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

Fricke, H.L.A.

Part II

Historical phonology and lexical innovations

CHAPTER 4

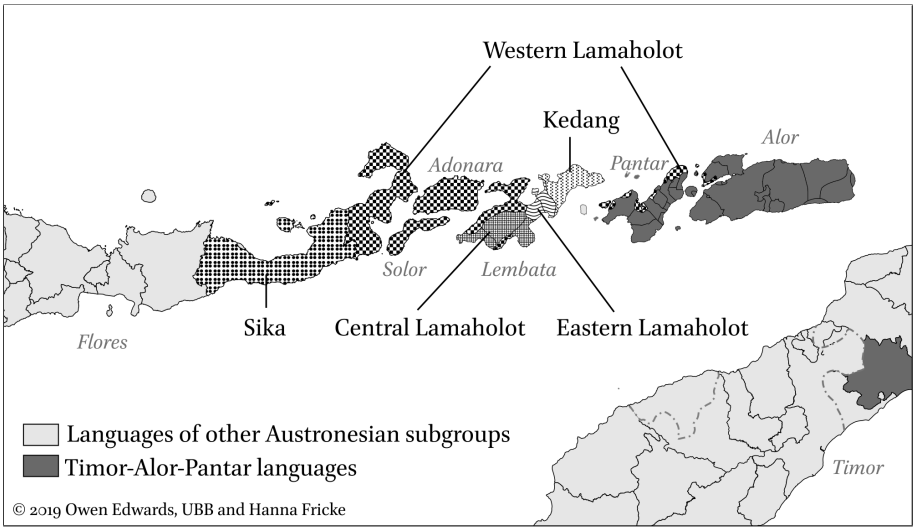
Introduction to Part II

4.1 The Flores-Lembata languages

4.1.1 Overview

The local languages spoken in the eastern part of Flores and in the Solor Archipelago are grouped together as the Flores-Lembata languages, displayed on the map in Figure 4.1. The Flores-Lembata languages are a subgroup within the Malayo-Polynesian branch of Austronesian. All neighbouring languages are also Austronesian, except for the non-Austronesian languages on the islands of Alor and Pantar to the east which belong to the Timor-Alor-Pantar family (cf. §1.3.1). In coastal areas of the islands of Alor and Pantar, an Austronesian language is spoken: the Western Lamaholot language Alorese (cf. §1.2.4.2). Towards the west of the Flores-Lembata languages, the Austronesian Central Flores languages are spoken (Elias 2018). On the island of Timor, southeast of the Solor Archipelago, the Austronesian Timor-Babar languages and Central Timor languages are found (cf. §1.3.1).

Figure 4.1: The Flores-Lembata subgroups and their linguistic context



Based on exclusively shared sound changes (cf. §5.3.1), I distinguish five Flores-Lembata subgroups which are listed in Table 4.1 in geographical order from west to east along with their abbreviations as used throughout this dissertation. In Chapter 5, I provide more details on these subgroups, the internal structure of the Flores-Lembata family, as well as on the classification of Flores-Lembata within Malayo-Polynesian.

Table 4.1: The Flores-Lembata subgroups

Subgroup	Abbreviation
Sika	SK
Western Lamaholot	WL
Central Lamaholot	CL
Eastern Lamaholot	EL
Kedang	KD

Based on lexical similarity or mutual intelligibility many more than five languages can be established.[1] Thus, each of the subgroups contain one or more

[1] Keraf (1978a) contains a lexicostatistical approach on the Lamaholot subgroups which distinguishes 17 languages based on the criterion of a lexical similarity of 80 % or higher.

languages.

Previous studies on the Flores-Lembata languages have not considered Central Lamaholot and Eastern Lamaholot as independent subgroups or branches of Flores-Lembata. The geographic areas of Central Lamaholot and Eastern Lamaholot were usually included in the so-called Lamaholot dialect chain or cluster (cf. Grangé 2015b) but remained linguistically undescribed. Most published research on individual varieties of the Lamaholot dialect chain so far has been conducted on varieties of the Western Lamaholot group.

In this dissertation, I use the term Lamaholot to refer to the three subgroups Western Lamaholot, Eastern Lamaholot and Central Lamaholot as a unit of closely-related subgroups that have been in close contact.[2] However, there is no evidence that Lamaholot, thus the three subgroups, forms an innovation-defined group within Flores-Lembata (cf. §5.3.2). The reasons for not abandoning the label and concept of Lamaholot encompassing the three subgroups as a whole is (i) the fact that the speakers of the three Lamaholot subgroups see themselves as belong to one socio-cultural unit opposed to their neighbours Kedang in the east and Sika in the west and (ii) the three subgroups have been in contact until today and share certain structural features that are not attested in Sika and Kedang, such as clause-final negation and an alienability distinction in the possessive construction (cf. Part III). Also lexically, they are more similar to eachother than to Sika and Kedang. In the following section, I introduce each of the Flores-Lembata subgroups in detail.

