Traces of language contact: The Flores-Lembata languages in eastern Indonesia

Fricke, H.L.A.

CHAPTER 6

Inheritance and innovation in the lexicon

6.1 Introduction

This chapter is about the lexicon of the Flores-Lembata languages and their ancestor Proto-Flores-Lembata (PFL). I show that PFL has a largely Austronesian vocabulary, as about 80% of my PFL reconstructions have a Proto-Malayo-Polynesian (PMP) source. However, the individual subgroups of Flores-Lembata underwent a considerable addition of vocabulary that cannot be traced back to PMP. The amount of additional vocabulary varies per subgroup. Larger amounts of additional vocabulary is found in Lamaholot, followed by Kedang, and then Sika with the smallest amount of lexical additions. In the Central Lamaholot subgroup, the non-Austronesian component amounts to more than 50%.

I propose that this new vocabulary is a lexical substrate that entered the subgroups due to contact with now extinct non-Austronesian languages. A lexical substrate is a layer of vocabulary from one or more substrate languages that are not spoken any more. Typically, the speakers of the substrate languages have shifted to new languages — here the proto-languages of the Flores-Lembata subgroups — and retained part of the lexicon of their original language. An alternative proposal could be to identify the newly added

vocabulary in the Flores-Lembata subgroups as (large scale) borrowing. In this case, one would not assume that non-Austronesian speakers shifted to the Flores-Lembata languages but that the speakers of the Flores-Lembata languages borrowed large amounts of vocabulary from a non-Austronesian source. There are two arguments that point to substrate rather than large scale borrowing. First, due to the genetic diversity of the population of the region that shows an almost equal Austronesian / non-Austronesian mix (cf. §1.3.2), there must have been a considerable amount of non-Austronesian speakers all over this region in the past. Nowadays all languages spoken in the region, except for the Timor-Alor-Pantar languages, are Austronesian. More likely than the extinction of a large amount of non-Austronesian speakers, which would have caused the genetic signal to be much weaker, is that the speakers of non-Austronesian languages shifted to Austronesian languages. Second, not only the lexicon but also the morpho-syntax of the Flores-Lembata languages shows non-Austronesian features as I will show in Part III of this dissertation. The transfer of morpho-syntactic features is a typical outcome of language shift preceded by a period of bilingualism (Muysken 2010:272). Borrowing grammatical features from an unrelated language is less likely.

This chapter is based on a systematic analysis of 422 lexeme sets collected from the Flores-Lembata wordlists stored in the Lexirumah database. A lexeme set is a set of related forms based on cognacy or borrowing. See §4.3 for more details on the methodology applied for this chapter and §4.4 for conventions in data representation.

Table 6.1 gives an overview of the results of this chapter. The lexeme sets are sorted by reconstructability to PFL according to three categories: (i) 210 sets that can be reconstructed to Proto-Flores-Lembata (PFL), (ii) 185 sets that cannot be reconstructed to PFL but have regular correspondences and (iii) 27 sets with irregular correspondences. Further, the three types of sets are assessed according to the availability of a PMP source. Most of the PFL reconstructions have a PMP origin. The lexeme sets of non-PMP origin are largely regular and with 248 sets in total, they outnumber those of PMP origin with 174 sets in total.

Table 6.1: Lexeme sets analysed for this chapter

	Total	PMP	no PMP source	Section
PFL	**210**	173	37	§6.2
Unreconstructible regular sets	**185**	-	185	§6.3
Unreconstructible irregular sets	**27**	1	26	§6.4
Total	**422**	174 (40%)	248 (60%)	

The lexeme sets that show regular sound correspondences but are classified as unreconstructible may ultimately be reconstructible to PFL but reflexes in some of the subgroups were lost or are missing from current data sources. More comprehensive data from the Flores-Lembata languages will likely show that a number of these lexeme sets are indeed reconstructible to PFL. However, it is also possible that these words are (early) borrowings that entered the Flores-Lembata languages after the break-up into subgroups but before the subgroup-defining sound changes occurred. These sets cannot be identified as late loans as they underwent the expected regular sound changes. The scenario of early borrowings would presuppose geographical separation of the subgroups, then contact resulting in lexical borrowing and only after new vocabulary had entered the subgroups, the regular sound changes occurred.

This chapter is divided into four sections. §6.2 presents the 210 lexeme sets that can be reconstructed to PFL according to the criteria explained in §4.3. §6.3 discussed the 185 unreconstructible lexeme sets which, nevertheless, show regular sound correspondences between subgroups. §6.4 discusses cases of lexeme sets with irregular correspondences. Some of these may be explicable by borrowing. §6.5 summarises the findings of the chapter and draws conclusions.

6.2 Proto-Flores-Lembata reconstructions

6.2.1 Overview

This section presents 210 cognate sets that can be reconstructed to Proto-Flores-Lembata (PFL). Most of these reconstructions (around 80%) have a

PMP source (§6.2.2), while a subset (around 20%) does not match to any known PMP form (§6.2.3).

6.2.2 PFL reconstructions with PMP sources

This section lists and discusses the 173 PFL reconstructions that have a PMP source. Out of these most have reflexes in all Flores-Lembata subgroups, meaning in Sika, Kedang and in at least on of the Lamaholot varieties (n=113). But there are also a few PFL reconstructions that only have reflexes in a subset of the Flores-Lembata subgroups. These are presented in separate tables. The PFL reconstructions in this section reflect the PMP form in a largely regular way. However, this does not exclude some minor irregularities at the level of individual phonemes, such as insertions of an additional phonemes (marked by a vertical line probably reflecting historical affixation), the loss of a single phoneme, sporadic metathesis or sporadic vowel changes.

Table 6.2 lists 113 PFL forms that are of PMP origin and are reflected with largely regular sound correspondences in all Flores-Lembata subgroups.

Table 6.2: PFL reconstructions and their PMP sources (n=113)

PFL	PFL meaning	PMP
*aku	‘1SG’	*i aku
*kami	‘1PL.EXCL’	*kami
*kita	‘1PL.INCL’	*kita
*hida	‘3PL’	*si ida
*tudu	‘accuse’	*tuzuq
*pəniki	‘bat’	*paniki
*vani/*blani	‘bee’	*wani
*manuk	‘bird; chicken’	*manuk
*m-paʔit	‘bitter’	*paqit
*mitəm	‘black’	*ma-qitəm
*puhun	‘blossom; flower’	*pusuŋ ‘heart; heart of banana’
*prupi/plupi	‘blow’	*upi
*vulu-k	‘body hair’	*bulu
*luri	‘bone’	*duRi
*vuhur	‘bow’	*busuR

PFL	PFL meaning	PMP
*(t)usu	'breast'	*susu
*mamaʔ	'chew'	*mamaq
*pipi/*klipi	'cheek'	*pipi
*ana(k)	'child; small'	*anak
*piliʔ	'choose'	*piliq
*hakay	'climb'	*sakay
*mai	'come'	*mai
*vatar	'corn; maize'	*batad 'millet; sorghum'
*lədav	'day; sun'	*qaləjaw 'sun'
*matay	'die'	*m-atay
*gali	'dig'	*kali
*bagi	'divide'	*baqagi
*ahu	'dog'	*asu
*-inu	'drink'	*inum
*mada	'dry; thirsty'	*maja
*pa-vari	'dry in sun'	*waRi
*kVan	'eat'	*kaən
*təlur	'egg'	*qatəluR
*mata	'eye'	*mata
*ama	'father'	*ama
*api	'fire'	*hapuy
*ikan	'fish'	*hikan
*təməla	'flea'	*qatiməla
*vuda	'foam'	*bujəq
*ləpət	'fold'	*lipət
*tuʔan	'forest'	*tuqan
*vua-n	'fruit; betelnut'	*buaq
*m-pənu-k	'full'	*pənuq
*bəli	'give'	*bəRay
*udu	'grass; bush'	*udu
*lima	'hand, arm, five'	*qalima
*kutu	'headlice'	*kutu
*dəŋər	'hear	*dəŋəR
*bərat	'heavy'	*(ma)bəRəqat

