snU   TagQ
'That's his house, it's awful, isn't it?'

(21) *?Ani   k"alar   nay   re,   lo?ir,   ni?*
3sPoss house   EM   DEM   beautiful(3snU)   TagQ
'That's his house, it's beautiful, isn't it?'

3.1.3 Genitive constructions

3.1.3.1 Genitive suffixes

There is a closed subset of nouns which must occur with a suffix that agrees with the possessor.[28] These inalienably possessed nouns include most body parts, most kinship terms, and many locational nouns, such as *yisan-di* 'underneath-3pG', *fifan-di* 'top-3pG', *yabal-di* 'inside-3pG', etc. The most common paradigm of genitive suffixes is as follows:

1s	2s	3s	1pe	1pi	2p	3p
-ŋu	*-m*	last vowel of root → *i*	*-ma*	*-da*	*-mi*	*-di*

Figure 10: Genitive suffixes

The following are some examples of the use of Genitive suffixes:

(22) *lima-ŋu*
hand-1sG
'my hand'

(23) *sabu-m*
grandparent-2sG
'your grandparent'

(24) *taʔar-ma*
ear-1peG
'our ears'

These suffixes undergo several morphophonemic changes which vary from dialect to dialect [see §2.5].

There are several variations in the third person singular, the most common of which is a change of the final root vowel (usually *a*) to *i*, thus: *sami* 'his father', or *sanin* 'his father-in-law' are formed from the roots *sama* and *sanan* respectively. Another fairly common 3s form is the suffix *-y*, which occurs after some (but not all) root final vowels (in the data this has always been the vowel *a*), e.g. *yaba-y* 'his leg'. There are a number of nouns that have irregular 3s possessed forms, for which see the paradigms in Figure 11 below.

The following are some examples of nouns that are inalienably possessed:

	1s	2s	3s	1pe	1pi	2p	3p
father	*sama-ŋu*	*sama-m*	*sami*	*sama-ma*	*sama-da*	*sama-mi*	*sama-di*
hand	*lima-ŋu*	*lima-m*	*limi*	(Plurals as above throughout.)			
elder sibling	*ʔaʔa-ŋu*	*ʔaʔa-m*	*ʔaʔi*				
in-law	*sanan-ŋu*	*sanan*	*sanin*				
ear	*taʔar-ŋu*	*taʔar*	*taʔir*				
tooth	*ŋeyan-ŋu*	*ŋeyan*	*ŋeyin*				
leg	*yaba-ŋu*	*yaba-m*	*yaba-y*				
skin	*ʔala-ŋu*	*ʔala-m*	*ʔala-y*				

---

[28] Occasionally some speakers use the 3s form of some of these nouns as a root which they then possess with the general possession word [see §3.1.3.2]. I attribute this to language change; some of the nouns that take genitive suffixes are beginning to be used with possession words, and the 3s form as the most common then takes the role of noun root. None of these nouns has been found to occur with the root alone, except in some cases where there is noun incorporation in a verb, such as *na-r-loy lima* [3sA-VR-hang hand] 'be empty handed' (cf. *lima-m* 'your hand'). Note that the verb is marked as having no object, by the *r-*.

Irregular 3s:

	1s	2s	3s
younger sibling	*kʷali-ʷu*	*kʷali-m*	*kʷel*
eye	*mata-ŋu*	*mata-m*	*may*
voice	*lar-ŋu*	*lar-m*	*ler*
spouse	*yakʷa-ŋu*	*yakʷa-m*	*yasi*

Irregular 2s:

	1s	2s	3s	1pe	1pi	2p	3p
child	*yana-ŋu*	*yana-w/-ʔu*	*yana-y*	*yana-ma*	*yana-da*	*yana-mi*	*yana-di*

(The plurals are followed by *kʷakʷay*, a plural classifier used for relationship terms.)

Figure 11: Paradigms of inalienably possessed nouns

Most of the variation in 3s forms is explicable:

(i)   The 3s suffix is *-i.*

(ii) If the root is consonant-final, the suffix metathesises with the final consonant.

(iii) *a* elides before *i.*

(iv) Certain nouns, such as *yaba-* are lexical exceptions to (iii).

(v)   In these cases, *i* resyllabifies to *y* following a root-final vowel.

(vi) Forms such as *kʷel* (root: *kʷali*) and *ler* (root: *lar*) are explicable in that, historically *'CaSi#* (where S is a sonorant, l,r,m,n) became 'CeS# in Dobel.[29]

Forms such as *may* 'eye-3sG' and *yasi* 'spouse-3sG' are still unaccounted for.

There are many nouns which occur almost exclusively in the third person, as they are possessed only by non-humans; most of these occur both in the singular and the plural. Some of these nouns are semantically equivalent to locative prepositions in English, such as *yabil* 'inside', literally 'the interior of'. The following are some examples of these and other nouns:

	Singular	Plural
'tail'	*suʔir*	*suʔar-di*
'hind leg'	*ʔukil*	*ʔukʷal-di*
'top'	*fifin/fufun*	*fifan-di*
'inside'	*yabil*	*yabal-di*
'underneath'	*yisin*	*yisan-di*

Figure 12: Inalienably possessed nouns occuring only in the third person

Thus the phrase 'in the house' is *kʷalar yabil*, literally 'the house's interior'.

---

[29] The *'CaSi#* forms are still preserved in the Manombai language of East Central Aru. Compare *ku'dari* the Manombai word for 'cassowary' with the Dobel word: *ʔu'der* 'cassowary'.

3.1.3.2 *General possession word*

The majority of nouns, unlike those in §3.1.3.1, may occur without a possessive or genitive construction, and do not take genitive suffixes. Possession of such nouns is marked, rather, by general possession words. These general possession words are unstressed when they precede a noun, but can stand alone as free possessive pronouns, functioning as predicates, with the meaning of 'mine', 'yours', 'his', etc., in which case they are stressed on the final syllable of the root. The following is a paradigm of the possession words.[30]

1s	2s	3s	1pe	1pi	2p	3p
ʔana	*ʔamu*	*ʔani*	*ʔama*	*ʔada*	*ʔami*	*ʔada*

Figure 13: General possession words

The 1pi form additionally requires the free pronoun *ʔita* to disambiguate it from the homophonous 3p form, which also often, but not always, takes the 3p free pronoun *ʔiri*.

The following possessed forms are illustrative:

(25) *ʔana   kʷalar*   'my house'
      1sPoss house

(26) *ʔami   letay*   'your (pl.) traditional boat'
      2pPoss traditional.boat

(27) *ʔa   ʔamu   turaʔu*   '*your* knife'
      2sPr 2sPoss knife

When the possession words function as predicates they act like non-active verbs and take the appropriate undergoer enclitics, the undergoer being the possessed noun, as may be seen in the following examples:

(28) *Kʷalar nay   saʔu   a'na-ɸ.*
      house   DEM   1sPr   3sPoss-3snU
      'This house is mine.'

(29) *Kʷoyar nay   ita   ʔa'da-ni.*
      dog   DEM   1piPr 1piPoss-3saU
      'This dog is ours.'

(30) *Kʷakʷa ne   Eka   ʔa'ni-ni.*
      child   DEM   Eka   3sPoss-3saU
      'That child is Eka's.'

(31) *Nor   wadi   ʔa'ni-ye.*
      coconut.tree DEM   3sPoss-3pU
      'These coconut trees are his.'

---

[30] Historically, the underlying root was probably *yaʔa*, the preposition 'to'. Thus *ʔani* 'his (3sPoss)' very likely derived from *yaʔa-ni* 'to him'. The only anomaly is *ʔana* 'my (1sPoss)', as one would expect **ʔaŋu* which does not occur.

(32) *ʔAni-ø.*
3sPoss-3snU
'It is his.'

(33) *ʔAmu-ye.*
2sPoss-3pU
'They are yours.'

The possession words may also take undergoer enclitics when they do not function as predicates if the possessed thing is old information in the discourse. In the following sentence the enclitic *-ye* '3pU' refers to ancestor spirits that have been mentioned earlier in the discourse:

(34) *Ta-faʔ   ʔaye   fuʔu   ʔada-ye   tu.*
1piA-use   Quant   other   3pPoss-3pU   also
'We use other people's ones too.'

3.2 Verbs

Verbs in Dobel may be divided into two major categories, active verbs and non-active verbs. This dichotomy is best described in terms of A, S and O (Dixon 1979: 61ff.; Andrews 1985: 98ff.), where A is a NP in a transitive sentence receiving treatment normally accorded to the Actor of a Transitive Verb, O is a NP in a transitive sentence receiving treatment normally accorded to the Undergoer of a Transitive Verb and S is the argument of a one-argument (intransitive) verb. In Dobel A is encoded by the Actor proclitics preceding the verb and O by the Undergoer enclitics following the verb, S when it is the semantic Actor (S_A) is encoded by the Actor proclitics, but when it is the semantic Undergoer (S_O) it is encoded by the Undergoer enclitics.[31] This is an example of a "split S phenomenon" (Dixon 1979: 71, 79ff; Andrews 1985: 147-150) where S_A is encoded in the same way as A, and S_O in the same way as O.[32]

The situation may be represented as in Figure 14 below:

	Actor Proclitic	Verb	Undergoer enclitic
Active Intransitive:	S_A	Active	
Active Transitive:	A	Active	O
Non-active Intransitive:		Non-active	S_O

Figure 14: Split-S in Dobel

---

[31] The clitic which encodes the syntactic subject must be present whether it is an Actor proclitic or an undergoer enclitic, whereas the Undergoer enclitic which encodes the syntactic object is only present when there is no NP_O.

[32] This is in contrast both to a nominative-accusative system, where S and A are encoded in the same way and O differently, and to an ergative-absolutive system, where S and O are encoded in the same way and A differently.

Figure 14 is laid out with the transitive clause in the middle so that it may be seen that the subject of an active intransitive clause is marked the same as the subject of a transitive clause, and that the subject of non-active intransitive clause is marked the same as the object of a transitive clause. The resultant three-way distinction is exemplified in sentence (35) for an active intransitive clause, in (36) for a transitive clause, and for (37) for non-active intransitive one:

(35) *ʔA-bana   ti.*   'He has left.'
       3sA-leave PERF
       S_A

(36) *ʔA-yokʷa-ni.*   'He sees it.'
       1sA-see-3saU
       A   O

(37) *ŋeŋan-ni*   'He is heavy.'
       heavy-3saU
            S_O

Thus, active verbs in Dobel comprise all two-argument (i.e. transitive) verbs and all one-argument (intransitive) verbs where S involves an Actor, as in (35) and (36) above. Non-active verbs are one-argument forms where S is an Undergoer, as in (37) above.

3.2.1 Active verbs

As can be seen from the above, active verbs have an obligatory Actor proclitic (unless they follow another active verb in a serial construction). The actor proclitic is obligatorily absent for non-active verbs.

3.2.1.1 The prefix r-

In some circumstances a prefix *r-* occurs immediately before the verb root of an active verb and after the Actor proclitic (which is then taken from the second allomorph set [see Figure 8, §3.1.2.2 above]). This occurs in three different circumstances, all of which involve reduced or low transitivity in the sense of Hopper and Thompson (1980). I shall label the *r-* function here Valency Reduction (VR).

3.2.1.1.2 Transitive verbs with r-: object deletion

A transitive verb may have the object deleted in order to shift the focus from the object to the activity to which the verb refers. When this happens the prefix *r-* is obligatory. This use may be illustrated by the following two sentences. In the first the focus is on the fact that the dog bites *people*, as opposed to, for example, other dogs. In the second sentence the focus is on the action of biting, so the object is deleted and the *r-* is inserted:

(38) *Kʷoyar ne   ʔa-ʔara   tamatu.*
       dog   that 3sA-bite   people
       'That dog bites people.'

(39) *K"oyar ne   na-r-?ara.*
    dog   that 3sA-VR-bite
    'That dog bites.'

Consider also the following conversation which I overheard:

(40) a. *M-dem   ya?*   'What are you doing?'
        2sA-do   what

    b. *?O-r-tutu.*   'I am pounding.'
        1sA-VR-pound

    c. *M-tutu   ya?*   'What are you pounding?'
        2sA-pound   what

    d. *?U-tutu   waway.*   'I'm pounding rice.'
        1sA-pound   rice

In this conversation the verb root *-tutu* 'pound' occurs with *r-* and without it. In (40b) it is the activity 'pounding' that is being asserted so *r-* is present, whereas in (40c) and (40d) it is the grammatical object that is in focus.

3.2.1.1.2 *Transitive verbs with* r-: *coreferential objects*

When the object (Undergoer) governed by a transitive verb is coreferential with the subject (Actor), that is a reflexive or a reciprocal meaning, the object is represented by an Undergoer enclitic, and the prefix *r-* is inserted. The verb root is also reduplicated. See the difference of meaning in the following three examples:

(41) *?A-dayar-ni.*
    3sA-hit-3saU
    'He hit him (someone else).'

(42) *Na-r-ddayar-ni.*
    3sA-VR-DUP-hit-3saU
    'He hit himself.'

(43) *Da-r-ddayar-ye.*
    3pA-VR-DUP-hit-3pU
    'They hit each other/themselves.'

What these two functions of *r-* seem to have in common is that they reduce the valency of the verb, either by deletion of the direct object, or by the direct object becoming coreferential with the subject.[33]

---

[33] For a relevant discussion of this see Comrie (1985:319-330).

3.2.1.1.3 *When* r- *is determined by the lexicon*

The third occurrence of the prefix *r-*, is a more or less non-productive one, in which the lexicon requires that certain active verbs always have *r-*. Most such verbs are intransitive, for example:

(44)   *na-r-la ʔa*
       3sA-VR-know
       'he knows'

(45)   *na-r-boy*
       3sA-VR-be.tired
       'he is tired'

(46)   *na-r-kʷo ʔu*
       3sA-VR-harbour
       'he makes harbour'

(47)   *na-r-tora*
       3sA-VR-throb
       'it throbs with pain'

(48)   *na-r-tir*
       3sA-VR-bathe
       'he bathes'

However there are two examples of transitive verbs:

(49)   *na-r-naw-ʔa*
       3sA-VR-teach-2sU
       'he teaches you'

(50)   *na-r-silay-ʔa*
       3sA-VR-pity-2sU
       'he pities you'

A small number of these are verbs formed from noun roots such as:

(51)   *na-r-sabu*   cf.   *sabu-m*
       3sA-VR-grandchild   'your grandfather'
       'he has grandchildren'

(52)   *na-r-fusi*   cf.   *fusi*
       3sA-VR-fruit   'its fruit'
       'it bears fruit'

(53)   *na-r-tubur*   cf.   *tubur*
       3sA-VR-stomach   'his stomach'
       'she is pregnant'

3.2.1.2 The prefix ser-

Active verb roots may also take the prefix *ser-*. The use of *ser-* results in a form analysable as a non-finite verb, that is a verb form which does not function as the predicate and does not carry person and number information.[34]

If this form is used to modify a noun, then the root is reduplicated [see §4.3.2.3.2]. The noun thus modified is often the word that would be the object governed by the verb in its finite form. However the precise relationship between the verb and the noun varies from occurrence to occurrence. Thus, *ʔay ser-wwaw* 'firewood' is formed from the root *-waw* 'burn, bake', literally meaning 'wood for burning'. However *siʔa ser-wwaw* 'baked fish' is formed from the same verb meaning 'fish which has been baked'. Other examples are:

(54) *ʔu-bana   ser-num*   cf.   *ʔa-num*
      1sA-go   NF-dive   3sA-dive
      'I'm going diving'   'he dives'

(55) *ʔuder   ser-llesi*   cf.   *ʔa-lesi*
      cassowary   NF-DUP-raise   3sA-raise
      'domesticated cassowary'   'he raises'

3.2.1.3 The prefix ʔin-

Some active verb roots take the prefix *ʔin-*. The prefix *ʔin-* only occurs on verbs with a root initial 't' which then elides. Thus *ʔin-* + *toran* → *ʔinoran*. The functions of *ser-* and *ʔin-* are very similar.

The prefix *ʔin-* occurs rarely, and therefore it is difficult to pin down its functions. The following are some of the ways it is used.

1. It is used to form nouns from active verbs:

ʔin- amaʔrer	'platform to stand on for sago washing'	cf.	*ʔin-* + (*nar-*)*tamaʔrer*	'stand'
ʔin-oran	'means of showing agreement (i.e. head-nodding, grunt, raising eyebrows)'	cf.	*ʔin-* + (*a-*)*toran*	'agree'
ʔin-ora[35]	'pain'	cf.	*ʔin-* + (*nar-*)*tora*	'throb with pain'
ʔin-alar[36]	'act of sitting'	cf.	*ʔin-* + (*a-*)*talar*	'sit'
ʔin-en[37]	'act of mourning'	cf.	*ʔin-* + (*a-*)*ten*	'cry'

---

[34] Occasionally the resulting form may be used as a noun, as in *ser-sifar* 'language' from the verb *na-r-sifar* 'speak'.

[35] Used in the compound phrase *sara si ʔinora* 'suffering', lit. 'illness and pain'.

[36] As in *ʔadem ʔinalar salafafi* 'he is sitting crosslegged'.

2.   It is used to form non-finite verbs from active verbs:

(56)   *Kʷakʷa   nay   na-diyala   ʔin-en.*
       child   DEM   3sA-stop   NF-cry
       'The child has stopped crying.'

(57)   *ʔO-r-boy   ʔin-alar.*
       1sA-VR-tired   NF-sit
       'I'm tired of sitting.'

3.   The resulting non-finite verb may be used as a modifier within the noun phrase, in which case the root is reduplicated.[38]

(58)   *yaba   ʔin-nen*
       song   NF-DUP-cry
       'mourning song'

3.2.1.4 Irregular verbs – portmanteau affixation

Certain high-frequency active verbs in Dobel act in an irregular manner, in that the Actor proclitic merges with the root, resulting in a one- or two-syllable word which incorporates both the prefix and the root. These verbs appear to have been formed from roots and *na*-type prefixes [see §3.2.1 above], even though the surface forms all have initial stress, except in the 2p form which has a two syllable underlying prefix.[39]
There are seven verbs of this type:

	1s	2s	3s	1pe	1pi	2p	3p
drink	*ʔon*	*mon*	*nan*	*man*	*tan*	*minan*	*dan*
sleep	*ʔon*	*mon*	*nen*	*men*	*ten*	*minen*	*den*
cause/do	*ʔom*	*mom*	*nam*	*mam*	*tam*	*minam*	*dam*
take/give	*ʔol*	*mol*	*nal*	*mal*	*tal*	*minal*	*dal*
complement[40]	*ʔoŋay*	*moŋay*	*naŋay*	*maŋay*	*taŋay*	*minaŋay*	*daŋay*
hold	*ʔodi*	*modi*	*nadi*	*madi*	*tadi*	*minadi*	*dadi*
take	*ʔoʔu*	*moʔu*	*naʔu*	*maʔu*	*taʔu*	*minaʔu*	*daʔu*

Figure 15: Irregular verbs with portmanteau actor proclitics

I suggest that an underlying root-initial vowel has merged with the prefix, and that this vowel deletes the vowel of the prefix in the surface form, except in the first and second persons singular where the *o* of the prefix deletes the root initial vowel. This is shown by the verb *nen* 'he sleeps', which is the only one not to have the vowel /a/ in the initial

[37] As in *ʔinen ttay* 'name of a traditional type of song' (lit. 'mourning of the sea').

[38] Since the *t* has elided, the reduplication is realised by a lengthening of the *n*.

[39] Thus *mi'naʔu* '2pA-take' [underlying form: *mina-aʔu*] contrasts with *'minaʔu* 'firearm'.

[40] The verb *naŋay* can be used to mark the complement either following a speech verb or certain other verbs which require a complement.

syllable. If it were not for *nen* we could say that the prefix retained its vowel in the portmanteau prefix throughout. Of course, *nen* could be explained as an exception to contrast with *nan* 'he drinks', but then you might expect them to contrast in the first and second persons also, which they don't. This vowel deletion is explicable in that Dobel does not allow vowel clusters. This might be formalised in the following rules:

$$a \rightarrow \emptyset/\_V$$
$$V \rightarrow \emptyset/o\_$$

So the following is my analysis of the underlying forms of these verbs, showing only 1s and 3s forms, as the other forms are predictable from these:

Surface	Underlying	Gloss
ʔon	*ʔo- +an*	'I drink'
nan	*na- +an*	'he drinks'
ʔon	*ʔo- +en*	'I sleep/marry'
nen	*na- +en*	'he sleeps/marries'
ʔoŋay	*ʔo- +aŋay*	'I say/do next'
naŋay	*na- +aŋay*	'he says/does next'
ʔom	*ʔo- +am*	'I cause/do'
nam	*na- +am*	'he causes/does'
ʔol	*ʔo- +al*	'I take/give'
nal	*na- +al*	'he takes/gives'
ʔodi	*ʔo- +adi*	'I hold'
nadi	*na- +adi*	'he holds'
ʔoʔu	*ʔo- +aʔu*	'I take'
naʔu	*na- +aʔu*	'he takes'

When a verb root is required onto which can be added the prefix *ser-* [see §3.2 above], then the 3s form is used as a root. Thus from the verb *nan* 'he drinks', the non-finite form *ser-nnan* may be formed.

3.2.2 Non-active verbs

The difference between non-active and active verbs is explained in section 3.2 above. Non-active verbs are verbs where the semantic Undergoer is the syntactic subject. In Dobel *Undergoer* is not only a semantic but also a syntactic category, which is marked in the language by the Undergoer enclitic. Non-active verbs in Dobel are intransitive verbs where the subject (S) is also the semantic Undergoer. These verbs take no Actor proclitics, and obligatorily take an Undergoer enclitic.[41] The subject may also be encoded by an optional NP$_S$ with a noun or pronoun head which precedes the verb.

The majority of non-active verbs encode semantic States, whereas active verbs encode Events. There are however some non-active verbs that encode semantic Events, such

---

[41] Except that, as was noted above, when the Undergoer is a 3s inanimate noun it is marked by zero.

doŋaluʔu-ni 'appear-3saU' and *kʷoy'tul-ni* 'dive-3saU'. This is presumably because Dobel considers the subjects of these verbs to be undergoers rather than actors. Syntactically the non-active verbs that encode Events are distinguished in that they cannot occur as modifiers in the NP, whereas non-active verbs that encode States can. This can be seen in the following examples where (62) is illegal:

(59) *Kʷoyar ne　　ŋaŋar-ni.*
　　　dog　　DEM　vicious-3saU
　　　'That dog is vicious.'

(60) *Kʷoyar　ŋŋaŋar-ni　　　　ne　　ʔa-kara　saʔu!*
　　　dog　　DUP-vicious-3saU　DEM　3sA-bite　1sPr
　　　'That vicious dog bit me!'

(61) *Kʷoyar ne　　doŋaluʔu-ni.*
　　　dog　　DEM　appear-3saU
　　　'That dog appeared.'

(62) **Kʷoyar　doŋalluʔu-ni　　　ne　　ʔa-kara　saʔu.*
　　　dog　　DUP-appear-3saU　DEM　3sA-bite　1sPr
　　　'That appearing dog bit me.'

