Zuzana Greksáková

TETUN IN TIMOR-LESTE: THE ROLE OF LANGUAGE CONTACT IN ITS DEVELOPMENT

· U C ·

UNIVERSIDADE DE COIMBRA

Faculdade de Letras

Zuzana Greksáková

TETUN IN TIMOR-LESTE:
THE ROLE OF LANGUAGE CONTACT
IN ITS DEVELOPMENT

Tipo de Trabalho	Tese de Doutoramento
Título	Tetun in Timor-Leste
Subtítulo	The Role of Language Contact in its Development
Autora	Zuzana Greksáková
Orientadora	Ana R. Luís
Orientador	Hugo C. Cardoso
Identificação do Curso	3º Ciclo - Doutoramento
Área Científica	Letras
Nome do Curso	Doutoramento em Linguística do Português
Data	2018

· U   C ·

UNIVERSIDADE DE COIMBRA

For John A. Holm (1943-2015)

ACKNOWLEDGEMENTS

I am also grateful to my family and friends and to all my colleagues I have met in the course of the last five years. Special thanks goes to Jonathan who has spent countless hours helping me build my analytical and statistical tools, making me fall in love with spreadsheets, graphs and colour schemes, and being my biggest fan.

ABSTRACT

From the first Chinese and Malay traders, to the Portuguese exploration of Southeast Asia in the sixteenth century and their colonization of Timor until now, the Tetun language has kept on acquiring many words and features from Portuguese, Malay and other languages to fill lexical gaps in certain semantic areas. At the same time, it has co-existed with Portuguese, Asian Creole Portuguese and Malay for many centuries. The main objective of this dissertation is to investigate the role of language contact in the continuing development of Tetun Prasa in Timor-Leste from a socio-historical perspective. The main focus is on Tetun Prasa's contact with Portuguese, Indonesian, and Asian varieties of restructured Portuguese and Malay at different historical periods while describing stages of its development in terms of not only lexis, but also phonology, morphology and syntax as far as this is possible. To this day, no extensive research has been conducted in the area of language contact in Timor-Leste, although various authors tried to put a label on Tetun Prasa, classifying it as a 'creole', 'pidgin', or 'hybrid language'. I believe it would be more fitting to consider the possibility of Tetun Prasa being a koine with heavy Portuguese lexical influence, especially when compared with the more conservative varieties of Tetun Terik and Tetun Fehan. Consequently, this work aims to show the influence of the language contact and the sociohistorical context on the development of Tetun Prasa as a distinct variety of Tetun. The research objectives include reviewing the available literature on Tetun, considering works written in different historical periods by mostly Portuguese, Indonesian, Dutch, Timorese and Australian authors; collecting natural-speech data and creating oral and written corpora of Tetun Prasa and an oral corpus of Tetun Terik; documenting and describing the actual state of Tetun Prasa, focusing on inter- and intra-speaker variation; and analysing the degree of influence of Portuguese (including Portuguese-based creoles), Indonesian (including Malay and Malay-based creoles), English and local Timorese languages on the phonology, morphology, lexicon, and syntax of Tetun Prasa.

Key words: Tetun Prasa, Tetun Terik, Timor-Leste, language contact, sociolinguistic history, restructured varieties of Portuguese and Malay, borrowings

RESUMO

  Desde os primeiros comerciantes chineses e malaios, à exploração portuguesa do Sudeste Asiático no século XVI através da colonização de Timor, até ao presente, a língua tétum continuou a adquirir muitas palavras e características do português, malaio e outras línguas, para preencher lacunas em certas áreas semânticas, coexistindo com o português, malaio e crioulo português asiático por muitos séculos. O principal objetivo desta dissertação é investigar o papel do contacto entre línguas no desenvolvimento contínuo do tétum-praça em Timor-Leste a partir de uma perspetiva socio-histórica. O foco central é no contacto do tétum-praça com português, indonésio e variedades asiáticas de português e malaio reestruturadas, em períodos históricos diferentes, enquanto é feita a descrição dos estágios do seu desenvolvimento em termos não só de léxico, mas também de fonologia, morfologia e sintaxe, tanto quanto possível. Até ao momento, não foi realizada investigação extensiva na área de línguas em contacto em Timor-Leste, embora vários autores tenham tentado rotular o tétum-praça, classificando-o como um "crioulo", "pidgin", ou "língua híbrida". Cremos ser mais adequado considerar a possibilidade do tétum-praça ser um koiné com grande influência lexical do português, especialmente quando comparado com variedades mais conservadoras, tais como tétum-térique e tétum-fehan. Consequentemente, este trabalho procura demonstrar a influência do contacto entre línguas e o contexto sócio-histórico no desenvolvimento do tétum-praça como variedade distinta do tétum. Os objetivos desta investigação incluem revisão da bibliografia disponível sobre tétum, considerando obras escritas em períodos históricos diferentes, em sua maior parte por autores portugueses, indonésios, holandeses, timorenses e australianos; coleta de dados de fala natural e criação de corpora oral e escrito de tétum-praça e corpus oral de tétum-térique, com foco na variação entre falantes, assim como na fala de cada falante; análise do grau de influência do português (incluindo crioulos de base portuguesa), indonésio (incluindo malaio e crioulos de base malaia), inglês e línguas timorenses locais, na fonologia, morfologia, léxico e sintaxe do tétum-praça.

Palavras-chave: tétum-praça, tétum-térique, Timor Leste, línguas em contacto, histórico sociolinguístico, variedades reestruturadas de português e malaio, empréstimos lexicais

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF MAPS

LIST OF IMAGES

LIST OF ABBREVIATIONS

Grammatical categories and syntactic structures

1S – 1st person singular (*ha'u*)

1PE – 1st person plural exclusive (*ami*)

1PI – 1st person plural inclusive (*ita*)

2S – 2nd person singular (*ó*)

2S.HON – 2nd person singular honorific (*Ita*)

2P – 2nd person plural (*imi*)

3S – 3rd person singular (*nia*)

3P – 3rd person plural (*sira*)

ANT – anterior marker

CONT – continuous

DET – determiner

FEM – feminine

MASC – masculine

NP – noun phrase

PFV – perfective marker

POSS – possessive

PPP – past passive participle

PRF – perfect marker

PRON – pronoun

PST – past tense

RC – relative clause

SVC – serial verb construction

TAM – tense-aspect-mood

VP – verb phrase

Languages

BCP – Bidau Creole Portuguese

CMP – Central Malayo-Polynesian

CP – Creole Portuguese

Ind. – Indonesian language

M. – Malay language

MacCP – Macau Creole Portuguese

MalCP – Malacca Creole Portuguese

OIN – Original Indonesian

PCMP – Proto-Central Malayo-Polynesian

PMB – Proto-Moributonic

Port. – Portuguese language

PT – Proto-Timoric

SEACP – Southeast Asian Creole Portuguese

SLM – Sri Lanka Malay

T. – Tetun

TB – Tetun Belu

TF – Tetun Fehan

TP – Tetun Prasa

TT – Tetun Terik

Recordings

Tetun variety

TD – Tetun recorded in Dili

TS – Tetun recorded in the Suai district

TV – Tetun recorded in the Viqueque district

Sex

F – female

M – male

Level of education

NA – not applicable (no education)

Pri – primary education (primary school)

PreSec – pre-secondary education (middle school/junior high school)

Sec – secondary education (high school)

Uni – tertiary education or professional course

Mother tongue

Bun – Bunak

Kem - Kemak

Mak – Makasae

Mam – Mambae

Mid – Midiki

Tet – Tetun (no specific variety)

TP – Tetun Prasa

TT – Tetun Terik

Wai – Waima'a

Others

CNRT – *Concelho Nacional de Resistência Timorense* (Timorese National Resistance Council)

DNE – *Direção Nacional de Estatística* (National Statistics Directorate)

G-L languages – grammar-lexicon languages

INL – *Instituto Nacional de Linguística* (National Institute of Linguistics)

L1 – first language/mother tongue

L2 – second language

NCSL – nonhybrid conventionalized second language

NSD – National Statistics Directorate

SLA – second-language acquisition

UN – United Nations

UNO – United Nations Organization

UNPD – United Nations Population Fund

UNTL – *Universidade Nacional Timor Loro Sa'e* (National University of Timor-Leste)

V-L languages – verbal-nominal system languages

VOC – *Vereenigde Oost-Indische Compagnie* (Dutch East India company)

Speakers

tx@ACB - 0014TD_ACB_21_M_Uni_Mak

tx@AFA - 0047TV_AFA_50_M_Sec_TT

tx@AGJ - 0033TS_AGJ_24_M_Sec_Bun

tx@AP - 0043TV_AP_68_M_NA_TT

tx@ATB - 0026TS_ATB_41_M_Sec_TT

tx@BB - 0019TS_BB_54_F_Sec_TT

tx@BE - 0005_0006TD_BE_30_M_Uni_TP_Mak

tx@CS - 0056TD_CS_50_F_Pri_Tet

tx@CSG - 0020TS_CSG_56_M_Uni_TT

tx@DLB - 0051TD_DLB_45_M_Uni_Kem

tx@FDP - 0014TD_FDP_19_F_Sec_Mak

tx@FJC - 0026TS_FJC_43_F_Sec_TT

tx@GJX - 0032TS_GJX_42_F_Uni_TT

tx@JA - 0018TS_JA_49_M_Uni_TT

tx@JBP - 0005_0006_0014TD_JBP_26_M_Uni_Mak

tx@JCP - 0044TV_JCP_58_M_Sec_TT

tx@JCV - 0046_0050TV_JCV_48_M_Uni_TT

tx@JLS - 0053TD_JLS_44_M_Uni_Wai

tx@JMB - 0014TD_JMB_31_F_Sec_Mak

tx@LAOB - 0057TD_LAOB_30_M_Uni_Tet

tx@LEG - 0014TD_LEG_20_F_Sec_Mak

tx@LLM - 0023TS_LLM_41_M_Uni_TT

tx@LP - 0045TV_LP_49_M_Uni_TT

tx@MA - 0023TS_MA_47_F_Uni_TT

tx@MAS - 0055TD_MAS_45_M_Uni_Mid

tx@MCA - 0017TS_MCA_29_F_Sec_TT

tx@MN - 0017TS_MN_44_F_Sec_TT

tx@MS - 0042_0049TV_MS_43_M_PreSec_TT

tx@MVJ - 0039TV_MVJ_51_M_Uni_TT

tx@NA - 0019TS_NA_54_M_Uni_TT

tx@NSG - 0054TD_NSG_51_M_Uni_TT

tx@OA - 0032TS_OA_56_F_Uni_TT

tx@OC - 0033TS_OC_31_M_Sec_Bun

tx@RASS - 0018_0020TS_RASS_49_M_Uni_TT

tx@TAB - 0008TD_TAB_55_M_PreS_Mam

1. INTRODUCTION

1.1 Introduction

A Portuguese speaker listening to a conversation in Tetun Prasa will immediately catch many words that sound familiar. From early Chinese and Malay traders stopping at Timor's shores, to the Portuguese exploration of Southeast Asia in the sixteenth century through their colonization of Timor until now, the Tetun language has kept on acquiring many words from Portuguese and Malay and other languages to fill lexical gaps in certain semantic areas, co-existing as it has with Asian Creole Portuguese and Malay for so many centuries. This process has resulted not only in lexical borrowing but also a certain degree of grammatical restructuring, which makes Tetun Prasa an interesting case of a contact language, yet by no means a creole language since so much of its original Austronesian vocabulary and syntax survives intact. Its borrowed vocabulary reveals that some of its contact with Portuguese was with restructured varieties such as the creole Portuguese of Malacca and Macau, as well as contact varieties of Malay (e.g. Ambonese Malay) until the modern Indonesian language was standardized as the official language, known as Bahasa Indonesia, in the twentieth century.