PFL	PFL meaning	PMP
*pida	'how many'	*pija
*bə-ləma[a]	'inside; deep'	*daləm
*una	'inside; house'	*qunəj 'pith of plant; core'
*viri	'left side'	*kawiri
*tave	'laugh'	*tawa
*ʔapur	'lime'	*qapur
*vivir	'lips'	*biRbiR 'lower lip'
*isi-k or *ihi-k	'meat'	*isi
*vulan	'moon'	*bulan
*ina	'mother'	*ina
*ili	'mountain'	*qilih
*vava	'mouth'	*baqbaq
*nadan	'name'	*ŋajan
*pusər	'navel'	*pusəj
*vəru	'new'	*baqəRu
*niduŋ/iduŋ	'nose'	*ŋijuŋ/*ijuŋ
*m-tuʔa	'old (people)'	*ma-tuqah
*əha	'one; alone'	*əsa
*uti	'penis'	*qutin
*ata	'person'	*qaRta 'outsider, alien people'
*vavi	'pig'	*babuy
*bayu	'pound'	*bayu
*veli	'price; bride price; expensive; buy'	*bəli
*udan	'rain'	*quzan
*uay	'rattan'	*quay
*vanan	'right side'	*ka-wanan
*m-tasak	'ripe'	*ma-tasak
*lalan	'road'	*zalan
*ramut	'root'	*Ramut
*layar	'sail'	*layaR
*m-pədu	'salty'	*qapəju 'gall' >*ma-pəju
*sama	'same'	*sama
*ənay	'sand'	*qənay
*garu	'scratch'	*garut

PFL	PFL meaning	PMP
*tahik	'sea'	*tasik
*pitu	'seven'	*pitu
*iu	'shark'	*qihu
*m-tidəm	'sharp'	*tazim 'whet'
*meya	'shy; ashamed'	*ma-həyaq
*ənəm	'six'	*ənəm
*ular	'snake'	*hulaR
*mətala	'star'	*mantalaq 'Venus'
*t<m>akav	'steal'	*takaw
*tai	'stomach; belly'	*tian
*vatu	'stone'	*batu
*mulur	'straight'	*lurus
*təvu	'sugarcane'	*təbuh
*naɲi	'swim'	*naŋuy
*luu	'tear'	*luhəq
*pulu	'ten'	*sa-ŋa-puluq
*m-kapal	'thick'	*ma-kapal
*rivu/*ribu	'thousand'	*Ribu
*təlu	'three'	*təlu
*panav	'walk'	*panaw
*kayu	'tree; wood'	*kahiw
*dʒua[b]	'two'	*duha
*uta	'vegetable; bean'	*qutan
*vaʔir	'water'	*wahiR
*apa	'what'	*apa
*budaʔ	'white'	*budaq
*aɲin	'wind'	*haŋin
*binay	'woman; sister'	*binay 'woman'
*sala	'wrong'	*salaq
*vadi	'younger sibling'	*huaji

[a] The prefix *b-* is a nominaliser in CL-Central Lembata (cf. §3.3.6.1).
[b] PFL *dʒ- < PMP *d- is an irregular reflex.

Table 6.3 lists the 24 PFL reconstructions that have a PMP form and show largely regular correspondences in the three Lamaholot subgroups and in

Kedang. Despite the absence of a reflex in Sika, this is enough evidence to classify these sets as cognate sets and reconstruct these words to PFL with the assumption that Sika has replaced these concepts with new words or a reflex is not attested in my dataset.

Table 6.3: PFL reconstructions without Sika reflex (n=24)

PFL	PFL Meaning	PMP source
*hakay	'ascend'	*sakay
*raya	'big'	*Raya
*tuno	'burn; grill'	*tunu
*tanem	'bury'	*tanəm
*doa[a]	'far; long'	*zauq
*pukət	'fishnet, fish trap'	*pukət
*kavil[b]	'fishhook'	*kawil
*əpat	'four'	*əpat
*paluk	'hit'	*palu
*k-silap	'lightning'	*silap 'sparkle; drizzle'
*təkek	'lizard'	*təktək
*a(m)pu	'mother's brother'	*əmpu 'grandparent/grandchild'
*nusu	'mouth'	*ŋusu
*kiput	'narrow'	*kiput
*garaŋ	'rough'	*garaŋ
*takut	'scared'	*takut
*kələm	'sky'	*kələm 'dark, overcast, obscure'
*diri	'stand'	*diRi
*lahe-k	'testicles'	*lasəR
*m-nipih-i	'thin'	*ma-nipis
*basa	'wash'	*basəq
*tani[c]	'weave'	*tənun
*kapik[d]	'wing'	*kapak
*tuun[e]	'year'	*taqun

[a] PMP *-au- > PFL *-oa- is an irregular change.

[b] Sika *kavir* 'fishhook' is related but has irregular initial *k =k rather than expected *k > ʔ/Ø.

[c] The vowel changes from PMP to PFL are irregluar.

[d] (i) Sika *kapik* 'wing' is related but has irregular initial *k =k rather than expected *k > ʔ/Ø. (ii) PMP *a > PFL *i is an irregular change.

[e] PMP *-aqu- > PFL *-uu- is an irregular change.

The 22 PFL reconstructions listed in Table 6.4 have reflexes in Sika and at least one Lamaholot variety but no reflex in Kedang. Nevertheless, these are cognate sets that can be traced back to PMP with regular correspondences. Therefore, these lexemes are reconstructed to PFL and it is assumed that Kedang replaced the respective concepts with new words or my database is missing a Kedang reflex for these words.

Table 6.4: PFL reconstructions without Kedang (n=22)

PFL	PFL meaning	PMP source
*modip	'alive, live'	*ma-qudip
*ʔavu	'ash, dust'	*qabu
*uma[a]	'garden'	*quma
*leba	'burden stick'	*lemba
*tani[b]	'cry'	*taŋis
*taʔi	'excrement'	*taqi
*puhun	'heart'	*pusuŋ 'heart; heart of banana'
*laki	'husband; male'	*laki
*gatər	'itchy'	*gatəl
*lotur	'knee'	*qulu tuhud
*siva	'nine'	*siwa
*meran	'red'	*ma-iRaq
*gəvalik[c]	'return'	*balik
*padi	'rice plant'	*pajay
*tali	'rope'	*talih
*plari/*kari	'run'	*lariw
*kulit	'skin'	*kulit
*g-nilu-k[d]	'sour'	*ŋilu
*ikur	'tail'	*ikuR
*m-panau	'tinea'	*panaw
*puki	'vagina'	*puki
*hapu	'wipe'	*sapu

[a] Kedang *lumar* 'garden' could be related.
[b] Intervocalic PFL *-n- < PMP *-ŋ- is irregular.
[c] PMP *balik > PFL *gəvalik is most likely PMP *b > *w > *v with the addition of a verbalising prefix *g-* (cf. §3.5.4).
[d] Kedang *kiru* 'sour' could be related.

The 14 PFL reconstructions listed in Table 6.5 have reflexes in two or more Lamaholot varieties but neither a reflex in Sika nor in Kedang. Nevertheless, these are cognate forms that can be traced back to PMP with regular correspondences. Therefore, these lexemes are reconstructed to PFL and it is assumed that Kedang and Sika replaced the respective concepts with new words or the reflexes have not been attested in my dataset.