3.2.3 Roots that occur as nouns and verbs

3.2.3.1 Active verb roots that also occur as nouns

A number of active verbs when stripped of their affixation are nouns; usually these nouns are semantically very closely related to the verb. These roots are not reduplicated. The following are some examples:

(63) *ʔa-tabay*　　　　'carry on shoulder'
　　　tabay　　　　　'pole for carrying'

(64) *ʔa-fola*　　　　　'suckle'
　　　fola　　　　　　'breast'

(65) *ʔa-fir*　　　　　'dream (verb)'
　　　fir　　　　　　'dream (noun)'

(66) *ʔa-lar*　　　　　'sail (verb)'
　　　lar　　　　　　'sail (noun)'

(67) *ʔa-feri*　　　　　'paddle (verb)'
　　　feri　　　　　　'paddle (noun)'

(68) *na-r-tubur*　　　'be pregnant'
　　　tubur　　　　　'stomach'

3.2.3.2 *Non-active verbs formed with verb-forming suffix* -y

Non-active verbs may be formed from certain noun roots by the addition of the verb-forming suffix *-y*. The resulting forms act the same as all non-active verb roots. Some examples follow:

(69)	*bbasa*	'mud'	*basa-y*	'muddy'
	kkʷura	'anger'	*kʷura-y*	'angry'
	lura	'oil'	*lura-y*	'smooth'
	mila	'fat (n.)'	*mila-y*	'fatty'
	sara	'illness'	*sara-y*	'sore'

3.3 Numerals

The unaffixed form of numerals are used for counting and are as follows:

1 *ʔe'tu, ye*		22 *ʔurafi ro na ro*	
2 *ro*		23 *ʔurafi ro na lay*	
3 *lay*		30 *ʔurafi lay*	
4 *ʔawa*		70 *ʔurafi dubu'yam*	
5 *lima*		90 *ʔurafi yera*	
6 *dubu*		99 *ʔurafi yera na yera*	
7 *dubu'yam*		100 *ratu ye*	
8 *ʔa'ro*		109 *ratu ye na yera*	
9 *yera*		115 *ratu ye na wur na lima*	
10 *wur*		120 *ratu ye na ʔurafi ro*	
11 *wur na ye*		167 *ratu ye na ʔurafi dubu na dubu'yam*	
12 *wur na ro*		200 *ratu ro*	
13 *wur na lay*		999 *ratu yera na ʔurafi yera na yera*	
20 *ʔurafi ro*		1000 *ribu ye/ rifin ye* [42]	
21 *ʔurafi ro na ye*		2222 *ribu ro na ratu ro na ʔurafi ro na ro*	

The unaffixed form of the numerals uses *wur* for ten in all combinations between ten and nineteen. For twenty, thirty, etc., up to ninety, *ʔurafi* is used.

3.3.1 *Numeral agreement*

Numerals, when they are used to count nouns, agree with those nouns. They agree, firstly, according to three categories, inanimate, animate (non-human), and animate (human), then, secondly, in the case of human nouns they take Undergoer enclitics which agree for person and number. Numeral agreement is the only instance in Dobel where there is a distinction between animate human and animate non-human. The numerals agree with the nouns they count as follows:

---

[42] *Rifin* was given in answer to elicitation of the word 'thousand', but in text I have always heard *-ribu*, a Malay borrowing which has probably all but replaced the original *rifin*.

(a)   When numerals enumerate *inanimate* nouns they take the unaffixed form which is also used for counting (see above).

(70)   *kʷalar   dubu*   'six houses'
        house   six

(71)   *turaʔu   wur   na   yera*   'nineteen knives'
        knife   ten   and   nine

(b)   When numerals enumerate *animate non-human* nouns the numerals 'two' to 'seven' take a prefix *ʔay-*, and *ʔaʔro* 'eight' and *yera* 'nine' take *ʔa-*. These prefixes are applied whenever the numeral word stands alone, or in a compound numeral when it follows the connector *na*. The numeral *ro*, 'two', becomes *-roy*, when prefixed with *ʔay-*. Numerals *ye-ni* 'one' and *ʔurafi* 'ten' are used for all animate nouns, human or non-human.

(72)   *kʷoyar   ʔay-dubu*   'six dogs'
        dog   AN-six

(73)   *yiram   ʔurafi   na   ʔa-yera*   'nineteen axes'
        axe   ten   and   AN-nine

The numerals as used with animate non-human nouns are thus as follows:

1 *ʔeʔtu-ni, ye-ni*	22 *ʔurafi ro na ʔay-roy*
2 *ʔay-roy*	23 *ʔurafi ro na ʔay-lay*
3 *ʔay-lay*	30 *ʔurafi lay*
4 *ʔay-ʔawa*	70 *ʔurafi dubuʔyam*
5 *ʔay-lima*	90 *ʔurafi yera*
6 *ʔay-dubu*	99 *ʔurafi yera na ʔa-yera*
7 *ʔay-dubuʔyam*	100 *ratu ye*
8 *ʔa-ʔaʔro*	109 *ratu ye na ʔa-yera*
9 *ʔa-yera*	115 *ratu ye na ʔurafi na ʔay-lima*
10 *ʔurafi*	120 *ratu ye na ʔurafi ro*
11 *ʔurafi na ye-ni*	167 *ratu ye na ʔurafi dubu na ʔay-dubuʔyam*
12 *ʔurafi na ʔay-roy*	200 *ratu ro*
13 *ʔurafi na ʔay-lay*	999 *ratu yera na ʔurafi yera na ʔa-yera*
20 *ʔurafi ro*	1000 *ribu ye/ rifin ye*
21 *ʔurafi ro na ye-ni*	2222 *ribu ro na ratu ro na ʔurafi ro na ʔayroy*

(c)   When numerals enumerate *human* nouns the numerals which stand alone, or follow *na* take an Undergoer enclitic, which agrees with the noun for person and number.[43] With human nouns as with animate non-human nouns the root *ro* 'two' becomes *roy*. The Undergoer enclitics used with numerals are the following:

---

[43] Of course, only the numeral *ye*, 'one', takes a singular Undergoer enclitic. *Ye-ni* is the 3s form. I have never heard *ye* with a 1s or 2s enclitic. (The word *ye* 'one' is homophonous with *-ye* '3pU', but there is no semantic connection.)

1s	2s	3s	1pe	1pi	2p	3p
-	-	-ni	-ʔama	-da	-ʔami	-ye

The following table shows the forms of the numerals 'two' to 'nine' for all persons:

		1pi	1pe	2p	3p
2		*roy-da*	*roy-ʔama*	*roy-ʔami*	*roy-ye*
3		*lay-da*	*lay-ʔama*	*lay-ʔami*	*lay-ye*
4		*ʔawa-da*	*ʔawa-ʔama*	*ʔawa-ʔami*	*ʔawa-ye*
5		*lima-da*	*lima-ʔama*	*lima-ʔami*	*lima-ye*
6		*dubu-da*	*dubu-ʔama*	*dubu-ʔami*	*dubu-ye*
7		*dubu'yam-da*	*dubu'yam-ʔama*	*dubu'yam-ʔami*	*dubu'yam-ye*
8		*ʔa'ro-da*	*ʔa'ro-ʔama*	*ʔa'ro-ʔami*	*ʔa'ro-ye*
9		*yera-da*	*yera-ʔama*	*yera-ʔami*	*yera-ye*

Compare the following examples with those above:

(74) *ʔodar   dubu-ye*
    woman six
    'six women'

(75) *tamatu   ʔurafi   na   yera-ye*
    person   ten   and   nine
    'nineteen people'

It is nearly possible to analyse numerals as non-active verbs of which the noun enumerated is the syntactic subject (Undergoer). The numerals that enumerate human nouns act in the same way as non-active verbs in that they obligatorily take the Undergoer enclitics. Those that enumerate inanimate nouns might also be analysed as non-active verbs, in that if the Undergoer of a non-active verb is 3s inanimate there is no Undergoer enclitic. However of course, except for *ye* 'one', the Undergoer with numerals would always be plural, and 3p inanimate Undergoers are marked with an enclitic with non-active verbs.[44] The category of animate non-human nouns does not fit into the morphological structure of non-active verbs. So while recognising that there are clear similarities between the morphology and syntax of non-active verbs and numerals, numerals have to be maintained as a separate word class.

3.3.2 *Classifiers*

As in many Austronesian languages, numerals often do not occur by themselves but with classifiers. Certain nouns require the presence of classifiers if they are to occur with numerals, and some nouns may occur with or without a classifier. Classifiers may be divided into two types, those that refer to the whole object and classify them into groups of objects, and those that refer to a part of an object or a group of objects. The form of the classifiers is that of inalienably possessed inanimate nouns. Figure 16 shows some of the classifiers that refer to whole objects. The meanings in brackets are the meanings when they occur as ordinary nouns and not as classifiers:

---

[44] Although in West Tarangan, and probably other Aru languages, the enclitic for inanimate nouns for both 3s and 3p is zero (Richard Nivens, pers. comm.).

Classifier	Type of objects	Examples of nouns classified	
kʷasir	'boats and villages'	*ʔalay*	'boat'
		letay	'sailing boat'
		fa'no	'village'
rakʷin (leaf)	'thin flat'	*ller*	'young leaf'
		fasir	'cloth'
faʔil	'thicker flat'	*rinatu*	'sago filter'
		ʔay ffin	'plank'
fusi (fruit)	'fruits, other'	*ʔa'llay*	'sp. lemon'
		nor	'coconut fruit'
		ler	'shouts, barks'
fatin (body)	'people, trees'	*tamatu*	'person'
		nor	'coconut tree'
yafir (shaft)	'long, pole-shaped'	*ada*	'trousers'

Figure 16: Classifiers for whole objects

The following examples illustrate the use of the classifiers above:

(76) *fa'no   kʷasir   lay*
village CLAS   three
'three villages'

(77) *fasir rakʷin   wur   na   ye*
cloth CLAS   ten   and   one
'eleven (layers of) cloth'

(78) *ʔay ffin   faʔil   ye*
plank   CLAS   one
'one plank'

(79) *nor   fusi   ro*
coconut   CLAS   two
'two coconuts'

(80) *nor   fatin   ro*
coconut   CLAS   two
'two coconut trees'

(81) *ada   yafir   ro*
trouser   CLAS   two
'two pairs of trousers'

The following classifiers refer to part of an object or groups of objects:

Classifier	Type of objects	Examples of nouns classified	
fin	'one side of something'	*lima-*	'hand'
waʔuy	'bunch (bananas)'	*muʔu*	'banana'
ʔayi	'head of grain'	*ʔoytela*	'maize'
		waway	'rice'
sali	'container full'	*sur*	'spoon'
		kʷoba	'basket'
tamin	'parallel part of a thing'	*yel*	'comb (tooth)'
		lola	'ladder (step)'
		buraʔu	'palm floor (plank)'

Figure 17: Classifiers for parts of or groups of objects

The following are some examples of the above:

(82) *taʔir   fin   ye*
ear   CLAS   one .
'one ear'

(83) *muʔu   waʔuy   ro*
banana   CLAS   two
'two bunches of bananas'

(84) *ʔoytela   ʔayi   ye*
maize   CLAS   one
'one corn on the cob'

(85) *sur   sali   lay*
spoon CLAS   three
'three spoonfuls'

(86) *buraʔu   tamin   ye*
palm-wood CLAS   one
'a plank of palm-wood flooring'

Some of the above classifiers may also occur without numerals, but if that is the case the numeral will be replaced by a quantifier, a demonstrative or another modifier as in the following examples:

(87) *rinatu   faʔil   nay*
sago-filter CLAS DEM
'this sago filter'

(88) *Mariri   fin   wumur*
Mariri   CLAS   east
'the east side of Mariri (Island)'

The word *rasa*, meaning 'crowd/large group', is a classifier which is used in this way with the noun *tamatu* 'person', or other nouns with humans as referents. It seems that this classifier is only used with demonstratives, and never with numerals:

(89) *tamatu   rasa   nay*
person   CLAS   DEM
'all these many people'

(90) *Oysabi   Yay   rasa   ne*
Koijabi native CLAS   DEM
'that crowd of Koijabi people'

3.4 Prepositions

Prepositions are used in Dobel to mark oblique phrases, and certain verbs idiosyncratically govern prepositional phrases rather than direct objects. Historically the prepositions in Dobel seem to have come from verbs; in fact most of the prepositions have cognate verb forms with a strong semantic relationship to the preposition. This is shown in the following Figure which shows all Dobel prepositions:

Preposition		Cognate verb	
ya?a	to, for, at, about		
?i	in, into, to	*?a-yi*	go
bana	from, since	*?a-bana*	come from, go
fay	out from, than	*?a-fay*	open
?am	away from	*?a-?am (fay)*	forbid
fo	with, during	*?a-fo*	carry
fel	along with	*?a-fel*	have/be with
fatu	through	*?a-fatu*	pass through
lola	around	*?a-lola*	go around
sak^{w}u	as far as, until		
saw	onto, (flying)		
tan	on (on top of)	*?a-tana*	cover
sira	onto (with force)	*?a-sira*	hit
feta	into, (splitting)	*?a-feta*	split
yay	onto, against	*?a-yay*	throw, plant

Figure 18: Prepositions and verbal cognates

It is sometimes difficult to tell whether we have a verb and a preposition or whether we have a serial verb construction with the actor proclitic occurring only on the first verb, this is particularly the case with forms such as *lola* 'around' and *feta* 'into, split' as in the following examples:

(91) *?U-samur   lola   fano.*
1sA-walk   around   village
'I walk around the village.'

(92) *?A-yak^{w}ur   feta   yaba-y.*
3sA-chop   into   leg-3sG
'He chopped and split his leg.'

In the process of the development of prepositions from verbs, the actor proclitics were lost.[45] However when a verb that governs a prepositional phrase has its valency reduced by the use of the prefix *r-*, the *r-* also occurs on the preposition with an actor proclitic. Thus the verb meaning 'plead to' or 'beg of', which always occurs with a preposition, *ʔa-mara fo*, may be detransitivised by adding *r-* to both the verb and the preposition, in which case the preposition also has an actor proclitic, and, because it too is 'detransitivised' has no complement:

(93) *ʔAma ma-r-mara   ma-r-fo   fara mina-r-sifar   fay   fa'no .*
    1pePr 1peA-VR-plead 1peA-VR-PREP that 2pA-VR-speak for   village
    'We plead that you speak on behalf of the village.'

cf. *ʔama ma-ma   ma-mara   fo-ʔa.*
    1pePr 1peA-come 1peA-plead PREP-2sU
    'We have come to beg of you.'

Also, when a verb and preposition are used reflexively or reciprocally and the prefix *r-* is consequently added, the verb and the preposition both have actor proclitics and the *r-*. The reduplication that is necessary in this construction [see §3.2.1.1.2], however, occurs on the preposition, not the verb, thus:

(94) *da-r-wuli da-r-yyaʔa-ye . . .*   cf. *da-wuli yaʔa-ɲu . . .*
    3pA-VR-say 3pA-VR-DUP-to-3pU   3pA-say to-1pU
    'they said to each other...'   'they said to me...'

3.5 Other word classes

3.5.1 Negators, dubatives and affirmers

There is a class of words that includes negators, dubatives and affirmers. These are words that say something about the truth or intent of a clause, whether it is true, perhaps true or not true. They occur within the phrase which functions as predicate of the clause. They may all (except the imperative/optative negator, *ʔokʷa'lay*) occur as proclauses (replacing a whole clause), as in (96), (102) and (103) below.

The general negator is *naʔu'da* 'no, not' which is often shortened to *nda*. There is also a temporal negator, *'nataʔa* 'not yet' (which may be shortened to *nata*). Then there is an imperative or optative negator *ʔokʷa'lay*, 'don't, may it not'. The dubative is *maysaʔa*. There are also two affirmers, *ʔoŋa'lay*, 'indeed', which affirms that a statement is true, and *ʔe*, 'yes' which answers a question in the affirmative.

(95) *Saʔu   nda   ʔo-r-laʔa.*
    I   not   1sA-VR-know
    'I don't know.'

---

[45] In some other Aru languages, including West Tarangan (Richard Nivens, pers. comm.), prepositions retain actor proclitics, and are thus even more like verbs than in Dobel.

(96)   *"M-yokʷa-n?"*   *"Nda."*
        2sA-see-3saU   No
        '"Can you see him?" "No."'

(97)   *'Nataʔa*   *ʔa-mul.*
        not.yet   3sA-return
        'He hasn't returned yet.'

(98)   *ʔOkʷa'lay*   *m-dayar-ni!*
        don't   2sA-hit-3saU
        'Don't hit him!'

(99)   *ʔOkʷa'lay*   *kʷusan*   *ʔa-fan.*
        don't   rain   3sA-fall
        'May it not rain!'

(100)   *Maysaʔa*   *ʔa-sula*   *ma'del.*
        perhaps   3sA-drunk   wave
        'Perhaps she is seasick.'

(101)   *ʔOŋa'lay*   *da-ma*   *ti.*
        indeed   3pA-come   PERF
        'They have indeed come.'

(102)   *"Sayi-ni."*   *"ʔOŋa'lay!"*
        tall-3saU   Indeed
        '"He's tall." "Yes, indeed!"'

(103)   *"ʔA*   *re?"*   *"ʔE!"*
        you   there   yes
        '"Is that you?" "Yes."'

3.5.2 Adverbs

There is a closed class of adverbs that qualify verbs. The following are examples of adverbs:

tu	'still, also, too, again'
mul	'again, back'
ʔay'ʔuy	'all over, completely'
kʷatan	'alone, empty handed'
fedi	'ahead, first'
ssel	'continuously'
sobi	'well'
toŋar	'truly, straight'
yala	'wrongly'

Examples illustrating the use of some of these adverbs follow:

(104)   *Bu'lay-ʔama*   *fel*   *kʷawul-ʔama*   <u>*tu*</u>.
        hungry-1peU   and   tired-1peU   also
        'We are hungry and we are also tired.'

(105) *Da-fo-ye*   *mul*   *yaʔa*   *fa'no.*
       3pA-bring-3pU   back   to   village
       'They brought them back to the village.'

(106) *Laran-ni*   *ʔay'ʔuy*   *yaʔa*   *kʷudu*
       red-3saU   completely because   blood
       'It was red all over because of the blood.'

(107) *Ta-ʔa*   *baw*   *wa*   *kʷatan.*
       1piA-eat   dry.sago   DEM   alone
       'We are eating this dry sago by itself.'

(108) *Saʔu*   *ʔu-fan*   *re*   *fedi.*
       1sPr   1sA-arrive   LOC   first
       'I got there first.'

(109) *Wursin*   *ʔa-tara*   *Ilafi*   *ssel.*
       Wursin   3sA-call   Ilafi   continuously.
       'Wursin kept on calling Ilafi.'

3.5.3 Locationals

The words that I am calling locationals are in fact both locative and temporal deictics. They are *wa*, which broadly means 'here' or 'now', and *re*, which broadly means 'there' or 'then'. A third locational *nno* 'away over there' only has a locative and not a temporal sense.

The locationals *wa*, *re*, and *nno* occur either as locative phrases (LP), which function as the head of locative clauses,[46] or in prepositional phrases (PP) in non-locative clauses. When they occur in locative clauses (i.e. as the main argument of a verb), locationals have an exclusively locative reference, but when they occur in PPs they may have either a locative or a temporal reference. The following examples illustrate the use of the locationals:

(110)   *men*   *re.*   'We slept there.'
        (1peA)sleep   there
        vPloc   LP

(111)   *ta-bana*   *nno.*   'Let's go away over there.'
        1piA-go   over.there
        vPloc   LP

(112)   *ʔa-kʷoy*   *yaʔa*   *wa.*   'It died here.'
        3sA-die   at   here
        vPintr   PP

---

46 These will be dealt with in Hughes (n.d.).

(113) *ʔa-kʷoy   fo   re.*   'He died at that time.'
       3sA-die   at/during then
       Vpintr   PP

It will be noted that in (112) the locational has a locative reference, but in (113) it has a temporal reference.

The locationals, *wa* and *re*, also function as clause-rank particles which mark the clause as either present time and present location (*wa*), or as remote time *or* remote location (*re*). The locational then occurs clause finally. The following examples illustrate their use:

(114) *Saʔu   ʔo-r-nau-ʔa   wa,   m-reɲin   fara   senaŋ-ʔa.*
       1sPr   1sA-VR-teach-2sU   LOC   2sA-listen so.that   happy-2sU
       'I'm teaching you now, listen so that you will be happy.'

(115) *Saʔu   billeŋay-ŋu   re,   ʔana   taffarbu-di   da-r-silay   yuʔu-ŋu.*
       1sPr   DUP-small-1sU LOC   1sPoss parent-PL   3pA-VR-love   INTENS-1sU
       'When I was small, my parents really loved me.'

(116) *Den   Sala   yasi   nay   ne   ʔa-dayar re.*
       Den   Sala   wife(3sG)   EM   DEM   3sA-hit LOC
       'It was Den Sala's wife banging back then.'

They act in the same way in relative clauses:

(117) *Kʷoyar ne,   saʔu   ʔu-wwuli-ni   re,   ʔa-kʷoy   ti.*
       dog   DEM I   1sA-DUP-say-3saU   LOC   3sA-die   PERF
       'That dog, that I talked about back then, has died.'

The use of these locationals functioning as particles in this way, as well as indicating present or remote time and/or location, also indicates that the action of the clause is in some way known and therefore given information. Indeed it could be argued that the primary function of the use of these particles is to background the clause, or at least the predicate, and that the choice of *wa* or *re* then give the additional remote-present information. In (114) and (115) the clauses with the locationals are in fact subordinate to the main clauses that follow them, so the whole clauses reflect information that is given. In (116) the information that someone was banging is known to all the hearers who actually heard the banging and wondered what it was. The new information is that it was Den Sala's wife who was doing it.[47] In this sentence the particle *re* marks both that the predicate is known information and that it is past.

3.5.4 Temporals

There is a closed class of temporals that function as the head of a temporal phrase. (A temporal phrase may also have a temporal noun as its head.) The following is a list of some temporals:

---

[47] In this case the agent is further foregrounded by being left dislocated with a pronominal construction filling the subject slot in the clause.

may're	'tomorrow'
yamayira	'the day after tomorrow'
ʔikʷu'san	'yesterday'
saʔaran	'now'
mayira	'daily, regularly'
ʔorama tu re wa	'earlier'
ʔorama	'just, just now'
ʔatu	'later'
ʔilaʔi	'often'

3.5.5 The question markers, ya and ba

Content questions are formed in Dobel by the use of the question markers *ya* and *ba*. The marker *ya* 'what?' is used to ask the identity of a thing or action as in (118)–(119). The question marker *ya* is also used to ask the reason for something as in (120).

(118) *ʔA*   *m-dem*   *ya?*
       you(sg.)   2sA-do   what
       'What are you doing?

(119) *Mi-ma,*   *mi-fo*   *ya?*
       2pA-come   2pA-travel.in   what
       'What did you travel in, when you came?'

(120) *M-ma*   *yaʔa*   *ya?*
       2sA-come   because   what
       'Why have you come?'

The marker *ba* forms a question by replacing a demonstrative or locative as in (121)–(122). When *ba* is replacing a demonstrative and its referent is a singular animate noun, or any plural noun, it takes undergoer enclitics as in (123)–(125).

(121) *Kʷalar*   *ba?*   cf. *kʷalar*   *nay*
       house   Qmk   house   DEM
       'Which house?'   'this house'

(122) *ʔA-bana*   *ba?*   cf. *ʔA-bana*   *nno*
       3sA-go   Qmk   3sA-go   LOC
       'Where is he going?'   'He is going there.'

(123) *Tamatu*   *ba-ni?*
       person   Qmk-3saU
       'Which person?'

(124) *ba-ni*   *ne?*
       Qmk-3saU   DEM
       'Who is that?'

(125) *turaʔu*   *ba-di*
       knife   Qmk-3pU
       'Which knives?'

In existential clauses[48] *ba* sometimes occurs twice in the same clause, once replacing a demonstrative and once a locative:

(126)   *Letay nay ba   ba?*   cf.   *Letay nay nno   tay're/nno*
       boat   EM   Qmk Qmk   boat   EM   DEM   shore/LOC
       'Where is the boat?'   'There is the boat, at the shore/over there.'

(127)   *Yana-ngu   k ʷak ʷay   nay   ba-di   ba?*
       child-1sG   CLAS-PL   EM   Qmk-3pU   Qmk
       'Where are my children?'

3.6 Clitics

There is a class of clitics. The actor proclitics cliticise to the verb word. The Undergoer enclitics cliticise to the verb phrase. They are described above in §3.1.2.2 and §3.1.2.3, respectively.

There are two other enclitics which both attach to the final phonological word of the clause, and which are both modifiers of the clause; that is they are clause constituents, not phrase constituents. These are·the perfective clitic *ti*, and the imminent clitic *da*, whose function is to mark the action of the clause as intending to be carried out immediately or before some other action that is in focus.

(128)   *ʔA-ma   ti.*
       3sA-come   PERF
       'He has come.'

(129)   *Ata'ler   wa   namataya ti.*
       vegetable   DEM   cooked   PERF
       'This vegetable is cooked.'

(130)   *ʔA-ma   da.*
       3sA-come   IMM
       'Let him come first (before something else that is in focus).'

(131)   *ʔU-bana   k ʷalar   da.*
       1sA-go   house   IMM
       'I'm going home now.'