In 2002, Timor-Leste's constitution made both Tetun and Portuguese the nation's official languages. Since 2004, the National Institute of Linguistics in Dili has been trying to promote the standardized grammar and orthography of Official Tetun for use in school and local public life. However, Tetun Prasa is still only partly standardized and even the mass media fail to use its standardized form. In terms of linguistic research, studies of the effect of language contact on Tetun Prasa's socio-historical development remain quite limited, although considerable research has been done by the historian Luís F. Thomaz, so it is possible to build on his work to document the development of Tetun Prasa more fully.

To map the influence of language contact, I divided my dissertation into nine chapters. In Chapter 1, I state the research questions, aims and objectives and initial hypotheses of this dissertation and give a short introduction to Austronesian languages of

Timor-Leste, discussing their typological division and the classification of Tetun. In Chapter 2, I present the sociolinguistic history of Timor, from the pre-colonial times and Portuguese Timor, to Indonesian occupation and independence, focusing on history, social context, linguistic attitudes and language planning. This chapter also includes a discussion on the various Tetun orthographies co-existing in Timor-Leste. Chapter 3 gives a short introduction to linguistic theories on language contact, koineization, codeswitching, second-language acquisition, and multilingualism. I also discuss why Tetun Prasa is not a creole language but rather a koine – a discussion that will be given more space in the Conclusions. The last part of Chapter 3 includes a general description of the methodology I used while collecting my data and analysing the corpora. In Chapter 4, I review the most important works on Timor-Leste, Tetun, based on the Chinese accounts dating back to the 13th century, the Portuguese accounts starting from the 16th century on, and the numerous international works written in the last two centuries. Chapters 5 to 8 focus on the structural and lexical features of Tetun Prasa in comparison to other varieties of Tetun, Portuguese and Malay and their respective creole languages. In Chapter 5, I discuss the limited inflectional and derivational morphology, in Chapter 6, I look at the syntactic structure of noun phrases, verb phrases, and clauses, in Chapter 7, I analyse the phonology of Tetun Prasa focusing on borrowed phonemes, and in Chapter 8, I examine the most salient evidence of the language contact, namely the lexicon. The last Chapter 9 is dedicated to Conclusions.

1.2 Aims and Objectives

The main aim of this dissertation is to investigate the role of language contact in the continuing development of Tetun Prasa in Timor-Leste from a socio-historical perspective. The main focus is on Tetun Prasa's contact with Portuguese, Indonesian, Malay and Asian varieties of restructured Portuguese and Malay. I hope to identify the degree of influence on the Tetun language at different stages of its development not only in terms of lexis, but also phonology, morphology and syntax insofar as this is possible.

To this day, no extensive research has been conducted in the area of language contact in Timor-Leste, although various authors tried to put a label on Tetun Prasa,

classifying it as a 'creole', 'pidgin', or 'hybrid language'. Consequently, this work aims to show the influence of the language contact and the sociohistorical context on the development of Tetun Prasa as a special type of contact language and a distinct variety of Tetun.

The main research questions are the following:

- To what extent have Portuguese (including Portuguese-based creoles), Indonesian, Malay (including Malay-based creoles), English and local Timorese languages influenced the morphology, syntax, phonology and lexicon of Tetun Prasa?
- How does the sociolinguistic background of the speakers of Tetun Prasa influence their idiolects?
- What kind of a contact language is Tetun Prasa?

To answer these questions, I set the following research aims and objectives:
- review the available literature on Tetun, considering works written in different historical periods by mostly Portuguese, Indonesian, Dutch, Timorese and Australian authors;
- collect natural-speech data and create oral and written corpora of Tetun Prasa and an oral corpus of Tetun Terik (a more conservative variety of Tetun spoken as a L1 in the south and west of Timor-Leste);
- document and describe the current state of Tetun Prasa, focusing on inter- and intra-speaker variation;
- conduct qualitative and quantitative analyses of Tetun Prasa and Tetun Terik to evaluate the extent to which Portuguese (including Portuguese-based creoles), Indonesian, Malay (including Malay-based creoles), English and local Timorese languages influenced the phonology, morphology, lexicon, and syntax of Tetun Prasa.

The initial hypothesis that I worked with was that Tetun Prasa is not a creole language but possibly a mixed language with a heavy Portuguese influence. I expected to find a clear split between the borrowed lexicon and native grammar, which would support

my hypothesis. There has been some work done on the extent of the Portuguese borrowings in Tetun Prasa, but without any relevant quantitive analysis. I expected to find overwhelming numbers especially in the written corpus. I also assumed that new structures might have emerged within the last 15 years since most of the grammars and dictionaries on Tetun Prasa have been written and, conversely, that some structures might have disappeared.

This dissertation aims to be an interdisciplinary work covering areas like general linguistics, contact linguistics, descriptive linguistics, language documentation, variationist sociolinguistics, history, second-language acquisition and language typology. However, its purpose is not to offer an exhaustive description of Tetun Prasa's grammar but only focus on the structures that have resulted from language contact.

1.3 Typological division

The linguistic situation in Timor-Leste is rather complex. In a country of 1.5 million, more than 20 languages are spoken and to this date, it is not clear how many languages there actually are. What we know is that they belong to two different language families: Austronesian and Papuan. Papuan languages (Bunak, Makasae, Fataluku and Makalero) are in the minority, and most of them are spoken in the eastern part of the island, except for Bunak, which is spoken in the southwest. Some languages, like Waima'a and Makuva (nearly extinct) were thought to belong to the Papuan family because of the heavy influence of the Papuan languages, but in reality, they do belong to the Austronesian branch (Lewis & Fenning 2013), together with many others, including Tetun, Mambae, Baikeno, Galolen, Kemak, Tokodede, Kairui, Midiki, Naueti. Sometimes, it is very difficult to determine what constitutes a separate language as opposed to what is just a variety of another language. For example, Idalaka has been referred to as a 'dialect continuum' of Idaté, Isní, Lakalei, and Lolei dialects. In the 2015 Census, the Timorese were able to choose their mother tongues from a list of 32 native languages, including the dialect continuum Idalaka. Officially, according to the Census results, Idalaka has 211 native speakers, much fewer than the respective dialects. In contrast, Idaté has 14,127 speakers, Isní 700 speakers, Lakalei 3,669 speakers and Lolei

1155 speakers (DNE[1] 2015). In this section, however, I only focus on Austronesian languages, and on Tetun in particular.

The Austronesian languages of Timor-Leste belong to the Central Malayo-Polynesian (CMP) subfamily (Blust 1993:243), as Figure 1 illustrates.
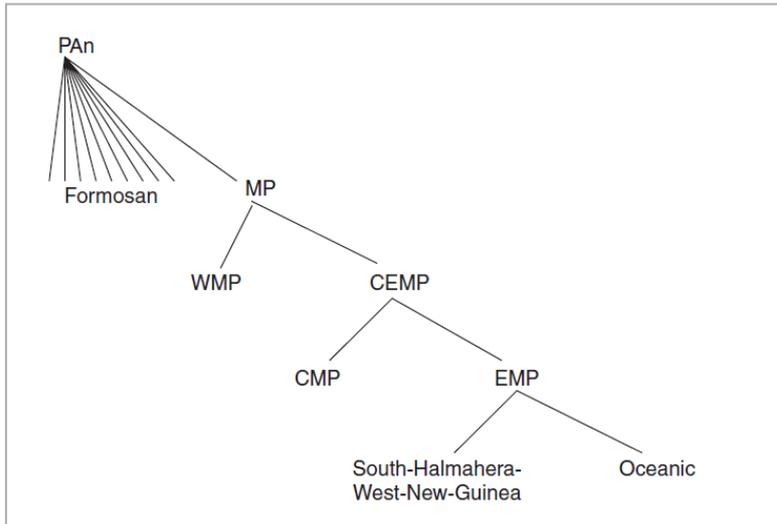
Figure 1 Classification of Austronesian languages (Adelaar 2005:9)

Various authors have attempted to further classify the Austronesian languages of Timor. Wurm and Hattori (1981-3) described a separate Timor Area group and Ross (1995:82) later also put Tetun in a separate Timor group. Van Engelenhoven (1995:17) further divided this group and put Tetun, together with Waima'a, into a South-East-Timor subgroup. Grimes et al. (1997:51-54) also included Tetun within a Timor Area group, together with 22 other languages (see Map 1). Hull (1993:vii-viii) first classified Tetun as a Central Malayo-Polynesian language (CMP) but later explored the possibility of the Celebic (Sulawesi) connection which would mean grouping Tetun with the Western Malayo-Polynesian group of languages (Hull 1998b). In general, there has been a lot of discussion about the existence of the Central Malayo-Polynesian family as well as its lower-level subgrouping (Adelaar 2005:26).