[2] The term Lamaholot itself probably goes back to a place name. The first part *lama* means 'place' (Keraf 1978a:7). It is found in many place names throughout the Solor archipelago. The second term *holot* means 'stick together' (Arndt 1937:3; Keraf 1978a:7) and according to Arndt (1937:3) in older forms of the language it also means 'human'. The word *holot* is cognate with the name Solor which refers to an island between Flores and Lembata and to the whole Solor archipelago covering Adonara, Solor and Lembata. The sound change from PFL *s > WL *h* is regular (cf. §5.2.3). So, the island name Solor represents the more archaic form of the name. The final *t* in *holot* replaced the *r* in the place name Solor. It could be related to the suffix *-t* found on attributes in Sika (cf. §8.3.3.1). So Lamaholot could mean something like 'the place of the Solorese' (cf. Hägerdal 2012:22). However, earlier sources, i. e. the Majapahit chronicles Nagarakertagama and the Pigafetti records from the Portuguese ship Victoria mention the name *Solot* with final *t* referring to the area west of Pantar and Lembata. Barnes (1982:409) suggests that the version of the name with final *t* is the original.

4.1.2 Sika

Sika (SK) is the westernmost language of the Flores-Lembata family, spoken by around 175,000 people spreading over an area of about 80 km from west to east in the eastern half of the island of Flores (Lewis and Grimes 1995:601), see map in Figure 4.1. The language is spoken in the whole Sikka regency, with its capital Maumere, as well as in the villages Hewa, Kokang, Wodong and parts of Pantai 'Oa in the southwestern edge of the Flores Timur regency.[3] According to the literature, there are three main dialects of Sika (Rosen 1986:40-41; Lewis 1988:9-10; Lewis and Grimes 1995:601): the dialect from the village"Sikka Natar" on the south coast, the dialect of "Krowe" (central mountain area of the Sikka regency) and the dialect of "Tana 'Ai" in the very east of the Sikka regency.

The oldest and perhaps most detailed description of the language is a grammar by the missionary Paul Arndt (1931) written in German. It covers a wide range of fields, it is however rather difficult to read for linguists today as example sentences are usually not translated and no wordlist is given. Other descriptions are Rosen (1986) on phonology, verb classes and verbal suffixes in Sika, Lewis and Grimes (1995), a very short sketch of Sika, and Bolscher (1982), an introduction to the Sika language written by a missionary in Indonesian for foreign priests coming to the area. Pareira and Lewis (1998) is a Sika-Indonesian dictionary. Lewis also published several anthropological works and articles on the Sika culture, especially on the Tana 'Ai area (Lewis 1988; Lewis 1998; Lewis 2006). The most recent publication is Fricke (2014a) on phonology, verbal inflection, possessive constructions and spatial language in the Hewa variety of Sika.

4.1.3 Western Lamaholot

Western Lamaholot (WL) is the biggest Flores-Lembata subgroup in terms of geographic space and number of speakers, covering an area of about 90 km from west to east, see map in Figure 4.1. Western Lamaholot varieties are spoken all over the Flores Timur regency, on the eastern tip of Flores with its capital Larantuka, the islands Solor and Adonara, as well as in coastal areas

[3] In the linguistic literature, the language Sika is usually written with a single *k*, while in official documents as well as in the name of the regency, Sikka is spelled with double *k* reflecting Dutch spelling from colonial times.

of the Lembata regency including its capital Lewoleba, and in a few villages, such as the village Ojang in the northeastern part of the Sikka regency. The Western Lamaholot varieties in the Sikka regency are called Muhang and are largely undescribed.[4] In addition, Alorese spoken in coastal areas of the islands of Alor and Pantar is linguistically part of Western Lamaholot although this group of speakers has been separated from their homeland for several hundreds of years and the language has diverged considerably since then (cf. §1.2.4.2). I come back to the history of Alorese at the end of this section. The total number of speakers of Western Lamaholot varieties must amount to at least 300,000, as the Flores Timur regency alone has about 240,000 inhabitants according to census data, not including the speakers on Lembata, Alor and Pantar.