4 Reduplication

Dobel, in common with other languages in Aru, has a complex reduplication system. Reduplication has both morphological and syntactic functions. It has also become lexicalised in certain words, where it has lost any syntactic or morphological function.

---

[48] Will be dealt with in Hughes (n.d.).

4.1 The phonological forms of reduplication

In Dobel reduplication occurs to the left of the stressed syllable. It is reduplication of only the initial consonant of the stressed syllable (which we will call C-reduplication). There is however a second reduplication-like form which started historically as reduplication, but has become completely lexicalised. This is reduplication of the initial consonant plus the vowel of the stressed syllable (CV-reduplication).

4.1.1 CV-reduplication forms

The CV-reduplication form, is very limited in Dobel. It occurs on nouns when the reduplication has become entirely lexicalised, as described below, and only when the initial syllable is stressed. In the Koijabi dialect CV-reduplication is manifest as the result of a historical reduplication of the first CV of the stressed syllable whether that syllable be a CV or CVC syllable. Although the stressed syllable is the one whose initial CV is reduplicated, in Koijabi the stress itself is not reduplicated,[49] thus the reduplicated word remains one phonological word (stress group) with the stress on the original stressed syllable of the root. Nouns that have this form cannot occur without 'reduplication', but they may occur with either CV- or C-reduplication.[50] Some examples of syllable CV-reduplication follow:

(132) k^wa-'k^wasa 'crocodile'
       ti-'tim   'sp. of bird'
       ʔa-'ʔaw   'butterfly'
       sa-'sar   'sand-fly'
       tu-'tun   'mosquito'

It will be noted that the above examples are all animals. In Dobel many animals names are 'reduplicated' (although not all). This may be the historical residue of a productive function of reduplication that commonly occurred with animal names. In any case these lexical items are now frozen forms of a reduplication which is no longer productive.

4.1.2 C-reduplication

All true productive reduplication and some lexicalised reduplication in Dobel is C-reduplication. C-reduplication is where only the initial consonant of the stressed syllable is reduplicated. In Dobel any consonant phoneme may occur as the onset of the syllable. As far as I am aware C-reduplication of this sort is not found in any of the other languages of

---

[49] In two other villages where the author has lived, Warjukur and Algadang, this reduplication is manifest as a reduplication of the whole syllable whether CV or CVC, and the stress is also reduplicated. Thus: /tun/ becomes /'tun-'tun/. It is a moot point whether the resulting form is two phonological words, or one phonological word with two primary stresses.

[50] All nouns that take CV-reduplication may instead have C-reduplication, however there are nouns (some even have initial stress) that take C-reduplication but may not take CV-reduplication.

the Aru Islands, with the possible exception of Lola, a language closely related to Dobel.[51] This reduplication of the initial consonant of the stressed syllable is manifest phonetically as a lengthening of the consonant. Phonologically it is gemination of the C onset of the *stressed* syllable, which then yields a derived syllable type $C_1C_2V(C_3)$, where C_1 is a copy of C_2; C_1 always occurs immediately preceding C_2, even if it is in the middle of a morpheme.[52] The following are some examples of C-reduplication showing the unreduplicated forms in parentheses:

(133)   'sago-pounder'   *yyak{}^wur*   *(yak{}^wur)*
        'sandy'   *ʔʔula*   *(ʔula)*
        'of the forest' *k{}^wu'bbol*   *(k{}^wu'bol)*
        'diving'   *ser-nnum*   *(ser-num)*
        'raw'   *mmata*   *(mata)*
        'third'   *llay*   *(lay)*
        'he/she/it does' *ʔa-ddem*   *(ʔa-dem)*

4.2 Lexicalised reduplication

The extant reduplication process results in the form described above where a geminate consonant forms the coda of the stressed syllable. The CV-reduplication forms above are no longer the outcome of a current reduplication process, but from a historical perspective are clearly the result of reduplication. All such CV- forms have become lexicalised. That is the CV-reduplication form is present in certain nouns simply as a requirement of the lexical item. However, in all cases where it occurs it is in free variation with C-reduplication, which is further evidence that we are dealing with reduplication here. Thus the forms given in (132) above have synonymous alternatives which result from C-reduplication:

        'crocodile'   *k{}^wa-'k{}^wasa*   *kk{}^wasa*
        'sp. of bird'   *ti-'tim*   *ttim*
        'butterfly'   *ʔa-'ʔaw*   *ʔʔaw*
        'sand-fly'   *sa-'sar*   *ssar*
        'mosquito'   *tu-'tun*   *ttun*

There are also many nouns that can have only C-reduplication. Any noun that has non-initial stress, if reduplicated must have C-reduplication, since CV-reduplication only occurs when the stressed syllable is initial. There are also several nouns with initial stress (which could thus conceivably have CV-reduplication) that have C-reduplication with a purely lexical function, but that may not take CV-reduplication.

---

[51] In other Aru languages corresponding functions are handled by CV- or CVC- reduplication, as well as some other more complex forms of reduplication.

[52] This Reduplication is in fact a separate morpheme, however, since C-reduplication occurs frequently in the middle of a morpheme (e.g. *k{}^wu'bbol* 'of the forest'), for convenience I will not show it hyphenated as separate, but simply written as a double consonant.

The lexicalised reduplication is contrastive with its absence, and there are several minimal pairs, for example:

(134) *ser*   'sp. of shell'   *sser*   'sp. of fruit'
       ?ay'law   'mangrove tree'   *?ay'llaw*   'papaya'
       ma'ŋay   'palm frond'   *ma'ŋŋay* [53] 'pleasant smell'

There are also a few examples of words from other word classes which are reduplicated for a purely lexical reason:

(135) *ssel*   'frequently'
       nno   'that (far)/ over there (far)'

The CV-reduplication form is probably a residue in the language of the outcome of the reduplication process used in Pre-Dobel. C-reduplication of this kind is an innovation of the Dobel-Lola sub-group of languages.[54] Also I have text examples of archaic Dobel preserved in songs and chants, which shows more widespread CV- and indeed CVC-reduplication. CV-reduplication (and other now extinct forms) were probably also used with the morphological and syntactical functions now restricted to C-reduplication. The form of CV-reduplication has been preserved solely where the reduplication has been lexicalised, in which case it co-occurs with forms analogous to C-reduplication. This is borne out by the fact, mentioned above, that there are several nouns with initial stress that may only occur with the lexicalised C-reduplication form. The reason there is no pattern determining which nouns with initial stress may occur as CV-reduplication forms, is probably because the CV-forms of these nouns happen not to have survived. All productive reduplication in Dobel is C-reduplication, hereafter simply referred to as reduplication.

4.3 The functions of reduplication

4.3.1 Morphological function: nominalisation of verbs

The nominalisation of a verb root is marked by reduplication. For active verbs, the resulting noun is usually the instrument with which the action of the verb is performed. Thus:

(136) *da-yak^w ur*   '3pA-pound'   *yyak^w ur*   'sago-pounder'
       da-fel   '3pA-mix (sago)'   *ffel* [55]   'mixing-spoon'
       da-siyar   '3pA-draw.line'   *ssiyar*   'stick for drawing lines'
       da-la?ur   '3pA-sweep'   *lla?ur*   'broom'

When a non-active verb is nominalised by reduplication, the nominalised form is used in a possessive construction with the noun which would have been the Undergoer (subject) of the verb. The form of the verb which is reduplicated is always the form that would be used

---

[53] Also *feri* 'paddle' and *fferi* 'shoulder-blade', although there is a semantic connection here because of the paddle-like shape of a shoulder-blade.

[54] See Hughes (1987:94ff) for the sub-grouping of Aru languages.

[55] When the phoneme /f/ is reduplicated the phonetic realisation varies ideolectically between a fortis lengthened voiceless bilabial fricative [ɸ:], and a fortis lengthened voiceless bilabial plosive [p:].

with a 3s inanimate Undergoer, that is with no Undergoer enclitic or with the final root vowel mutated to *i.* Thus:

(137) *Daba   ne   sayi.*
       Canoe   DEM   long
       'The canoe is long.'

(138) *Daba·   ne   ʔani   ssayi   re   lay.*
       Canoe   DEM   3sPoss DUP-long fathom three
       'The canoe's length is three fathoms.'

Nouns thus formed are also used in a construction which marks intensification of the predicate. There are very few examples in the data of nominalised active verbs that act like this (139), whereas the construction is common with nominalised non-active verbs (140). The construction to intensify the predicate requires the use of a noun, and therefore the verb is nominalised; a nominalised verb is substituted for a finite verb. The construction is also marked by sharp rising intonation on the nominalised verb.

(139) *ʔAni   ssamur.*   cf. *ʔA-samur*
       3sPoss DUP-walk .   3sA-walk
       'He walks a lot.'   'He walks.'

(140) *Kʷoyar ne   ani   lloʔir*   cf. *Kʷoyar ne   loʔar-ni.*
       dog   DEM 3sPoss DUP-good(3snU)   dog   DEM   good-3saU
       'What an attractive dog!'   'That dog is attractive.'

4.3.2 *Syntactic functions of reduplication*

The syntactic function of reduplication may be generalised as marking background information which acts as a modifier. In nearly all cases the reduplicated element acts as a modifier within the noun phrase, in which case the modifier may be a numeral, a non-finite verb, a noun or a relative clause. These will each be dealt with below. There is one case, however, where reduplication is used to mark the modifier of the predicate verb. This is described in the following section.

4.3.2.1 *Reduplication of non-active verbs to form adverbial phrases*

In Dobel an adverbial phrase of manner may be formed using the verb *nam*[56] 'do' and a reduplicated non-active verb. This is best illustrated by examples; the non-active verbs in (141) are shown as they are used adverbially in (142) - (145).

(141) *sin*   'big'
       bari   'big'
       lakʷay   'fast'
       butemuy   'slow, late'

---

[56] The paradigm for this verb is *ʔom, mom, nam,* etc. [see §3.2.1.4].

(142) *Mo-r-sifar   mom   ssin.*
2sA-VR-speak (2s)do DUP-loud
'Speak loudly.'

(143) *M-dem-ø   mom   bbari.*
2sA-make-3snU (2s)do DUP-big
'Make it large.'

(144) *ʔA-samur nam   llakʷay.*[57]
3sA-walk (3s)do DUP-fast
'He walks fast.'

(145) *ʔU-samur ʔom   buttemuy.*
1sA-walk (1s)do DUP-slow
'I walk slowly.'

4.3.2.2 Ordinal numbers: numerals as modifiers within the noun phrase

Ordinal numbers are formed by the reduplication of the root of the numeral. As with the cardinal numbers, the ordinal numbers agree with the noun they qualify for animacy. Ordinal numbers differ from cardinal numbers in that ordinal numbers that agree with animate nouns have Undergoer enclitics.[58] With ordinal numbers, no distinction is made between human and non-human animate nouns, thus ordinal numbers that agree with animate nouns are always prefixed with *ʔay-* or *ʔa-*.

While the use of the prefixes *ʔay-* or *ʔa-* with ordinal numbers distinguishes them morphologically from verbs, their syntactic role is similar to that of verbs in that both may function as the predicate of relative clauses opened with the demonstrative *ne* functioning as a relative pronoun. Moreover the presence of the Undergoer enclitic *-ni* with animate nouns and its absence with inanimate nouns is the same as the 3s usage for non-active verbs. [See also the discussion in §3.3, Numerals].

With compound ordinal numbers the reduplication occurs only on the final word of the numeral. Compare the following examples:

(146) *labun   ro*   *labun   ne   rro*
garment   two   garment   REL DUP-two
'two garments'   'the second garment'

(147) *mayira ʔurafi dubu*   *mayira ne   ʔurafi ddubu*
day   ten   six   day   REL ten   DUP-six
'sixty days'   'the sixtieth day'

---

[57] The non-active verbs *lakʷay* and *butemuy* may also be used as adverbs without *nam*. In this case they have no reduplication, e.g. *ʔa-samur lakʷay*.

[58] It may be noted that 3s inanimate Undergoers are not marked on verbs (except in some cases by a vowel mutation with non-active verbs). Since ordinal numbers always qualify 3s nouns, there are never any agreement markers on ordinal numbers that qualify inanimate nouns, and the enclitic for ordinal numbers that qualify animate nouns is always *-ni* '3saU'.

(148) *kʷoyar   ʔay-lay*   *kʷoyar   ne   ʔay-llay-ni*
        dog   AN-three   dog   REL   AN-DUP-three-3saU
        'three dogs'   'the third dog'

(149) *fulan   ʔurafi na   ʔay-roy*   *fulan   ne   ʔurafi na   ʔay-rroy-ni*
        month   ten   and AN-two   month   REL ten   and AN-DUP-two-3saU
        'twelve months'   'the twelfth month'

(150) *ʔodar   ʔawa-ye*   *ʔodar   ne   ʔay-ʔʔawa-ni*
        woman   four-3pU   woman   REL   AN-DUP-four-3saU
        'four women'   'the fourth woman'

(151) *tamatu   ʔurafi*   *tamatu ne ʔurrafi-ni*
        person   ten   person REL DUP-ten-3saU
        'ten people'   'the tenth person'

4.3.2.3 Verbs as modifiers within the noun phrase

Reduplication of a verb root signals that it is being used as the modifier in a descriptive noun phrase. Non-active verbs, in that the vast majority of them encode semantic States, commonly function as modifiers within the noun phrase, but active verbs, which generally encode semantic Events, may also function as modifiers in their non-finite forms.

4.3.2.3.1 Non-active verbs as modifiers within the noun phrase

When a non-active verb functions as the predicate of a clause it is not reduplicated. The use of reduplication on a non-active verb marks it as modifier in a noun phrase rather than as the predicate of a clause. In other words, verbs are reduplicated when used attributively and not when they are used predicatively. Thus the following example is a noun phrase rather than a full clause:

(152) *ʔamu   ser'tay   mmaray*
        your   clothes   DUP-dry
        'your dry clothes'

Compare the following, where 'S' marks the subject and 'P' the predicate:

        S   P
(153) [*ʔAmu ser'tay*]   [*maray*].
        your   clothes   dry
        'Your clothes are dry.'

The use of the Undergoer enclitics is the same when the verb is a modifier in the noun phrase as it is when the verb is the predicate of a clause. In the above example the noun modified is a 3s inanimate noun, and therefore there is no Undergoer enclitic on the verb. When the noun is not 3s inanimate there will be an Undergoer enclitic, thus:

(154) *tamatu ssoba-ye   wadi*
        person DUP-good-3pU   DEM
        'these good people'

This function of reduplication is a syntactic one in that it marks the verb as modifier in a noun phrase, rather than predicate of the clause.

4.3.2.3.2 *Non-finite active verbs as modifiers within the noun phrase*

Non-finite active verbs may also function attributively within the noun phrase. As explained in §3.2.1.2 above, the prefix *ser-* may be added to the verb root to form non-finite verbs. When non-finite verbs of this kind are used as modifiers within the noun phrase, the root is reduplicated. If the non-finite verb is not a modifier within a noun phrase the root is not reduplicated. Thus:

(155)   *tamatu*   *ser-nnum*   cf.   *Ma,*   *ta-bana ser-num.*
        person   NF-DUP-dive   come   1piA-go NF-dive
        'a person who goes diving'   'Let's go diving.'

(156)   *si?a*   *ser-wwaw*
        fish   NF-DUP-burn
        'baked fish'

(157)   *tabul*   *ser-llesi*
        animal   NF-DUP-raise
        'domesticated animals'

(158)   *?ubu*   *ser-kkʷubu*
        stone   NF-DUP-sharpen
        'sharpening stone'

(159)   *si?a*   *ser-nniw* [59]
        fish   NF-DUP-fly
        'flying fish'

(160)   *kʷar*   *ser-ttir*
        water   NF-DUP-bathe
        'bathing water'

There is another structure whereby the reduplicated verb root of active verbs acts as the modifier in the noun phrase. This structure is less common, and it may be that all such uses are in frequently used lexicalised phrases such as the following:

(161)   *?abal*   *nniw*
        ship   DUP-fly
        'aeroplane'

(162)   *tamatu*   *ssamur*
        person   DUP-walk
        'person who walks a lot'

---

[59] 'Flying fish' may be either *si?a nniw* or *si?a ser-nniw*.

4.3.2.4 Nouns as modifiers within the noun phrase

When a noun functions as a modifier of another noun within the noun phrase, the noun functioning as modifier is reduplicated. The modifying noun applies the features, or a particular feature, of itself to the noun it qualifies.

(163) *bala   ʔʔula*
       soil   DUP-sand
       'sandy soil'

(164) *tamatu   kʷu'bbo'*
       person   DUP-forest
       'a person of the forest'

(165) *tamatu   ssiʔa*
       person   DUP-fish
       'fisherman'

(166) *siʔa   kkʷar*[60]
       fish   DUP-fresh.water
       'fresh-water fish'

(167) *siʔa   ttay*
       fish   DUP-sea.water
       'sea-water fish'

(168) *kʷoyar   tammatu*
       dog   DUP-person
       'a dog that likes people'

It will be noted that the semantic relationships signalled by this construction vary considerably: if we call the head noun X, and the modifying noun Y, the following relationships are signalled in the examples above: in (163) X is like Y in substance, in (164), (166) and (167) X is (often) found in place Y, in (165) X catches Y, and in (168) X likes Y.

4.3.2.5 Reduplication in relative clauses

Relative clauses in Dobel are marked with the use of a demonstrative functioning as a relative pronoun. When a core argument, that is the subject or object, of a clause is relativised, the predicate verb of the relative clause is reduplicated. There is a strong resemblance between this role of reduplication and that mentioned above where verbs act as modifiers in the noun phrase. Here the whole clause acts as a modifier in the noun phrase, but it is the verb that carries the reduplication. The exception to this reduplication rule is where the subject of a transitive clause with a noun phrase as object is relativised, as will be seen below. If an oblique argument is relativised then the preposition which governs it is reduplicated. Each of these cases is discussed below.

---

[60] The phoneme /kʷ/, when reduplicated, is realised as a lengthened labialised velar plosive [k:ʷ].

4.3.2.5.1 Relativisation of the subject of intransitive verbs

When the subject of an intransitive verb is relativised, the verb which is the predicate of the relative clause is reduplicated. All kinds of intransitive active verbs are included in this pattern, whether basic intransitives, with or without the prefix *r-*, or derived intransitives (with the prefix *r-*).[61]

(169)   *Tamatu ne ʔa-lla   re   ʔa-mul   ti.*
        person REL 3sA-DUP-run LOC 3sA-return PERF
        'The person who ran (away) has returned.'

(170)   *Tamatu ne na-r-ttara   nay niʔa kʷalar.*
        person REL 3sA-VR-DUP-call EPr DEM   house
        'The person who is calling is in the house.'

4.3.2.5.2 Relativisation of the subject of transitive verbs

Whereas relativising the subject of intransitive verbs always causes the verb of the relative clause to be reduplicated, relativising the subject of transitive verbs does not always do so. Whether or not the verb is reduplicated in a transitive clause when the subject is relativised depends on whether the *object* of the clause is a noun phrase or an Undergoer enclitic. If the object of the clause is a noun phrase, then the verb is not reduplicated when the subject is relativised, but if the object is an Undergoer enclitic then the verb is reduplicated when the subject is relativised. This is illustrated in the sentences below:

(171)   *Kʷoyar ne ʔa-ʔara tamatu de   re   ʔa-kʷoy ti.*
        dog   REL 3sA-bite person DEM LOC 3sA-die PERF
        'The dog that bit <u>those people</u> has died.'

(172)   *Kʷoyar ne ʔa-ʔara iri   re   ʔa-kʷoy ti.*
        dog   REL 3sA-bite 3pPr LOC 3sA-die PERF
        'The dog that bit <u>them</u>$_{(NP)}$ has died.'

(173)   *Kʷoyar ne ʔa-ʔʔara-ye   re   ʔa-kʷoy ti.*
        dog   REL 3sA-<u>DUP</u>-bite-3pU LOC 3sA-die PERF
        'The dog that bit <u>them</u>$_{(encl.)}$ has died.'

Thus, when the subject of a transitive clause with an Undergoer enclitic is relativised, the clause behaves the same way as an intransitive clause whose subject is relativised, whereas a transitive clause with a full noun phrase as object behaves differently when its subject is relativised.

---

[61] While it is possible for the subject of non-active verbs to be relativised, it is very rare as the same meaning is achieved by the non-active verb functioning as a modifier in the noun phrase, as in 4.3.2.3.1.

It is worth noting that when constituents of the causative clause [62] using *nam* are relativised, they act in the same way as for transitive clauses, except that the *nam* does not carry the reduplication, but rather the verb that is its complement.

(174)   *ʔodar   ne   nam-di   da-r-ttir...*
       woman   REL   3sA.cause-3pU   3pA-VR-<u>DUP</u>-bathe...
       'the woman who is bathing them$_{(encl.)}$...'

(175)   *ʔodar   ne   nam   yana-y   na-r-tir...*
       woman   REL   3sA.cause   child-3sG   3sA-VR-bathe
       'the woman who is bathing her child...'

4.3.2.5.3 *Relativisation of the object of transitive verbs*

When the object of a transitive clause is relativised, the object itself is marked in the clause by an Undergoer enclitic on the verb phrase. The relativisation of the clause is marked by the use of the relative pronoun and by reduplication of the verb root. When 3s inanimate nouns are used as relativised objects they have no Undergoer enclitic. The following pairs of examples illustrate this function of reduplication:

         P   O
(176) a. [*ʔA-dem*]   [*kʷalar   bbari-ø.*]
        3sA-make   house   DUP-big-3snU
        He is making a big house.'

         S   P
    b. [*Kʷalar* [*ne   ʔa-ddem-ø*]]   [*bari-ø*]
       house   REL   3sA-DUP-make-3snU   big-3snU
       'The house which he is making is big.'

         P   O
(177) a. [*ʔA-ka*]   [*siʔa   lloʔar-ni.*]
        3sA-eat   fish   DUP-good-3saU
        'He is eating a good fish.'

         S   P
    b. [*Siʔa* [*ne   ʔa-kka-ni*]]   [*loʔar-ni.*]
       fish   REL   3sA-DUP-eat-3saU   good-3saU
       'The fish which he is eating is good.'

Locative clauses with intransitive locative verbs as the head, which have a specific locative goal, are treated like transitive clauses in that the locative has a direct object-like relationship to the verb. The locative of such clauses is relativised in the same way as the object of a transitive clause.

---

[62]   This is dealt with fully in Hughes (n.d.).

```
        S                                      P
(178)   [ʔAdera    [ne      m-ttalar-ni]]      [mona-ni.]
        chair      REL      2sA-DUP-sit-3saU   old-3saU
        'The chair on which you are sitting is old.'

   cf.  M-talar    ʔadera.
        2sA-sit    chair
        'You are sitting on a chair.'
```

As with the relativisation of subjects, if the object of a causative clause is relativised, although the object is syntactically the object of the causative verb *nam*, the reduplication is on the root of the verb which is the complement of *nam*.

```
(179)   Yiram ne   ʔom-ni         ʔa-ffan        re,  koytul-ni   ʔoŋa'lay.
        axe   REL  1sA.cause-3saU  3sA-DUP-fall   LOC  sink-3saU   indeed
        'The axe, which I dropped then, did indeed sink.'
```

4.3.2.5.4 Relativisation of the object of a preposition

If the noun governed by a preposition is relativised it is marked in the relative clause by an Undergoer enclitic on the preposition, which is in turn reduplicated. As explained above, 3s inanimate nouns are not marked by an enclitic. Compare the following examples:

```
        P         O           Oblique
(180) a. [ʔA-soʔa   [lala]      [saw   yaba-y.]
        3sA-spill  hot.water   on    foot-3sG
        'He spilt hot water on his foot.'

        S                                          P
   b. [Yaba-y [ne   ʔa-soʔa   lala      ssaw-ø]]   [saray-ø.]
      foot-3s  REL  3sA-spill hot.water DUP-on-3snU  sore-3snU
      'His foot, that he spilt hot water on, is sore.'
```

4.3.2.6 Some thoughts on reduplication in relative clauses

The lack of reduplication when the subject of a transitive clause with a noun phrase as object is relativised raises interesting questions. Although Dobel is not an ergative language, if it were the case that when the subject of *all* transitive verbs is relativised there was no reduplication on the verb, then we could say that when the *absolutive* (subject of intransitive verbs or object of transitive verbs) is relativised the verb is reduplicated, but when the *ergative* (subject of transitive verbs) is relativised the verb is not reduplicated. However, as we have seen, the relativisation of the subject of a transitive verb whose object has cliticised to the verb phrase also causes the verb to be reduplicated, thus such a subject acts in the same way as the subject of an intransitive clause. As we have seen, the argument O may occur as a full NP, a full pronoun, or an undergoer enclitic. Each of these is a full clause constituent, but it is when the surface occurrence of O is a full phonological word that the verb is reduplicated in the relative clause if the subject is relativised.