---

[1] Direcção Nacional de Estatística (National Directorate of Statistics).
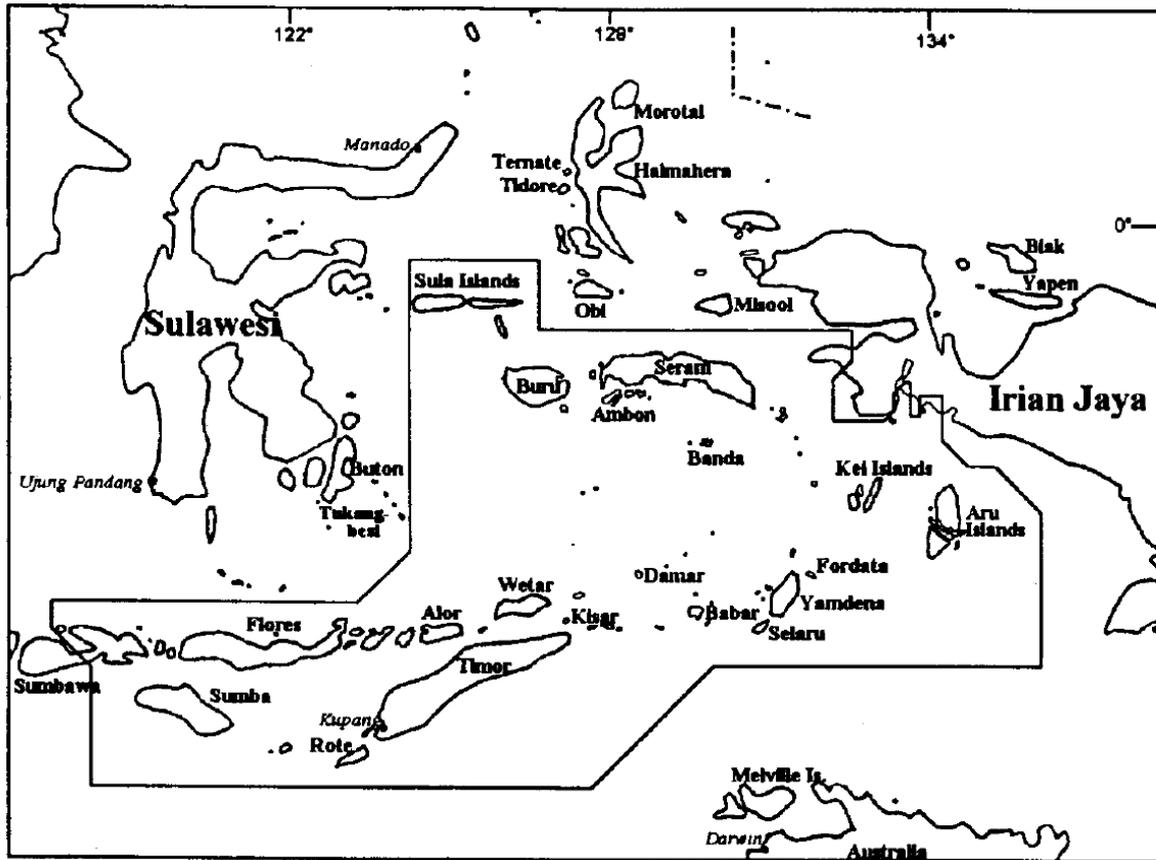
Map 1 Central Malayo-Polynesian languages in Eastern Indonesian (Grimes et al. 1997:7)

Regardless of the official classification, Austronesian languages share a lot of features. Most of them belong to the group of preposed possessor languages, which contrasts with the group of symmetrical voice languages. The differences are summarized in Table 1:

Symmetrical voice languages	Preposed possessor languages
Symmetrical voice alternations	No or asymmetrical voice alternations
Postposed possessor	Preposed possessor
No alienable/inalienable distinction	Alienable/inalienable distinction
Few or no differences between narrative and equational clauses	Clear-cut differences between narrative and equational clauses
Person marking only sporadically attested	Person marking prefixes or proclitics for S/A arguments
Numerals/quantifiers precede head	Numerals/quantifiers follow head
Negators in pre-predicate position	Clause-final negators
V-initial or SVX	V-second or -final

Table 1 Characteristic features of symmetrical voice and preposed possessor Austronesian languages (Himmelmann 2005:175)

Klamer and Ewing (2010:10) tried to pinpoint specific features typical for the Austronesian languages of East Nusantara[2]. In Table 2, I summarise these features and add my observation about their presence in Tetun Terik (TT) and Tetun Prasa (TP).

	Specific features listed by Klamer & Ewing (2010:10)	My observation about features present in TT and/or TP
Phonology	*Prenasalized vowels*	in TT and TP

	Roots are generally CVCV
	
	 - *dispreference for homorganic consonant clusters*
	 - *dispreference for closed syllables, creation of open syllables*
		in TT and native words in TP
		    - in TT but not in TP loanwords
		    - not attested

	Metathesis	not attested
Morphology	*No productive voice system on verbs*	in TT and TP
	Agent/subject indexed on verb as prefix/proclitic	in TT but lost in TP
	Morphological distinction between alienable/inalienable nouns	in TT but fossilized in TP
	Left-headed compounds	in TT and TP
	Inclusive/Exclusive distinction in pronouns	in TT and TP
Syntax	*Verb-object order*	in TT and TP
	Prepositions	in TT and TP
	Gen-noun	in TT and TP
	Noun-numeral order	in TT and TP
	Clause-final negators	not attested
	Clause-initial indigenous complementizers	in TT and TP
	Absence of a passive construction	in TT but possibly developing in TP
	Formally marked adverbial/complement clauses	in TT and TP
Other	*Parallelisms without stylistic optionality*	in TT but lost in TP

Table 2 Characteristic features of Austronesian languages in East Nusantara (Klamer & Ewing 2010:10) and their presence in TT and TP

---

[2] A geographical area extending from Sumbawa, through the islands of the Indonesian East Nusa Tenggara province (Komodo, Flores, Solor islands, Alor-Pantar islands, Sumba, and Timor), Maluku, including Halmahera, the Bird's Head peninsula of West Papua province, and Sulawesi. Papuan and Austronesian languages spoken in this area share many linguistic features which made Klamer and Ewing (2010:1) refer to East Nusantara as a linguistic area or a *Sprachbund*.

As is clear from the table above, there is a difference between Tetun varieties but again, there is little consensus on how big the gap actually is. Should these varieties be considered different languages or dialects of a single language? As we know, the boundary is often fuzzy, and theoretically, all dialects can develop into separate languages, given enough time and the right social circumstances. During this transition process, it is impossible to draw a dividing line between 'possible to understand' and 'impossible to understand' (Thomason 2001:2).

The existence of different dialects of Tetun was already noted in the earliest works on this language. Silva (1889) distinguished between the dialect of Dili and the dialect of the interior of the island. Dores (1907) also noted that there was a distinct dialect of Tetun spoken in Dili.

In 1952, Sá (1952a:23) argued against calling all native languages 'dialects' and adopted the linguistic view on the division of languages into dialects and subdialects. He agreed with the previous authors (Silva 1889, Dores 1907) that Tetun as a language has several varieties with Tetun Terik[3] being a subdialect of Tetun.

Morris (1984:x) was more precise delimitating the various dialects of Tetun. He identified four of them and classified them as follows:

- Tetun Loos[4]: spoken in the south-central region of Timor-Leste (in Soibada and Somoro kingdoms). The author used this dialect as the basis of oral conversational Tetun in his dictionary.
- Tetun Terik: spoken in NW of Timor-Leste and NE of West Timor. The author believed it was closely related to Belu dialect.
- Tetun Belu: spoken in SW of Timor-Leste and SE or West Timor. The author argued that these two dialects (Terik and Belu) were often regarded as a different language from Tetun Loos because of different definitions of individual words, but the syntax was the same.
- Tetun Dili: the dialect taught to the Portuguese and other people in need of a common language for commerce. The author believed it was simpler

---

[3] In his other work, *Timor* (1952b), Sá referred to the two varieties of Tetun as Tetun popular (popular Tetun = Tetun Prasa) and Tetun clássico (classical Tetun = Tetun Terik). See also section 4.2.3.6.
[4] In Tetun, *loos* means 'right, correct'.

in grammar and regarded it as a lingua franca in Timor-Leste in the Portuguese times, but suggested it might be replaced by Indonesian, which did not prove to come true.

We find these denominations in many later works too, but they often refer to different areas, which makes it difficult to adopt one single classification. For example, Hull (1993: viii-ix) identified the same Tetun-speaking areas but gave them different names:

- Tetun Terik/Tetun Loos: Tetun spoken in SE region of Timor-Leste (Soibada and Viqueque) and its western variety spoken from coast to coast;
- Tetun Belu: SW dialect spoken around the border;
- Tetun Prasa/Dili/Maka: spoken in Dili.

However, most of the authors prefer the basic Tetun Prasa/Tetun Terik division. Some decided to refer to these two varieties as two languages, while others consider them to be merely two different dialects.

Thomaz (1974:294-295) identified Tetun (also called Tetun Loos or Tetun Terik) and its four dialects as a separate language, different from Tetun Prasa/Tetun Dili.

Hull and Eccles (2005:xvi-xvii) distinguished two main varieties of Tetun: Tetun Prasa (spoken in Dili and as a L2 in most of the country) and Tetun Terik (a collection of rural varieties of Tetun). The authors also noted that they were not interested in the variety of Tetun spoken in West Timor and that the reference to Tetun Belu only applied to the dialects of Balibó and Suai. On the other hand, Tetun Belu in Hull (1999c:x) refers to the Western Tetun spoken in West Timor.

Van Klinken (1999:3) studied the Fehan dialect of Tetun spoken in West Timor. She divided Tetun into four dialects (see Map 2):

- East Tetun: spoken on the southern coast of Timor-Leste;
- Dili Tetun: spoken around Dili on the northern coast;
- Foho dialect: northern dialects spoken around the border;

- South Tetun: southern dialect spoken around the border with two sub-dialects: Fehan, spoken in the southern agricultural plain, and Suai/Kamanasa spoken in the region of Suai.