Table 6.5: PFL forms without reflexes in Sika and Kedang (n=14)

PFL	PFL meaning	PMP source
*sika	'chase away'	*sika
*buŋa/*puŋa	'flower'	*buŋa
*(kə)namuk	'fly' (n.)	*ñamuk 'mosquito'
*tuma	'louse on clothing'	*tumah
*ta(ke)	no; not	*taq
*bukat	'open'	*bu(ŋ)kas
*mula	'plant'	*mula
*(v)uvuŋ[a]	'ridge'	*bubuŋ
*hira	'salt'	*qasiRa
*tudu	'sleep'	*tuduR
*ipe	'teeth'	*(n)ipən
*baŋun	'wake up'	*baŋun
*an	'what'	*anu
*muav	'yawn'	*ma-huab

[a] Sika *puvun* 'ridge' could be related.

6.2.3 PFL reconstructions without PMP sources

Table 6.6 lists 37 regular PFL reconstructions that, based on the current stage of knowledge, do not go back to a PMP form. If a related or resemblant form is known to also occur in regional Austronesian languages outside of the Flores-Lembata family, this is indicated in the last column with "Flores" meaning the Austronesian languages of Flores, "Timor (AN)" meaning the in the Austronesian languages of Timor, "Timor (TAP)" meaning in the Timor-Alor-Pantar languages of Timor, and "Alor-Pantar" meaning in the Alor-Pantar languages on the islands of Alor and Pantar. I do not consider

the possible occurrences of the lexemes in languages outside of the East Nusa Tenggara and Timor-Leste region. Further research on the lexicon of the languages in this area and beyond will probably increase the number of these regionally spread items. Currently, 14 out of 37 lexeme sets listed here are also found outside of the Flores-Lembata family. The remaining 23 reconstructions may be considered as exclusive innovations of PFL.

Table 6.6: PFL reconstructions without PMP sources (n=37)

PFL	PFL meaning	Regional spread
*təmisi	'ant'	
*dasan	'ask; report'	
*muku	'banana'	Flores, Timor (AN), Timor (TAP), Alor-Pantar
*təmayuŋ	'bedbug'	Flores, Timor (AN)
*giki	'bite'	Flores, Timor (AN), Timor (TAP), Alor-Pantar
*vəki	'body'	Flores
*tena	'canoe'	
*laku	'civet cat'	Flores, Timor (AN), Alor-Pantar
*rusu or *ruhu	'coral reef'	
*pati	'cut'	Flores, Timor (AN)
*gurit	'dig'	
*bao	'float'	
*lodoŋ	'fall down; descend'	
*voda-k	'fat'	Flores
*pə-vunu	'fight'	
*napu-k	'flat; stream; river'	
*pau[a]	'mango'	Flores, Timor (AN)
*motoŋ	'marungga'	Alor-Pantar
*osan	'mat'	
*k<n>əpuŋ/*həpuŋ	'mosquito'	
*kəmeruŋ	'rice ear bug'	Timor (AN)
*(n)ubak	'stream; river'	
*vura	'sand'	
*labur	'shirt'	Flores, Maluku

PFL	PFL meaning	Regional spread
*kpali-k/*kwali-k	'shoulder'	
*kamak	'skin; bark of tree'	
*kə-melu	'smooth'	
*m-potaŋ	'spit' (v.)	
*(k)rəvun	'sweat'	
*səru-k	'sweet'	
*alis	'tendon'	Flores
*kera	'turtle'	Flores, Timor (AN), Alor-Pantar [b]
*ale	'waist'	
*hogo	'wake up'	
*gəbi/*gnəbin	'wall'	Flores
*(l)oyor	'wave; sea'	
*nora	'with'	Flores, Timor (AN)

[a] Could be related to PWMP *qambawaŋ 'manggo'.
[b] PCEMP *kəRa or *keRa 'turtle'.

6.2.4 Irregular reflexes in individual subgroups

In the following, I discuss instances of regular PFL reconstructions that are linked to cognate sets which contain unexpected changes in individual subgroups. These sets have been listed and counted already in the tables above because they can be reconstructed to PFL. Most irregularities appear in Kedang and the Lamaholot varieties. The Sika reflexes are largely regular. This observation is in line with the fact that there is more additional non-PMP vocabulary attested in Lamaholot and Kedang than in Sika as will be shown in §6.3 below. Both, irregularities in inherited words, such as in discussed in this section, and the additional non-PMP vocabulary, as discussed below, may both point to the historic presence of speakers of unrelated languages, especially in the Lamaholot and Kedang areas, that ultimately switched to Lamaholot and Kedang varieties.

Table 6.7 show the reflexes of the first person pronouns in the FL languages. While the Sika reflexes are completely regular, the Kedang and Lamaholot reflexes show several irregularities, highlighted in bold.

Table 6.7: Irregularities in the first person pronouns

PMP	*i aku	*kami	*kita
PFL	*aku	*kami	*kita
SK	*aʔu*	*ʔami*	*ʔita*
WL	***go***|*ʔe*	*kame*	***t****ite*
CL	***go***|*ne*	*kame*	***t****ite*
EL	***go***|*ʔe*	*ame*	**g***ite*
KD	***ko*** / ***ɛʔi***	*e* / ***ke***	*te*
	'1SG'	'1PL.EXCL'	'1PL.INCL'

For the 1SG pronouns, it appears that the Lamaholot varieties and Kedang underwent irregular *k > *g* and lost the initial vowel *a*. The lowering of the final vowel *u > *o* is regular in WL and EL but not in CL and Kedang (cf. §5.2.7). The subsequent change of *g > *k* in Kedang to gain *ko* '1SG' is a regular change (cf. §5.2.1). For Kedang two variants for 1SG are given: *ɛʔi* as a general pronoun and *ko* as an emphatic pronoun (Samely 1991a:69). It is unclear if the general pronoun *ɛʔi* comes from PFL *aku as two irregular vowel changes would remain unexplained.

The reflexes of PFL *kami appear largely regular. The Kedang pronoun *e* '1PL.EXCL' can be explained by the loss of the second syllable and the regular changes of PFL *k > *ʔ*/*Ø* and *-a > *e*. However, the Kedang variant *ke* has an irregular retention of PFL *k = *k*.

The Sika reflex of PFL *kita is regular. The Kedang reflex is also regular when assuming the loss of the initial syllable. However, in the Lamaholot varieties, an irregular change of the initial consonant is observed. CL and WL undergo *k- > *t-* and EL undergoes *k- > *g-* in their reflexes of PMP *kita. The initial *t* in CL and WL could be explained by sporadic assimilation of the initial consonant to that of the medial consonant.

Table 6.8 lists examples of sporadic lenition of PFL *b > *v* attested in the languages of Flores-Lembata. The same type of lenition in different lexical items has been attested already on a higher level in Proto-Bima-Lembata (cf. §5.5). Here again a few lexemes appear to start a new wave of lenition. Two of the three sets have a PMP source. I also include one set without PMP source because it shows the same pattern of lenition.

Table 6.8: Sporadic lenition of PFL *b > *v*

PMP	*Ri**b**u	***b**inay	-
PFL	*ri**b**u/ *ri**v**u	***b**inay ‘female; sister’	#e**b**el
SK	*ri**v**u*	***w**ine|ŋ*	-
WL (AD)	*ri**b**u*	-	*ə**v**ər|ət*
WL (LWI)	*ri**b**u*	-	*ve**v**el|əŋ*
WL (LWL)	*ri**b**u*	***b**ine?*	*ve**v**e*
CL (KK)	*ri**b**u*	***b**inadʒ* *k|**w**inadʒ* ***b**ine*	*e**v**el*
CL (LR)	*rib*	[...]	*e**v**el*
EL	[...]	[...]	*e**b**lə*
KD	*ri**b**u*	***b**ine|n*	*e**b**el*
	‘thousand’	‘woman’	‘tongue’

For the concept ‘thousand’, it is also possible that PFL had *rivu, thus lenition going back to an earlier stage, and the Kedang and Lamaholot varieties borrowed the Malay word *ribu* ‘thousand’.