Relative clauses tend to encode information which is old information and also backgrounded information that is not on the event line. Also modifiers in the noun phrase

usually encode given information, for example when Dobel verbs are used attributively within the noun phrase they encode given information, whereas when they are used predicatively in the main clause they tend to encode event-line information. In Dobel reduplication of the modifier in the noun phrase marks the fact that the information is backgrounded. Compare the following examples, where in (181) the information that the person is good is old information, and the new information is that he has died, whereas in (182) new, foregrounded information is that he is good.

(181)   *Tamatu   ssoba-ni   ne   ?a-kʷoy   ti.*
        person   DUP-good-3saU   DEM   3sA-die   PERF
        'That good person has died.'

(182)   *Tamatu   ne   soba-ni.*
        person   DEM   good-3saU
        'That person is good.'

If we return to the relative clauses, we may note that intransitive relative clauses where the subject is relativised, are simply giving known information about the head of the noun phrase. The same is true of transitive relative clauses, where the object is relativised and is marked in the relative clause as an enclitic. In the case of transitive clauses where the subject is relativised, the information is still only backgrounded information *if* the object is an Undergoer enclitic, as in (173) above. However, if the object is a full pronoun or a full noun phrase, as in (172) and (171) respectively, then that object is *marked* as being foregrounded information. In the case of a full pronoun, it is old information, but marked as foregrounded; in the case of a full noun phrase it is new information or information that is brought back into focus in the discourse *and* it is foregrounded. Because of this marking of the information as foregrounded, it would be contradictory to mark the clause as backgrounded by reduplication. This accounts for the lack of reduplication in relative transitive clauses where O is a full phonological word.

4.3.2.7 *Other syntactic uses of reduplication*

There are certain other areas where reduplication is used, which are not cases of modification.

Firstly, the verb is reduplicated after the construction *?e fuy fay* 'not want' as in:

(183)   *?E-di   fuy   fay   da-rreɲin-ni.*
        desire-3pU   finish   from   3pA-DUP-hear-3pU
        'They don't want to listen to him.'

In this case the foregrounded information is the 'not wanting', rather than the hearing.

Secondly, the verb is reduplicated in subordinate temporal clauses such as the following:

(184)   *Jon   orama na   ?a-bbana   nama na-r-tom   tamatu ne.*
        Jon   just CONJ   3sA-DUP-leave then   3sA-VR-meet person DEM
        'Jon had just left when he met that person.'

Thirdly, the verb is reduplicated in the cleft construction with *ya?a* translated as "It is X who...".

(185)   *Tuwan   nay   fufun   ya ʔa   ʔa-yyilatu-ye.*
        Lord   the.one top   that   3sA-DUP-choose-3pU
        'It is God who chose them.'

(186)   *Ne   nda   ʔa-yi,   saʔu   ya ʔa   ʔu-yyi.*
        He   neg   3sA-go   1sPr   that   1sA-DUP-go
        'He didn't go, it is I who went.'

Fourthly the verb is reduplicated after the construction *nama nay wa na*, a construction that assumes that the information is known information, and it links it to some new information in a contrast relationship:

(187)   *Nal   tufu   wur   ari,   nama   nay   wa   na*
        (3sA)take   generation ten   approximate, then   EM   DEM   and

        ʔa-kk ʷoy   wa.
        3sA-DUP-die   LOC
        'He lived to see about ten generations and now he has died/and here he's died now (as we know).'

Fifthly, the verb may be reduplicated when the verb is a complement of another verb:

(188)   *Wa   nda   na-r-la ʔa   ʔa-kk ʷoy,   nak ʷain*
        seem   NEG   3sA-VR-know   3sA-DUP-die,   that's.why

        ʔa-yi   na-r-fayyiri.
        3sA-go   3sA-VR-DUP-tell
        'It seems he didn't know he had died, that's why he went and told.'

In this case, the fact that the man had died is common knowledge, and therefore background information, so it is reduplicated.

As I have said, the above examples are not cases of modification within the nominal or verb phrase. However, reduplication can seen to be consistently marking elements which encode information which is semantically backgrounded. While individually it may not constitute strong evidence, collectively they support my conclusion that the underlying function of reduplication is to mark non-event-line, backgrounded information.

References

Andrews, Avery, 1985, The major functions of the noun phrase. In Shopen, ed., 1:62-154.

Comrie, Bernard, 1985, Causative verb formation and other verb-deriving morphology. In Shopen, ed., 1985, 3:309-348.

Dixon, R.M.W., 1979, Ergativity. *Language* 55:59-138.

Hopper, Paul J. and Thompson, Sandra A., 1980, Transitivity in grammar and discourse. *Language* 56:251-399.

Hughes, Jock, 1987, The languages of Kei, Tanimbar and Aru: a lexico-statistic classification. In Soenjono Dardjowidjojo, *NUSA, Linguistic studies of Indonesian and other languages in Indonesia*, Vol. 27, 71-111. Jakarta.

1989, A phonology of Dobel. In Wyn D. Laidig, ed., *Workpapers in Indonesian Languages and Cultures*, Vol. 7., 43-88. Pattimura University and Summer Institute of Linguistics, Ambon.

1995a, Dobel introduction. In Tryon, ed., Comparative Austronesian dictionary: an introduction to Austronesian studies, 5 vols., 637-650. Berlin: Mouton de Gruyter.

1995b, Dobel wordlist. In Tryon, ed., Comparative Austronesian dictionary: an introduction to Austronesian studies, 5 vols. Berlin: Mouton de Gruyter.

n.d., A grammar of Dobel. Manuscript.

Shopen, Timothy, ed., 1985, Language typology and syntactic description, vol. 1: Clause structure, vol. 3: Grammatical categories and the lexicon. Cambridge: Cambridge University Press.

7 *A phonological description of Fordata*

CRAIG H. MARSHALL

1 Introduction

Fordata is an Austronesian language included among the Central Malayo-Polynesian languages (Blust 1978) and is located in the eastern Indonesian province of Maluku. Fordata is spoken mainly in the northern part of the Tanimbar archipelago (see Map 1). It is the vernacular on the islands of Fordata, Larat, Lutur, Nuswotar, Labobar, Teneman, Molo, Maru and north-western Yamdena, along with one island, Sera, in the southern part of the Tanimbar archipelago. The language most closely related is Kei, spoken approximately 100 miles away in the Kei Islands.

In addition to the above areas, the larger trade centers of Saumlaki and especially Ambon now have sizeable communities of Fordata speakers. The current total number of speakers is approximately 23,000–25,000.

The data for this paper were collected by the author during residences totalling twelve months between 1988 and 1990 in the Fordata language area. The majority of that time was spent in the village of Romean, on the island of Fordata. All the villages on Fordata have been visited by the author, along with several on Larat, West Yamdena, and Sera in order to gather word lists, do sociolinguistic surveys, and ascertain dialect boundaries.

1.1 Previous linguistic work

Among previous linguistic research efforts, the first writings mentioning Fordata were done by Catholic missionaries. One sketch by Riedel (1886) largely consisted of ethnographic observations. Geurtjens (1928) wrote a paper on idioms in Tanimbar, using the work of P. Drabbe, also a Catholic priest. Drabbe's grammar sketch (1926) and dictionary (1932) have given this author helpful insights into the language. In his sketch of Fordata grammar, Drabbe briefly discusses his orthographic decisions, which include morpheme breaks and compounding. Drabbe (1926:2) concluded that there were five vowels and fifteen consonants, overlooking the glottal phoneme, and also states that word stress is penultimate on the root, regardless of any affixation. More recently McKinnon's

book, *From a shattered sun* (1992), which describes the Fordata alliance and exchange system, contains a small amount of linguistic information, much of it being quotes from Drabbe. A native Fordata-speaking teacher at Pattimura University in Ambon, J. Vatkaat, has written two unpublished manuscripts dealing with aspects of Fordata phonology (1984) and language use (1989). Vatkaat's phonology paper is mainly a segmental analysis of the phonemes and distributional statements regarding consonants and vowels. He did not include any comments on the morphophonemics. These two papers have been of some use in this phonological analysis. Brief mention of Fordata is also included in several comparative studies (Blust 1978, Collins 1982, Hughes 1987).

1.2 Economic considerations

People in the Fordata speaking area are mainly subsistence-level farmers and fishermen. The cash crop base of the Fordata area has traditionally been *copra* (dried coconut), from which coconut oil is made. In recent years, several factors have been changing the importance of *copra*. Market prices have dropped dramatically for *copra*. New markets in topshells, sea cucumbers, shark fin, bird nests, fruit, and dried fish have also opened up. Improvements in long-range transportation have brought in many more ships to the district ports to trade in these commodities. This broader economic base, along with greatly increased government development programs, has linguistic implications (§1.4) as well. The greater amount of available cash has enabled many families to send most of their children to high school or higher on other islands. The great majority of young people who finish high school (a high percentage in the Fordata area) will not return to the village, but will look for work in larger commercial centers. More Fordata people are now travelling to other parts of Maluku and Indonesia.

1.3 Religious considerations

Almost 100 percent of ethnic Fordata people call themselves Christian. The majority belong to the *Gereja Protestan Maluku* (Moluccan Reformed Church), with small groups of Catholics and Pentecostals. The villages of Labobar and Karatat are mostly Muslim, being comprised of people from other ethnic groups, such as Geser and Bugis, who have been migrating into the area for several generations. These Muslim villages historically have taken on Fordata as their *lingua franca*, although language use and attitudes are currently poor.

1.4 Language use

Although Fordata is the vernacular of the area, there is increasing influence from Ambonese Malay, the trade language of central and south Maluku. The domains in which Fordata is used seem to be stable at this point in time. These domains are the home, in the field, on the trails, and traditional law (*adat*). Standard Indonesian and Ambonese Malay are used in schools, government (community) meetings, and most church activities. Most parents will speak mixed Fordata and Malay to children. For this reason, Fordata people do not become proficient in all areas of the vernacular until the mid to late teen years. Those who go off to

high school normally will not become fluent in using Fordata in anything but surface conversations. Many of the young people in this category could be considered semi-lingual, not knowing any language really well.

1.5 Dialect considerations

Fordata people state that most of the villages speak the same language, with lexical and intonational differences. They all point to Sera as being the most divergent. Time that I have spent in many Fordata villages bears this out, along with corroborating evidence from word lists (Hughes 1987) and sociolinguistic surveys. I propose that there are four dialects: *Romean, Sofyanin, Molo* and *Sera*, with Sera being the most divergent. Figure 1 below[1] shows the phonological variations between the dialects:

Romean	*Sofyanin*	*Molo*	*Sera*
[#ʔV]	[#ʔV]	[hV]	[#ʔV]
[VʔV]	[VʔV]	[VʔV]	[V:]
[VhV]	[VhV]	[VhV]	[V:]
[ai]	[eː],[ɛː]	[ai]	[eː], [ɛː]
[aʔi]	[eʔe]	[aʔi]	[eː]
[au]	[ɔː]	[au]	[ɔː]
[aʔu]	[ɔʔɔ]	[aʔu]	[ɔː]
[ea]	[eː], [ɛː]	[ea]	[eː], [ɛː]
[ahɪ,i]	[ɛhɛ]	[ahɪ,i]	[eː]
[ahu]	[ɔhɔ]	[ahu]	[ɔː]
[Va#]	[V:#]	[Va#]	[V:#]

Figure 1: Variations between the four Fordata dialects

[1] Abbreviations and symbols used in this paper:

σ	syllable	C	consonant	S₁	final syllable of root.
*σ	stressed syllable	CAUS	causative	S₂	penultimate syllable of root
[Âant]	anterior	CAUS, NOM, experiencer		S₃	antepenultimate syllable
[ßcor]	coronal	DUP	reduplication	Sd	ultimate syllable of a disyllabic
[-cont]	continuant	GEN	genitive		genitive enclitic suffix
[+nas]	nasal	incl	inclusive	SF	sentence form
[-son]	sonorant	INTRANS (active)		Sm	monosyllabic enclitic
[+syll]	syllabic	INTRANS intransitive		Sn	pre-penultimate syllables
[-syll]	non-syllabic	lab	labial	Sp	penultimate syllable
[+voice]	voiced	lat	lateral	STATE (accidental or unspecified	
1pe	first person plural exclusive	nas	nasal		causer)
1pi	first person plural inclusive	NOM	nominaliser	STATE (in process)	
1s	first person singular	PART	particle	STATE (progressed or finished)	
2p	second person plural	PL	plural	STATE (progressive, on going)	
2s	second person singular	QNT	quantity	V	vowel
2s	second person singular	RCP	reciprocal	vd.	voiced
3s	third person singular	s.o.	someone	vl.	voiceless

Intonation pattern differences also exist between the dialects, and even between villages, with Sera again being the most divergent.
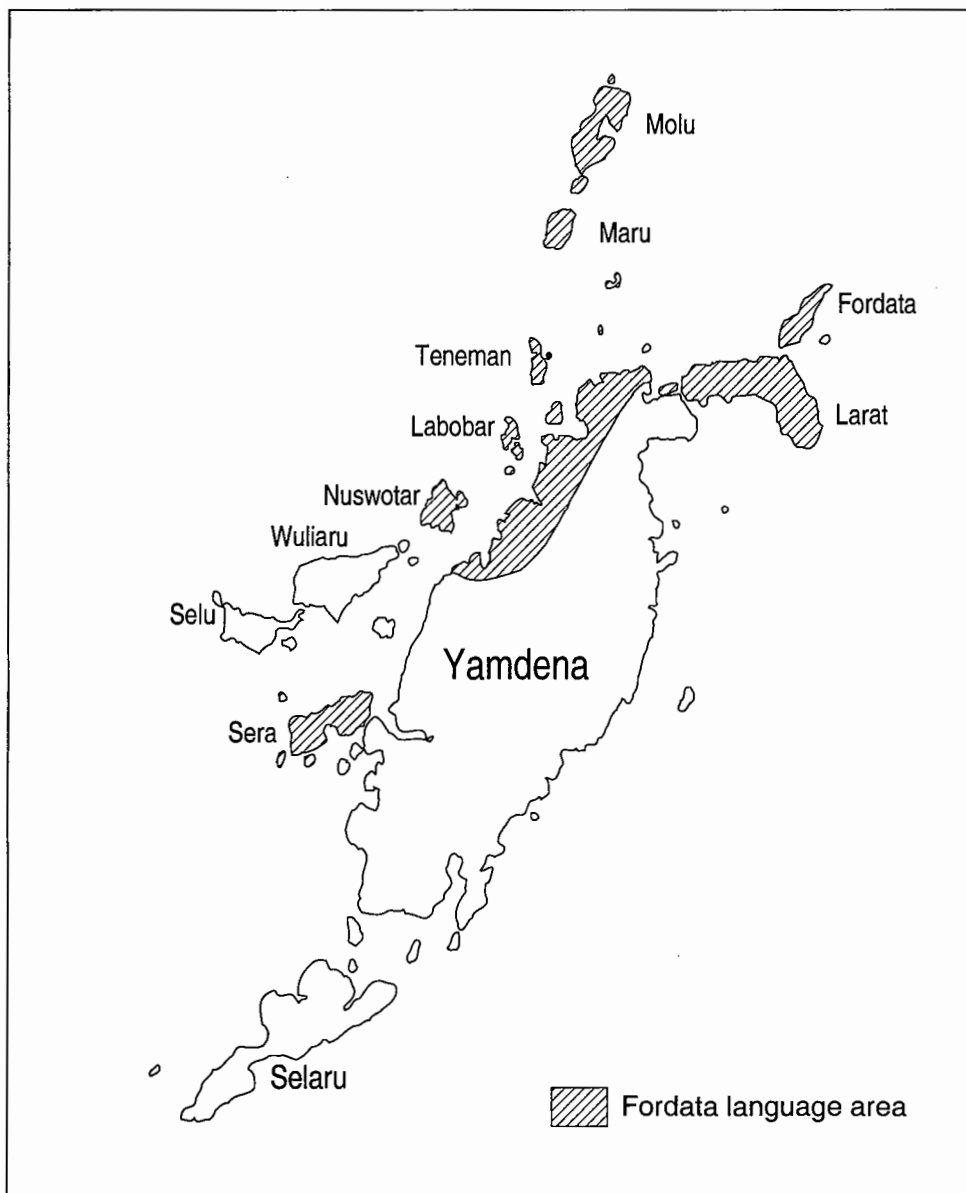
The Tanimbar Islands, showing the Fordata language area

2 Segmental phonology

The Fordata phoneme system has five vowels and sixteen consonants.[2] Stress, which falls on the penultimate syllable of the root, is completely predictable and therefore is not marked in this paper, except in the §4 on word stress. Contrasts between consonants and between vowels are illustrated by minimal pairs in Appendix A.

2.1 Consonantal segments

		Labial	Apical	Laminal	Dorsal	Glottal
Stop	vl.		t		k	ʔ
	vd.	b	d			
Fricative	vl.	f	s			h
	vd.	v				
Nasal		m	n		ŋ	
Liquid:	Lat		l			
Trill			r			
Semivowel		w		y		

Figure 2: Fordata consonants

In Figure 2 above, the points of articulation are labelled as such to show a good symmetrical distribution. In the apical position, /t/ is phonetically dental and /d/ and the other apicals are alveolar. Laminal /y/ is justified as a separate phoneme because of distributional limitations (§3.1). The glottals /h/ and /ʔ/ also display differences in distribution and behaviour (§3.1).

2.2 Description of consonants

2.2.1 Stops

The two stops that are conspicuously missing are the voiceless labial [p] and the voiced dorsal [g].[3]

The phoneme /t/ is a voiceless apico-dental stop and is found in all positions, for example:

(1) [titi]   /titi/   'girl'
     [temar]   /temar/   'bamboo'
     [lotar]   /lotar/   'defend'
     [tetar]   /tetar/   'hop quickly'
     [manut]   /manut/   'bird'

---

[2] Fordata speakers also make use of the following loan phonemes: /c/, /p/, and /j/. These are borrowed from Malay and are not included in the inventory of inherited phonemes.

[3] There are several examples of PAn *p becoming *f* in Fordata. **p*a* 'CAUS' becomes *fa* and **paRi* 'stingray' becomes *fari*.

[balit]	/balit/	'left'
[tniri]	/tniri/	'kingfish'
[tranan]	/tranan/	'south'
[taltelun]	/tal-telu-n/[4]	'in threes'
[fatvelar]	/fa-ta-velar/	'turn s.t. over'

The phoneme /k/ is a voiceless dorsal stop, occurring in all positions, for example:

(2)
[kidə]	/kida/	'friend'
[kesi]	/kesi/	'food for a journey '
[lokat]	/lokat/	'every'
[kikur]	/kikur/	'tail'
[nafraik]	/na-fraik/	'he blows s.t.'
[n̩lorak]	/na-lorak/	'he cuts s.t.'
[kba:]	/kbaa/	'board'
[kyavu]	/kyavu/	'dust'
[n̩siklabir]	/na-si-ka-labir/	'he lies'
[katkatan]	/kat-katan/	'scorpion'

The phoneme /ʔ/ is a glottal stop appearing only intervocalically. Its occurrence is not, however, predictable in this intervocalic position. Phonetically, glottal stops occur also before all vowel-initial words, but are not contrastive in this position (§3.3). Note the following examples:

(3)
[ʔaʔan]	/a'a-n/	'his sibling'
[ʔaʔit]	/a'it/	'centipede'
[muʔu]	/mu'u/	'banana'
[sŋaʔut]	/snga'ut/	'fast'
[n̩siʔik]	/na-si'ik/	'he looks at s.t.'
[daʔut]	/da'ut/	'rain'

The phoneme /b/ is a voiced bilabial stop.[5] It is found word-initially and -medially, but never word-finally in normal speech in word roots,[6] for example:

(4)
[n̩belat]	/na-belat/	'he fences'
[bobən]	/boban/	'casket'
[yabi]	/yabi/	'skill'
[kaban]	/kaban/	'cone shell'
[branə]	/brana/	'male'
[blaŋar]	/blaŋar/	'coral trout'
[nabla:n]	/na-blaan/	'he coughs'
[rabrubi]	/rabrubi/	'yellowfin tuna'

---

[4] Hyphens denote morpheme breaks, either for derivational or inflectional affixes or for reduplicated morphemes.

[5] Reflexes of PAn *p in Fordata are interesting, as the labials /b/ and /f/ represent historical *p, for example *bana* 'go' is from *panaw and *fana* 'arrow, shoot arrow' is from *panaq.

[6] Non-nasal labials, e.g. /b f v/ do not occur word-finally in normal speech. However, in fast speech, final vowel deletion (§6.7.1.2) will permit these non-nasal labials to occur in word-final position.

The phoneme /d/ is a voiced apical stop, occurring in all positions,[7] as demonstrated in the following examples:

(5) | [dalaŋ] | /dalaŋ/ | 'very' |
     | [dekut] | /dekut/ | 'orphan' |
     | [kidə] | /kida/ | 'friend' |
     | [didə] | /dida/ | 'our (incl.)' |
     | [wahad] | /waha-d/ | 'our (incl.) face' |
     | [yatad] | /yata-d/ | 'our (incl.) liver' |
     | [nakduvul] | /na-ka-duvul/ | 'he sank' |
     | [dakdokun] | /dak-doku-n/ | 'DUP-sit-PART' |

2.2.2 Fricatives

The phoneme /f/ is a voiceless labio-dental[8] fricative and does not occur word-finally in normal speech, for example:

(6) | [fenu] | /fenu/ | 'turtle' |
     | [faliak] | /faliak/ | 'expression' |
     | [nfofə] | /na-fofa/ | 'he floats' |
     | [lafar] | /lafar/ | 'starvation' |
     | [fteni] | /fteni/ | 'parrot fish' |
     | [tfuri] | /tfuri/ | 'triton shell' |
     | [naftofi] | /na-f-tofi/ | 'he's doing wash' |
     | [fatfitun] | /fat-fitu-n/ | 'in sevens' |

The phoneme /s/ is a voiceless apical fricative occurring in all word positions, as can be seen in the following examples:

(7) | [sula:r] | /sulaar/ | 'sailfish' |
     | [suhut] | /suhut/ | 'disease' |
     | [tesin] | /tesin/ | 'original' |
     | [maresən] | /maresan/ | 'chile pepper, spicy' |
     | [nlabas] | /na-labas/ | 'he chases s.t.' |
     | [kalumus] | /kalumus/ | 'mud' |
     | [slaru] | /slaru/ | 'corn' |
     | [sŋa:t] | /sŋaat/ | 'spear' |
     | [salsolə] | /sal-sola/ | 'fish spear' |
     | [saksakan] | /saksakan/ | 'wave' |

The phoneme /h/ is a voiceless glottal fricative appearing only intervocalically,[9] for example:

---

[7] The phoneme /d/ word-finally has a limited occurrence, appearing only as a 1pi possessive suffix.

[8] The inner lower lip to teeth is the specific point of articulation for both /f/ and /v/.

[9] One word, *hi'i* 'truly, really', has been recorded in which /h/ is initial. This word was not found in Drabbe's (1932) dictionary. As previously mentioned in §1.5, speakers of the Molo sub-dialect add an /h/ to word-

(8)   [suhut]   /suhut/ ·   'disease'
       [ʔihin]   /ihi-n/   'his flesh'
       [vavahan]   /va-vahan/   'mud'
       [vahi]   /vahi/   'paddle'

The phoneme /v/ is a voiced labio-dental fricative. Among some of the younger speakers of Fordata, the contrast with /w/ has been lost and /v/ has been replaced by /w/.[10] It occurs word-initially and -medially but not word-finally in normal speech, for example:

(9)   [vata]   /vata/   'female'
       [vinik]   /vinik/   'peel'
       [vavu]   /vavu/   'pig, boar'
       [nanavut]   /na-navut/   'he is drunk'
       [vrekə]   /vreka/   'type of lizard'
       [vra:]   /vraa/   'hermit crab'
       [savsovu]   /sav-sovu/   'hot sauce'
       [varvuru]   /varvuru/   'small bore bamboo'

2.2.3 Nasals

The phoneme /m/ is a voiced labial nasal appearing in all word positions, though it has a limited occurrence in word-final position, appearing only as 2sg. genitive suffix.