In her later works, Williams-van Klinken kept the division simple. In Williams-van Klinken et al. (2002a:1), the authors distinguished between two main varieties, Tetun Dili and Tetun Terik, while recognizing the various dialects of the latter.
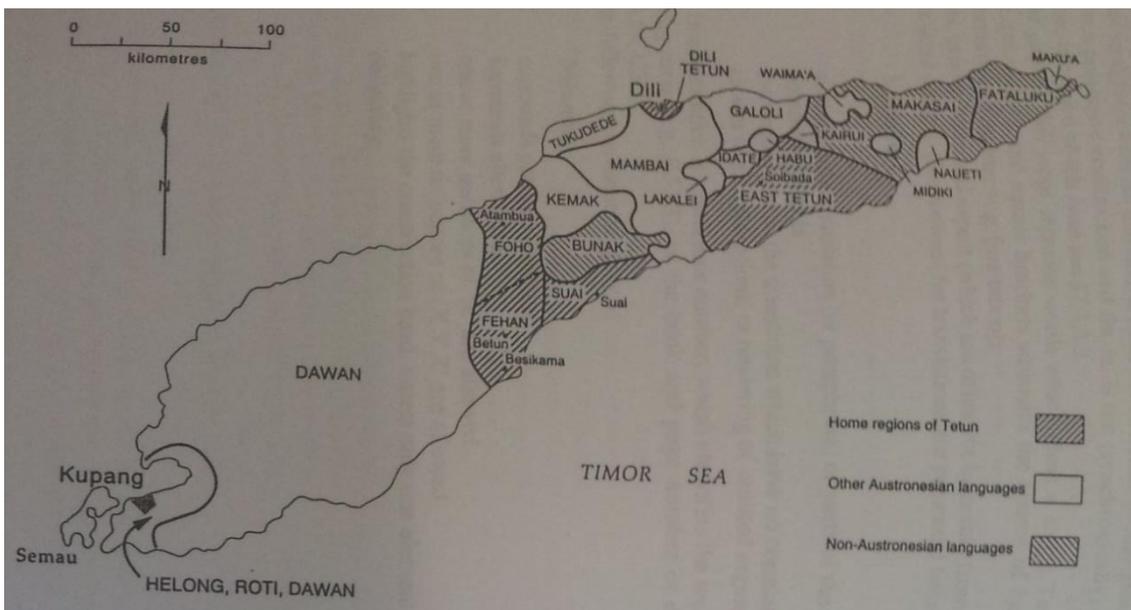
Map 2 Tetun dialects according to van Klinken (1999:xxx)

The latest classification was attempted by Albuquerque (2011:61), as shown in Figure 2, who tried to unify the two previous proposals: that of Williams-van Klinken et al. (2002a) and that of Hull and Eccles (2005). Albuquerque (2011) also made clear the distinction between Tetun Prasa spoken as a second language (Tetun Lingua Franca) and spoken as a first language in Dili (Tetun Dili). According to him, Tetun Lingua Franca does not have any native speakers and is used as an L2, in diglossic situations and as a lingua franca in various districts. It also lacks recent vocabulary that refers to modern concepts but presents the influence of Indonesian. On the other hand, Tetun Dili is the mother tongue of the Dili population, which has been greatly influenced by Portuguese and contains a lot of recent vocabulary borrowed from Portuguese and English (Albuquerque 2011:65).
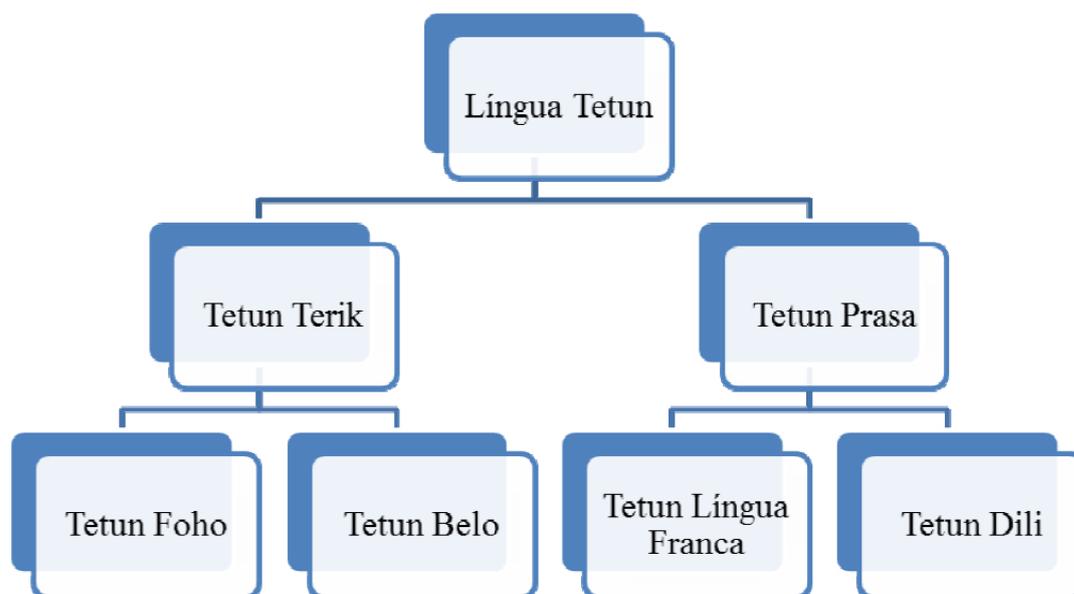
Figure 2 Classification of Tetun according to Albuquerque (2011:61)

However, Albuquerque (2011) did not present any data to support this classification so I decided to treat native and non-native speakers of Tetun Prasa together. This was due to the fact that most of my non-native informants have been living in Dili for a long period of time and many have slowly shifted to Tetun Prasa, which they would often also use at home with people from the same district and of the same mother tongue. On the other hand, I have noticed 'atypical' grammatical and syntactic structures in native speakers of Tetun Prasa as well (see section 6.1.3 on Possession), which proves again that there is a high level of inter-speaker variation (see section 3.2).

Albuquerque's division of Tetun Terik is also rather unclear. He uses Tetun Foho to refer to the native language of the Viqueque district. He seems to base his division on van Klinken (1999), but as you can see on Map 2, Tetun Foho is spoken on the northern coast around Atambua. Following the classification of Hull and Eccles (2005), he opted for Tetun Belo to refer to the language spoken in the wide region around the border with Indonesia. He also argued that this sub-variety has been largely influenced by Indonesian (Albuquerque 2011:65), although I propose that the influence has been rather limited (see section 8.6.1).

In Table 3, I provide an overview of the different approaches to the classification of Tetun varieties and dialects in the literature.

Author(s)	Classification of Tetun
Silva (1889)	• dialect of Dili • dialect of the interior of the island
Dores (1907)	• dialect of Dili • dialect of the interior of the island
Sá (1952a, 1952b)	• Tetun Terik/classical Tetun • Tetun Dili/popular Tetun
Thomaz (1974)	• Tetun/Tetun Terik/Tetun Loos • Tetun Dili/Tetun Prasa
Morris (1984)	• Tetun Loos • Tetun Terik • Tetun Belu • Tetun Dili
Hull (1993)	• Tetun Terik/Tetun Loos • Tetun Belu • Tetun Prasa/Dili/Maka
van Klinken (1999)	• East Tetun • Dili Tetun • Tetun Foho • South Tetun (dialects of Tetun Fehan and Tetun Suai/Kamanasa)
Hull & Eccles (2005)	• Tetun Prasa • Tetun Terik • Tetun Belu (dialects of Balibó and Suai)
Williams van Klinken et al. (2002a)	• Tetun Dili • Tetun Terik
Albuquerque (2011)	• Tetun Terik (subvarieties of Tetun Foho and Tetun Belu) • Tetun Prasa (subvarieties Tetun Lingua Franca and Tetun Dili)

Table 3 Overview of different classifications of Tetun Terik and Tetun Prasa in the literature

Because of this conflicting nomenclature, I decided to rely on my own evidence and follow my own classification (see Figure 3). As my main focus is Tetun Prasa, I did not conduct an extensive research into other varieties. However, because I thought that it was important to make comparisons with Tetun Terik, I have collected my data also in two Tetun Terik-speaking areas, in Suai and Viqueque. I noticed that there were some differences between these two subvarieties, so in some cases, I make a distinction between Tetun Suai and Tetun Viqueque. In cases where this distinction is not necessary or possible, I will use Tetun Terik to refer to the more conservative variety of Tetun spoken as L1 in the south of Timor-Leste and in the border region with Indonesia.

As mentioned above, I use Tetun Prasa to refer to the L1 variety spoken in Dili and L2 variety spoken as a lingua franca in most of Timor-Leste. All of my data were collected in Dili and include L1 and L2 speakers. Although I had to deal with a great deal of inter-speaker variation, I have not found enough evidence to make claims about subvarieties of Tetun Prasa.

The last variety of Tetun I have considered in this thesis is Tetun Fehan, documented in van Klinken (1999). Although the author claims Tetun Fehan is a dialect of South Tetun, which also includes Tetun Suai, I decided to treat this variety separately. We will see just how different these two varieties are, especially in the chapters on morphology (Chapter 5) and syntax (Chapter 6).
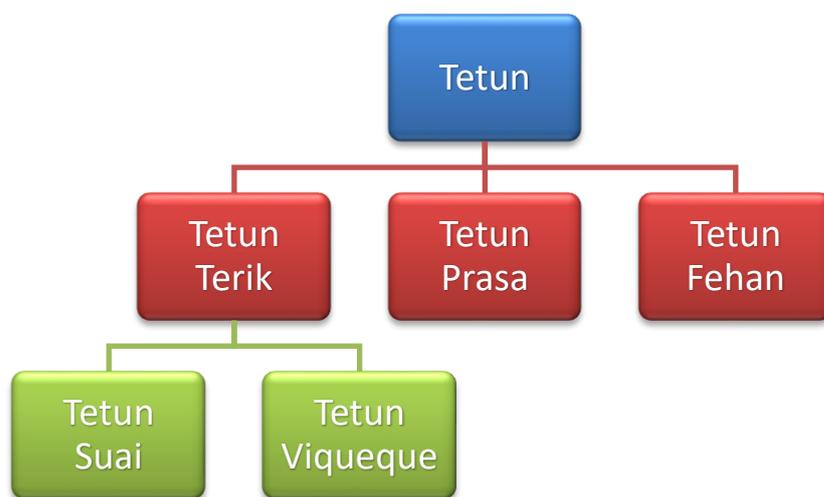
Figure 3 Division of Tetun used in this thesis

To summarize, I will use the word Tetun to refer to the official language of Timor-Leste in general, especially in contexts where division between varieties in not important, or when certain information is related to the language as such, especially in the historical context, considering we can only really speak of Tetun Prasa after 1769 when the capital was moved to Dili. In all other cases, in which language contact caused variation between the Tetun varieties, I will make the necessary distinction.

One last distinction that should be made is between low and high registers of Tetun. My oral corpus can be characterized as casual, informal speech. On the other hand, my written corpus, consisting of newspaper articles and official press releases, can be classified as high register. Press Tetun was described by Williams-van Klinken (2002:5)

as controlled by the modern educated elite which also means extensive Portuguese influence. Portuguese loans often make up to 50% of the word count and, although most of them are widely known among the public, many are still not understood by speakers who are not proficient in Portuguese. Apart from the borrowed lexicon, Tetun used in the press often makes use of Portuguese structures, too, e.g. Portuguese plurals, and number and gender agreement in NPs of Portuguese origin (see sections 6.1.2 and 6.1.4). The fact that the media resort to Portuguese loanwords more often is caused by various factors: most journalists were educated during the Portuguese or Indonesian times and are familiar with technical vocabulary, and, at the same time, the use of Portuguese loans allows them to avoid Indonesian borrowings. Portuguese is preferred over Indonesian – it was the language of the resistance movement and it is the language of the Timorese elite. The biggest issue here is, however, the lack of qualified translators since newspaper articles are, often literally, translated from Portuguese, English and Indonesian (Williams-van Klinken 2002:5-6). In the last ten years, there has been a certain status quo in the Timorese daily newspapers: news covering national affairs is published in Tetun, but articles on international affairs, sport, economy and culture are still taken from Portuguese, Indonesian or English media in their original language.