Western Lamaholot varieties have a long history of linguistic and ethnographic research. The first linguistic description of Western Lamaholot varieties is Arndt (1937), a grammatical description by a German missionary using information from several western varieties but excluding those on Lembata (Arndt 1937:3). The grammar is written in German and has a structure based on a grammar of a European language. For modern-day use, it is unfortunate that the language examples are not always translated and never glossed on a word-by-word basis. However, it represents an early record of the language which, in some cases, gives information on older stages of the language. In the 1970s, two linguists, native to different parts of the Lamaholot area, published grammatical descriptions of Western Lamaholot varieties in Indonesian. Fernandez (1977) is a short description of the Western Lamaholot variety spoken around the mountain Ile Mandiri close to the town of Larantuka on Flores. Keraf (1978a) describes the morphology of the Western Lamaholot variety spoken in the village of Lamalera on the south coast of Lembata. The exceptional aspect of Keraf's work is his lexical survey on 33 Lamaholot dialects of all three subgroups, which he includes in the appendix of his work. Contemporary research on Lamaholot has involved linguists of various origins. Nishiyama and Kelen (2007) is a brief description of the variety of the villages Lewoingu and Lewolaga located in the eastern part of Flores. Nagaya (2011) represents a grammatical description of one of the westernmost Lamaholot varieties spoken around the

[4] Most likely the variety Keraf (1978a) labels as Pukaunu, located in the western edge of the Flores Timur regency, is similar or identical to Muhang.

mountain Lewotobi located on Flores. Klamer (2011) is a sketch grammar of Alorese, the WL offspring spoken on Alor and Pantar. Grangé (2015b) describes split intransitivity in East Adonara Lamaholot and compares several Lamaholot varieties in terms of their lexicon. Grangé (2015b:47) proposes eastern Adonara as the homeland of Lamaholot, which most likely means the homeland of the Western Lamaholot subgroup because his hypothesis is based only on data from Western Lamaholot varieties.[5] Kroon (2016) is a descriptive grammar of the variety of Lamaholot spoken on the island of Solor. Akoli (2010) and Michels (2017) are Master theses comparing different aspects in several Western Lamaholot varieties. Pampus (1999) is an extensive dictionary of the Lewolema variety of WL with translations into Indonesian and German.

Vatter (1932), Arndt (1938) and Arndt (1940) are early records of cultural aspects in the Western Lamaholot area. Vatter (1932) is not restricted to the Western Lamaholot area but also contains information on the wider region of the whole Solor and Alor Archipelago. Barnes (1987) is a study on weaving practices among the Lamaholot people, which also takes the Central and Eastern Lamaholot areas into account. Barnes (1996) is an ethnographic study on the whale hunting community of Lamalera on Lembata. Kohl (1998) is an extensive ethnographic description of the Lewolema community in the eastern tip of Flores. Rappoport (2014) is a musicological study on songs in Tanjung Bunga, the northeastern tip of Flores.

A more extensive note on Alorese is in order here. According to historical and ethnographic sources, speakers of Alorese emigrated from the Lamaholot area to Pantar around the year 1300 (cf. §1.2.4.2), and later spread further to Alor (Klamer 2011:16). It has been suggested that Alorese is a dialect of Lamaholot by Stokhof (1975:9) but based on lexical divergence — only 55% lexical similarity between Alorese and WL varieties could be attested — and morphological simplification compared to WL varieties, Alorese should be considered a separate language (Klamer 2011:24).[6] This percentage of lexical similarity between Alorese and WL indeed shows that Alorese is lexically as different from other WL varieties as these are differ-

[5] Grangé (2015b:48) himself suggests that more research on the Lamaholot varieties on Lembata is needed to confirm or reject his proposal of Eastern Adonara as the Lamaholot homeland.

[6] The WL varieties used for comparison to Alorese are Lewoingu, Solor and Lamalera (Klamer 2011:118-127).

ent from Central Lamaholot and Eastern Lamaholot. Keraf (1978a:Appendix VI) calculates 55% lexical similarity between Western Lamaholot, Central Lamaholot and Eastern Lamaholot resepctively. In contrast, according to Keraf (1978a), Sika and Kedang only show 30% of lexical similarity to all Lamaholot varieties. But the shared sound change of Proto-Flores-Lembata (PFL) *r > *ʔ* in Western Lamaholot and Alorese (cf. §5.2.5 and §5.3.1.2) and the shared innovation of the clause-final negator Proto-Western Lamaholot (PWL) *hala (< PMP *salaq 'wrong') > Alorese *lahé* (cf. §10.3.4.4) points to a common ancestor for WL and Alorese: Proto-Western Lamaholot (PWL). Therefore, Alorese can be considered an offspring of PWL which has undergone independent and more drastic changes compared with other varieties of WL. This divergence can be explained by the fact that Alorese has been geographically separated from other varieties of WL for about 700 years, as well as by longstanding contacts between speakers of Alorese and speakers of Alor-Pantar languages.