(10)   [maŋun]   /maŋun/   'guest'
       [manovən]   /manovan/   'rat'
       [kamɛrən]   /kameran/   'late'
       [kamat]   /kamat/   'testicle'
       [liməm]   /lima-m/   'your (sg.) arm'
       [ninim]   /nini-m/   'your (sg.) gum'
       [m̩ŋaʔun]   /mŋa'un/   'true'
       [m̩nɛrən]   /mneran/   'hunting tools'
       [marma:r]   /marmaar/   'north'
       [makmukur]   /makmukur/   'round'

The phoneme /n/ is a voiced apical nasal occurring in all word positions, for example:

(11)   [nara]   /nara/   'star'
       [naⁱt]   /nait/   'wind'
       [binən]   /binan/   'plate'
       [lunə]   /luna/   'carry'

---

initial vowels (instead of [ʔ]). Historically, Molo speakers came from the other sub-dialect areas. In the Sera dialect, /h/ has disappeared altogether.

10 The vast majority of speakers who are fluent in speaking Fordata do use /v/. The degree to which this loss of contrast occurs varies from village to village. In Larat (the town) and outside of the Fordata language area (Ambon, Saumlaki, etc.) this neutralisation is much more pronounced (my own observations). Young speakers also are more likely to produce /w/ instead of /v/. This may be due to influence from Malay, which does not have the voiced labio-dental fricative phoneme /v/. The /f/ and /v/ in Malay words (phonetically speaking) are replaced by /v/ in Fordata (§8).

[lawan]	/lawan/	'appearance'
[latan]	/latan/	'stick'
[tnabar]	/tnabar/	'stomping dance'
[ŋneə]	/ŋnea/	'hot'
[faŋnanaŋ]	/fa-ŋnanaŋ/	'cause to remember'
[banbanan]	/ban-bana-n/	'journey'

The phoneme /ŋ/ is a voiced dorsal nasal and occurs in all word positions, which can be seen in the following examples:

(12)	[ŋeən]	/ŋean/	'garden fence'
	[ŋorə]	/ŋora/	'full'
	[suŋu]	/suŋu/	'pierce s.t.'
	[noŋə]	/na-oŋa/	'it lights'
	[dalaŋ]	/dalaŋ/	'very'
	[weəŋ]	/weaŋ/	'deceive'
	[ŋgrovə]	/ŋrova/	'grass'
	[ŋtoən]	/ŋtoan/	'black'
	[ŋatŋatul]	/ŋat-ŋatul/	'gecko'
	[nafŋai]	/na-f-ŋai/	'he clears (a garden)'

2.2.4 Liquids

The phoneme /l/ is a voiced apical lateral occurring in all word positions, for example:

(13)	[lanu:n]	/lanuun/	'land, ground'
	[liŋat]	/liŋat/	'forest hut'
	[amfalak]	/am-falak/	'we (incl) say'
	[sula:r]	/sulaar/	'sailfish'
	[nail]	/na-ail/	'he dives'
	[ņwatɪl]	/na-watil/	'he jumps'
	[slaru]	/slaru/	'corn'
	[vlunur]	/vlunur/	'type of tree'
	[salsolal]	/sal-solal/	'slippery'
	[siklabɪr]	/si-ka-labir/	'to lie'

The phoneme /r/ is a voiced apical trill. It occurs in all positions, and word-finally can at times can be heard in free variation with the voiced alveolar flap [ɾ], for example:

(14)	[ralan]	/rala-n/	'his insides'
	[renən]	/rena-n/	'his mother'
	[ņmɛrət]	/na-merat/	'he is clean'
	[mara:n]	/maraan/	'light (weight)'
	[livur]	/livur/	'village'
	[ņfanaʔur]	/na-fa-na'ur/	'he fishes'
	[fra:]~[fɾa:]	/fraa/	'iron wood'
	[brana]	/brana/	'male'
	[marmuri]	/marmuri/	'young'
	[nakreti]	/na-kreti/	'it is shallow'

2.2.5 Semivowels

The phoneme /w/ is a high back non-syllabic semivowel appearing word-initially and word-medially, as is illustrated in the following examples:

(15) [wahan]   /waha-n/   'his face'
      [wa?ar]   /wa'ar/   'root'
      [?awan]   /awa-n/   'his spouse'
      [dawan]   /dawan/   'big'

The phoneme /y/ is a high front non-syllabic semivowel and occurs word-initially. This semivowel occurs in non-initial positions within a morpheme,[11] but not intervocalically, for example:

(16) [yaha]   /yaha/   'dog'
      [yai]   /yai/   'shark'
      [fyawan]   /fyawa-n/   'its price'
      [ŋyebə]   /na-yeba/   'it lights'
      [yatyatak]   /yat-yatak/   'top of a roof'
      [yadyadu]   /yad-yadu/   'shivering'

The phonemes /w/ and /y/ are consonants that occur in the non-nucleus position of a syllable. Since CC onsets can occur in the stressed syllable and at the boundaries of the stressed syllable and a preceding syllable, and sequences of only two vowels are allowed, the following phonemic representations are proposed:

(17) /bwea/   [bweə]   'crocodile'
      /bwaa/   [bwa:]   'maggot'

(18) /lyawan/   [lyawan]   'clear'
      /ilyaan/   [?ɪlya:n]   'tomorrow'
      /u-yadu/   [uyadu]   'I shiver'
      /na-yeba/   [ŋyebə]   'it illuminates'

2.3 Vowel segments

	Front	Central	Back
High	i		u
Mid	e		o
Low		a	

Figure 3: Vowels

2.4 Vowels: description

Fordata has five vowels. The following is a description of these five vowels.
The phoneme /i/ is a high tense[12] front unrounded vowel, for example:

---

[11] The phoneme /y/ fills the medial slot only in reduplicated forms.

(19) [ʔiə]   /ia/   'he/she/it'
      [ʔiɲi]   /iɲi/   'lingua wood'
      [fitik]   /fitik/   'lightning'
      [liək]   /liak/   'other'
      [diti]   /diti/   'female of nobility'
      [vaʔi]   /va'i/   'garden'

It occurs in all positions, except preceding /r/ or /l/, where its lax variant [ɪ] occurs, for example:

      /i/ _ [-tense]/ ___ C
                          [+son]
                          [-nas]

(20) [ʔɪrə]   /ira/   'they'
      [ʔɪrit]   /irit/   'thin rope'
      [lahɪr]   /lahir/   'superlative'
      [ņbahɪr]   /na-bahir/ 'he pays'
      [basɪl]   /basil/   'true'
      [ņwatɪl]   /na-watil/ 'he jumps'

The phoneme /e/ is a mid tense front unrounded vowel.

(21) [ʔevun]   /evu-n/   'his stomach'
      [ʔetal]   /etal/   'piece'
      [ņleal]   /na-leal/ 'he is'
      [kedən]   /kedan/   'a little'
      [te]   /te/   'or'
      [ne]   /ne/   'that'

It occurs in all positions,[13] except preceding /r/, where its lax variant [ɛ] occurs,[14] for example:

      /e/ → [-tense] / ___   C
                            [+son]
                            [-nas]
                            [-lat]

(22) [sɛrə]   /sera/   'sago'
      [lɛrə]   /lera/   'sun'
      [ņmɛrat]   /na-merat/ 'he's clean'
      [wɛrin]   /weri-n/   'its fin'

---

[12] The tense and lax distinctions make no claims of tenseness of musculature in Fordata, but are conventional terms for close and open vocoids.

[13] There are a few examples of /e/ occurring word-finally that have been discovered so far: [ne] /ne/ 'that' and /ma/ + /ne/ 'will', which is frequently heard as /ma'nene/ or shortened to /me/, and [te] /te/ 'or'.

[14] In a few instances, [ɛ] has been recorded preceding /l/ and /n/. In all of these instances, the form was shortened either as a result of rapid speech (vowel elision or colloquial shortening). The word [weən] is frequently spoken [wɛn], [vɛrin] as [vɛn]. The word /wahal/ becomes [wɛl] depending on its grammatical position in the sentence. In summary, /e/ becomes [ɛ] before /l/ in a mono-syllabic morpheme.

The phoneme /a/ is a low lax central unrounded vowel occurring in all positions, for example:

(23)	[ʔara]	/aru/	'viper'
	[ʔahu]	/ahu/	'village'
	[blaman]	/blaman/	'deep'
	[nail]	/na-ail/	'he dives'
	[ʔivar]	/ivar/	'news'
	[n̥foar]	/na-foar/	'he is stupid'
	[ŋkaʔa]	/na-ka'a/	'he knows'
	[nara]	/nara/	'star'

Its allophone [ə] occurs when the preceding stressed syllable contains a non-low vowel and the following consonant is either /n/ or /k/. The allophone [ə] also occurs in word-final syllables with no coda.[15] I posit the following rule for [ə]:

$$/a/ \rightarrow ə \;/\; \underset{[-low]}{V} \quad \underset{n,k}{C} \underline{\quad} (C) \,\#$$

(24)	[ʔalemən]	/aleman/	'heavy'
	[voən]	/voan/	'spider'
	[n̥fonək]	/na-fonak/	'he hides s.t.'
	[faŋofək]	/fa-ŋofak/	'turn over'
	[lɛrə]	/lera/	'sun'
	[kidə]	/kida/	'friend'
	[favɪrə]	/favira/	'split bamboo'
	[n̥lekə]	/na-leka/	'he falls'

The phoneme /u/ is a high tense back rounded vowel and is found in all word positions, as illustrated by the following examples:

(25)	[ʔubun]	/ubu-n/	'his grandchild / grandfather'
	[ʔuru]	/uru/	'spoon'
	[dubɪl]	/dubil/	'arrow'
	[vulən]	/vulan/	'moon, month'
	[taul]	/taul/	'small bucket'
	[daʔuk]	/da'uk/	'brother / sister in-law'
	[vutu]	/vutu/	'ten'
	[vatu]	/vatu/	'rock'

The phoneme /o/ is a mid tense back rounded vowel appearing word-initially and word-medially, for example:

(26)	[ʔovən]	/ovan/	'night'
	[ʔoar]	/oar/	'river'
	[boku]	/boku/	'some'
	[naŋtoən]	/na-ŋtoan/	'he is black'

---

[15] The lax allophone [ə] does occur in certain grammatical positions at the end of a clause or sentence; e.g., [ə mə], [ə nə] and [ə bə], which are particles plus a conjunction. When the a morpheme occurs with none of the above conjunctions, or when the conjunctions do not have a preceding [a], the [a] is tense.

All word-initial vowels are initiated with a phonetic glottal. There is no contrast with its absence, so the glottal stop in this position is not phonemic. That the glottal stop is omitted when the stem is prefixed is a further reflection of its purely phonetic role in initial position, for example:

(27)   [myail]   [mi]+[?ail]   'you all dive'
         [rotu]   [ra]+[?otu]   'they make/do s.t.'

Note that in both examples the resulting vowel cluster is resolved. See the discussion on verb prefixing in §6.1.

2.5 Vowel sequences

In Fordata each vowel forms a syllable peak or nucleus. Thus, sequences of like and unlike vowels cross syllable boundaries. The syllable will be more thoroughly discussed in §5.1, and 5.2. There are no dipthongs in Fordata.

(28)   [naⁱt]   /nait/   'wind'
         [romeən]   /romean/   'Romean village'
         [bweə]   /bwea/   'crocodile'
         [voən]   /voa-n/   'it's smell'
         [?oar]   /oar/   'stream, river'

To interpret [naⁱt] as /nayt/ would introduce a new CV pattern (§5.1). Thus, the above examples, plus words such as [voən] 'it's smell', with sequences of vowels in either order, are best interpreted as vowel sequences: /ia/, /ai/, /au/, /ea/, /oa/, and /ua/.[16] The section on stress (§4) will further clarify the decision to interpret vowel sequences as segments instead of diphthongs.

The following illustration shows the relationship of vowels, syllables and stress, which in Fordata is on the penultimate syllable of the word (99 percent of the time this is the root; §4.1). The syllable is represented by [σ] and the stressed syllable by [*σ].

(29)   *σ   σ
                 / | | \
               (C) V   V (C)
                | | | |
                    o   a   '2s'
               t   u   a   t   'palm brandy'
               w   e   a   n   'like, as'
               t   a   u   l   'bucket for drawing water'

Geminate vowel sequences in Fordata are manifested by phonetic length and like non-geminate vowel sequences, occur across syllable boundaries with the first vowel carrying the stress. Geminate sequences are found on the phonemes /a/, /i/, and /u/.

---

[16] Several other Central Malayo-Polynesian languages, such as Buru (Grimes 1991), Roma (Lee Steven, pers. comm.), and Meher (John Christensen, pers. comm.) also permit vowel sequences.

(30)
```
      σ        *σ   σ
      |   /    |    |   \
     (V) (C)   V    V  (C)
      |   |    |    |    |
          b    a    a         'where'
      a   r    a    a         'jellyfish'
               u    u    ŋ     'sarong'
          n    u    u    ŋ     'coconut'
      i   ŋ    i    i    r     'foul smell'
```

The following examples illustrate the contrast between single vowels and geminate vowel sequences:

(31) [la'laːr]   /lalaar/   'peninsula'
      ['lalar]   /lalar/   'heated coconut milk'

(32) ['ʔalan]   /ala-n/   'its gills'
      [ʔa'laːn]   /alaa-n/   'his arm'

(33) [raː]   /raa/   'outrigger'
      [ra]   /ra/   'plural'

(34) [maː]   /maa/   'come'
      [ma]   /ma/   'so that'

(35) [ma'raːn]   /maraan/   'light (weight)'
      [maran]   /maran/   'type of cloth'

(36) [niːŋ]   /niiŋ/   'smile'
      [niŋ]   /niŋ/   'my'

(37) [naf'niːt]   /na-fniit/   'he spits'
      ['fnitə]   /fnita/   'ladder'

(38) [ʔi'ŋɪːr]   /iŋiir/   'bad smell'
      ['ʔiŋi]   /iŋi/   'iron wood'

(39) ['fuːn]   /fuun/   'bottom'
      ['funu]   /funu/   'banana cluster'

(40) ['luːn]   /luu-n/   'his tear'
      ['lunə]   /luna/   'carry'

(41) ['vuːt]   /vuut/   'hair'
      ['vutu]   /vutu/   'ten'

These phonetically long vocoids are phonemically two homogeneous vowels, and as with all vowels, form individual syllable peaks. Evidence for this analysis can be found in four areas.

First, contrast between [a] and [aː] and [i] and [iː] has clearly been shown in examples (31–38) above; thus the contrastive nature of vowel length can be readily established. Second, CV patterns (non-geminates containing mid vowels) described in sections §3.2 and §3.3 permit two-vowel sequences to be interpreted as separate syllabic segments, for example V.V. The third area of evidence is found in word stress. Stress is always

penultimate, thus the following words with phonetically long vowels must be analysed as two adjacent vowels, each in separate syllables, for example:

(42)   [la'la:r]   /lalaar/   'peninsula'
       [ma'ra:n]   /maraan/   'light (weight)'
       [la'la:n]   /lalaan/   'long time ago'

In the examples above, if phonetic vowel length is analysed as being phonemically short, or treated as one syllable, the resulting words would have ultimate stress, thus introducing a new stress pattern.

Lastly, if stems with phonetically long vowels are analysed as short vowels and they undergo reduplication, the resulting forms are incorrect, for example:

Analyzed as V (yields incorrect form)			Analyzed as V.V (yields correct form)		
Structure of root **CV:(C)**	Expected under reduplication **CVC.CVC**		Structure of root **CVV(C)**	Expected under Reduplication **CV.CVV(C)**	
bu:k	*bakbuk	*buuk*	'suck'		*babuuk*
la:r	*larlar	*laar*[17]	'sail'	*lalaar*	
ŋai	*ŋaiŋai	*ŋai*	'clear'		*ŋaŋai*

Figure 4: Reduplicated words with V.V sequence

For a full discussion on reduplication, see §7.

3 Distribution

3.1 Consonants

The phonemes /b v f m d t s n l r k h ŋ ' w y/ are distributed as follows:

(a) All consonants, with the exception of /h/ and /'/, occur morpheme-initially.

(b) All consonants, with the exception of /y/, occur intervocalically.

(c) All consonants except the labials /b f v/ and glottals /h/ and /'/ occur morpheme-finally in normal speech. As mentioned previously in §2.2.2 and §2.2.3, the occurrence of /d/ and /m/ morpheme-finally is limited to marking nouns for 1pi and 2s possession.

(d) Consonant clusters occur across syllable and morpheme boundaries. Geminate clusters are not allowed. Glottals /h/ and /'/ do not occur in either the first or second slot of a consonant cluster. Semivowels /w/ and /y/ do not occur in the initial position of a CC cluster. Labials and sonorants comprise most of the second position cluster options.

---

[17] The loss of historical *y from *layaR 'sail' accounts for the Fordata form *laar* with homogeneous vowels.

Clusters do occur within a derived syllable type (§5.2.1, §6.7.1.1) at the beginning of a word. The first consonant in a complex (CC) onset can be filled only by an obstruent. Any consonant (except /h/ and /'/) can fill the second position.

3.2 Vowels

Any vowel, with the exception of /o/, can occur in any position. There are, however, restrictions on which vowel sequences can co-occur. The phonemes /i e a o u/ all occur initially in the vowel sequence. Only /i/, /a/ and /u/ occur in the second position, as /i/ and /u/ occur only when following /a/, or when the vowels form a geminate sequence, as in /ii/ and /uu/.

	i	e	a	o	u
i	*dii* 'base'		*fiaŋ* 'time'		
e			*wear* 'water'		
a	*nait* 'wind'		*araa* 'jellyfish'		*raun* 'true'
o			*voa-n* 'smell-its'		<u> </u> [18]
u			*ruat* 'tide'		*ruun* 'sea cow'

Figure 5: Vowel sequences

3.3 Consonant and vowel combinations

The consonants /b d t k v f s m n ŋ l r/ co-occur with all vowels. In syllable-final position, however, /e/ is preceded only by /t/ and /n/ (see §2.4 for the discussion of /e/). The phoneme /h/ has a limited distribution and never occurs with mid vowels. All vowels occur preceding /'/, but only /a/, /i/ and /u/ can occur following /'/. However, in word-initial position, all vowels can follow the purely phonetic appearance of [ʔ] (see §2.4).

Co-occurrence restrictions exist also with semivowels. The phoneme /y/ is followed only by non-high vowels and /w/ can precede any vowel except /u/.

Co-occurrence restrictions between the onset and vowel peak of a syllable are shown as follows:

---

[18] The vocative *-mou* 'call a dog' is the only example found to date with this vowel sequence.

Onset	Vowel Peak
b	i e a o u
v	i e a o u
f	i e a o u
m	i e a o u
w	i e a o
d	i e a o u
t	i e a o u
s	i e a o u
n	i e a o u
l	i e a o u
r	i e a o u
y	e a o
k	i e a o u
h	i a u
ŋ	i e a o u
ʔ	i e a o u

Figure 6: Co-occurrence restrictions

4 Stress

Word stress is described using a metrical grid (cf. Goldsmith 1990:169ff). Stress falls on the penultimate syllable of phonological words (being defined as a root plus any affixation) and is not affected by prefixes, infixes or by extrametrical monosyllabic enclitics, such as the phrase level plural marker *ra* and the particle *a*. The notion of extrametricality states that some languages ignore certain sequences or specific morphemes in building the metrical structure of a word.

(43)
```
                  x
         (x)      x        x       (x)
    (C)   V - (C) V   (C)  V  -(C)  V
     |    |   |  |     |   |   |    |
              v   a    h   i              'paddle'
              v   a    h   i    r    a    'paddles'
              v   a    h   i         a    'paddle-PART'
     n    a   v   a    h   i              '3s-paddles'
              l   i    m   a    n         'hand-3sGen'
```

Some monosyllabic genitive enclitic suffixes are also extrametrical when the final consonant of the root is /r/,[19] in that they also do not participate in stress assignment.

---

[19] All roots that require a genitive enclitic suffix are vowel final, except for a very small number of nouns that end in *r*.

(44)   x
             x   x   (x)
          s u m a r -a n   'stomach-3sGen'
          b u b u r -a m   'chest-2sGen'
             e . a r -a ŋ   'ear-1sGen'

Stress is not phonemically significant, so it will only be written in phonetic renderings when word stress is focused.

4.1 Stress with disyllabic genitive enclitic suffixes

The following figure shows both the mono and disyllabic genitive enclitic suffixes.

	1s	1pi	1pe	2s	2p	3s	3p
normal	-ng -k[20]	-dida -d	-mami	-m	-bira	-n	-rira -r
root ends in /r/	-ang	-ad		-am		-an	

Figure 7: Genitive enclitic suffixes

When disyllabic genitive enclitic suffixes are added to the root, the main stress shifts from the root to the new penultimate syllable.

(45)   x
             x   x
          y a n a   'child'

(46)   x
             x   x   x   x
          y a n a -r i r a   'child-3pGen'

(47)   x
             x x   x   x
          w a a -d i d a   'place, position-3piGen'

The plural enclitic *ra*, when added after a disyllabic genitive suffix, still has no effect on the word stress.

(48)   x
             x   x   x   x   (x)
          a w a -b i r a -r a   'spouse-2pGen-PL'

---

[20] The genitive enclitic suffix -*k* 'my' has a very restricted use, only appearing with the root *yana*- 'child' and *ina*- 'like, desire'.

4.2 Stress with vowel loss

There is a general pattern in Fordata to drop vowels before a syllable that takes primary stress (see §6.7.1.2 for a more complete discussion). In the example below, the vowel loss is from a non word-final geminate vowel sequence before a stressed syllable.

(49)

```
                x
        x   x   x
    s u l a   a r        'sailfish'
        m e l a          'nobility'


                x
    x   x       x   x
    s u l a r .me l a     'sailfish of nobility'
```

The combination of more than two roots does not change the pattern of the penultimate syllable in a stress group taking the primary stress.

(50)

```
                x
        x   x   x
    t a m a t a            'person'
    i s a   a              'one'
        w a t a n          'only'


                    x
    x   x   x  x    x   x
    t a m a t .i s a .w a t a n   'only one person'
```

Where there is more than one stress group in a grammatical phrase, the phrase stress occurs on the final stressed syllable of the phrase.

(51)

```
                x
    x   x   x
    t a m a t a            'person'
        d a w a n          'big'
    i s a   a              'one'
        w a t a n          'only'


                                x
                    x           x
    x   x       x   x   x   x   x   x
    t a m a t .d a w a n .i s a .wa t a n   'only one big person'
```

5 Larger phonological structures

The phonological word, defined earlier in §4 as a root plus any affixation, can be further clarified as: where a person-subject marking prefix or derivational prefix is present, this defines the onset of the word. If both prefixes are present, the subject-person prefix is always the first prefix. Similarly, a genitive enclitic suffix, when present, defines the coda of a phonological word.

5.1 Syllables

There are four canonical syllable types represented in Fordata underlying forms. These are all derived from the abstract (C)V(C).

(52)   V   *a*   'BE'
       VC   *al*   'to, for'
       CV   *ma*   'so that'
       CVC   *wel*   'NEG'

These four canonical syllable types are organised around the following notation:

$$
\begin{array}{ccccccc}
 & & x & & & & \\
x & x & x & x & x & x & x \\
(S_n) & (S_3) & S_2 & (S_1) & (S_m) & (S_d) & (S_p)
\end{array}
$$
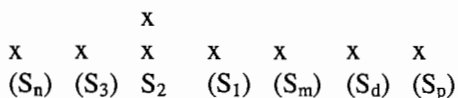
S_m is the monosyllabic enclitic (either the genitive suffix, plural marker, or the particle *a*), which does not affect stress. S_d is the ultimate syllable of a disyllabic genitive enclitic suffix. Only with this suffix does the stress shift to the penultimate syllable (S_m position) of the suffix. The plural marker enclitic *ra* is the only member of S_p and it does not affect stress. S_1 is the final syllable of the root. S_2 is the penultimate syllable of the root and always takes stress if disyllabic S_d is not present. S_3 is the antepenultimate syllable of the root. S_n are pre-penultimate syllables.