Apart from Press Tetun, there are two other high registers: Ritual Tetun Terik and Church Tetun. Church Tetun in used by the Catholic church in oral and written forms, especially on formal occasions (e.g. masses, funerals). According to Williams-van Klinken (2002:3), it emerged in the 19th century with the establishment of the *colégio de Soibada* in 1898, located in a Tetun Terik-speaking area, which has heavily influenced its form. It shows preference for Tetun Terik forms although it has incorporated many Potuguese borrowings. From the structural point of view, Church Tetun uses simple juxtaposition for adnominal possession and has retained the native phoneme /w/ and the glottal stop. Interestingly, Church Tetun does not employ subject marking on verbs. Ritual Tetun Terik, which can be characterized by poetic parallelism, special vocabulary (honorifics) and pervasive metaphor is, on the other hand, only used by the traditional ritual specialist *Na'i Lia*, or *Lia Na'in* (Williams-van Klinken 2002:2-3). Before the 19th century, Timorese literary culture was purely oral, and these specialists had the skills to tell stories, recite poetry and relate long verses about various subjects, straight off their heads. They would recite for hours, repeating important points as often as possible so the

message of the poet would get through. In villages, the *katuas* (older men) would tell stories to the children to lead them to adopt the behaviour that is expected of their clan and to know their place in life (Morris 2003).

2. SOCIOLINGUISTIC HISTORY

The history of the island of Timor prior to the arrival of the Dutch and the Portuguese is still very unclear. Very few documents were preserved from the earliest days of the Portuguese administration, mostly due to a big fire in 1866 which spread across the capital Dili all the way to the *Palácio do Governo* where the documentation was stored (Oliveira 1950:122-123). The same applies to the history of Tetun. We can only hypothesize where the language originated and how it spread across the whole eastern part of the island, becoming the lingua franca and ultimately the official language of the first democracy of the 21st century – the Democratic Republic of Timor-Leste (Thomaz 2002).

In this chapter, I will look at the earliest mentions of Timor before the arrival of the Europeans (section 2.1) as well as the Portuguese presence in Southeast Asia prior to the colonization of Timor itself (section 2.3). I will describe the first Portuguese settlements in Lesser Sunda (section 2.4) and the subsequent establishment of Lifau as the capital of the Portuguese colonial administration in Timor (section 2.5). In 1769, the capital was moved to Dili and in 1859 the border between the Portuguese Timor and the Dutch East Indies was formally declared after the Treaty of Lisbon (section 2.6). The last 200 years of the Portuguese presence and their attitude towards the language policy are described in section 2.7. Section 2.8 deals with the sociolinguistic situation during the Indonesian occupation and I will also discuss the language attitudes and the language policy in the post-independence times (section 2.9). The last section 2.10 is dedicated to the discussion on the official orthography of Tetun Prasa, its history and development, as well as other competing orthographies of Tetun still used in various registers.

2.1 First accounts

The first mentions of Timor can be found in Chinese and Malay works, although they are usually very scarce. These were mostly made by traders coming to Timor, possibly as early as the rule of the Tang dynasty (618 to 907 A.D.), long before the Portuguese arrived (Thomaz 1974:244). The Chinese businessmen mostly established

ports on the western and north-western coast of Timor but artefacts of early Chinese presence were also found in the east. According to Morris (2003), the early Chinese presence can be documented by the indigenous lunar calendar which is similar to the Chinese, by the Timor pony which has Asian origins and existed in Timor before the Portuguese, and by the musical instruments which are Asian in design and sound. There is also a Chinese record of the *liurai*[5] of Besa Kama (the old Belu capital) who "paid a yearly tribute to China before the Portuguese Dominicans were on the scene in 1566" (Morris 2003).

The first Malay account dates back to the 14th century and there is a lot of linguistic evidence that the Malay traders had a significant influence on Timorese languages in the following centuries. Malay spread as a trade language across eastern Indonesia at the beginning of the 15th century when the traders started to visit the islands of Tidore, Ternate and Ambon to buy spices. Later in the 15th century, these traders coming from the Central Moluccas turned their attention also to Timor in search of sandalwood and beeswax. They spoke a restructured variety of Malay, namely Ambonese Malay, which is still spoken currently (Hull 2005a:84). When the Portuguese arrived in Ambon, they found that the local Malay variety had spread among the linguistically diverse population as a lingua franca. As Adelaar and Prentice (1996:683) put it: "it was a literary language and a language of religious instruction in Muslim communities, and it was a contact language between locals and foreigners (Malay, Javanese and Macassarese) spice traders." When the Portuguese arrived in Ambon, they also used Malay to preach Roman Catholicism, however, when the Dutch took over the Moluccas, they replaced Catholicism with Calvinism but continued to use Malay in churches and Protestant schools (Adelaar & Prentice 1996:683)[6].

2.2 Timorese society before and after the arrival of Europeans

The original inhabitants of the western part of the island were Atonis, a Melanesian people who still form the majority in West Timor and speak 'Timorese'

---

[5] Traditional Timorese ruler.
[6] For more discussion on the earliest literature, see 4.2.1 and 4.2.2.

(now officially called Uab Meto, or, pejoratively, Dawan). Together with the Belu people, who spoke Tetun, they formed a single Waiwiku-Wehale monarchy. Pigafetta, who visited Timor in 1522, confirmed the importance and spiritual supremacy of this kingdom.

At the beginning of the 18th century, the island was divided into two spheres of the influence: the land of Bellos, ruled by the Belu people in the east, and the land of Servião, ruled by Sonba'i[7] in the west (see Map 3). These were then further divided into smaller kingdoms, ruled by local *liurais*. Although they spoke different languages, all kingdoms in the west recognized the supremacy of the Sonba'i rule. The eastern kingdoms, on the other hand, were often fighting each other and there was very little political cohesion (Villiers 1985:576). However, their attitude towards the Portuguese was different. The people from Servião were hostile towards the Portuguese administration but the Belus were faithful (Sá 1961:232). The two parts of the island were also divided by religion: Atonis were Protestants and Belus were Roman Catholics.
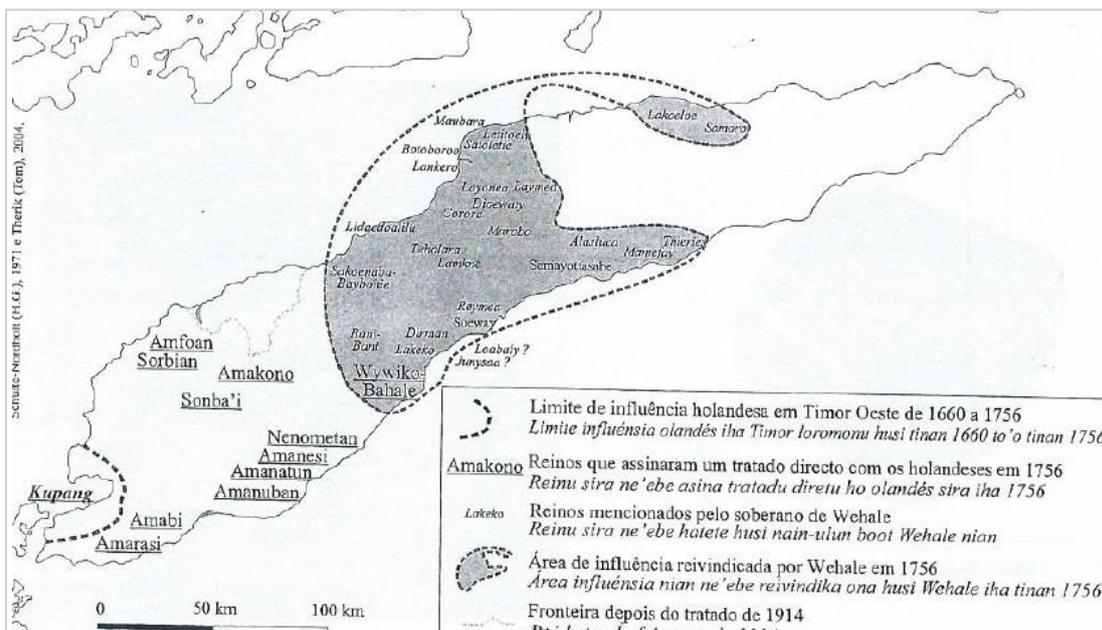
Map 3 Waiwiku-Wehale kingdom in the 18th century (Durand 2010:44)[8]

The society was based on a class system: there were the cattle keepers (*lutun*), the slaves (*atan*), the common people (*ema rai*), and the nobility and royalty (*dato*). Inherited

---

[7] It was an Indonesian dynasty that ruled over various parts of West Timor until the mid-20th century. Sonba'i allied with both the Portuguese and the Dutch with frequent attempts to break independent from the colonial rule.
[8] Legend: Limits of the Dutch influence in West Timor between 1660 and 1756; Kingdoms that signed the direct treaty with the Dutch in 1756; Kingdoms mentioned by the ruler of Wehale; Area of the influence demanded by Wehale in 1756; Border after the Treaty of 1914.

upper class positions included doctors (*matan dook*), sorcerers (*buan*) and keepers of sacred relics (*malulik*), kings (*liurai*) and warriors (*asu'uain*). It was the *liurais* who dealt with the foreign traders. They gained more power with the sale of sandalwood, which, consequently, led to long and bloody wars. As they wanted more and more land that grew sandalwood, their greed led to conflicts with other kingdoms as well as to rebellions against the colonial masters. In the 16th and 17th centuries, the Timorese *liurais* had a reputation of being very violent and warlike (Morris 2003).

Regarding the spread of Tetun itself, several hypotheses have been proposed, some more plausible than the others. Thomaz (1974:212-213) rejected both of the earlier hypotheses: firstly, that Tetun was diffused by missionaries from the Soibada mission[9] founded in 1898, and secondly that Tetun was diffused as a lingua franca from Dili when the capital was transferred there from Lifau in 1769. Neither of these hypotheses bears close examination: there is evidence that Tetun started to diffuse well before 1898[10] and that Dili and its rural suburbs used to be a Mambae[11]-speaking region. Thus, Thomaz (2002:72-73) formulated his own hypothesis, according to which it was the tribe of Belus that spoke Tetun and then expanded from its original area prior to the 17th century and came to dominate the eastern part of Timor. The Belu people thus became a noble social class and Tetun spread over the territory they ruled. When the Portuguese arrived in the 16th century, Tetun was already of great importance.