For the concept ‘tongue’, no PFL form can be reconstructed because this set cannot be traced back to PMP and Sika does not have a related word. It only appears in Kedang and the Lamaholot varieties. As a change of *b* > *v* is more likely than *v* > *b*, as the same change is attested in other sets, I assume that the original form was *ebel* ‘tongue’ and the the Western and Central Lamaholot varieties underwent lenition. In some varieties, an additional *v* is added before the initial vowel. Sporadic insertion of *v* before vowels is not uncommon in the Flores-Lembata languages. Another example is the Central Lamaholot form for ‘bow’ from PMP *busuR > PFL *vuhur ‘bow’. As PFL *h is lost in Central Lamaholot, *v* is inserted in between the two medial vowels resulting in CL-Kalikasa *vuvor* ‘bow’. Also in Sika *vaten* ‘liver’ from PFL *ate-n < PMP *qatay ‘liver’, the initial *v* is inserted before the vowel.

The two sets ‘salt’ and ‘how much’ in Table 6.9 have irregular forms in some varieties that can possibly be explained as borrowings, marked with an arrow (→). Both concepts are related to the market place. From the Cent-

ral Lamaholot and Eastern Lamaholot forms, PFL *hira 'salt' can be reconstructed. The unexpected Western Lamaholot reflexes containing *s* are most likely loans from Central Flores languages which have *siʔe* 'salt'. PFL *pida (< PMP *pija) 'how much' is regularly reflected in most subgroups, as the correspondences of SK *r* - KD *y*/Ø - WL *r* are regular reflexes of PFL *-d- < PMP *-j-. The Central Lamaholot forms are irregular, as PFL *-d- < PMP *-j- is normally reflected as Central Lamaholot *-dʒ-*. It is possible that Central Lamaholot speakers borrowed *pira* 'how much' from their Western Lamaholot neighbours.

Table 6.9: Possible borrowings (→) in individual subgroups

PMP	*qasiRa	*pija
PFL	*hira	*pida
SK	-	*pira*
WL (AD)	→*siʔa*	*pira*
WL (LWL)	→*siʔa*	*pira*
WL (BL)	→*sia*	[...]
WL (PD)	→*sia*	*pira*
CL (KK)	*ira*	→*pira*
CL (LR)	*ira*|*r*	→*pira*
EL	*hira*	[...]
KD	-	*pie*
	'salt'	'how much'

Table 6.10 shows reflexes of PFL *tuʔan (< PMP *tuqan) 'forest' and instances of resemblant forms in the Lamaholot varieties that cannot be explained by regular sound changes. The Lamaholot forms underwent unexpected voicing of the initial plosive *t.

Table 6.10: Irregular forms for 'forest' in Lamaholot

PMP	*tuqan
PFL	*tuʔan
SK	*tuʔan*
WL (LWL)	***d**uã*
CL (KK)	***d**uan*
EL	-
KD	*tuen*
	'forest'

6.2.5 Discussion

Table 6.11 summarises the features of the PFL vocabulary reconstructed for this dissertation. The left-most column categorises the spread of the lexeme sets. The Lamaholot subgroups are grouped together as LH here as they are located in the centre of the Flores-Lembata family. Kedang and Sika are at the edges. LH thus means one or more Lamaholot subgroups. For the last category of PFL reconstructions that only contain reflexes in Lamaholot varieties, this means that reflexes are attested in at least two Lamaholot subgroups. Lexeme sets that neither have related forms in Sika and Kedang nor a PMP source are not reconstructible to PFL. These are discussed in §6.3.

Table 6.11: PFL reconstructions (n=210)

	PMP source	no PMP source	Total
SK - (LH) - KD	113	37	150
LH - KD	24	-	24
SK - LH	22	-	22
LH	14	-	14
Total	173 (81%)	37 (19%)	210 (100%)

The 210 PFL reconstructions are to a great extent of Austronesian origin, for 81% of them there is a known PMP source. About one fifth of the PFL

vocabulary remains of unknown origin. PFL, as a descendant of PMP, has thus replaced about 20% of the vocabulary for the concepts in this study since PMP times. PMP was spoken around 4000 years ago (Pawley 2005).

When selecting only basic vocabulary forms from the sample, around 124 PFL forms remain.[1] Out of these only 13% are not of PMP origin. This lower percentage of non-PMP vocabulary in PFL basic vocabulary compared to the whole database confirms that lexical replacement in basic vocabulary is less likely to occur than in other parts of the vocabulary.

The PFL vocabulary which is not of PMP origin could be regarded as a non-Austronesian lexical substrate in PFL. However, at the current stage of research, it is not entirely clear if the set of lexical items in PFL that do not trace back to PMP (listed in §6.2.3) can be part of a substrate because it is unknown how much of this vocabulary traces further back to an earlier ancestor of PFL. In §6.2.3, I have shown that about 30% of the non-PMP vocabulary in PFL has related forms elsewhere in the region. As this number is based on an initial survey, more in-depth systematic investigation into the

[1] About half of the lexeme sets in the database have been classified as denoting basic concepts. Often one basic concept is expressed by two or more lexeme sets. The classification as basic concepts is based on the Leipzig-Jakarta Basic Vocabulary list (Tadmor et al. 2010:238-241) with my own extensions, concerning in particular regionally relevant concepts. In total, the following 192 concepts have been classified as basic for the purpose of this study: 1pl excl; 1pl incl; 1sg; 2sg; 3pl ; 3sg; all; ant; ash, dust; back; banana; bathe; betel vine; big; bird, chicken; bite; bitter; black, dirty; blood; blow; body, self; body hair; bone, seed; breast, milk; burn, shine; child, small; cloud, fog; coconut; come; cry; cut, kill; day, sun; deaf; die; dog; dream; drink; drop, fall from above; dry, thirsty; ear; eat; egg; eight; excrements; eye; fall from above, descend; fall over; far, long; fat; fingernail; finished; fire; fish; flat, below, river; flower, blossom; fly; fly (n.); flying fox; foot, leg; forehead; forest; four; fruit, betelnut; full; give; go; good; grass, bush; hair; hand, arm, five; head; headlice; hear; heart; heavy; here; hide; hillwards, above; hit; horn; hot; house; how much, how many; how?; hungry; inside, deep; inside, liver, house; itchy; knee; knife; know; laugh; leaf; lie down (non-human); liver; man; many; meat, flesh ; meeting house; moon, market; mosquito; mother; mountain; mouth; name; narrow; navel; near; neck; needle; new; night; nine; no, not; nose; old; one, alone; person; pound; price, bride price, expensive, buy; rain; rat; rattan; red; rice; road; roof rafter; root; rope; round; run; salt; sand, soil; say; say; sea, wave; see; seven; short; sick, painful; sit; six; skin, bark of tree; sky; sleep, lie down; smoke; snake; soil; spit; stand; star; stomach, belly; stone; storage house, barn ; suck; sugar palm; sugarcane; sun; sweet; swim; tail; teeth; ten; that; thatch for roofing; thatched roof; thick; this; thousand; three; tie; tongue, say; tree, wood; two; vomit; wake someone up; wake up; walk; wash, bathe; water; what; where; white; who; wide; wife, husband; wind; wing; woman, sister; yellow; yesterday; younger sibling.

lexicon of the languages of the region and even beyond may shed light on how far this vocabulary can be traced back. Some of it may even ultimately go back to PMP. It is possible that with further research, the number of PFL reconstructions without PMP source becomes so small that one could account for it by lexical replacement that naturally occurs in any language for different reasons, such as avoidance of homophony, semantic change, borrowing and invention of new words.