5.2 Syllable patterns

Monomorphemic trisyllabic roots are uncommon. Most words with more than two syllables can be shown to be compounds, "frozen" forms or loans. The syllable in Fordata always has a vowel as the syllable peak with its associated consonants. As mentioned above, the general shape of the syllable is (C)V(C). A list of the various shapes, with examples, of a phonological word is found in Appendix B.

A set of ordered rules (adapted from Grimes 1991:55) provides a framework for understanding how syllables are constructed and modified under certain conditions (resyllabificaton) in Fordata. The following rules associate consonants with syllable peaks, building from left to right.

Consonants (C) immediately preceding a syllable peak (V) are associated with that peak.

Rule 1

$$
\begin{array}{ccc}
{}^*\sigma \quad \sigma & & \sigma \quad {}^*\sigma \quad \sigma \\
/\ |\ \ /\ | & \rightarrow & /\ |\ \ /\ | \quad /\ | \\
C\ V\ C \ \ V\ C & & C\ V\ C\ C\ V \quad C\ V\ C
\end{array}
$$
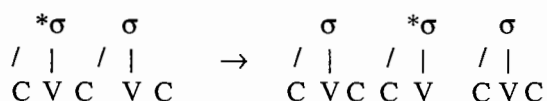
Any unassociated consonants immediately following a syllable peak are then associated with that preceding peak.

Rule 2

$$
\begin{array}{ccc}
{}^*\sigma \quad \sigma & & \sigma \quad {}^*\sigma \quad \sigma \\
/\ |\ \ /\ |\ \backslash & \rightarrow & /\ |\backslash\ /\ | \quad /\ |\backslash \\
C\ V\ C \ \ V\ C & & C\ V\ C\ C\ V \quad C\ V\ C
\end{array}
$$
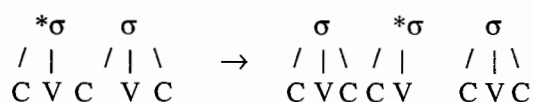
If a syllable peak is followed by another syllable, consonants immediately to the right of the syllable peak are ambisyllabic (cf. Clements and Keyser 1983:36). Therefore, instead of having discrete syllable boundaries, the consonants in this environment are associated with both peaks. The impact of ambisyllabicity will be discussed further in section §6.7.2.

Rule 3

```
  *σ   σ              σ   *σ    σ
 / | \ / | \    →    / | \ / | \ / | \
 C V C  V C          C V C C  V  C V C
```
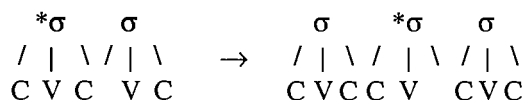
The following examples show the patterns of association with each of the three rules.

(53)

```
        1                    2                      3
      *σ σ                 *σ σ                   *σ σ
     / | |               / |  | \               / |  | \
   (C) V V (C)         (C) V  V (C)           (C) V  V (C)
       | | |             |  |  |  |             |  |  |  |
     o  a a                                                   'wood'
        i a                                                   '3s'
     t  e a                                                   'faeces'
     v  a i                                                   'language'
                       w  e  a  r                             'water'
                       v  u  a  r                             'mountain'
                       ŋ  u  u  r                             'sand'
                          u  u  ŋ                             'sarong'
                          i  i  k                             'dimple'
                          o  a  r                             'steam'
```

(54)

```
        1                  2                    3
     *σ   σ              *σ   σ               *σ    σ
    / | / |             / | / | \            / | V | \
  (C) V C V           (C) V C V C          (C) V C V  C
   |  | | | |           |  | | | |           |  | | |  |
                                             l  i m a  n    'hand-3sGen'
                                             k  i k u  r    'tail'
                                             r  a f a  t    'thatch'
                                             u  l u  n      'head-3sGen'
                                             a  l a  n      'gill'
                                             i  t a         '1pi'
                                          y  a h a          'dog'
                                          f  e n u          'turtle (generic)'
```
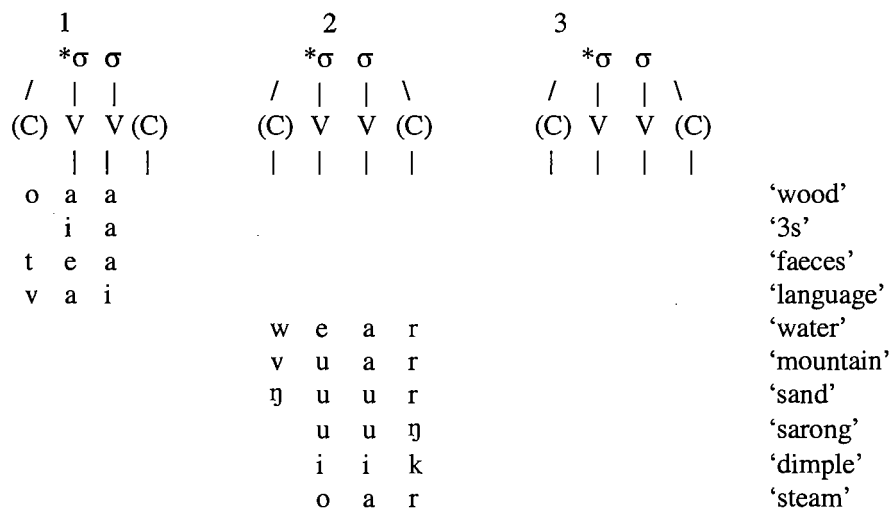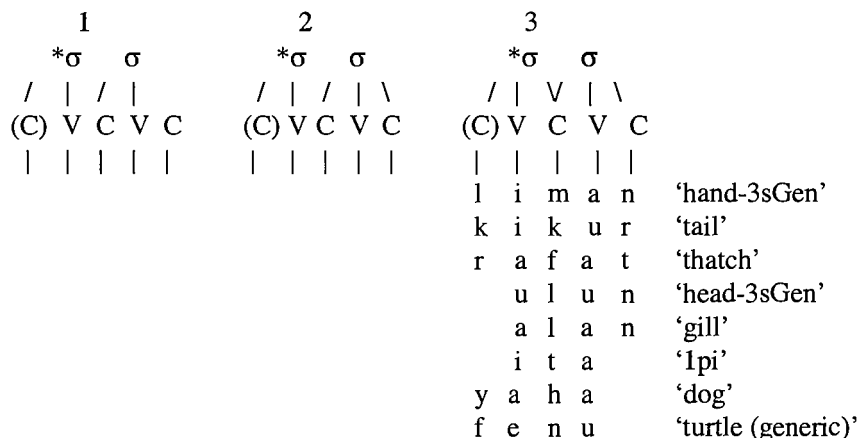
The addition of monosyllabic enclitic suffixes results in the following associations:

(55)   1   2   3
      *σ   σ   σ   *σ   σ   σ   *σ   σ   σ
     / | / | / | / | / | / | / | / | / |
     (C) V C V -C V   (C) V C V - C V   (C) V C V - C V
      | | | | | | | | | | | | | | | | | |
                                                   d   u d   u   r a   'cockroach PL'
                                                   f   i d   u   r a   'gate PL'

(56)   1   2   3
      *σ   σ   σ   *σ   σ   σ   *σ   σ   σ
     / | / | \ | \   / | / | / | \   / | / | \ / | \
     C V C V C - V C   C V C V C - V   C   C V C V C - V C
     | | | | | | | | | | | | | | | | | | |
       'mouth-3sGen'   'chest-1sGen'   s u m a r   a n
       'chest-1sGen'   b u b u r   a ŋ

Stress shifts with disyllabic genitive enclitic suffixes:

(57)   3
             σ   σ   *σ   σ
            / | V | \ / | V |
            C V C V - C V C V
            | | | | | | | | |
            r e n a - d i d a   'mother-1piGEN'
               i n a - b i r a   'desire, like-2pGEN'

Consonant clusters occur in some trisyllabic roots and when roots are partially reduplicated.

(58)   3
             σ   *σ   σ
            / | V / | V | \
            C V C C V C V C
            | | | | | | | |
            k a l b u   u r   'fly'
               i l y a   a n   'tomorrow'
               a l n u r a n   'giant trevally'
            l a b l o b a ŋ   'help'
            k a t k a t a n   'scorpion, pincer'

5.2.1 Disyllabic roots

Disyllabic roots that have not been modified[21] conform to the following pattern:

(59)
```
            *σ     σ
        /  |  V  |  \
       (C)  V (C)  V (C)
```

The following example illustrates disyllabic root patterns:

(60)

'VV	*aa*	'wood'
'VVC	*ear*	'tongue'
'VCV	*ita*	'1pi'
'CVV	*buu*	'large legume'
'CVVC	*viiŋ*	'pull s.t. in'
'CVCV	*leta*	'cross over s.t.'
'CVCVC	*bisak*	'blow s.t.'

Where initial CC clusters are found in surface forms of disyllabic roots, there is a good possibility that these CC clusters probably result from an historical loss of antepenultimate vowels (§6.7.1.1).

(61)

fnita	'ladder'
mnelat	'adolescent female'
bwaa	'maggot'

5.2.2 Monosyllabic roots

Most monosyllabic roots follow the syllabic shape of CV.[22] Many conjunctions and particles are monosyllabic in their full forms. Only one monosyllabic root with the shape V has been observed to date (*a* 'copula, focus').

(62)

ma	'so that, then'
na	'and'
ba	'then, so'
ta	'because'
al	'to, in order to'

5.2.3 Polymorphemic words

Lexical roots combine with affixes and clitics and with other roots in compounding to form polymorphemic words. In following sections, polymorphemic words will be illustrated and discussed.

---

[21] Modified by affixation, compounding, etc.

[22] One monosyllabic root, *wel* 'NEG', with a CVC pattern has been observed. It is highly probable that *wel* is a contraction of *wahal* 'NEG'.

5.3　The shape of enclitic suffixes of a phonological word

The shape of both (S_p) and (S_d) are always CV. The syllable that is represented by (S_m) has a shape of (C)V(C). The morphemes in S_m are very limited, for example

(63)　　*a*　　　　　　　　'PART, singular, focus'
　　　　ra　　　　　　　　'PL'
　　　　aŋ, am, an　　　　　'1s, 2s, 3sGEN'

5.4　The shape of the final syllable of canonical roots

The final syllable (S_1) of a canonical root has a (C)V(C) shape.

(64)　　*aa*　　　　　　　　'wood'
　　　　beat　　　　　　　'type of mango'
　　　　nara　　　　　　　'star'
　　　　betan　　　　　　'rotten'

5.5　The shape of the penultimate syllable of the root

The shape of the penultimate syllable (S_2) of the root is (C)V.[23] This syllable takes word stress unless there is a disyllabic genitive enclitic suffix.

5.6　The shape of pre-penultimate syllables

The pre-penultimate syllables S_3 and S_n have shapes of (C)V(C). The sources of these syllables are from a) prefixes and combinations of prefixes, b) historical trisyllabic roots, c) compounding, and d) loans. Most of the S_3 and S_n syllables that occur with prefixes, loans or historical trisyllabic roots have the vowel /a/ as a peak.

(65)　　**Prefix**
　　　　ta-otu　　　　　[totu]　　　　　'1pi-do'
　　　　na-rata　　　　[nrata]　　　　'3s-go'
　　　　mu-fa-doku　　[fwadoku]　　'2s-CAUS-sit'

(66)　　**Historical trisyllabic**
　　　　kalbuur　　　　'fly (insect)'
　　　　daŋweʔu　　　'hermaphroditic'
　　　　ilyaan　　　　'tomorrow'
　　　　abaa　　　　　'who (PL)'

(67)　　**Loan**
　　　　kareda　　　　'church'
　　　　valada　　　　'Dutchman or Holland'

---

[23] When conditions cause resyllabification, CCV, CVC, and CCVC syllable shapes are derived (§6.7.1.1, §6.7.2).

6 Morphophonemic processes

Several processes bring about changes in the underlying forms of roots and words in Fordata. These changes occur across syllable, morpheme, and word boundaries. Possibly the most significant process involves the deletion of syllable peaks in the antepenultimate syllable and root or word-medial and final positions and resulting resyllabification.

6.1 Subject-marking prefix on the verb

All verbs (transitive, intransitive, stative or active/non-active, etc.) are obligatorily marked for subject. The general rule is that verb roots with simple labial consonant onsets[24] take markers from **Set 1**, simple non-labial consonant onsets take **Set 2**, vowel-initial (V) roots take **Set 3**, and roots with complex onsets (CC) take **Set 4**. **Set 4** prefixes are the underlying forms.[25] The following table displays each subject-person marker in each set:

Verb root begins with	1s	2s	3s	1pi	1pe	2p	3p
Set 1 C [+lab]	u-	-w-	n-	t-	am-	-y-	r-
Set 2 C [-lab]	u-	m-	n-	t-	am-	m-y-	r-
Set 3 V	$\emptyset$[26]	m-	n-	t-	am-	my-	r-
Set 4 CC	u-	mu-	na-	ta-	ama-	mi-	ra-
Abstract form	u-	mu-	na-	ta-	ama-	mi	ra-
Historical source	*(ak)u	*(ka)mu	*(i)a	*(ki)ta	*(k)ami	*(ki)miu	*(si)Da

Figure 8: Subject-person marking prefix paradigms

6.1.1 Paradigm rules

I propose the following rules to enhance the discussion of the four types of subject-marking prefixes. **Rules 1** through **7** are crucially ordered and apply disjunctively:

Rule 1 states that *u* is deleted preceding a root-initial vowel, which applies only to **Set 3**, for example:

       Rule 1. $u \rightarrow \emptyset /$___+V

(68)	mu-ahu	*mahu*	'2s-go.(in a direction)'
	mu-etan	*metan*	'2s-cut.down.s.t.'

---

[24] **Sets 1** and **2** could be combined into one set. In separating them, however, the important distinction between labial and non-labial consonant onsets and how they affect surface forms can more easily be seen.

[25] This set of subject marking prefixes is adapted from Coward and Coward's analysis (in this volume) of Selaru, which is structurally similar to Fordata.

[26] In Drabbe's (1932) material and in personal communication with a few of the very old (80 years plus) Fordata speakers, there was a 1s prefix (*u-*) marking all verbs, no matter what the initial CV pattern was; *uail 'I dive' was an acceptable form at that time.

Rule 2 applies only to **Set 2** and states that *u* is deleted when following a non-syllabic phoneme and preceding a non-syllabic non-labial phoneme simple onset:

Rule 2. $u \rightarrow \emptyset$ / [-syll] ___ [-syll] [+syll]
                                              [-lab]

(69)   mu-rafat   *mrafat*   '2s-put.thatch.on.s.t.'
        mu-suŋu   *msuŋu*   '2s-puncture.s.t.'

Rule 3 applies to **Sets 1, 2,** and **4** and states that high vowels become non-syllabic following a non-syllabic phoneme and preceding a root-initial vowel or an optional non-syllabic non-palatal onset, i.e., everything except /y/:

Rule 3. [+syll] → [-syll] /[-syll] ___ +   ([-syll])[+syll]
            [+hi]   [-pal]

(70)   mi-doku   *mydoku   '2p-sit'
        mu-yari   *mwyari   '2s-rest'

Rule 4 states that a semivowel is deleted preceding a semivowel (abbreviated by the character S). **Sets 1** and **2** are affected by this rule, which is as follows:

Rule 4. S → $\emptyset$ /___ S

(71)   *mw-yari   *myari*   '2s-rest'
        *my-yatak   *myatak*   '2p-cap.the.top.(usually a thatch roof)

Rule 5 applies to **Sets 1** and **2** and is a metathesis rule, where in a sequence of consonant–semivowel–consonant, the second and third positions metathesise, for example:

Rule 5. C S C → 1 3 2
            1 2 3

(72)   *my-doku   *mdyoku   '2p-sit'
        *mw-fedan   *mfwedan   '2s-kill.s.t.'

Rule 6 is very limited in that it applies only to one stem-initial phoneme. It states that /v/ becomes non-continuant following a labial consonant, and is summarised in the following manner:

Rule 6. v → [-cont] /   C ___
                            [+lab]

(73)   *mw-vahi   *mbwahi   '2s-paddle'
        *my-visal   *mbyisal   '2p-destroy.s.t.'

Rule 7 states that /m/ is deleted preceding a labial consonant that was the second consonant in a complex onset of the initial syllable of a word. **Rule 7** applies only to **Set 1,** and is written as follows:

Rule 7. m → $\emptyset$ / #___ C
                            [+lab]

(74)   *mbwahi   *bwahi*   '2s-paddle'
        *mbyisal   *byisal*   '2p-destroy'

The final rule, **Rule 8**, applies to **Sets 1, 2,** and **4** and states that /a/ is deleted in the following contexts: (1), following a non-syllabic onset and preceding an utterance-initial non-syllabic phoneme with a different point of articulation, e.g. *non-geminate* consonants and semivowels; (2), preceding a vowel-initial root.

Rule 8. $a \rightarrow \varnothing$ / ([-syll]) ___ ([-syll]) [+syll]

$\phantom{\textbf{Rule 8.} \ a \rightarrow \varnothing \ / \ ([-syll])}$ [α pl] [ß pl]

(75)	na-data	*ndata*	'3s-come'
	ta-otu	*totu*	'1pi-do'

Examples of words from each of the four sets are given with all of the subject-person prefixes represented. All eight rules will be tested on word derivations to see which ones apply. These examples follow in the order of the prefixes in Figure 8.

6.1.2 Root-initial labial consonant

In this section, three words are chosen. The verb *falak* 'say' shows how the labials /f/, /m/, and /b/ in a root-initial position react to the subject-person marking prefixes. The verb *vara* 'carry' illustrates **Rule 6** and *welat* 'stop by' illustrates **Rule 4**.

(76)　*falak* 'say'

Rule	1s	2s	3s	1pi	1pe	2p	3p
	u-falak	mu-falak	na-falak	ta-falak	ama-falak	mi-falak	ra-falak
1							
2							
3		mwfalak				myfalak	
4							
5		mfwalak				mfyalak	
6							
7		fwalak				fyalak	
8			nfalak	tfalak	amfalak		rfalak
SF	*ufalak*	*fwalak*	*nfalak*	*tfalak*	*amfalak*	*fyalak*	*rfalak*

In example (74) above, the first rule to apply is **Rule 3**, where high vowels lose their syllabicity. The 2s and 2p paradigms undergo metathesis in **Rule 5**, then the /m/ is deleted preceding labials word-initially in **Rule 7**. Finally, in **Rule 8**, /a/ is deleted before roots with a simple onset.

(77)　*vara* 'carry'

Rule	1s	2s	3s	1pi	1pe	2p	3p
	u-vara	mu-vara	na-vara	ta-vara	ama-vara	mi-vara	ra-vara
1							
2							
3		mwvara				myvara	
4							
5		mvwara				mvyara	
6		mbwara				mbyara	
7		bwara				byara	
8			nvara	tvara	amvara		rvara
SF	*uvara*	*bwara*	*nvara*	*tvara*	*amvara*	*byara*	*rvara*

(78)   *welat* 'stop by'

Rule	1s	2s	3s	1pi	1pe	2p	3p
	u-welat	mu-welat	na-welat	ta-welat	ama-welat	mi-welat	ra-welat
1							
2							
3		mwwelat				mywelat	
4		mwelat				myelat	
5							
6							
7							
8			nwelat	twelat	amwelat		rwelat
SF	*uwelat*	*mwelat*	*nwelat*	*twelat*	*amwelat*	*myelat*	*rwelat*

6.1.3 Root-initial non-labial consonants

There are also three examples in this section. The verb *doku* 'sit' illustrates the basic non-labial consonant, *toruŋ* 'agree' illustrates the failure of **Rule 8** due to the presence of geminate consonants, and *yatak* 'cover' shows the unusual morphology of /y/-initial roots:

(79)   *doku* 'sit'

Rule	1s	2s	3s	1pi	1pe	2p	3p
	u-doku	mu-doku	na-doku	ta-doku	ama-doku	mi-doku	ra-doku
1							
2		mdoku					
3						mydoku	
4							
5						mdyoku	
6							
7						dyoku	
8			ndoku	tdoku	amdoku		rdoku
SF	*udoku*	*mdoku*	*ndoku*	*tdoku*	*amdoku*	*dyoku*	*rdoku*

(80)   *toruŋ* 'agree'

Rule	1s	2s	3s	1pi	1pe	2p	3p
	u-toruŋ	mu-toruŋ	na-toruŋ	ta-toruŋ	ama-toruŋ	mi-toruŋ	ra-toruŋ
1							
2		mtoruŋ					
3						mytoruŋ	
4							
5						mtyoruŋ	
6							
7							
8			ntoruŋ		amtoruŋ	mtyoruŋ	rtoruŋ
SF	*utoruŋ*	*mtoruŋ*	*ntoruŋ*	*tatoruŋ*	*amtoruŋ*	*mtyoruŋ*	*rtoruŋ*

In example (54), the /a/ deletion (**Rule 8**) is blocked by the presence of identical consonants on either side in 1pi (geminates are not allowed). Metathesis takes place with the 2p prefix, but the /m/ is not deleted, since the following consonant is not labial.

(81)   *yatak* 'cover'

Rule	1s	2s	3s	1pi	1pe	2p	3p
	u-yatak	mu-yatak	na-yatak	ta-yatak	ama-yatak	mi-yatak	ra-yatak
1							
2		myatak					
3							
4							
5							
6							
7							
8			nyatak	tyatak	amyatak		ryatak
SF	*uyatak*	*myatak*	*nyatak*	*tyatak*	*amyatak*	*miyatak*	*ryatak*

The 2p prefix in example (79) does not undergo **Rule 3**. The phoneme /y/ is the only non-syllabic not to undergo this rule. If *i* became non-syllabic and **Rule 4** deleted one of the semivowels, the *m* would retain syllabic qualities, thus causing a pronunciation like the 2s form, which is incorrect for 2p.

6.1.4 Complex onsets in verb roots

Only one example is needed to illustrate prefixing with complex root onsets, since none of the rules apply, for example:

(82)   *slavat* 'trap'

Rule	1s	2s	3s	1pi	1pe	2p	3p
	uslavat	muslavat	naslavat	taslavat	amaslavat	mislavat	raslavat
1							
2							
3							
4							
5							
6							
7							
8							
SF	*uslavat*	*muslavat*	*naslavat*	*taslavat*	*amaslavat*	*mislavat*	*raslavat*

6.1.5 Vowel-initial roots

The **Set 4** paradigms in Figure 8 are illustrated in this section. **Rules 1, 2, 3**, and **8** apply to vowel-initial roots, and VVV sequences are not allowed.

(83)   *ail* 'dive'

Rule	1s	2s	3s	1pi	1pe	2p	3p
	u-ail	mu-ail	na-ail	ta-ail	ama-ail	mi-ail	ra-ail
1	ail						
2		mail					
3						myail	
4							
5							
6							
7							
8			nail	tail	amail		rail
SF	*ail*	*mail*	*nail*	*tail*	*amail*	*myail*	*rail*

6.1.6 Variant verbs

There are some exceptions, which comprise approximately one percent of the lexicon, to the above rules. They are prefixed with the form that is associated with a complex onset. There do not seem to be any phonological processes affecting these forms. The underlying forms, along with the expected and actual ones are shown in the examples below. These particular roots behave the same with any of the person-number prefixes.

Underlying	Expected	Actual	Gloss
na+fiɲin	*nfiɲin	*nafiɲin*	'he plays'
mi+rata	*mryata	*mirata*	'you all go'
ra+terak	*rterak	*raterak*	'they choke'
na+rawan	*nrawan	*narawan*	'it rusts'
ta+navut	*tnavut	*tanavut*	'we are drunk'

Figure 9: Irregular verbs

Further investigation is necessary to adequately account for these type of forms.

6.2 Consonant insertion

At the levels of phonological phrase and phonological word, the insertion of a phonetic transitional epenthetic voiced stop with the same articulation point as the preceding nasal occurs when the nasal is followed by /r/. The phone [d] is inserted between /n/ and /r/, while [b] is inserted between /m/ and /r/ and then [g] between /ŋ/ and /r/. On the phrase level, the voiced stop is inserted between phonological words. On the word level, the insertion occurs between the prefix and the root.