2.3 Portuguese presence in Southeast Asia

The Portuguese established themselves in Southeast Asian region with the capture of Malacca in 1511. The influence of Malacca in Southeast Asia was undeniable; however, the relations with Timor were not that intense. The ships would stop in Timor in search of sandalwood, but they were more interested in clove and nutmeg from the Moluccas and Banda islands. This was due to the fact that Europe was already getting

---

[9] The town of Soibada lies in the central Timor-Leste, in an area where Tetun Terik is spoken as a first language.
[10] In 1889, Silva (1889:i) already commented on the fact that Tetun was the most wide-spread language spoken in the Portuguese Timor.
[11] Mambae is a native Austronesian language that is still spoken outside of Dili all the way to the south coast of Timor. It is the second largest ethnic group in Timor-Leste, right after Tetun.

sandalwood from India, from the area around Mysore and thus the main consumer of the Timorese sandalwood was China. At the beginning of the 16th century, the sandalwood trade was mostly in the hands of Malaccan merchants, Portuguese, and Asians who carried out business directly with the Timorese ports. The goods were shipped via Malacca and then reexported to China and India (Villiers 1985:595). But from the second half of the 16th century, the trade was conducted directly from China (Thomaz 1974:224, 2002:86).

When the Portuguese were expelled from Malacca by the Dutch and pushed out of Moluccas where they faced indigenous revolts, they disembarked in Macassar, Sulawesi. The sultanate had a very favourable attitude towards foreign merchants and maritime trade and the city became a commercial entrepôt in the early 17th century (Souza 1986:88-92). Later, the trade was diverted to Macau.

From the mid-16th century until the 18th century, Portuguese was a lingua franca in the whole East. It was spoken by all missionaries regardless of their nationality. Most of them learned this language in Batavia (present-day Jakarta) where the Catholic community numbered around 4000 people in 1713 (Menezes 1992:228-229). This is interesting because the Portuguese were never really interested in conquering the hinterlands of ports they occupied. They controlled the coastal fortified settlement but never tried to establish an empire on land. In the Indonesian archipelago, the only control over ports they had was in Tidore, Ternate, Ambon and Solor. Their main intention was to gain control over the Indian Ocean by occupying the strategic points (like Malacca and Macau) which gave them advantage in the competitive Asian trade (Kartodirdjo 1970:176-180).

2.4 Portuguese settlements in Lesser Sunda

Portuguese presence in the Timor region can be divided in four distinctive phases. The first one started when the Portuguese landed and built settlements in the Lesser

Sunda[12] region. Until 1556, their presence was exclusively commercial (Thomaz 1985:318). These Portuguese establishments had usually three main forms: *feitorias* 'trading posts', *fortalezas* 'fortified strongholds' and *cidades* 'urban settlements'. However, only two establishments in the SE Asia in the 16th and 17th century could pride themselves on having the status of *cidades* – Macau and Malacca (Baxter 1996:300).

The second phase can be referred to as commercial and religious (Thomaz 1985:319). In 1561 three Dominican priests (Fr. António da Cruz, Fr. Simão das Chagas, and Fr. Aleixo) left Malacca to establish the first missions in the Lesser Sunda (Sá 1961:IX). They were based on the islands of Solor, Timor and Ende[13] where they started their mission of evangelization. In 1566, the Portuguese built a fort in Solor, but it was not inhabited by Portuguese nationals. The Dominicans established a seminary there teaching Portuguese and Latin and in 1606 they also established a seminary in Larantuka to teach Christian students (Baxter 1996:311). The population was made up of the offspring of Portuguese soldiers, sailors and sandalwood traders who came from Malacca and Macau and intermarried with native or 'creole' women[14].

These people of a mixed descent became to be known as Topasses[15] (Boxer 1947:1). This was typical for all Portuguese settlements in Asia in the 16th and 17th

---

[12] The Lesser Sunda islands include Bali, Lombok, Sumbawa, Flores, Sumba, Timor, Alor archipelago, Barat Daya islands and Tanimbar islands, divided into four Indonesian provinces and the independent Timor-Leste.

[13] Dominicans and Jesuits divided their regions of influence. While the Dominicans settled on the islands of Lesser Sunda, the Jesuits founded their missions in the Moluccas, including Ambon island (Teixeira 1961a:193).

[14] There were almost no Portuguese women accompanying the Portuguese soldiers and merchants on their voyages. The limited number of these European women were outcasts (convicts or prostitutes) who were shipped to India. So the Portuguese maintained relationships, often out of wedlock, with local native women, daughters of the converted natives or they married Eurasian women, the offspring of mixed marriages, who had some European blood, social, political and economic status and were raised in homes with Christian religion and Portuguese language (Tomás 2009:54-56).

[15] According to Dalgado (1936:346), this "term was employed in the seventeenth and eighteenth centuries as synonymous with mestizo to denote those who claimed to be Portuguese descendants, spoke Portuguese, affected the Portuguese style of dress, professed the Catholic faith and served ordinarily as soldiers in the army". The origin of the word has been widely discussed and there are numerous explanations for its etymology. Some people believe the word was derived from Hindustani *topi* 'hat' or *topivala* 'one who wears hat', which "used to be a distinguishing mark, at times honourable, at others opprobrious" (Dalgado 1936:346-347), or from Turkish/Persian/Hindustani *topchi* 'a gunner'. The most plausible, however, is the Dravidian *tupassi* (Boxer 1947:1), Tamil *tuppási* < Neo-Aryan *dubhāśí* or *dobāśí* < Sanskrit *dvibhāṣya* 'bilingual, interpreter'. Topasses spoke two languages and originally served as interpreters, hence the name. The Portuguese would also refer to the local interpreters in the East as *línguas* (Sá 1961:261). Topasses in East Nusa Tenggara were also known by the Portuguese as *Larantuqueiros*, named after the city of Larantuka, and as *Zwarte Portugesen* 'Black Portuguese', who were Christian, spoke no Dutch and there was no Portuguese blood in their veins, by the Dutch. They were later also known as *mardijkers* or *mardikas*

century. Since the lack of manpower was a continuous problem for the Portuguese crown, the population of these establishments usually consisted of a couple of Portuguese nationals, so called *reinóis* (either priests, merchants or possibly soldiers), *casados* (Portuguese married to indigenous women), many *mestiços*, native Christians and slaves. Some of these establishments, just like the one in Solor, did not have almost any ties to Portuguese administration and were often under loose Portuguese jurisdiction (Baxter 1996:300).

When the Dutch captured Solor in 1613, Larantuka (on the island of Flores) became the centre of the Portuguese influence. It continued to be the headquarters of the Portuguese throughout the 17th century, although the main port for the sandalwood trade was Lifau in Timor (Boxer 1947:7). At the beginning of the 17th century, more priests were asked to come and serve on these islands. Only a few of them were Dominicans originating from Portugal; many of them came from India and some were of mixed descent (Teixeira 1961a:193). The priests found it difficult to convert the local population. In a letter to a provincial in India, Father Baltasar Dias expressed his thoughts on the language and people of Timor: "The language is somewhat similar to Malay. (...) The people of Timor are the most stupid found in these parts. They worship nothing, nor do they have idols and they do everything the Portuguese ask them to do." (Teixeira 1961a:390).

Since the priests received very little support from the government of Portuguese India, they got involved in business, for which they were resented by the Jesuits. To make ends meet, their missionary work went hand in hand with the sandalwood trade for the next 150 years. The more successful their missions were, the more profitable the trade became (Villiers 1985:573-590). However, their big rival was the Dutch East India company (Vereenigde Oost-Indische Compagnie – VOC). In 1659, it offered the Portuguese communities on Solor, Flores and Timor islands a localized truce, but the Portuguese rejected it, and, in return, the VOC attacked the Portuguese in Makassar and successfully eliminated their commercial competition (Souza 1986:111).

'men exempted from taxes' although this name was later applied to all 'foreign Christians' (Tomás 2009:60).

2.5 Portuguese settlement in Lifau and the Topasses

There was still no permanent Portuguese settlement on the island of Timor in the 17th century although there were 22 churches and ten Dominican missionaries stationed here in 1640. In 1641, Father António de S. Jacinto baptized the queen of Mena kingdom and this event marks the 2nd period in the history of these islands. A year later, António Dias, the representative of the Portuguese government, together with his army, the queen of Mena, kings of Lifau and Amanubau attacked and defeated the king of Servião (Teixiera 1961a:96-97). Soon the missionaries managed to convert more kings and queens to Catholicism and their faith gained a more prominent role. In 1646, a Vicar General of Timor was named.

But the Portuguese still had a difficult position trying to settle down in Timor. They had to fight the Topasses, the local kings but especially the Dutch. In 1653, the Dutch captured the Portuguese fort of Kupang and turned it into their headquarters. They also captured the fortress Panakoekang in Makassar and made the local king sign a treaty, according to which he had to expel all the Portuguese from his country: 110 of them went to Siam and Batavia, 530 to Macau, and 120 landed in Timor (Teixeira 1961b:99-100). But when an English navigator, William Dampier, passed by Timor in 1699, he wrote: "The residents of Lifau speak Portuguese and are Catholics. (…) However, I did not meet more than three white people and two of them were priests." (cited in Durand 2010: 60-61).

The third phase of the Portuguese presence started with the first permanent settlement in Timor which was founded by António Coelho Guerreiro in 1702. The seat of the government was transferred from Larantuka[16] to Lifau (Boxer 1947:9) and the Portuguese presence was thus not only commercial and religious, but also political (Thomaz 1985:318). Guerreiro, who served as a Portuguese colonial administrator between 1702 and 1705, considered his main and first task to win over the local kings (*liurais*), so he adopted one of the colonial assimilation strategies: convert the *liurais* to the Catholic faith, and give them Catholic names and Portuguese aristocratic and military titles. This way, the *liurais* would swear loyalty to the Portuguese crown, which, at the

---

[16] Larantuka was part of the Portuguese colonial empire until 1859 when it was ceded to the Dutch. It is believed that part of the local population left for Timor (Baxter 1996:311).

same time, would let them rule over their people in the traditional way (Hull 1999b:60). However, he also had to face the Topasses, who were known for their constant conflicts with the Portuguese administrators and refused to acknowledge the Portuguese rule. The local kings also seemed be more loyal to the them than to the Portuguese (Villiers 1985:576).

The Topasses were represented by two families: de Hornays and de Costas. The two families fought each other for power and supremacy, until the mid-18th century, when they reached an arrangement in which the power would rotate between them.