6.3 Unreconstructible regular lexeme sets

6.3.1 Overview

This section discusses 185 lexeme sets that cannot be reconstructed to PFL but which show regular correspondences between the subgroups in which they occur. These sets cannot be reconstructed to PFL because related forms neither occur in Sika and Kedang, nor is there a PMP source which could justify a PFL reconstruction. The regularity of the related forms in the set makes it possible to reconstruct a hypothetical form that could be a PFL form if the missing forms in Sika and/or Kedang would be found. These hypothetical reconstructions are marked with a hash tag (#). The lexeme sets discussed here may ultimately be reconstructible to PFL when more data becomes available but they could also be early borrowings that entered a subset of the Flores-Lembata subgroups (cf. §6.1).

The lexeme sets are organised in three categories: sets without a reflex in Sika (§6.3.2), sets without a reflex in Kedang (§6.3.3) and sets with neither a reflex in Sika nor in Kedang (§6.3.4). For each category, I provide a list of hypothetical reconstructions with hash tag (#) and their regional spread as far as this information is available to me. The same categories of regional spread as for the PFL reconstructions without PMP source in §6.2.3 apply.

6.3.2 Lamaholot-Kedang lexeme sets

Table 6.12 contains 73 Lamaholot-Kedang (LH-KD) regular but unreconstructible lexeme sets. Out of these, 19 sets are regionally more wide-spread, thus are attested in at least one language outside of the FL subgroup. The great

majority (54 out of 73 sets) is only attested in Kedang and the Lamaholot varieties.

Table 6.12: Shared lexemes in LH and KD without PMP source (n=73)

#Lamaholot-Kedang	Meaning	Regional spread
#soloi	'answer' (v.)	
#gəter	'ask question'	
#bovoŋ	'bark'	
#həbu[a]	'bathe'	
#malu	'betel vine'	Timor (AN), Timor (TAP)
#puur	'blow'	Flores, Timor (AN), AP
#papi	'burn; clear land'	
#letuʔ	'close' (v.)	
#kova[b]	'cloud; fog'	
#korok	'chest'	
#tapu	'coconut'	
#hekan	'condition; time; garden'	
#mudəŋ	'correct; the following'	
#bəpap	'crocodile'	Alor-Pantar
#belu	'cut; kill'	Flores
#sedu	'dance'	
#klebit	'deaf'	
#butu	'eight; bunch; group'	Flores, Timor (AN), AP [c]
#gokal	'fall over'	
#bəka	'fly'	
#lei	'foot, leg'	
#(kəne) breuŋ[d]	'friend'	
#neʔi	'give'	Timor (AN)
#gedi	'go up; ascend'	
#dikə-n[e]	'good; person'	
#vurek	'gravel'	
#tava[f]	'grow; stem'	
#pohiŋ	'help'	
#vuok	'hole'	
#vetak	'house; barn'	

#Lamaholot-Kedang	Meaning	Regional spread
#nara bone gaku	'how'	
#kverak	'jackfruit'	Alor-Pantar
#kudul	'knee'	
#lolo	'leaf'	
#ləpa	'leaf; sheet; lontar leaf'	
#benehik	'light (not dark)'	
#(kutu) kihan	'louse eggs'	
#kabe	'man; husband; person'	
#rai-k^{g}	'many'	
#tudak	'narrow'	
#dahe-k	'near'	
#vuli	'neck'	Alor-Pantar
#batul	'needle'	Alor-Pantar
#payam	'papaya'	
#volar	'ridge'	
#vadək	'rope'	
#doru^{h}	'rub; wipe'	Alor-Pantar
#taʔu	'salt'	
#bota(n)	'sand'	
#kəburak	'scabies'	Flores
#kuluk	'seed'	Alor-Pantar
#durum	'sell'	
#saur	'sew'	Timor (AN), Alor-Pantar
#məkul	'short'	
#tobe	'sit'	
#təguʔ	'skewer'	
#molan	'sorcerer'	
#gala(r)	'spear'	Flores
#təmidu^{i}	'spit'	Timor (AN)
#bəta	'split'	
#tubak	'stab'	
#(kə)boti	'stomach; belly'	
#kebaŋ	'storage house; barn'	Alor-Pantar
#pola	'sugar palm'	

#Lamaholot-Kedang	Meaning	Regional spread
#soŋa	'tie'	
#ebel	'tongue'	
#(bela) bayan	'treaty'	Alor-Pantar
#deko	'trousers'	Flores, Timor (AN), AP
#ləvu	'village'	
#luaŋ	'vomit'	
#hamu	'wipe; sweep'	Timor (AN)
#kumas	'yellow'	
#evian	'yesterday'	

[a] Central Lamaholot *ləbo* 'bathe' could be related.

[b] Sika *kova* 'cloud' could be related but would involve an irregular retention of PMP *k = Sika *k*. This lexeme set might trace back to PMP *awaŋ 'atmosphere, space between earth and sky' with an insertion of initial *k*- and an irregular change of PMP *a > PFL *o.

[c] PCEMP *butu 'group, crowd, flock, school, bunch, cluster'

[d] Sika *deuŋ* 'friend' could be related but would involve an irregular correspondence of Lamaholot/Kedang *br*- and Sika *d*-.

[e] The set #dikə-n could derive from PMP *diqaq 'good' with an irregular change of PMP *-q- > PFL *-k- before *ə*. However, as also the change of PMP *-aq > PFL *-ə in this word remains unexplained, PFL *dikə 'good; correct' might also be unrelated to PMP *diqaq. The original meaning of this set is probably 'good; correct'. The word 'good' is combined with another word for 'person', i.e. PFL *ata, such as still in used for example in Central Lembata *ata dikən* 'person'. This was probably done as an opposition of members of another group that were enemies. Over time, also the second part of the compound acquires the meaning 'person'. However, in some subgroups, such as for example in Kedang and Eastern Lamaholot, both meanings 'good' and 'person' are retained. In Alorese, a reflex of PFL *dikə means 'right side'.

[f] Eastern Lamaholot *nava* 'stem' could be related.

[g] #rai 'many' could trace back to PMP *Raya 'big'.

[h] Western Lamaholot *doruk* 'rub; wipe' could be related but would involve an irregular retention of PFL *r = WL *r*.

[i] This could be related to PWMP *qizuR 'saliva; spittle'.

6.3.3 Lamaholot-Sika lexeme sets

Table 6.13 lists 41 lexeme sets with regular reflexes in at least one Lamaholot (LH) variety and in Sika (SK). 18 out of 41 sets have related forms in other languages of the region. 23 out of 41 lexeme sets are only attested in Sika

and the Lamaholot subgroups.

Table 6.13: Shared lexemes in LH and SK without PMP source (n=41)

#Lamaholot-Sika	Meaning	Regional spread
#supel	'arrow'	Flores, Alor-Pantar (?)
#baka	'bite'	Flores
#(sə)mei	'blood'	
#nahi	'breath'	Flores
#ihere	'close' (v.)	
#kobu	'crocodile'	
#gasik	'count'	Timor (AN)
#kəbehar	'cuscus'	
#baŋak	'flow'	Flores
#-ai	'go'	
#voloŋ	'hill; ridge'	Flores
#tara	'horn'	
#(raʔi) etan	'know'	Timor (AN)
#blavir	'long; far'	
#koli	'lontar palm'	Flores, Alor-Pantar
#(meiŋ)ʔətan	'meat'	
#təker	'narrow'	Flores
#lusir	'needle'	
#guman	'night'	Timor (AN), Alor-Pantar
#dʒəma	'night, time unit'	
#pehan[a]	'other'	Flores
#likat	'oven'	Flores
#əpak	'palm of hand; footprint'	
#pahat	'plant yam'	Flores
#tubu	'pull'	
#gide	'pull'	
#gualok	'round'	
#madi	'say'	Flores
#kəmekot	'scorpion'	
#buʔu	'short'	Flores
#blara	'sick; painful'	

#Lamaholot-Sika	Meaning	Regional spread
#tuʔay[b]	'sleep'	
#nuhi	'smoke'	Flores, Timor (AN)
#pemek	'squeeze'	Alor-Pantar
#robak	'stab'	
#hukut	'think; remember; miss'	
#kleka[c]	'thunder'	
#papa lele	'trade'	
#puʔu	'wash'	Flores
#kəsako	'whisper'	
#ledan	'wide'	

[a] Kedang *palan* 'other' could be related.
[b] Kedang *tɛʔɛl* 'sleep' could be related.
[c] CL-Kalikasa *kələgor* 'thunder' could be related but would require an irregular change of the last syllable #ka to Kalikasa *gor*.