4.1.4 Central Lamaholot

Central Lamaholot (CL) varieties are spoken in the central and southwestern parts of the island Lembata, see map in Figure 4.1. Most parts of this area are mountainous and roads to smaller villages are not always in good conditions. This makes it hard to reach the villages. The estimated number of speakers amounts to around 17,000 people based on census data of the Lembata regency. The Atadei district in the central mountains of Lembata is the only district which is completely Central Lamaholot speaking. Other districts that are partly CL speaking are Wulandoni, Nagawutung and Nubatukan.

Central Lamaholot has been largely undescribed. Until recently, the 200 items wordlists of eight Central Lamaholot varieties in Keraf (1978a) were the only published information on Central Lamaholot. The grammar sketch of the CL variety Central Lembata in Chapter 3 of this dissertation is the first detailed description of a Central Lamaholot variety. There are two recent works of Masters students that discuss varieties of Central Lamaholot. Akoli (2010) describes aspects of the Lewokukung variety, based on a wordlist of 200 basic items and a transcribed and translated folk story. Kraußе (2016) is a brief grammar sketch on the CL variety spoken on the Atadei peninsular of Lembata, named Atadei Painara or Eastern Atadei.

4.1.5 Eastern Lamaholot

Eastern Lamaholot (EL) is spoken in a small area of about 20 km in the eastern part of the island Lembata, see map in Figure 4.1. Speakers of Eastern Lamaholot are mainly found in the Lebatukan district, but probably also partly in the Nubatukan district to the west or the Omesuri district to the east. In the west of this area Western Lamaholot and Central Lamaholot varieties are spoken. Towards the east the language is adjacent to Kedang. The number of speakers amounts up to 8,000 people based on census data. Until today, Eastern Lamaholot remains largely undescribed. The only materials available on Eastern Lamaholot are two 200 items wordlists collected by Keraf (1978a) of the varieties Lewoeleng and Lamatuka.

4.1.6 Kedang

Kedang (KD) is spoken by approximately 29,000 people at the easternmost tip of Lembata in 28 villages surrounding the volcano Ujolewung (Samely 1991a:1), see map in Figure 4.1. The area is part of the Lembata regency which includes the whole island of Lembata where, besides Kedang, also several different Lamaholot varieties are also spoken. Samely (1991a) is the only published grammatical description of Kedang. This grammar is focused on the phonetics and phonology of the language. Samely and Barnes (2013) is a comprehensive dictionary resulting from Samely's data collection and many decades of anthropological fieldwork carried out by Barnes. Several anthropological publications about Kedang by Robert Barnes and Ruth Barnes are available, such as Barnes (1974), an ethnographic study of the Kedang people, or Barnes (2005) about the innovation of weaving practices in Kedang.

4.2 Language sample and data sources

The historical reconstruction and lexical comparison carried out in Part II of this dissertation is based on lexical data from 46 Flores-Lembata varieties, listed in Table 4.2. The lexical items are taken from published dictionaries and, to large extents, from the LexiRumah database which collects comparative wordlists from various sources at https://lexirumah.model-ling.eu/lexirumah/ (Kaiping and Klamer 2018; Kaiping et al. 2019). In the list below, I

first give the abbreviation of the variety I use in this dissertation, then the ISO 369-3 code, the name of the variety and in the last column the sources I used for the lexical data. The varieties are sorted by subgroup. The names of the varieties are taken from the data sources. No claims are made as to whether these varieties represent definable languages or dialects within the given subgroup. In most cases, the names are identical with village names, other place names or mountain names. As some larger units of several varieties have been defined as languages, I indicate those language names in brackets after some of the varieties in the list.