But the Portuguese crown was not directly involved in the governance of the fort in Lifau, which was, in fact, built by Macau. Macau had always been very active in the sandalwood trade and by the end of the 17th century, the colony's main source of revenue was sandalwood, wax, gold and slaves (Villiers 1985:596). It would buy sandalwood from the Lesser Sunda islands and then resell it in Makassar and Batavia, or bring it back to Macau to sell it there. By the end of the 17th century and at the beginning of the 18th century, Timor was the only profitable market in the South China Sea since the Macau traders managed to minimize the penetration of the VOC and China into the Timor market (Souza 1986:181). In 1716, it was prohibited by the senate that the Macau boats buy sandalwood from other islands in 'Insulindia', the region of maritime Southeast Asia. They could only go to Timor since they had a monopoly on sandalwood there. As a consequence, a significant number of people from Macau lived in Larantuka and there were matrimonial relationships between the locals and the people from the colony (Fernandes 2000:13).

Soon, the Macanese felt the need to take Lifau fully under their control. However, this decision was met with resistance from the local Topasses. And so, the locals led by Francisco de Hornay joined forces with other *liurais* and decided to drive the Portuguese out once and for all (Boxer 1947:13).

Not surprisingly, the number of the Portuguese living in Timor has always been very low. The Portuguese community mostly consisted of a couple of civil servants, military, missionaries and outcasts (especially military men). The main reason was a huge fluctuation. Businessmen would only stay for the dry season (end of March to end of August) and since they were only interested in the sandalwood trade, they didn't feel the

need to settle down nor explore the interior of the island. D. Pedro Miguel de Portugal, the viceroy of India, affirmed that in 1750 there were only 7-8 Portuguese on the island, not counting the numerous missionaries, whose presence was, as I mentioned above, also not widely efficient (Miranda & Serafim 2001:242-244). But their mission wasn't entirely in vain. In 1747, the Dominicans founded a seminary in Manatuto, which was during the first centuries of the Portuguese presence in Timor the only educational institution (Matos 1974:194). Vaquinhas (1883) also commented on the relatively high literacy rate among the local chiefs and their auxiliaries, which he attributed to the good work of missionaries: "Do que não resta duvida alguma, é que nos annos de 1680 até 1760 todos os regulos e seus principaes, com pequenas excepções, sabiam bem ler e escrever; instrucção esta ministrada pelos verdadeiros e bons missionarios (…)."[17] (Vaquinhas 1883:280).

2.6 Portuguese Timor

The Portuguese, led by Antonio José Telles de Menezes, abandoned Lifau and on the 10th of October 1769 landed in Dili:

"(...) in this unhealthy and malarial site, which had nevertheless the advantage of being pretty far removed from the area controlled by de Hornay, da Costa, and the Toepassen around Lifao and Larantuka, he founded the new capital of Portuguese Timor where it has remained (at times somewhat precariously) ever since." (Boxer 1947:16).

According to Hajek (2000b:402), one thing that the Portuguese took with them was Tetun, which ended up replacing the local Mambae in Dili and its surroundings. In most places around the world, where the Portuguese built their settlements, Portuguese became the lingua franca. As Sá (1961:XVI) put it:

"Nestes centros de presença lusitana, o português penetrou nos dialectos locais, adaptou-se-lhes convenientemente, e, assim, mais ou menos acrioulado, acabou por fixar-se aí como língua franca da localidade. São exemplos Goa, Damão e

---

[17] "What does not leave any doubts is that between 1680 and 1760 all the kings and their chiefs, with small exceptions, could read and write well, and this instruction was ministered by truthful and good missionaries (...)." (my translation)

Diu; Cochim, Ceilão e Pegu; Malaca e Macáçar; Solor e Timor; Amboino, Ternate e Macau."[18]

Indeed, the population of Solor and the Topasses of Timor did speak some kind of a creole Portuguese. And when the capital was transferred from Lifau to Dili, two distinct groups of the *moradores*[19]*,* who were native Timorese, moved with it and formed these communities: one that originated from Flores (Sikka community) and one that originated in Solor (Bidau community) and spoke their own version of Creole Portuguese (Baxter 1996:311). Also, based on the observation of François-Etienne de Rosily, a French captain, in 1772 the local elites (consisting of 30 *liurais* of the neighbouring kingdoms to the Portuguese administrative base) spoke Portuguese (Thomaz 1985:320). The Portuguese language was especially prevalent in three areas: church, administration and military (Baxter 1996:312).

But the colonization of Timor was slow, with a small amount of Portuguese willing to come and settle down. Garcia (1901:9) explained why:

"Para colonizar Timôr, seria difficil estabelecer uma corrente emigratoria dirigida directamente da metropole. São demoradas e dispendiosas as communicações, desconhecida quasi por completo a colonia e ha falta absoluta de incentivos efficazes, que interessem o emigrante para demandar terras tão longiquas."[20]

It made more sense for the Portuguese from Macau to take part in the administration. The province of Macau and Timor was created by the Decree of 20th September 1844, but to the dismay of Macau. Timor was never made an integral part of Macau and was always seen by the East Asian colony as a poorer cousin who took

---

[18] "In these centres of Portuguese presence, the Portuguese language penetrated into the local dialects, adapted itself to them conveniently, and, thus, as a more or less creole-like language became stable as a lingua franca of the area." (my translation)

[19] Originally, *morador* used to describe the Portuguese, mostly traders, who settled down on the island. Later, it was used for the local soldiers, who were not part of the military but paid various services to the Crown. For example, they guaranteed order and maintained certain surveillance over the offices and residences of the authorities (Sá 1961:184).

[20] "It would be difficult to colonize Timor and establish a migration flow directed directly from the metropole. The communication is costly and lengthy and it is almost completely unknown to the colony. There is absolute lack of effective incentive that would interest an emigrant to demand to go to such far-away places." (my translation)

advantage of its revenues (França 1897:209-212). The reservations of Macau were supported by a British explorer Alfred R. Wallace:

> "The Portuguese government in Timor is a most miserable one. Nobody seems to care the least about the improvement of the country. And at this time, after three hundred years of occupation, there has not been a mile of road made beyond the town [Dili], and there is not a solitary European residence in the interior." (Wallace 1869: 307).

The fourth and the last phase of the Portuguese presence in Timor started with the administration of governor José Celestino da Silva (1894-1908), who led violent but victorious campaigns against the kingdoms in the interior of the island. The local kings lost some of their powers and were subjected under the colonial administration (Thomaz 1985:320). Subsequently, the Portuguese rule decided to impose greater political and military rule over the whole territory and also transformed the originally subsistence economy into an agricultural export-oriented system, which caused many localized uprisings (Cabral & Martin-Jones 2008:152).

Menezes (1992:232) recounted a story that he witnessed in Dili port in 1969 when a Dutch ship arrived from Singapore and nobody on board could communicate with the customs staff. Luckily, there was an officer from the Porto brigade who was a Capeverdian. He spoke the Capeverdian Creole and the sailors replied in the 'dialect' of Malacca. In the end, they could easily understand each other and ended up going for a walk in the city together. At that time, Portuguese wasn't the lingua franca of the region anymore – it had been replaced by Malay. However, local varieties of Portuguese creoles were still alive.

2.7 Language policy in the Portuguese times

Although the Portuguese crown never paid much attention to educating the Timorese population and to developing the colony, some saw Timor as the greatest Portuguese colonial achievement. "Conquistada não pelas armas mas pela cruz, não pela violência mas pela caridade dos missionários de Portugal, Timor é a sua conquista mais

gloriosa."[21] (Fernandes 1931, cited in Menezes 1992:222). But even after centuries of the Portuguese presence in Timor, their attempts to introduce their language were minimal and concentrated only on the Timorese elite (for example, between 1859 and 1863 a school was built exclusively for sons of *liurais*).

The Church was the main institution of education operating several primary schools, seminaries and teacher training institutions (Nicolai 2004:42) but the attempts of the Church were later halted by the dissolution of religious orders at the end of the Portuguese civil war in 1834, and the use of Portuguese decreased. The Dominican seminaries and convents disappeared and the clergy was decimated to two Goan priests (Thomaz 1985:320). It was not until after 1874 that the missions in Timor were reorganized under the Diocese of Macau. After that, numerous schools and churches were newly built (Baxter 1996:312).

On the other hand, the Portuguese used Malay as a trade language in Timor until the mid-19th century, which is documented by the numerous Malay borrowings, many of Chinese and Dutch origin, in Tetun and other local languages. The Portuguese language became Timor-Leste's second language only after 1860[22], when the Portuguese government in Dili developed into a real colonial administration. Twenty years later, Malay could be hardly heard in Portuguese Timor (Hull 2005a:85-86). This is how Anna Forbes commented on the lack of Malay in Timor in 1887 (cited in Hull 1999b:58): "It is strange to hear no Malay in Timor. This language is heard otherwise all over the civilized archipelago; but natives here must learn the language of the [Portuguese] possessors if they will have any contact with them."

After the Treaty of Lisbon of 1859, the new Portuguese overseas language policy was to introduce the Portuguese language as a lingua franca. This however, required an organized effort, which was present in the former Portuguese colonies in Africa but was absent in Timor. The Portuguese administration was business-oriented, and it was the Catholic church that was responsible for the dissemination of the faith and the Portuguese

---

[21] "Conquered not with arms but with a cross, not with violence but with charity of the Portuguese missionaries, Timor is their most glorious conquest." (my translation)
[22] In 1859, the Portuguese and the Dutch signed the Treaty of Lisbon, which established the border between the two colonies: Portuguese Timor and Dutch West Indies.

language, although most of the missionaries used Tetun as the language of catechism and prayers.

In 1898, the Church founded a *colégio* in Soibada, the first secondary school in Timor, which could be considered one of the most important events in the cultural history of Timor in those times. It was run by the Jesuits until 1910 and it was aimed at educating professors and catechists who would teach religion, but also give instruction in reading and writing. Most of the students went on to become part of the cultural elite, either pursuing a career of a catechist or in public administration (Thomaz 1985:321). By the beginning of the 20th century, there were 20 schools across Timor providing education to the locals, and the Portuguese language was firmly established in Dili (Baxter 1996:312).

However, until 1915, there was no systematic planning in education, which was still only accessible to a minority – the children of *liurais* and catechists. The idea was to educate a native Timorese elite that could serve in the colonial administration. It was only between 1916 and 1938, during the administration of Filomeno da Câmara, Teófilo Duarte and Álvaro Fontoura, that there were attempts to adapt the education to local conditions and necessities, aimed at general and professional agricultural education. In 1912, Filomeno da Câmara suggested that the primary education be carried out in Tetun. He even ordered that the educational manuals be printed out for the students not only in Tetun but all local languages which would allow children to learn how to read in their native languages and lead to an easier acquisition of Portuguese. Based on this direct appeal of Câmara, Manuel Mendes Larangeira published in 1916 (reprinted in 1932) *Cartilha-Tetun* (see section 4.2.3.4) (Cardoso 2017:12-13). But the mother-tongue based education was never implemented – the language of instruction continued to be Portuguese (Thomaz 1985:321). In 1915, the first official school was opened in Dili, followed by many others across the island, although they were never as widespread as missionary schools. Unfortunately, during the Japanese occupation (1942-1943), the majority of schools were destroyed.