6.3.4 Lamaholot lexeme sets

Table 6.14 lists 71 lexeme sets with regular reflexes in at least two Lamaholot (LH) subgroups but no reflexes in Sika and Kedang. 16 out of 71 sets have a possibly related forms in other languages of the region. 55 out of 71 lexeme sets are only attested in the Lamaholot subgroups.

Table 6.14: Shared lexemes in LH varieties without PMP source (n=71)

#WL-CL-EL	Meaning	Regional spread
#əvan	'accuse'	
#tapan	'answer'	Timor (TAP)
#svaol	'all'	
#knaru	'back'	
#navak	'body'	
#ravuk	'body hair'	Timor (AN)
#esari nai	'breathe' (v.)	
#hopi	'buy'	
#kiri	'comb'	Alor-Pantar (PAP *kir (Robinson 2015))

#WL-CL-EL	Meaning	Regional spread
#oli	'come; arrive'	
#suda	'command; order' (v.)	
#bisu	'cook'	
#kluok	'cooked rice; uncooked rice'	
#vekan	'divide'	
#knavi	'door'	Alor-Pantar (?)
#ləŋat	'fall from above'	
#gəni	'fight'	
#vahak	'finished'	
#lerek	'flat; below'	
#kənito	'forehead'	
#alus	'good'	
#pehen	'grasp; hold'	
#madu	'grasshopper'	
#latar	'hair'	
#kote	'head'	Timor (AN)
#soroŋ	'hide'	
#dani	'hit (drum)'	
#umaŋ	'hole'	
#plati/kati	'hot'	
#maluv	'hungry'	
#bati	'hunt'	
#gekay	'laugh'	
#səmekiŋ	'left side'	
#loit	'let go'	
#pavaŋ	'lie' (position for things)	
#kleak[a]	'light (weight)'	
#kmoruŋ	'locust'	
#vuda	'lungs'	Alor-Pantar
#elam	'meat; flesh'	
#vətəm[b]	'millet'	Flores
#vala	'mud'	Alor-Pantar
#niləŋ	'necklace'	
#magun	'old people'	

#WL-CL-EL	Meaning	Regional spread
#toʔu	'one'	
#gesak	'other'	
#glasa	'play'	
#nakiŋ	'promise'	Alor-Pantar
#vidu	'pull'	Flores
#magar	'rack above hearth'	
#tue	'return'	
#(a)luŋu	'river; stream'	
#bua	'sail' (v.)	
#sodam	'smell'	Timor (AN)
#m<an>akap	'sorcerer'	
#pərino	'spit'	
#piʔuk	'squeeze'	
#puka	'stem'	Flores
#mopa	'straight'	
#kebol	'sugar palm'	
#luvak	'sun'	
#blolo/golo	'tall'	Alor-Pantar
#luʔo	'thatch for roofing'	
#tnakar	'thatched roof'	
#pənəŋe	thick	
#prəvak	thick	
#petən	'think; miss'	
#məna	'vagina'	Flores
#rio	'wake someone up'	
#ga(ne)	'where'	Alor-Pantar, Timor (TAP)
#henaku	'who'	Timor (AN)
#ugadak	'wound'	

[a] Sika *heak* 'light (weight)' and Kedang *ʔahaʔ* 'light (weight)' could be related to #kleak.
[b] Kedang *vereʔ* 'millet' could be related to #vətəm.

6.3.5 Discussion

In the previous section, 185 sets of related lexemes that cannot be reconstructed to Proto-Flores-Lembata (PFL) have been presented. These lexeme

sets have forms in at least two subgroups of Flores-Lembata and the sound correspondences between the lexical items are regular. The sets cannot be reconstructed to PFL because they do not fullfill the criteria set out in §6.1. In short, the criteria for reconstructibility to PFL are the occurrence of reflexes at least in Sika and Kedang, or alternatively, having a PMP source. The 185 lexeme sets examined here only occur in a subset of the Flores-Lembata languages and do not have a known PMP source.

Table 6.15 provides an overview of the numbers of unreconstructible vocabulary sets that are found. There is a set of vocabulary that is attested in Lamaholot varieties and in Kedang (LH-KD), a set that is attested in Lamaholot varieties and in Sika (LH-SK) and a set of vocabulary that is only attested in the Lamaholot varieties (LH). For the three Lamaholot varieties this means that a related form is attested in at least one of the three Lamaholot subgroups when also shared with Sika or Kedang (category LH-SK or LH-KD), but attested in at least two subgroups when not shared with Sika or Kedang (category LH).

Table 6.15: Unreconstructible regular lexeme sets

	no PMP source
LH - KD	73
LH - SK	41
LH	71
Total	185

There are two ways to explain this unreconstructible vocabulary. First, it is possible that the remaining subgroups lost the reflexes so that ultimately the lexemes could be reconstructed to PFL. However, the ultimate origin of the words would still remain unclear, as they do not appear to be of Austronesian origin (because they lack a PMP ancestor form). Second, it is possible that the words entered the subgroups after the split of Flores-Lembata into subgroups but before the regular sound changes occurred in the individual subgroups. It is not possible that these words were added to the lexicon of the Flores-Lembata subgroups after all subgroup defining sound changes had occurred because then, the regular correspondences between the words in the subgroups could not be explained.

The second option appears more realistic because for the first option one has to assume that a huge amount of vocabulary, especially in Sika and Kedang, has been lost or is not attested in the dataset. This is very unlikely. I thus argue that the new vocabulary entered the subgroups before the subgroup defining sound changes occurred. New lexical items of unknown origin can indicate that the speakers of the language invented new words or that they borrowed the words from an unknown source. Considering the large amount of new lexical items appearing in the Flores-Lembata subgroups, the invention of such a large amount of new words appears very unlikely. So my hypothesis is that the vocabulary must have come from at least one substrate language. Of course, this does not exclude the possibility that for some of the sets the missing forms in either Sika or Kedang will still be found and the set will thus be reconstructible to PFL.

One may propose that the shared lexical items in several Flores-Lembata subgroups are evidence for mid-level subgroups within the Flores-Lembata family. However, as no phonological evidence for such mid-level subgroups could be found (cf. §5.3.2), shared lexical items alone are very weak evidence for subgrouping. I rather suggest that the shared lexical items in Kedang and Lamaholot, Sika and Lamaholot and among the Lamaholot varieties result from contact with the same substrate language(s).

In fact, three lexical substrates can be proposed. A western substrate that affected Sika and the Lamaholot varieties (§6.3.3), a central substrate that only affected the Lamaholot varieties (§6.3.4) and an eastern substrate that affected Kedang and the Lamaholot varieties (§6.3.2). As Lamaholot is located in the middle, it has been affected by all three substrates. Whether these three substrates were actually three different languages or just represent three different selections of vocabulary from the same language cannot be decided from the present data.

6.4 Irregular lexeme sets

In my dataset, there are 27 irregular lexeme sets that cannot be reconstructed to PFL because they do not show regular sound correspondences between the forms in the different subgroups. Nevertheless, the sets contain forms that are resemblant and their similarity cannot be ignored. Some of these sets can possibly be explained as late loans that entered the subgroups after

the sound changes had occurred but others remain unclear. I discuss possible loans in §6.4.1 and unclear sets in §6.4.2.