	'already big'	'she gives birth'	'you hear'	'tall weeds'
Underlying Form	*dawan+roak*	*na+rali*	*mu+renar*	*ŋrova*
C-insertion	[dawan droak]	[ndrali]	[ṃbrenar]	[ŋgrova]

Figure 10: Transitional epenthetic consonant

The following rule accounts for the consonant insertion:

$$\emptyset \rightarrow C \quad / \quad C \quad \underline{\quad} r$$

[+voice]	[+nas]
[-cont]	[αant]
[-son]	[ßcor]
[αant]	
[ßcor]	

6.3 [a] Epenthesis

Consonant collocational restrictions described in §3.1 prevent a phonemic three consonant (CCC) sequence or geminate consonants from occurring, either within the morpheme or inter-morphemic or between separate words. The insertion phonetically of an epenthetic [a] provides a transition between these restricted CV shapes, for example:

(84)   *u-viiŋ*   *slavat*   *ini*   → [uvi: ŋa slavat ini]
         1s-pull.in fish.trap this
         'I'm pulling in this fish trap'

(85)   *vulan na-yeba*   → [vuləna nyebə]
         moon 3s-shine
         'the moon is shining'

(86)   *ia na-siʔik kabal*   → [iən siʔika kabal]
         3s 3s-see   ship
         'he sees a ship'

6.4 Infixing

A distributional restriction on the complex consonant onsets involves a nominalising infix /-Vn-/ which occurs only following the phonemes /t/, /s/, and /v/. This infix always derives a trisyllabic word, such as *soba* → *sanoba* (87). The antepenultimate vowel loss (§4.2, §6.7.1.1) accounts for the forms in the examples below.

	Verb		**Noun**	
(87)	*soba*	'honor'	*snoba*	'ceremony'
	surak	'advise'	*snurak*	'advice'
	sibi	'plug up'	*snibi*	'plug, cork'
(88)	*vuat*	'load'	*vnuat*	'load, cargo'
	vilu	'envelop'	*vnilun*	'presence of death'
(89)	*tebar*	'sink'	*tnebar*	'Tanimbar'
	tabar	'stomp'	*tnabar*	'dance' (lit. 'the stomping')

6.5 Root-final /n/ becoming [+back]

Root-final apical /n/ in Fordata becomes dorsal when the root is prefixed with the causative marker *fa-*.[27]

(90)   Root   With causitive prefix

dawan	'big'	*fa-dawaŋ*	'glorify one's self'
lolin	'good'	*fa-loliŋ*	'correct s.t.'
maraan	'light'	*fa-maraaŋ*	'lighten s.t.'
teman	'try'	*fa-temaŋ*	'taste s.t.'
tinemun	'whole'	*fa-tnemuŋ*	'make whole'

6.6 The phoneme /t/ becoming [+nas]

In a limited number of examples, verb root-initial /t/ becomes [+nas] when certain derivational prefixes are applied. The phoneme /t/ has been observed to take on nasal qualities, changing from a voiceless apical stop to an apical nasal, when prefixed with the causative *fa-*, with the detransitiviser *f-* and with the stative *s-*.

Root		**With derivational prefixes**	
na-tuba	'he sleeps'	*nfanuba*	'he puts s.o. to sleep'
na-tuful	'he removes s.t.'	*nafnuful*	'it is removed'
na-tobur	'he jumps'	*nasnobur*	'he jumps into s.t.'
na-tabar	'he stomps s.t.'	*nasnabar*	'he tramples s.t.'

Figure 11: /t/ → /n/

Other *t*-initial verb roots are also prefixed as above, but with no change to the /t/. [28]

(91)   **Stem**   **With derivational prefix**

na-teman	'he tries s.t.'	*nfatemaŋ*	'he tastes s.t.'
na-tofi	'he washes s.t.'	*naftofi*	'he washes'

6.7 Resyllabification

The repackaging of the shape of a word is directly influenced by the deletion of syllable peaks (vowels). Vowels are deleted in the antepenultimate syllable of a word and in medial and final positions. This vowel loss brings about a resyllabification of the word(s). Cliticisation is also discussed.

---

[27] The prefix *fa-* is from PAn *pa- 'causative'.

[28] One plausible explanation for this is that the nasal *n* results from the coalescence of the object prefix *na-* '3s' with the root initial *t*. This hypothesis, however, has a weakness in that it proposes that agents with these verbs can only be third person singular (*na*). In the following sentence, *Ira rfanuba kasikoʔu ra.* 'They put the children to sleep.' the agent is obviously plural, but *n* still occurs.

6.7.1 Deletion of syllable peaks

6.7.1.1 Antepenultimate vowel deletion

Phonetically, many examples of consonant clusters are found root-initially in the same syllable (§5.2.1). There is a possibility that in Fordata this probably results from a historical loss of antepenultimate vowels. This loss of antepenultimate vowels is prevalent in eastern Indonesia (Blust 1990; C. Grimes 1991). Thus historical CVCVCV(C) is reduced to CCVCV(C).

(92)	**PAn/PMP**	**Fordata**	**Gloss**
	*kuRita <	*krita*	'octopus'
	*baRanay <	*brana*	'male, courageous'
	*tabuRi <	*tfuri*	'conch, triton shell'
	*SabaRat <	*varat*	'west'
	*qasawa <	*awa-n*	'spouse-3sGEN'
	*qaniCu <	*nitu*	'ghost, ancestral spirit'

Where consonants are normally retained, *S and *q are lost.

A second environment that could produce initial CC clusters is with a trace of a historical prefix:[29]

$$*CV\text{-}CVCV(C) > CCVCV(C)$$

Words such as those in the following examples could have been reduced in this manner.

(93)	*mŋaʔun*	'true'
	kmia	'urine'
	fraa	'iron wood'
	kmaat	'white (hair)'
	ŋraa	'dirty'
	slaru	'corn'
	blawat	'long'
	tmaʔan	'metal'

It is interesting to note that the apical consonants *d*, *l*, *n* and *r* and the semivowels *w* and *y* do not occur at the beginning of a root in a CC cluster. This leaves only *v* as the exception to the above examples, in that each word in (92) could have been reduced from one of the historical prefixes.

(94)	*vluri*	'group of ten (fish or pigs)'
	vreki	'large iguana lizard'

Subject-marking prefixes (§6.1) and derivational prefixes clearly show the antepenultimate vowel deletion.[30]

---

[29] The current set of prefixes include: *fa-* 'CAUS, NOM, experiencer'; *ba-* 'STATE'; *ka-* 'STATE (in process)'; *-an-* 'NOM'; *si-* 'RCP, INTRANS'; *sa-* 'INTRANS (active)'; *ta-* 'STATE (accidental or unspecified causer); *ma-* 'STATE (progressed or finished)'; *ŋa-* 'STATE (progressive, on going)'.

[30] One notable exception to this antepenultimate vowel deletion is with the causative prefix *fa-*. Probably in order to distinguish the causative from the intransitive experiencer *f-*, the full form of *fa-* is retained even in antepenultimate positions.

(95)   na-etal   *netal*   '3s-cross s.t.'
        na-surak   *nsurak*   '3s-advise'
        ta-watil   *twatil*   '1pi-jump'

(96)   ama-ba-lafar   *amablafar*   '1pe-STATE-hunger'
        na-ka-idat   *nakidat*   '3s-STATE-crack'
        na-ta-vadil   *natvadil*   '3s-STATE-open'

A secondary pre-antepenultimate vowel deletion can also occur when several prefixes are used.

(97)   ama-si-fa-bobar   *amsifabobar*   '1pe-RCP-CAUS-fear'
        na-ka-ma-saat[31]   *nkamsaat*   '3s-PROCESS-STATE-tear'

6.7.1.2 Word-medial and final vowel deletion

Vowel or syllable peak deletion occurs preceding the syllable that takes word, phrase or clause stress. Vowel deletion takes place after any consonant except *ʔ*, *h*, *w*, and *y* in rapid speech. The following word may begin with a vowel or consonant.

(98)   x   x
        x x   x   x x
        waa-n   'place-3sGEN   *w a n l o l i n*   'heaven'
        lo lin   'good'

(99)   x
        x x   x
        vavu   'pig'   x   x x
        naŋan   'jungle'   *vav naŋan*   'wild pig'

(100)   *tamata*   'person'
        isaa   'one'
        tamat isaa 'one person'

(101)   *na-tuba*   '3s-lie.down'
        lufa   'forget'
        ntub lufa   'he sleeps (deeply)'

Words ending with two vowel segments may also drop the final vowel.

(102)   *U-rea yaha.*   →   [ure yaha]
        1s-see dog
        'I see the dog.'

The following example illustrates the exceptional instances where word-final vowel elision does not occur:

---

na-fa-dawan [nfadawaŋ]   '3s-CAUS-big'
na-f-renar   [nafrenar]   '3s-INTRANS-hear'

[31] When *ka* 'STATE (in process)' and *ma* 'STATE (progressed)' are joined, they derive verb meaning something done purposely.

(103)   *U- tunu   yaha   isaa.*   →   [utun yaha isaa]
        1s-shoot   dog   one/a
        'I shot a dog.'

6.7.1.3 Compounding

Compounding in Fordata is a very productive way of combining lexical roots. Compounds are characterised by tight semantic and grammatical cohesion. Compounds are also semantically recognisable as the sum of its parts and the members of the compound can be productively replaced by other forms.

Closely related to compounding is what Grimes (1991:72) calls lexicalisation. When a combination of lexical roots are frozen in form and semantics, the resulting form is often only vaguely related to the roots by cultural association. The parts also will not normally be productively replaced by other forms. Several of the morphemes in the lexicalised compounds are no longer used.

Both compounding and lexicalisation involve the deletion of antepenultimate syllable peaks where permissable.

(104)   **Nominal compounding**
        a.duan   (aa duan)   'wood.master = tree worm'
        a.etal   (aa etal)   'wood.piece = branch'
        afa.m.tahan   (afa ma taha-n)   'thing.to.carry-n = food'
        af.vunun   (afa vunun)   'thing.reef = shellfish'
        surat.ralan   (surat rala-n)   'letter.inside = Bible'

(105)   **Verbal compounding**
        ban.yaha   (bana yaha)   'go.dog = go pig hunting'
        ban.oraŋ   (bana oraŋ)   'go.follow = go following'
        tub.lufa   (tuba lufa)   'lay.forget = sleep'
        keak.teri   (keak.teri)   'tie.hold = tie s.t. up'

(106)   **Lexicalised forms**
        sera.waʔ   'sago.? = marlin species'
        daŋ.walu   '?.eight = eel species'
        sakramat   'pumice stone'
        sul.kaan [32]   '?.stem = praying mantis'

6.7.2 Resyllabification

In sections §6.7.1.1 and §6.7.1.2 above, the various processes that cause the loss of syllable peaks or vowels are illustrated. The following are examples of resyllabification.

---

[32] Several compounds begin with *sul-* and have the idea of flying, e.g. *sulaar* 'sailfish', *sulfaan* 'manta ray', and *sulbaar* 'small bird species'.

(107)   Antepenultimate vowel loss within a word.
CV.CV.CVC#   →   CCV.CVC

(108)   Loss of syllable peak (word-medially) before a stressed syllable across word boundaries.
CVC.CV.VC#CV.CV#   →   CVC.CVC.CV.CV

(109)   Loss of syllable peak (word-finally) before a stressed syllable across word boundaries.
CV.CV#CV.CVC#   →   CVC#CV.CVC

(110)   Loss of syllable peak (word-finally) before a stressed syllable across word boundaries and in the antepenultimate position within the word.
CV.CV.CV#CV.CV   →   CCVC#CV.CV

The restructuring of syllables, as illustrated above, can be explained using traditional notions of discrete boundaries between syllables. This, however, would require complicated rules and much discussion. By defining the three rules of association and syllabification (§6.2), the loss of syllable peaks and the resulting association of the remaining consonants with other peaks is clearly and concisely explained. Simplicity and economy are maintained by defining certain segments as being ambisyllabic. Thus, following the loss of a syllable peak, remaining lines of association will determine the resulting syllabification.

(111)   No ambisyllabic segments

```
    *σ σ                *σ
   / | | \             / | \
  (C) V V (C)         (C) V (C)
   |  | |  |           |  |  |
      a a          a.    d a w a n      'large tree'
   v  a i          v a.  d i d a        'our language'
   f  a a          f a.  k u k u        'young mango'
   d  i i          d i.  m a t a n      'anus'
   r  u u n        r u  n. d a w a n    'big sea cow'
```

(112)   Ambisyllabic segments

```
    *σ      σ           *σ
   / | V  |            / | \
  (C) V (C) V         (C) V  C
   |  | |  |           |  |  |
      a f  a            a f. v u n u n   'shellfish'
   n  i t  u         n i t. r a t a n    'type of spirit'
   v  u t  u         v u t. r u a        'twenty'
   v  a v  u         v a v. y a n a t    'piglet'
```

In prefixed roots, syllables are resyllabified in the following manner:

```
(113)        σ    *σ   σ              *σ   σ
           / | V | V | \          / / | V | \
           C V-C V C V (C)        C C V C V (C)
           | | | | | | |          | | | | | |
           m u- k  a n  a k       m- k  a n  a  k      '2s-joke with s.o.'
           n a- l  a b  i r       n- l  a b  i  r      '3s-deceive s.o.'
           t a- l  e t  a         t-  l  e t  a        '1pi-cross over water'
           r a- m  e l  a         r-  m  e l  a        '3p-grow'
```

With the loss of the antepenultimate syllable peak in a word and word-final vowel, the skeletal tier is rearranged as follows:

```
(114)   σ   *σ   σ   *σ   σ              *σ      *σ    σ
      / | V | V | / | V | \          / / | \  / | V | \
      C V-C V C V #C V C V (C)        C C V C #C V C V (C)
      | | | | | | | | | | |          | | | | | | | | |
      n a- t u n u #f e  d a n        n- t  u n #f e  d a n
      '3s-shoot kill'                 'He shot and killed s.t.'
```

Repackaging of the skeletal tier also occurs on the grammatical phrase and clause level. In the three examples (115–117) below, the subject-person markers *mu-* '2s', *ra-* '3p', and *na-* '3s' are seen in the underlying forms as prefixed on the verb root, and phonetically as resyllabifying with the preceding word.

(115) V.V CV   V.VC

O a mu-doku naʔa ini.   [ʔo amdoku naʔa ʔini]

2s 2s-sit at   here

'You sit here.'

(116) V. CV CV   VC VC

I ra ra-tabar.   [ʔɪ r artabar]

3p   3p-stomp

'They are dancing (lit. stomping).'

(117)   CV CV CVCVC   CVC

Rala -n   lalau ma   na-falak...   [ralan lalau man falak..]

inside-3sGEN sad   and 3s-say

'His heart (lit. insides) was sad and he said...'

In example (118) below, the *m* in *mnelat* is phonemically syllabified as the third syllable when the word is in isolation, for example *m.ne.lat*, but phonetically it resyllabifies with the preceding noun head, such as *vata* 'woman, female'.

(118) CVCV   CCVCVC   CVCVC CVCVC

       v a t a   m n e l a t   [v a t a m n e l a t]

       girl   adolescent

       'adolescent girl'

6.7.3 *Cliticisation*

In this final discussion of this section, I will look into the process of cliticisation. This process may be a factor in vowel deletion. I need time with a native Fordata speaker to gather more data to substantiate cliticisation in Fordata. In his description of Buru stress, Grimes (1991:52) states:

> Cliticisation of lexical roots causes loss of stress from lexical roots that become cliticised to other lexical roots. By this I mean the normally stressed syllable no longer carries greater force, higher pitch or slight lengthening of the vowel.

Along with the loss of stress from lexical roots, cliticisation is further defined (for the Buru language) as the loss of the final vowel of a root as well as signalling tighter semantic cohesion with the root(s) to which it cliticises in the phrase (Grimes 1991:69). Another way of describing this cliticisation process is to say that cliticised roots are restrictive, i.e. descriptive or independent. Uncliticised roots are non-restrictive, i.e. semantically bleached or compounds.

The following examples illustrate the loss of stress of a lexical root and the loss of a syllable peak when it becomes restrictive:

(119)   UNRESTRICTED

```
            x
        x       x
    (C) V (C) V (C)
     |   |  |  |  |
     w   a     a  n        'place-3sGen'
     l   o     l  i  n     'good'
```

(120)   RESTRICTED

```
                x
        x           x   x
    w a n       l o l i   n        'heaven, good place'
```

(121)   UNRESTRICTED

```
            x
        x       x
    k a l b u       u r        'fly'
            n i t u             'corpse, demon'
```

RESTRICTED

```
                    x
        x       x       x
    k a l b u r n i t u        'small fly species'
```

(122)   UNRESTRICTED

```
        x
        x   x
    v a v u                'pig'
    n a ŋ a n              'forest'
```

[RESTRICTED]

```
        x
x       x    x
v  a  v  n  a  ŋ  a  n        'feral pig'
```

The aspect of cliticisation that needs to be further researched is the semantic side. I have found only a few examples of vowel loss resulting in a semantically bleached word or phrase. This needs more research to determine it's significance, if any, in Fordata.

(123)   [UNRESTRICTED]

```
        x            x
        x    x       x    x
v  a  t  u      r  a  l  a  n        'inside of a rock'
```

[RESTRICTED]

```
                x
x       x    x
v  a  t  r  a  l  a  n              'rocky soil'
```

(124)   [UNRESTRICTED]

```
            x         x
        x    x    x    x    x
k  a  r  e  d  a      r  a  l  a  n  'inside of a church'
```

[RESTRICTED]

```
            x
x       x    x
k  a  r  e  d  r  a  l  a  n        'week'
```

7 Reduplication

7.1 Phonological forms

The Fordata reduplication process is mainly a simple syllabic reduplication, where the stressed syllable and the onset of the unstressed syllable are reduplicated, with the exception of full reduplication, where the stem is copied intact. I have found that only disyllabic stems undergo the reduplication process. Furthermore, the stressed syllable in these disyllabic stems can have only V or CV shapes. Only disyllabic stems are involved in reduplication. The reduplication patterns have been observed to follow the same patterns in both the Fordata and Sera dialects. I will analyse reduplication in Fordata using an autosegmental approach. Marantz (1982) proposes that stem may be affixed (in Fordata, prefixed) with the appropriate C–V skeletal template, the entire phonemic melody is copied over the C–V skeleton, and then associated to the eligible C and V slots available in the template. Marantz describes this kind of association as "phoneme-driven", meaning that the phonemes of the stem take priority over the C–V template. Thus, it may be the case that not all of the C–V template is manifested in a given instance. The alternative to a phoneme-driven process can be characterised as "template driven", according to which priority is given to the C–V template so that not all of the phonemes of even the first part of the stem

are necessarily manifested.[33] The following example could be either phoneme driven or template driven.

Affixation	**Melody Copying**	**Association**
f o l a t	f o l a t + f o l a t	f o l a t + f o l a t
| | | | |	| | | | |	| | | | | | | |= falfolat 'door'
CVC+CVCVC	CVC+CVCVC	CVC CVCVC
|	|	|
a	a	a
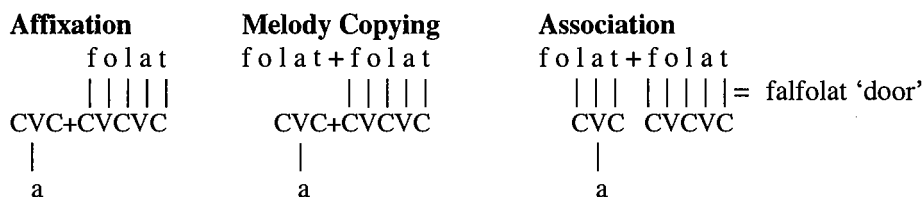
Figure 12: Affixation, copying and association

The direction of matching phonemes to segmental slots in the above diagram proceeds from left to right. The initial affixation process of the syllabic reduplication identifies CVC as the reduplicative template, with V being pre-associated (specified) with the vowel *a*. Then the melody copying step positions the template *CaC* to select features from a copy of the root. In the final step, the *f* associates with the initial C of the reduplicative CVC template. The *o* associates with the V, but by convention is overridden by the pre-associated *a*. The *l*, finally, associates with the second C in the reduplicative CVC template and the reduplicative prefix *fal* is derived.

As shown above in §6.7.2, the medial C in multisyllabic words is ambisyllabic. This furthers the notion of reduplication in Fordata as being syllable reduplication.

The following Figure of a vowel-initial root being reduplicated clearly shows a phoneme-driven copying process:

Affixation	**Melody Copying**	**Association**
u m a t	u m a t + u m a t	u m a t + u m a t
| | | |	| | | |	| | | | | = *amumat*
CVC+VCVC	CVC VCVC	CVC VCVC 'cooked salted fish'
|	|	|
a	a	a
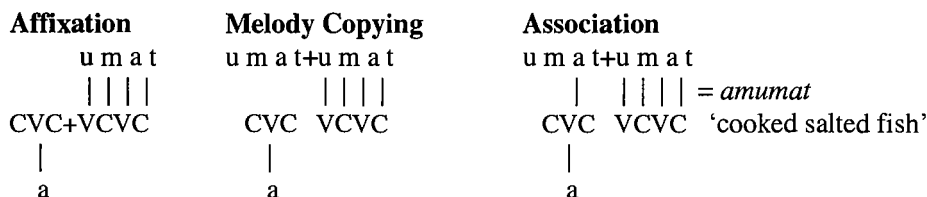
Figure 13: Phoneme–driven process

Marantz (1982:447) states:

> The association of phonemic melodies and C–V reduplicating affixes is "phoneme-driven" in the sense that, for each phoneme encountered linking from left to right or from right to left, the association procedure scans along the skeleton to find a C–V slot eligible for association with the phoneme.

As shown in Figure 12, the affixation process identifies CVC as the reduplicative template. The melody-copying step then positions the template *CaC* to select features from the copied root. In the final step, the association scans the CV template for eligible slots. The first eligible position for the word-initial *u* is the V in the CV template (the pre-associated *a* takes priority over whatever vowel from the phoneme melody that was associated, in this case *u*). Since the medial C is ambisyllabic, it associates with the final C position of the reduplicative template. Any unassociated phonemes or CV positions are then discarded.

---

[33] For examples of template-driven reduplication in Tagalog, see French (1988).

If the process associating the CV skeleton to the phonemes were a template-driven process, the following example would occur:

Affixation	**Melody Copying**	**Association**
umat	umat+umat	umat+umat
CVC+ VCVC	CVC VCVC	CVC VCVC = *matumat
a	a	a
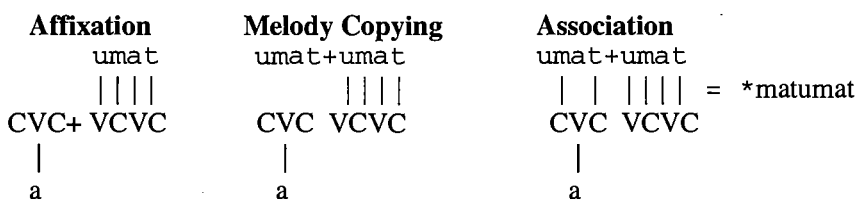
Figure 14: Template-driven process

The template drives the association, therefore the surface form would be **matumat*, because the initial C of the template would select the *m* of the reduplicated stem as the starting point of the association.

Since it is the stressed syllable that is reduplicated, the onset of the reduplicated *CaC* morpheme in Fordata will always match the onset of the stressed syllable, as seen in Figure 12, the example of *fal-folat* 'door'. If the stressed syllable has no prenuclear margin, the reduplicative prefix will be an open syllable with the pre-associated *a* as its nucleus, for example *am-umat* 'cooked salted fish' from Figure 13 above. The coda of the reduplicative morpheme is always the same as the reduplicated syllable, e.g., the *l* in *folat* becomes the coda of the reduplicated morpheme. There are surface constraints which apply after the reduplication process. As mentioned previously in §3.1, the consonants *h*, *w* and *ʔ* cannot occur in the first slot of a cluster, and geminate clusters are not allowed.