Table 4.2: The Flores-Lembata varieties in the language sample of Part II

Abbr.	ISO 369-3	Variety	Sources
Sika (SK)			
HEW	ski	Hewa	Keraf 1978b; Fricke 2014b, Klamer 2015a
KR	ski	Krowe	Pareira and Lewis 1998
MM	ski	Maumere	Fricke 2014b
TA	ski	Tana 'Ai	Lewis 1995
Western Lamaholot (WL)			
AB / AL	aol	Alor Besar (Alorese)	Moro 2016a
BN / AL	aol	Baranusa (Alorese)	Moro 2016b
MS / AL	aol	Munaseli (Alorese)	Moro 2016c
PD / AL	aol	Pandai (Alorese)	Moro 2016d
AD	adr	Adonara	Klamer 2015c
BT	adr	Botun	Keraf 1978b
DL	adr	Dulhi	Keraf 1978b
HRW	adr	Horowura	Keraf 1978b
KO	adr	Kiwangona	Keraf 1978b
LMK	adr	Lamakera	Keraf 1978b
WT	adr	Watan	Keraf 1978b
WW	adr	Waiwadan	Keraf 1978b
BL	lmr	Belang	Keraf 1978b
LH	lmr	Lamahora	Keraf 1978b

Abbr.	ISO 369-3	Variety	Sources
LML	lmr	Lamalera	Keraf 1978b
MD	lmr	Merdeka	Keraf 1978b
ML	lmr	Mulan	Keraf 1978b
WK	lmr	Wuakerong	Keraf 1978b
BM	slp	Bama	Keraf 1978b
BP	slp	Baipito	Keraf 1978b
LG	slp	Lewolaga	Keraf 1978b
LWI	slp	Lewoingu	Klamer 2015d
LWL	slp	Lewolema	Keraf 1978b; Pampus 1999
RE	slp	Ritaebang	Keraf 1978b
TJ	slp	Tanjung	Keraf 1978b
WB	slp	Waibalun	Keraf 1978b
IA	ila	Ile Ape	Keraf 1978b
LTB	lwt	Lewotobi	Keraf 1978b
PK	-	Pukaunu	Keraf 1978b
Central Lamaholot (CL)			
KK	lvu	Kalikasa (Central Lembata)	Fricke 2015a; Fricke 2019
LK	lvu	Lewuka	Keraf 1978b
LWK	lvu	Lewokukung	Keraf 1978b
IL	lmj	Imulolo	Keraf 1978b
LP	lmj	Lewopenutu	Keraf 1978b
LWT	lmj	Lewotala	Keraf 1978b
MR	lmj	Mingar	Keraf 1978b
LR	lmf	Lerek (Atadei Painara)	Fricke 2015b
PN	lmf	Painara (Atadei Painara)	Keraf 1978b
Eastern Lamaholot (EL)			
LMT	lmq	Lamatuka	Keraf 1978b
LE	lwe	Lewoeleng	Keraf 1978b
Kedang (KD)			
LB	ksx	Leubatang	Klamer 2015b

Abbr.	ISO 369-3	Variety	Sources
LW	ksx	Leuwayang	Samely 1991b; Samely and Barnes 2013

In parts of Chapter 5, I also use lexical data from other Austronesian languages outside of the Flores-Lembata subgroup. These languages and their sources are provided in Table 4.3 hereafter. For these languages no abbreviations are used in this dissertation.

Table 4.3: Other Austronesian languages used for comparison in Part II

ISO 369-3	Language	Sources
Central Flores and Palu'e		
-	Proto-Central Flores (PCF)	Elias 2018
ljl	Lio	Elias 2017b
end	Ende	Aoki and Nakagawa 1993
end	Nga'o	Elias 2017e; Elias 2017d
xxk	Keo	Baird 2002
nxe	Nage	Elias 2017c
nxg	Ngada	Arndt 1961
ror	Rongga	Arka et al. 2007
ple	Palu'e	Donohue 2003
Western Flores		
mqy	Manggarai	Verheijen 1967
kvh	Komodo	Verheijen 1982
Bima and Sumba-Hawu		
bhp	Bima	Ismail 1985
xbr	Kambera	Onvlee 1984; Klamer 1998
hvn	Hawu	Grimes et al. 2008; Jonker 1908
Timor (AN)		
-	Proto-Rote-Meto (PRM)	Edwards in prep

The Proto-Malayo-Polynesian (PMP) sounds and reconstructions in Part II are from Blust and Trussel (2010). For the purpose of this dissertation, I take the PMP forms by Blust and Trussel as they are without evaluation their evidence as this is not feasible within the scope of this work. However, it has to be noted that some of the PMP forms are based on comparably small sets of data with sometimes a geographical bias towards the western part of the Austronesian language family. Therefore, it is possible that with further research in this field and increasing availability of lexical data, some PMP form may have to be revised. It also needs to be taken into account when reading the PMP forms in this dissertation that they are based on a much larger set of languages than the languages in the subgroup of Flores-Lembata. Blust and Trussel (2010) also reconstruct lower-level forms, such as Proto-Central-Malayo-Polynesian (PCMP) and Proto-Central-Eastern-Malayo-Polynesian (PCEMP). However, as the existence of these subgroups is still disputed (cf. §1.3.1), these forms are not taken into account systematically. Nevertheless, if such a PC(E)MP reconstruction is known for a cognate set discussed here, I indicate this.