6.4.1 Possible loans

This section discusses 13 lexeme sets with irregular sound correspondences that can most likely be explained by borrowing. Table 6.16 lists three lexemes that are most likely loans from Malay.[2] These words are considered loans because there is a clear Malay source and they did not undergo the regular sound change of PFL *r > *ʔ* in Western Lamaholot. However, in the sets #ritik 'drizzle' and #rusa 'deer', the sound changes PFL *s > *h* in SK and WL, as well as the change of PFL *k > *ʔ* in Kedang can be seen. These sound changes are not subgroup defining and occur in more than one subgroup. In the set #soroŋ 'give', the change of PFL *s > *h* has not occurred. It has been observed that PFL *s > *h* is sometimes incomplete (§5.2.3). However, as the unexpected *s* occurs in all subgroups here, it is more likely that #soroŋ 'give' is a more recent loan from a time when *s > *h* did not occur any more. In case of the incomplete sound change, some variation between varieties would be expected.

Table 6.16: Malay loans in the Flores-Lembata languages

Malay	*rintik* 'speckle' #ritik	*rusa* 'deer' #rusa	*soroŋ* 'push; shove' #soroŋ
SK	-	*ruha*	*soroŋ* 'serve; stretch out hands'
WL (LWI)	(*kite*)***r***ite|ŋ	***r***uha	*soroŋ*
WL (MS)	***r***ik	*ruha*	-
CL (KK)	*rətik*	*rusa*	-
EL	[...]	[...]	*sorõ*
KD	*riti?*	*ruha*	*soroŋ*
	'drizzle'	'deer'	'give'

There are four other instances of a missing sound change of PFL *r > *ʔ* in Western Lamaholot listed in Table 6.17. For these cases, the loan hypothesis

[2] The Malay meanings are taken from Stevens and Schmidgall-Tellings (2004).

is less easy to prove because there are no known sources for these potential loans. All four lexeme sets are not reconstructible to PFL but they occur in several Lamaholot subgroups, similar to the substrate sets discussed in §6.3.4. However, they are missing regular correspondences in Western Lamaholot. Therefore, I propose that they entered the Flores-Lembata languages, or at least Western Lamaholot, after the sound change of PFL *r > *ʔ*.

Table 6.17: Lexeme sets that did not undergo PFL *r > WL *ʔ*

	#turu	#kromi	#raaŋ	#krogoŋ
SK (KR)	-	-	-	*ʔruguŋ*
WL (AD)	*təʔuru*	*krome*	*rãː*	-
WL (LWL)	*nurãʔ*	*kərome*	*ra*	*krogoŋ*
WL (PD)	*nuroŋ*	*kmore*	[...]	-
CL (KK)	*turən*	*kromi*	*raːŋ*	*krogoŋ*
CL (LR)	*turən*	[...]	*raŋa*	-
EL	[...]	[...]	[...]	[...]
KD	-	-	-	-
	'dream' (v.)	'rat'	'voice'	'skinny'

Similar to #soroŋ 'give' from Malay *soroŋ* 'push; shove', there are six other lexeme sets with unexpected *s* occurring in four subgroups Sika, Western Lamaholot, Kedang and Eastern Lamaholot as listed in Table 6.18. These could be loans that came into the languages after the sound change of *s > *h* in these subgroups. Central Lamaholot varieties regularly retain *s = *s* with the exception of CL-Lerek, as can be seen in the data in the table below. This sound change of Proto-Central-Lamaholot (PCL) *s > *h* in Lerek must be of a much more recent date than the change of PFL *s > *h* in the other Flores-Lembata subgroups. For some of these words in Sika and Kedang, there is also an unexpected *k*. This suggests that the words entered the language also after that sound change had occurred.

Table 6.18: Lexeme sets that did not undergo PFL *s > WL/KD/SK *h*

	#soka	#pasak	#sadok	#si(n)oŋ	#səgat	#soga
SK (HEW)	*soka*	*pasak*	*sadok*	-	-	-
SK (KR)	*soka*	*pasak*	*sadok* 'kick ball'	*sinoŋ*	-	-
WL (LWI)	*soka*	*pasak*	-	*sioŋ*	-	-
WL (LWL)	*soka*	*pasak*	*sadok*	*siõ*	*səgat*	*soga*
CL (KK)	*soka*	*pasak*	*sadok*	-	*səgat*	*soga*
CL (LR)	-	*pahak*	[...]	-	*həgət|əŋ*	-
EL	*sokə*	*pasa*	*sado*	-	-	-
KD	*soka*	*pasaʔ*	-	-	-	-
	'dance'	'shoot'	'fist'	'smell'	'stab'	'hold'

Related forms to the set #soka also occurs in languages further west, such as in Palu'e with the word *tfoka* 'dance' and in Bima with *soka* 'dance'. But no language of origin can be determined. The set #pasak is related to the word *pasa* 'shoot' in several Central Flores language and Proto-Central Flores (PCF) *pasa 'shoot' can possibly be reconstructed. Thus, a Central Flores language could be the donor for this lexeme set. The set #si(n)oŋ 'smell' could be related to Malay *cium* 'smell' but the intervocalic *n* would then remain unexplained. The remaining three sets (#sadok, #səgat, and #soga) are more problematic and it remains unclear if they can be explained by borrowing. No related forms are known and they only appear in a small subset of the Flores-Lembata languages.

6.4.2 Lexeme sets with unexplained correspondences

In this section, I discuss 14 lexeme sets that have obvious similarities but no regular correspondences across the subgroups. Only one out of the sets can be traced back to PMP, namely PMP *susu 'breast'. All others are of unknown origin and only occur in a subset of the Flores-Lembata subgroups. The lexeme sets of unknown origin could possibly be part of the lexical substrate discussed in §6.3 but for some reason, they did not undergo regular sound changes after entering the subgroups. Table 6.19 lists two examples of irregular initial correspondences.

Table 6.19: Unexpected initial correspondences

PMP	-	*susu
	#bərəkət/dəkət	? PFL *(t)usu
SK	-	*uhu*
WL (AD)	***d**əkət*	***t**uho*
WL (LWI)	***b**ə**r**əkə*	***t**uho*
WL (BL)	***r**əkət*	[...]
WL (PD)	***d**əkek*	***t**uho*
WL (MS)	***d**əkɛk*	***t**uho*
CL (KK)	***r**əkət*	***t**usu*
CL (LWT)	***b**ə**r**əkət*	[...]
CL (LR)	***r**əkət|ən*	***t**uho|r*
EL	***r**aʔe*	[...]
KD	-	***t**uʔu*
	'sharp'	'breast'

The lexeme set #bərəkət/dəkət 'sharp' is of unknown origin and only occurs in the Lamaholot varieties. There is alternation of (*bə*)*r* and *d* in onset position. The alternation does not align with the subgroups. A possible explanation for this set is to analyse *dəkət* 'sharp' as the base form which is nominalised with the prefix *b-*, attested as a nominaliser in CL-Central Lembata (cf. §3.3.6). As a following step one would assume the change of *b*(*ə*)*d* > *b*(*ə*)*r* ~*r*.

The lexeme set meaning 'breast' traces back to PMP *susu 'breast'. In the reflexes, the initial PMP *s is lost in Sika, while in Kedang and the Lamaholot varieties, it is replaced by *t*. The replacement of *s by *t* is sporadically found in other words as well, such as for example WL-Adonara *təratu* 'one hundred' < PMP *sa-ŋa-Ratus 'one hundred'. This does not provide enough evidence for this form to be reconstructed to PFL with either *usu or *tusu for instance.

Table 6.20 lists examples of unexpected irregular correspondences of consonants in intervocalic position. All the sets are of unknown origin.