Consider the following basic forms with their reduplicative counterparts:

dudur	'go down hill'
lahur	'to be wild mannered'
lawan	'to roll s.t.'
koʔu	'small'

	Affixation	**Melody Copying**	**Association**
(125)	dudur	dudur+dudur	dudur+dudur
	CVC+CVCVC	CVC CVCVC	CVC CVCVC = *daddudur
	a	a	a
(126)	lahur	lahur+lahur	lahur+lahur
	CVC+CVCVC	CVC CVCVC	CVC CVCVC = *lahlahur
	a	a	a

	Affixation	**Melody Copying**	**Association**
(127)	lawan	lawan+lawan	lawan+lawan
	CVC+CVCVC	CVC CVCVC	CVC CVCVC = *lawlawan
	a	a	a

(128)   k o ʔ u   k o ʔ u+k o ʔ u   k o ʔ u+k o ʔ u
             | | | | | | | | | | | | | | = *kaʔkoʔu
          CVC+CVCV   CVC   CVCV   CVC   CVCV
             | | |
             a   a   a
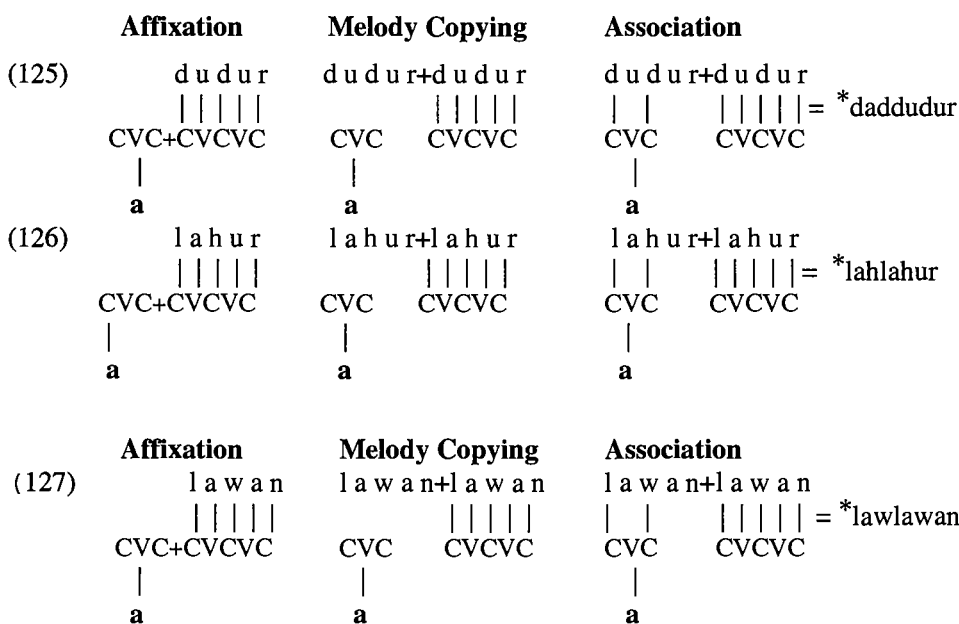
The four examples above all go through the same three-step derivational process as explained for the reduplicated forms in Figures 12 and 13. The reduplicated forms then would be subject to the following two general rules, which are reiterations of the collocational restrictions stated in §3.1:

Rule 1:

Geminate reduction $C \rightarrow \varnothing\ /\ \_\_\_$ C
                    $[\alpha\,\text{pl}]$ $[\alpha\,\text{pl}]$

Rule 2:

Cluster restriction $\begin{Bmatrix} h \\ \text{ʔ} \\ w \end{Bmatrix} \rightarrow \varnothing\ /\_\_\_$ C

Below, the four incorrect forms in examples (122) and (123) undergo **Rules 1** and **2** to arrive at the correct surface forms:

	Reduplication form		**Rule 1**
(129)	*daddudur*	_	da-dudur 'downslope'

			Rule 2
(130)	*lahlahur	_	la-lahur 'crazy'
	*lawlawan	_	la-lawan 'weaving tool'
	*kaʔkoʔu	_	ka-koʔu 'small, less'

Roots with the structures CVCVC and CVCV satisfy the C–V template by reduplicating the first CVC:

CVCVC:

(131)	_fal_-folat	'door'
(132)	_kat_-katan	'scorpion'
(133)	_lalolak_ [34]	'glasses'
(134)	_ma_-mumak	's.t. steamed with coconut milk'
(135)	_mar_-merat	'cleanliness'

CVCV:

(136)	_lab_-leba	'stick to carry s.t. on one's shoulder'

---

[34] The ruduplicated forms are *lallolak* and *mam-mumak* become *lalolak* and *ma-mumak* after application of geminate reduction. Forms such as *lalolak*, which do not have a hyphen between the reduplicated morpheme and the stem are frozen forms.

(137)   *dar-diri*   'pig hunt'

(138)   *wal-walu-n*   'in eights'

(139)   *larlora*   'watermelon'

(140)   *tavteva*   'top shell'

Bases with the patterns CVVC and CVV reduplicate only the CV of the template, which was predicted by McCarthy and Prince (1986:94), who stated that the template will not skip the second V in a CVVC stem to complete the entire CVC skeletal structure by using the second C. This is illustrated below in Figure 15:

Affixation	**Melody Copying**	**Association**
k e a r	k e a r+k e a r	k e a r+k e a r
| | | |	| | | |	| | | | |= *kakear*
CVC+CVVC	CVC +CVVC	CVC CVVC 'hole'
|	|	|
a	a	a
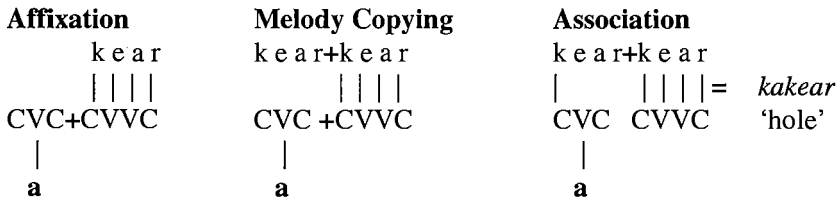
Figure 15: Reduplicating CVVC stems

The phoneme driving the association must be fully satisfied. There is no V for the *a* in *kear* to associate with, so the association stops and the unused phonemes and CV slot are discarded.

CVVC:

(141)   *ka-kear*   'well, hole'   *karkear

(142)   *lalaar*   'peninsula'   *larlaar

(143)   *ba-buuk*   's.t. to smoke' *bakbuuk

(144)   *ŋa-ŋiar*   'white'   *ŋaṃiar

CVV:

(145)   *sa-saa-n*   'in ones'

(146)   *ra-roa*   'far'

(147)   *lalau*   'sad'

(148)   *ŋa-ŋai*   'rubbish after clearing a garden'

Stems having the pattern VCVC associate only with the VC segments of the template:

VCVC:

(149)   *am-umat*   'cooked salted fish'

(150)   *al-aliŋ*   'different'

Full reduplication is found in Fordata, but with limited use. To date, few examples have been found. In full reduplication, the stem is copied as normal and the template will be the C–V pattern of the stem. There are no pre-associated phonemes. Stress also seems to be evenly distributed on both morphemes. Below is an example of the full reduplication process:

Affixation	**Melody Copying**	**Association**
l e b a	l e b a+l e b a	l e b a +l e b a
| | | |	| | | |	| | | | | | | | = *lebaleba*
CVCV+CVCV	CVCV CVCV	CVCV CVCV 'canoe'

Figure 16: Full reduplication

Forms fully reduplicated with the patterns CVCV, CVVC and VCVC have been recorded: [35]

CVCV:

(151)   *teva-teva*   'top shell'

(152)   *leba-leba*   'canoe'

VCVC:

(153)   *amar-amar*   'every day'

(154)   *ovan-ovan*   'every night'

7.2 Syntactic functions

One of the main syntactic functions of reduplication in Fordata is the nominalization of verbs. The following are derived from transitive and intransitive verbs:

(155)   *fal-folat   dawan*   *na-folat*
       DUP-close   big   'he closes (a door)'
       'big door'

(156)   *Tan-tunu-n   ra   r-   si-   karas.*   *u-tunu*
       DUP-roast   PL   3p   RCP   bite   'I roast (s.t.)'
       'The fire ants bite.'

(157)   *Ia   ni   dak-doku-n   na?a ini.*   *ta-doku*
       3s   3sGEN   DUP-sit-n   at   here   'we sit'
       'Its position is here.'

(158)   *Var-vara   aleman.*   *ra-vara*
       DUP-carry   heavy   'they carry (s.t.)'
       'heavy responsibility'

Another function of reduplication is attributive use of verbs (usually non-active) to function as modifiers in a NP, rather than as the predicate of a verbal clause. The obligatorily marked Predicative verb forms are shown on the right:

---

[35] The usage of some of these forms varies from village to village, for example, *tevateva* 'topshell' is rendered *tavteva* in several villages. *Lebaleba* is an old word meaning 'canoe' and has been for the most part replaced by another term.

	Attributive	**Predicative**
(159)	*wanat val-vuli-n* rice DUP-red 'red rice'	*na-vuli* 'he/it is red'
(160)	*ŋuur ŋa- ŋiar* sand DUP-white 'white sand'	*na-ŋiar* 'he is white'
(161)	*Mata-n ra was-wosu.* eye-3sGEN PL DUP-tired 'His eyes are tired.'	*na-wosu* 'he is sleepy'
(162)	*Afa ovi mak-mukur.* thing these DUP-round 'These things are round.'	*na-mukur* 'he/it is round'
(163)	*afa fan-fonak* thing DUP-hide 'secret thing'	*na-fonak* '3s-hide (s.t.)'

Nouns can also be reduplicated to modify other nouns:

(164)	*ian sar-sira-n* fish DUP-salt 'salted fish'	*sira* 'salt'
(165)	*Mataŋ lav-lova.* eye-1sGEN DUP-fog 'My eye is cloudy.	*lova* 'fog'

When following a verb it functions as a modifier:

(166)	*ia na-kaʔa kad-kedan* 3S 3S-know DUP-little 'He knows a little bit.'	*kedan* 'little, a little'

Cardinal and ordinal numbers are not reduplicated except for special purposes. The function of a reduplicated number to derive a quantity 'by x number' or 'in x numbers':

(167)	*tamata tal-telu-n* people DUP-three 'people in threes (or three by three)'	*i-telu* 'QNT-three'

A distributive function, 'each Noun', which is little used, is signified by full reduplication:

(168)	*vai amar-amar* language DUP-day 'every day language'	*amar* 'day'
(169)	*Ovan-ovan u- dava oa.* DUP- night 1s-look.for 2s 'Every night I looked for you.'	*ovan* 'night'

On a fair amount of reduplicated forms with no post-nucleus coda, a word-final *n* is added. In Indonesian, the prefix and suffix combination *ber- + -an* means 'in X numbers'. With numbers in Fordata, there is a very possible correlation, however both nouns and adjectives are also found with this added *n*. To the right are the cardinal numbers and the verb forms.

Numbers:

(170)   *sa-saa-n*   *isaa*
        DUP-one   'one'
        'by ones'

(171)   *vat-vutu-n*   *vutu*
        DUP-ten   'ten'
        'by tens'

Nouns:

(172)   *ban-bana-n*   *u-bana*
        DUP-go   'I go'
        'journey, path'

(173)   *mat-mata-n*   *ta-mata*
        DUP-dead   'we're dead'
        'death'

(174)   *lak-leka-n*   *na-leka*
        DUP-fall   'he fall'
        'the falling'

Adjectives:

(175)   *val-vuli-n*   *na-vuli*
        DUP-red   'he is red'
        'red or reddish'

(176)   *sar-sira-n*   *na-sira*
        DUP-salt   'he salts something'
        'salty'

8 Loan words

Many loan words from Malay and Ambonese Malay have been 'Fordatanised' to conform with the phonological patterns of Fordata. With some words, points of articulation will remain the same or close, but the voicing will change. Others will see the articulation point move and voicing remain the same.

Dorsals lose voicing. The examples below also show the apical affricate [j] becoming a stop.

(177)   *gereja*   Malay   'church'
        kareda   Fordata   'church'

        gaji   Malay   'salary'
        kadi   Fordata   'salary, payment'

Voiced labial stops become fricative and voiceless labial stops become voiced or become voiceless labial fricatives.

(178) *Balanda*   Ambonese Malay   'Holland, Dutch person'
       Valada   Fordata   'Holland, Dutch person'

(179) *pasiar*   Malay   'stroll'
       basiar   Fordata   'stroll'

       pendeta   Malay   'pastor (Protestant)'
       bandita   Fordata   'pastor (Protestant)'

       perintah   Malay   'order, command'
       fareta   Fordata   'authority, command'

Other loans are reanalysed as follows:

(180) *baca*   Malay   'read'
       basa   Fordata   'read'

(181) *marinyo*   Ambonese Malay   'village crier'
       mariɲun   Fordata   'village crier'

(182) *fikaris*[36]   Ambonese Malay   'female pastor candidate'
       vikaris   Fordata   'female pastor candidate'

---

[36] In standard Indonesian, it is written *vikaris*, but pronounced [fikaris].

Appendix A

**Consonant
Contrasts**
b/f/v/w
b/f

[nabat]	/nabat/	'seed, seedling'
[nafat]	/na-afat/	'he cuts'
[bana]³⁷	/bana/	'go'
[fana]	/fana/	'shoot (an arrow)'

b/v

[bara]	/na-bara/	'it is swollen'
[vara]	/na-vara/	'he carries'
[votu]	/votu/	's.o. returns'
[botun]	/botu-n/	'his nape'
[yabar]	/yabar/	'large bat'
[yavar]	/yavar/	'type of tree'

b/w

[baba]	/baba/	'younger male'
[waba]	/waba/	'season closed for harvest or hunting'
[belat]	/na-belat/	'he fences s.t.'
[welat]	/na-welat/	'he stops off'

f/v

[fa:n]	/faan/	'fish bait'
[va:n]	/vaan/	'cave'
[vra:]	/vraa/	'hermit crab'
[fra:]	/fraa/	'iron wood'
[fahi]	/na-fahi/	'he passes gas'
[vahi]	/na-vahi/	'he paddles'
[tfuri]	/tfuri/	'parrot fish'
[tvuri]	/tvuri/	'helmet shell'

f/w

[fa:n]	/faan/	'bait'
[wa:n]	/waa-n/	'its place'

---

³⁷ Both *bana* and *fana* are historically from *p (*panaw* 'go, leap', *panaq* 'arrow, shoot arrow').

[faha]	/faha/	'buy'
[wahan]	/waha-n/	'his face'

<center>v/w</center>

[ʔavan]	/ava-n/	'his aunt'
[ʔawan]	/awa-n/	'his spouse'

[vərin]	/verin/	'for'
[wərin]	/weri-n/	'his wing, fin'

[va:n]	/vaan/	'cave'
[wa:n]	/waa-n/	'its place'

[vaʔar]	/vaʔar/	'clay'
[waʔar]	/waʔar/	'mangrove, root'

<center>h/ʔ</center>

[yaha]	/yaha/	'dog'
[yaʔa]	/yaʔa/	'I'

[vahi]	/vahi/	'paddle'
[vaʔi]	/vaʔi/	'garden'

[waʔa]	/waʔa/	'tidal crab'
[wahan]	/waha-n/	'his face'

[vaʔan]	/vaʔa-n/	'his kidney'
[vahan]	/vaha-n/	'it's liquid'

<center>k/ʔ</center>

[taku]	/taku/	'hit'
[taʔu]	/taʔu/	'lunar-tail cod'

[boku]	/boku/	'some, part'
[boʔu]	/boʔu/	'female genitals'

<center>t/d/r</center>
<center>t/d</center>

[datar]	/datar/	'bark skirt'
[dadar]	/dadar/	'dizzy'

[tokə]	/toka/	'iguana'
[dokə]	/doka/	'buttock'

		t/r	
[nditi]	/na-diti/		'it drips'
[ndiɾi]	/na-diɾi/		'he stands'
[fafaʔat]	/fa-faʔat/		'in fours'
[fafaʔaɾ]	/fa-faʔaɾ/		'walking'
[nata]	/nata/		'following, next'
[naɾa]	/naɾa/		'star'
[tenən]	/tena-n/		'his body'
[ɾenən]	/ɾena-n/		'his mother'

		d/r	
[nadan]	/nadan/		'catfish'
[naɾan]	/naɾan/		'enough'
[daʔut]	/daʔut/		'rain'
[raʔut]	/raʔut/		'bamboo cage'

		m/n	
[yamam]	/yama-m/		'your (sg.) father'
[yanam]	/yana-m/		'your (sg.) child'
[mahin]	/mahin/		'giant clam'
[nahin]	/nahin/		'knife'
[melə]	/mela/		'nobility'
[nelən]	/nela-n/		'his neck'
[mata]	/mata/		'dead'
[nata]	/nata/		'next'

		l/r	
[vɛlin]	/veli-n/		'it's property'
[vɛrin]	/verin/		'for'
[lolak]	/na-lolak/		'he visits'
[loɾak]	/na-loɾak/		'he cuts'
[nail]	/na-ail/		'he dives'
[naiɾ]	/na-aiɾ/		'he teaches'
[ŋu:l]	/ŋuul/		'coconut oil'
[ŋu:ɾ]	/ŋuuɾ/		'sand, beach'

	b/m	
[binan]	/binan/	'plate'
[minan]	/mina-n/	'it's body grease'
[balit]	/balit/	'on the left'
[malit]	/malit/	'laugh'
[bomə]	/boma/	'to, for'
[momə]	/moma/	'tidal crab'
[nabat]	/nabat/	'seed, seedling'
[namat]	/namat/	'type of sea cucumber'

	n/ŋ	
[nu:r]	/nuur/	'coconut'
[ŋu:r]	/ŋuur/	'sand, beach'
[dedən]	/dedan/	'dark'
[dedəŋ]	/dedaŋ/	'sing'
[ʔanar]	/anar/	'white cockatoo'
[ʔaŋar]	/aŋar/	'coral'
[ʔini]	/ini/	'this'
[ʔiɲi]	/iɲi/	'lingua wood'

	w/y	
[wadu]	/wadu/	'a catch of fish'
[yadu]	/yadu/	'shiver'
[wahan]	/waha-n/	'his face'
[yaha]	/yaha/	'dog'
[watan]	/watan/	'only'
[yatan]	/yata-n/	'his liver'
[waʔa]	/waʔa/	'tidal crab'
[yaʔa]	/yaʔa/	'I'

Vowel Contrasts
i/e

[ʔɪrə]	/ira/	'they'
[ʔɛrə]	/era/	'call'

[riə]	/ria/	'grass'
[reə]	/rea/	'find'
[livur]	/livur/	'village'
[levur]	/levur/	'burning'
[sɪrə]	/sira/	'salt'
[sɛrə]	/sera/	'tapioca'
[ʔimən]	/iman/	'thorn'
[ʔemən]	/eman/	'loin cloth'
[ni]	/ni/	'his'
[ne]	/ne/	'that'

u/o

[lurak]	/na-lurak/	'he cuts'
[lorak]	/na-lorak/	'he lets out'
[sukə]	/na-suka/	'he lifts'
[sokə]	/na-soka/	'he rubs'
[vulən]	/vulan/	'moon'
[volən]	/volan/	'appointed time'
[vutuk]	/na-vutuk/	'he pulls (hair)'
[votuk]	/na-votuk/	'it appears'

Appendix B

Syllable Patterns

S

V	*a*	'PART, singular'
VC	*al*	'to, for'
CV	*ma*	'so that'

SS

V.V	*oa*	'2s'
V.CV	*ahu*	'village'
V.CVC	*ahir*	'basket'
VC.CV	*aksa*	'alone'
CV.V	*tea*	'faeces'
CV.VC	*wear*	'water'
CV.CV	*kesi*	'food'
CV.CVC	*kikur*	'tail'
CCV.CV	*fteni*	'parrotfish'
CCV.CVC	*blaŋar*	'coral trout'

S.SS

V.CV.CV	*itelu*	'three'
V.CCV.CVC	*ablawat*	'emperor fish'
V.CV.CVC	*u-wahak*	'he deceives'
VC.CV.CVC	*u-ɲanaŋ*	'I remember'
VC.CVC.CV	*alnuran*	'horse-eye jack'
CV.CV.CV	*na-kaʔa*	'he knows'
CV.CV.VC	*sulaar*	'sailfish'
CV.CCV.CVC	*na-blafar*	'he is hungry'
CVC.CV.CVC	*ŋatŋatul*	'gecko lizard'

SS.SS

V.CVC.CV.CVC	*u-fa-ɲanaŋ*	'I remind s.o.'
CV.CV.CV.CV	*na-fa-leka*	'he drops s.t.'
CV.CV.CV.CV	*serawaʔi*	'marlin'
CV.CV.CV.CVC	*Nalalahan*	'slow'

SSS.SS

V.CV.CV.CV.CV	*u-fa-malola*	'I straighten'
V.CV.CV.CCV.CVC	*u-fa-makrakat*	'I tickle s.o.'
CV.CV.CV.CV.CV	*mi-si-ŋarahi*	'you all fight'
CV.CV.CV.CV.CVC	*ta-si-fa-bobar*	'we scare each other'

References

Blust, Robert, 1978, Eastern Malayo-Polynesian: A subgrouping argument. *Proceedings of the Second International Conference on Austronesian Linguistics, fasc.* 1, 181–234. Canberra: Pacific Linguistics.

　1990, Central and Central-Eastern Malayo-Polynesian. Paper presented at the Maluku Research Confrerence: Honolulu.

Clements, G.N. and S.J. Keyser, 1983, *CV Phonology*. Cambridge: MIT Press.

Collins, James T., 1983, Linguistic research in Maluku: a report of recent fieldwork. *Oceanic Linguistics* 21 (1–2):73–150. Honolulu: University of Hawaii.

Coward, David and Naomi Coward, 2000, A phonological sketch of the Selaru language. In Charles E. Grimes, ed., *Spices from the east: papers on languages of eastern Indonesia*, 11-56. Canberra:Pacific Linguistics.

Drabbe, P., 1926, *Spraakkunst der Fordaatsche taal*. Batavia, Dutch East Indies: Albrecht and Co.

　1932, *Woordenboek der Fordaatsche taal*. Bandung, Dutch East Indies: A.C. Nix and Co.

French, Koleen M., 1988, *Insights into Tagalog: reduplication, infixation, and stress from nonlinear phonology*. Dallas: Summer Institute of Linguistics and University of Texas at Arlington.

Geurtjens, H., 1928, Over Tanimbareesch taaleigen. Unpublished manuscript.

Goldsmith, John A., 1990, *Autosegmental and metrical phonology*. Oxford: Basil Blackwell.

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

Hughes, Jock, 1987, The languages of Kei, Tanimbar and Aru: a lexicostatistic classification. *NUSA* 27:71–111. Jakarta: Universitas Atma Jaya.

Hughes, Jock and Katy Hughes, 1989, A phonology of Dobel. *Workpapers in Indonesian Languages and Cultures, vol. 7*. Ambon: Universitas Pattimura–Summer Institute of Linguistics.

Marantz, A., 1982, Re reduplication. *Linguistic Inquiry* 13 (3): 435–482.

McCarthy, John and Alan Prince, 1986, Prosodic morphology. Unpublished manuscript.

McKinnon, Susan, 1991, *From a shattered sun*. Madison: The University of Wisconsin Press.

Nivens, Richard, 1993, Reduplication in four dialects of west Tarangan. *Oceanic Linguistics* 32 (2): 353–388.

Riedel, Gerard, 1886, De sluik en kroesharige rassen Selebes en Papua. Martinus Nyhoff.

Selkirk, Elisabeth, 1982, The syllable. In Harry van der Hulst and Norval Smith, eds, *The structure of phonological representations,* Part II, 337–383. Dordrecht: Foris.

Travis, Ed, 1988, The Kei language: a phonological description. Unpublished manuscript, The Summer Institute of Linguistics and Pattimura University.

Vatkaat, J., 1984, Analisis fonem segmental bahasa Fordata. Unpublished manuscript, Universitas Pattimura.

  1989, Fungsi bahasa Fordata dan bahasa Melayu Ambon, pada para remaja dan anak-anak di kecamatan Tanimbar Utara, Maluku Tenggara. Unpublished manuscript, Universitas Pattimura.

Wurm, S.A. and S. Hattori, eds, 1981, *Language atlas of the Pacific area.* Canberra: The Australian Academy of the Humanities in collaboration with the Japan Academy.