4.3 Methodology

To answer the research questions for Part II, 46 wordlists of Flores-Lembata varieties (cf. §4.2) were analysed using the online tool EDICTOR (etymological dictionary editor) at http://edictor.digling.org. The wordlists are accessible through the lexical database LexiRumah but originally come from various sources as indicated in the database and in the tables in §4.2. The wordlists contain between 200 and 600 lexical items each. In total they cover 607 different concepts of basic as well as special vocabulary. For some concepts, additional information from dictionaries, also listed in §4.2, was added.

From these wordlists, I collected over 400 lexeme sets. I define a lexeme set as a set of related words that appear across languages. These sets are often cognate sets which means that they trace back to a reconstructible proto-form in a proto-language, such as Proto-Flores-Lembata (PFL) or/and Proto-Malayo-Polynesian (PMP). An example of a cognate set is the set for the concept 'seven' in Table 4.4. However, a lexeme set can also be a set of related forms that cannot (yet) be reconstructed to a common proto-form,

such as in the set for 'tongue' in Table 4.4. Such an unreconstructible form is marked by a hashtag (#) (cf. §4.4). The unreconstructibility of sets could be due to missing cognates, unexplained irregularities or borrowing.

Table 4.4: Examples of lexeme sets

	'seven'	'tongue'
PMP	*pitu	-
PFL	*pitu	#ebel
SK	*pitu*	-
CL (KK)	*pito*	*evel*
WL (LWI)	*pito*	*veveləŋ*
EL	[...]	*eblə*
KD	*pitu*	*ebel*

I establish a set when a lexeme occurs in at least two of the Flores-Lembata subgroups (cf. §4.1). Occasionally, my database contains lexemes that are only found in one Flores-Lembata subgroup but these are only considered if they go back to a PMP form.

The lexeme sets used for Proto-Flores-Lembata (PFL) reconstructions and those sets of similar forms that cannot be reconstructed to PFL, are listed in Appendix B. In the appendix, the sets are sorted alphabetically according to concept. Each set contains a PFL reconstruction or a potential reconstruction marked by a hashtag, as well as all related forms in individual languages that are attested in the LexiRumah database and occasionally in dictionaries. The language varieties in the appendix are marked with the abbreviations listed in Table 4.2.

The lexeme sets, potentially cognate sets, are analysed in several ways to answer the research questions in §1.4, replicated here. The first two questions are addressed in Chapter 5 and the last three questions are answered in Chapter 6.

(1) How are the Proto-Malayo-Polynesian (PMP) proto-sounds reflected in the Flores-Lembata languages?
(2) What is the evidence for (i) subgroups within Flores-Lembata, (ii) Flores-Lembata as a subgroup, and (iii) Bima-Lembata as a higher-

level subgroup including Flores-Lembata and other Austronesian languages?
(3) Which Flores-Lembata lexical items are inherited from an Austronesian ancestor?
(4) Which Flores-Lembata lexical items are not inherited from an Austronesian ancestor?
(5) Which Flores-Lembata lexical items can be reconstructed to Proto-Flores-Lembata?

To answer question (1) on how the Proto-Malayo-Polynesian (PMP) sounds are reflected in the Flores-Lembata languages, I used the comparative method (cf. Campbell and Poser 2008) to establish regular sound correspondences in lexeme sets that go back to PMP forms. The results of this analysis are shown in the first part of Chapter 5.

To answer question (2) on evidence for subgroups, I identified exclusively shared sound changes from the regular correspondences established before. On the basis of regular sound changes as shared innovations, five Flores-Lembata subgroups could be established. To provide evidence for Flores-Lembata and higher level subgroups, other Austronesian languages in the region but outside of the Flores-Lembata group had to be taken into account. As the LexiRumah database also contains lexemes from Austronesian languages of Timor and Flores, for the same concepts as the Flores-Lembata wordlists, I could use this database for the purpose of finding evidence for exclusively shared sound changes in Flores-Lembata and in Bima-Lembata, with the addition of a few other sources (cf. Table 4.3). The evidence for subgroups is provided in the second part of Chapter 5.

Chapter 6 examines the Flores-Lembata lexicon according to PMP and non-PMP origin of lexeme sets and their reconstructability to PFL. To answer questions (3) and (4) which concern the origin of the Flores-Lembata lexemes, I evaluated each set according to the presence or absence of a related PMP form. Lexeme sets with regular sound correspondences and a PMP form were considered to be of PMP origin, thus going back to an Austronesian ancestor. Lexeme sets without related PMP form were considered of non-PMP origin.