Table 6.20: Unexpected intervocalic correspondences

	#bəCo(l)	#uduk/uruk	#loka	#dihe	#nudəp
SK	*boʔu*	-	-	-	-
WL (AD)	*beto*	-	*loʔok*	-	*nudərət*
WL (LWI)	-	-	*loʔo*	-	*ude*
WL (LWL)	*beʔo* / *bəso* / *beto*	-	*loʔok*	-	*udet*
WL (LML)	*bəso*	*uruk*	[...]	*die*	[...]
WL (WB)	*beto*	-	[...]	-	[...]
WL (BL)	*bəsol*	[...]	[...]	*didi*	[...]
WL (PD)	*beta*	[...]	-	-	*ude|k*
CL (LR)	*bedʒo*	*uduk*	*lokaŋ*	-	*udəm*
CL (KK)	-	*uduk*	*loka*	*didʒi*	*nudəp*
CL (LWT)	*bəsol*	-	[...]	*dihe*	[...]
EL	*bəso*	udu	[...]	-	[...]
KD	-	*uruʔ*	-	-	-
	'come'	'push'	'let go'	'blow'	'heel of foot'

Table 6.21 shows instances of similar looking pairs in two different subgroups. The initial or medial consonant alternates with zero. Table 6.22 shows three sets with words that obviously look related but no clear pattern can be determined. For the set #dʒeta 'hillwards', I also consider the possibility that these are not all cognates of the same set. There could be #reta to account for the words in Sika and WL-Waibalun (WB) and #dʒ(a)e to account for the forms in CL-Kalikasa and WL-Lewolema.[3]

[3] Proto-Central-Flores *d(ʒ)eta 'above' could be related. (Elias 2018) suggests that this form comes from PMP *i atas 'above'. However, the change of PMP *i a > PCF *d(ʒ)e does not appear to be regular.

Table 6.21: Instances of consonant alternating with zero

LH-KD	#(g)iu	#me(r)it	#(d)anen	#(m)are
WL (LWI)	-	-	-	-
WL (LML)	-	[...]	-	-
WL (BL)	-	[...]	-	-
WL (PD)		-	-	*mare*
CL (KK)	*giu*	*mərit*	*danen*	-
EL	-	[...]	[...]	*are*
KD	*iu*	*mɛi?*	*anen*	-
	'cook'	'knife'	'(uncooked) rice'	'smell'

Table 6.22: Resemblant sets with unclear pattern

PMP	(?) *kalawaq	-	-
PFL	*kalaka	#(kə)bukal	#dʒe(ta)
SK	(*kuku*) *raka*	*bulak*	*reta*
WL (AD)	-	*kəbukare*	[...]
WL (LWL)	-	-	*rae*
WL (WB)	[...]	[...]	*reti* [b]
WL (PD)	*klake*	-	[...]
CL (LR)	*laka* (*borit*)	*kəbukal*	*re*|*dʒe* [a]
CL (KK)	*lak* (*borit*)	*kəbukal*	*dʒae* 'hillwards' / *dʒe* 'upwards'
EL	[...]	[...]	[...]
KD	-	*ʔebal*	-
	'spider'	'butterfly'	'hillwards; above'

[a] Kraußе 2016:126
[b] Akoli 2010:59

6.5 Conclusions

In this chapter, I have shown that while Proto-Flores-Lembata (PFL) has a largely Austronesian vocabulary, the lexicon of the individual Flores-Lemba-

ta subgroups has been considerably influenced by non-Austronesian substrate languages. In the following, I first discuss the PFL lexicon and then the lexicon of the individual subgroups.

About 81% of the Proto-Flores-Lembata (PFL) vocabulary that I reconstructed has an Austronesian, i. e. Proto-Malayo-Polynesian (PMP), source (§6.2). PMP was presumably spoken 4,500-3,500 years ago in the northern Phillipines, and the first Austronesian speakers arrived about 3,500 years ago in the area where the FL languages are spoken today (Klamer 2019; Pawley 2005). This suggests a very fast spread of PMP speakers and languages through the whole Indonesian archipelago. As PFL is not a direct descendant of PMP, and its closest relatives of the Bima-Lembata family (cf. §5.5) are spoken in the area, Proto-Bima-Lembata (PBL) could have been spoken by the first Austronesian speakers on Flores and beyond. It is unknown how much time may have passed between the times of PBL (possibly 3,500 years ago) and the times of PFL. It may have been one or two millennia. If this estimation is correct, a time span of about 1,500 to 2,000 years between PMP and PFL can be proposed. As only about 19% of the PFL reconstructions are of non-PMP origin, one can propose a lexical replacement rate of 19% for the time span between PMP and PFL. In contrast to a stronger substrate hypothesis for some of the individual Flores-Lembata subgroups — discussed in more detail in the following paragraph —, it remains unclear if these 19% of non-PMP vocabulary in PFL (13% when considering only basic vocabulary) can be attributed to substrate influence. Further research into the regional spread of this vocabulary is needed.

As illustrated in Table 6.23, in the individual Flores-Lembata languages that are spoken today, the percentage of Austronesian vocabulary drops further to between 62% in the Sika variety of Hewa and 47% in the Central Lembata variety of Central Lamaholot. When only examining basic vocabulary, the PMP percentages are about 10% higher for each language. See footnote (1) earlier in this chapter on the selection of basic vocabulary in this study. In the table, each subgroup is represented by one variety as I have not observed significant variation between the varieties of one subgroup regarding the distribution of PMP versus non-PMP vocabulary.

As the time span between PFL and the present-day Flores-Lembata languages is the same for all, it can be concluded that since PFL times the Lamaholot subgroups underwent the biggest increase in lexical replacement, followed by Kedang and then Sika. Considering the whole database, Central

Lamaholot with 53%, Western Lamaholot with 51% and Eastern Lamaholot with 46% have the highest percentages of non-PMP vocabulary, followed by Kedang with 43%, and then Sika with 38%. This suggests more non-Austronesian influence in the Lamaholot subgroups than in Sika and Kedang.

Table 6.23: Lexemes of PMP / non-PMP origin in individual languages

	Whole database			Basic vocabulary		
	PMP	non-PMP	Total	PMP	non-PMP	total
SK (HEW)	**62%**	**38%**		**75%**	**25%**	
	136	84	220	91	31	122
WL (LWI)	**49%**	**51%**		**61%**	**39%**	
	134	142	276	89	58	147
CL (KK)	**47%**	**53%**		**57%**	**43%**	
	158	175	333	97	73	170
EL (LMT)	**54%**	**46%**		**62%**	**38%**	
	69	59	128	57	35	92
KD (LB)	**57%**	**43%**		**64%**	**36%**	
	131	97	228	78	44	122

The comparably low percentage on non-PMP vocabulary in Eastern Lamaholot is most likely influenced by the small number of non-basic vocabulary that is known for this subgroup. Less than one third of the 128 EL lexical items are non-basic vocabulary, while out of the 333 CL lexical items almost half is non-basic. Due to the different absolute numbers of lexical items available per subgroup, the percentages are not entirely comparable.

About 19% of the non-PMP vocabulary was already present in PFL and is inherited into the individual languages. To this non-PMP vocabulary of PFL, Sika has added about 18% of non-PMP vocabulary after having split from PFL, while in Central Lamaholot, an addition of more than 30% of new vocabulary is attested. Much of the non-PMP vocabulary is shared between Kedang and the Lamaholot varieties (cf. §6.3.2), and another part of it only among the Lamaholot varieties (cf. §6.3.4).

In §6.3.5, I have proposed that the Flores-Lembata languages have been influenced by a lexical substrate after the split of PFL. The Lamaholot sub-

groups located in the centre were affected most and thus have the largest amount of lexical replacement. A larger amount of lexical replacement can also indicate that the contact was more intense or of longer duration. I have pointed out that a number of the lexical items that do not go back to PMP are also found in other Austronesian and non-Austronesian languages of the region, but this is still the minority.