For a cognate set to be reconstructible to Proto-Flores-Lembata (PFL), as asked for in question (5), the following criteria must be fulfilled. First, the sound correspondences between the reflexes in different subgroups have to

be regular. The regularity is determined based on the regular correspondences established in Chapter 5. Second, there are two possible conditions that lead to a PFL reconstruction. (i) If the lexeme set can be traced back to a PMP form (attested in Blust and Trussel 2010), then the set is always reconstructed to PFL, or (ii) if no related PMP form is known, the set must be attested in at least Sika and Kedang, then the set can also be reconstructed to PFL. This means that a form that appears in only one or two Flores-Lembata subgroup and has a PMP form is reconstructed to PFL. However, if no PMP form exists, only items that are attested in Sika and Kedang are reconstructed to PFL. Sika and Kedang are the two Flores-Lembata subgroups that are geographically the furthest apart and therefore, the occurrence of related forms in these two languages points to inheritance from Proto-Flores-Lembata (PFL) rather than to diffusion after the split of the family. The possibility of Sika and Kedang forming a subgroup within Flores-Lembata (and therefore sharing vocabulary) has been ruled out (cf. §5.3.2). In the majority of non-PMP sets, the lexeme is not only attested in Sika and Kedang but also in at least one Lamaholot variety. In the sample, there are only 6 sets of related lexical items that are attested in Sika and Kedang but no known reflexes in the Lamaholot variety. Five out of these six go back to Proto-Malayo-Polynesian which suggests that the Lamaholot varieties have replaced these concepts with new words.[7] A lexeme shared by several subgroups, but not fulfilling the criteria just set out, is not reconstructed to PFL. For example, if a set of forms with regular correspondences is currently only known to be found in Kedang and Lamaholot, and does not trace back to PMP, I do not consider it reconstructible to PFL because I cannot rule out diffusion between adjacent subgroups. For such a set, a potential reconstruction is established and marked with a hashtag (#) and a subgroup abbreviation, such as LH-KD for Lamaholot-Kedang.

The following limitations apply to the analysis in Chapter 6. For reconstructions and regularity judgements in this chapter, I focus on consonants, as in the scope of this study, an equally systematic analysis of the vowel reflexes is not feasible. Nevertheless, I take into account that a sporadic change

[7] The Flores-Lembata cognate sets in my sample that only have reflexes in Sika and Kedang but in none of the Lamaholot varieties are: PFL *m-tida ‘sharp’ < PMP *tazim, PFL *tave ‘laugh’ < PMP *tawa, PFL *udu ‘grass; bush’ < PMP *udu, PFL *vivir ‘lips’ < PMP *biRbiR ‘lower lip’, PFL *posok / *blosok ‘rub; wipe’ < PMP *usuq (?) and PFL *ipoho? ‘breathe’ with no known PMP source.

between phonetically more similar vowels, such as *i* > *e*, is more likely than a change between very different vowels, such as *a* > *u*. Thus, a possible lexeme set with a very unlikely vowel change would not be considered reconstructible to PFL. In this chapter, a systematic research on whether the reconstructions of the non-Austronesian Flores-Lembata lexicon are also attested in other Austronesian languages of the area or in the non-Austronesian Timor-Alor-Pantar (TAP) languages was not within the scope of this dissertation. Nevertheless, connections to forms in other languages were indicated when known. It is also possible that some of them turn out to be actually of Austronesian origin if more comparative research is done on Austronesian vocabulary.

4.4 Data representation

4.4.1 Transcription conventions

PMP phonemes, as well as occasional Proto-Austronesian (PAN) forms, are transcribed as in Blust and Trussel (2010). Most of the symbols used by Blust and Trussel (2010) are equivalents to IPA. However, the grapheme <j> in the PMP reconstructions represents [g], [ɣ] or [gʲ], the grapheme <z> represents a voiced palatal affricate [dʒ] or [ɟʝ], the grapheme <R> represents [r], the grapheme <y> represents [j] and the grapheme <e> represents schwa [ə] (Ross 1992; Wolff 2010; Blust 2013:245, 554, 588, 601). To allow a differentiation between schwa [ə] and the unrounded front vowel [e] on lower levels, I retranscribe all PMP *e with *ə. In all other cases, I keep the transcriptions in Blust and Trussel (2010). The non-IPA symbols in the PMP forms are listed in Table 4.5.

My own reconstructions of PFL are given using IPA symbols. Only for the palatal approximant [j], I use the symbol <y> instead of its IPA symbol [j] to avoid confusion with a voiced palatal affricate [dʒ]. All reconstructions are marked with an asterisk <*> in front of the word and given in normal font. Individual phonemes are highlighted in bold to show correspondences. If no PFL form can be reconstructed but a set of similar forms is attested, I give a form that could potentially be a PFL reconstruction based on reversing the sound changes they undergo in the subgroups. I mark these unreconstructible forms with a hash tag (#), instead of an asterisk (*). A form marked by

thus refers to a lexeme shared by several subgroups but which, based on current knowledge, is not reconstructible to a proto-language.

Table 4.5: Non-IPA symbols in PMP forms in Blust and Trussel (2010)

Blust and Trussel 2010	IPA symbol	This dissertation
<j>	[g] / [ɣ] / [gʲ]	<j>
<z>	[dʒ] / [ɟʝ]	<z>
<R>	[r]	<R>
<e>	[ə]	<ə>
<y>	[j]	<y>

Reflexes of entire words or phonemes that are attested in present-day languages are given in italic and transcribed in phonemic IPA. Again with the exception of the palatal approximant [j], which I represent as <y> to avoid confusion with a voiced palatal affricate [dʒ]. The symbol *w* is retranscribed as *v* for the Flores-Lembata languages, as it is realised as a voiced fricative [v] or approximant [ʋ] in all languages of Flores-Lembata. The vowels *ɛ* and *æ* in some Lamaholot sources are retranscribed as *e*, as they are not phonemic. The same is done for *ɔ* and *ɜ* which are retranscribed as *o* and schwa *ə* respectively for the same reason. To make the transcriptions of Kedang lexical data from two different sources consistent, I retranscribe Samely (1991a)'s <e> [ɛ] and <è> [æ] as *e* and *ɛ* respectively and I also retranscribe Klamer (2015b)'s [w] as *v*. In addition, all Kedang [ɔ] are retranscribed as *o* as there is no phonemic contrast between [o] and [ɔ]. Note that the words in the lexeme sets in Appendix B are given according to the sources and are not retranscribed.

4.4.2 Organisation of tables

Tables with sets of related lexemes in this part are composed according to the following conventions. For an example of a lexeme table, see Table 5.4 in Chapter 5 or Table 4.4 in §4.3. Each column lists reconstructions and their reflexes for one concept. The first rows show reconstructions. Then in the subsequent rows, the respective reflexes in the five Flores-Lembata subgroups are provided. The concepts given in the last row of the tables are the

concepts associated with the synchronic meanings of the reflexes. If these do not match with the reconstructed meanings or if the meaning of one lexical items in the set has changed, this is indicated. The subgroups are given in the left-most column. The varieties of one subgroup sometimes differ considerably in their lexicon and for some items, PMP reflexes are only found in a subset of varieties. Therefore, I indicate the variety in brackets if necessary. See §4.2 for a list of varieties and their abbreviations. A dash in the table means that the concept is expressed by a new lexical item that is not cognate. The symbol [...] means that no data is available on this concept for the given language. A vertical line (|) marks a (historic) morpheme boundary in a reflex, mostly to show fossilized morphology or that a certain part of the word is added later and not reconstructible to PFL or PMP. In contrast, a hyphen marks a morpheme boundary signalling active morphology or a morpheme boundary in reconstructions. Angle brackets < > mark infixes. Morpheme boundaries of an infix are marked differently to other morpheme boundaries to avoid confusion with two affixes. A dash in stem-final position means that person suffixes are usually attached to this word. A questions mark (?) before a form means that it is not entirely clear if this form belongs to the set. Lexical items given in square brackets [ ] have been identified as lexical replacements or borrowings.

Chapter 6 contains tables with lists of reconstructions based on cognate sets. The reconstructions are given alphabetically according to the Flores-Lembata meaning and with a PMP source if applicable. The meaning of the PMP source is only indicated when diverging from the PFL meaning. If no PFL form can be reconstructed following the criteria set out in 4.3, I give a form with a hash tag (#) that could potentially be a PFL reconstruction (cf. §4.4.1). For some lexemes two forms are reconstructed. If they are separated by a slash (/), I assume that there was variation between the two forms in the proto-language. This variation is often related to (historic) affixation. If the two forms are separated by 'or', I mean that it is not possible to decide on the actual proto-form based on the current data. Parts of reconstructions may be represented in brackets to indicate that there was probably variation between two forms. An example is PFL *(v)uvuŋ 'ridge' which was probably realised as PFL *vuvuŋ and PFL *uvuŋ.