After World War 2, but especially in the 50s and 60s, Portugal started paying more attention to the social development of its colony. Their main objective was to 'civilize' the local population by introducing the Portuguese way of life, including a proper command of Portuguese:

"The 2nd fundamental goal of our struggle in education is: that everyone has to speak Portuguese! If there are prayers? Pray in Portuguese. If there is discussion, discuss in Portuguese. If there is a real need to curse, then curse in Portuguese! If we need to understand each other, then let's understand each other in Portuguese!" (translation of Grade 1973:219 cited in Hajek 2000b:403).

The change of policy was also connected to the anti-colonial moods in the Portuguese colonies in Africa, which made Portugal fear that the same sentiment might start developing also in Timor (Hajek 2000b:402-403).

New primary schools and *internatos* were founded all across the country and Portuguese was strongly enforced even outside of the classroom. All education was done in Portuguese, although the majority of teachers were Timorese (Thomaz 1977:536). First secondary schools started to emerge in the 50s – in 1952 the famous *Liceu* in Dili was founded, followed by the seminary in Dare and a technical school in Dili in 1965. However, it was only in the last years of the Portuguese colonization that the number of students in schools increased from 28% in 1970-71 to 51% in 1972-74 to 77% in 1973-74. According to the population census in 1970, 90,8% of the population older than 10 years old was illiterate (Thomaz 1985:321-322).

An important role was played by the *mestiço* population. Back in the 16th century, Portuguese crown introduced an official policy in Portuguese Goa whose aim was the production of mixed-race offspring between European men and native women, which would result into *mestiço* population. Although we can't speak of such a policy in the 20th century, the *mestiço*, or Luso-Timorese population became the ruling class of Timor and often adopted Portuguese as their first language (Baxter 1996:314). In 1970, there were 1939 *mestiços* living in Timor (Thomaz 1977:539). On the contrary, those without European ancestry, but who adopted Catholicism, Portuguese language, European names and thought of themselves as Portuguese, were called *assimilados* 'assimilated' (Hull 1999b:58-59).

Even so, as Sá (1961:XX) concluded, the Portuguese effort was not all that successful: "Mas não vá pensar-se, por isso, que hoje, passando quatro séculos de domínio

nosso, Timor fale o português. Não!"[23] The author further discussed the two reasons why the locals did not feel the need to learn the language of the colonizer: they did not fear Portuguese power and they did not give in to the Portuguese attempts to 'civilize' them (Sá 1961:XXI), although by the end of the colonial rule, around one quarter of the population had been converted to Catholicism.

At the same time, there has always been widespread multilingualism (villagers also often spoke the languages of neighbouring settlements) and the spread of Portuguese could never disrupt the multilingual repertoire of the speakers; on the contrary, it would be an expansion of it. Indeed, there is no evidence that the introduction of Portuguese put local languages in peril or that locals would shift exclusively to Portuguese (Hajek 2000b: 401-402).

This partly changed after the events of 1974-1975, when a rapid political change in Portugal (also known as the Carnation Revolution) led to its gradual withdrawal from the Portuguese overseas colonies[24]. In April 1974, Timorese political parties were legalized. And just like in other former colonies, many new political parties sprung up and fought for power. The party that came out victorious of this short civil war in 1975 was Fretilin (*Frente Revolucionária de Timor-Leste Independente* – Revolutionary Front of the Independent Timor-Leste), joined by military forces of Falintil (*Forças Armadas da Libertação Nacional de Timor-Leste* - The Armed Forces for the National Liberation of East Timor). Fretilin was a huge advocate of Tetun as a national language and, as is further discussed in section 2.10, it had the ambition to consolidate Tetun orthography and promote literacy in Tetun. In 1975, they produced a handbook *Rai Timor, Rai ita nian* (The Land of Timor is Our Land) which incorporated some ideas about literacy and teaching approaches (Cabral & Martin-Jones 2008:156). According to Hajek (2000b:404), there were local volunteer programs, that would target young and older rural population using simple Tetun readers.

---

[23] "But do not go thinking that today, after four centuries of our dominance, Timor speaks Portuguese. No!" (my translation)

[24] On June 17th, 1975, Portugal issued a decree about the decolonization of Timor-Leste which foresaw the election for National Assembly for October 1976 (Durand 2010:116-117).

2.8 Sociolinguistic situation during the Indonesian occupation

Unfortunately, most of these projects came to a halt with the Indonesian invasion in December 1975, although "during these first years of occupation, Fretilin continued to run schools and literacy programmes while hiding in the mountains and forests" (Nicolai 2004:43). The Suharto government justified the annexation by arguing that the left-wing Fretilin was trying to turn Timor-Leste into a communist state and that the independence tendencies might spread over to other parts of Indonesia (for example, West Papua). These arguments won Indonesia support of part of the international community, including the USA, and on December 7, 1975 Indonesia invaded Timor-Leste. This led to an extremely violent offensive and it took Indonesians only a few months to capture some of the major towns, including Dili. By the 1980s, Fretilin and Falintil forces suffered huge casualties and went into hiding. They created a clandestine resistance movement with its first leader Xanana Gusmão. The international community supported the invasion (Australia, Japan, Canada) and/or supplied weapons (USA, UK) to Indonesia, opposing the UN resolution of December 12, 1975 which condemned the invasion.

The 24 years of Indonesian occupation did not only cost the lives of one third of the population but also had an impact on the sociolinguistic situation. Indonesia had a clear objective, namely fast cultural, political and linguistic assimilation. Before the invasion, only a third of the population was Catholic while the rest preserved their animist beliefs. However, the Indonesian government officially recognizes only six religions and animism is not one of them. After mass conversions[25], now more than 90% of the population reports to be Catholic. At the same time, the Catholic church, the only Timorese institution left, was a constant critic of the Indonesian invasion. The Timorese turned to the Church for protection and some Timorese priests even joined the resistance fighters in the jungle (Steele 2000).

The use of Portuguese was prohibited while the local languages were 'tolerated'. Indonesians made sure that everyone learned their language as soon as possible and in

---

[25] According to the 1970 census, only 25% of the population of 610,000 were Catholics (Thomaz 1977:499). Based on the Indonesian 'Blasphemy Law' adopted in 1965, it is required that all citizens identify themselves with one of the six official religions: Islam, Catholicism, Protestantism, Hinduism, Buddhism and Confucianism. Until 2006 it was obligatory to list an official religion on identity cards (*Kartu tanda penduduk* – KTP) (Crouch 2012:4). Atheism or animism are not recognized, and blasphemy is illegal.

order to achieve that, they invested a lot of money in education and new schools. While the Portuguese introduced a Western colonial model of education and, until the 70s, focused only on educating the elite, Indonesians made education available to masses and used it as a tool to conquer the Timorese (Nicolai 2004:30). Indonesians focused on the quantity and primary education became obligatory for everyone. Indonesian language was the only language of instruction (English was taught as a foreign language) and teachers were brought from Indonesia. Although the quality of the education was rather low, Indonesia introduced the concept of *Pendidikan untuk Semua* (Education for All) and by 1985 almost every village had a primary school (Nicolai 2004:43-44), as shown in Figure 4[26].

Access to education	1975	1999
Number of primary schools/number of children attending	47 primary schools 10,500 students	788 primary schools 167,181 students
Number of junior secondary schools/ number of children attending	2 junior secondary schools 315 students	114 junior secondary schools 32,197 students
Number of senior secondary schools/ number attending	None	54 senior secondary schools 18,973 students

Figure 4 Access to education during Portuguese and Indonesian times (United Nations 2000 cited in Nicolai 2004:44)

Since Indonesian is an Austronesian language and is typologically more similar than Portuguese to native Austronesian languages of Timor, the process of 'Indonesianisation' was rather fast. According to Hajek (2000b:405), the number of people able to speak Indonesian almost doubled within 10 years, from 30% in 1980 to 60% in 1990. However, the numbers also include non-Timorese residents. The fact that 40% population did not speak Indonesian in 1995 can be attributed to both the limited access to formal education among the older population and a form of passive resistance towards the occupation (Leach 2017).

The only other allowed language was Tetun. The Indonesian government wanted to eradicate Portuguese from every aspect of life: media, official administration, schools

---

[26] The numbers cited in Nicolai (2004) and based on UN statistics from 2000 seem to be much lower than the numbers, taken from official statistics, cited by Thomaz (1985) who noted that in 1973-74 77 % of children were enrolled in school.

and the church. In 1981, the local Catholic Church, with the support of the Vatican, managed to install Tetun as a liturgical language. Tetun also spread quite quickly for another reason. A large portion of the population was displaced or had to flee and hide in the mountains, outside of their sociolinguistic area, and had to live and fight alongside people speaking different languages. Tetun was the vernacular language in the urban areas as well as in the mountains (Cabral & Martin-Jones 2008:155) and became a unifying element for the common Timorese people, although statistics on the number of speakers in 1975 and 1999 are missing.

The only place where the use of Portuguese was encouraged was in the Timorese resistance movement that had been fighting for independence. Written communication among the members was carried out in Portuguese and together with English, Portuguese was also used in international and diplomatic environments. However, Portuguese was used in tandem with Tetun. While Portuguese was used for the formal internal political affairs (policy documents, reports, messages for the Timorese diaspora, correspondence with the international community), Tetun was used for informal communication (letters, announcements, records) (Cabral & Martin-Jones 2008:165). At the same time, Fretilin continued to pursue some of their educational and welfare programs, like the literacy campaign focused on Tetun-medium education based on the handbook *Rai Timor, Rai ita nian*. Literacy was also promoted through songs and poems which were written in both Tetun and Portuguese (Cabral & Martin-Jones 2008:156). Indonesian was, however, used as a lingua franca in two regions: in the Fataluku-speaking area in the very east of Timor-Leste and in the Oecussi enclave located in West Timor.

Such an intense 'Indonesianisation' would probably lead to language death in many other countries, but the local Timorese languages showed remarkable resilience (Hajek 2000b:406). The only language that has been pronounced nearly extinct by Ethnologue is Makuva, an Austronesian language spoken around the town of Tutuala (Lewis & Fenning 2013).

2.9 Language attitudes and language planning in independent Timor-Leste

It was only after the Santa Cruz massacre on November 12, 1991 that the international community and the public worldwide became fully aware of the atrocities the Indonesian military had been committing in Timor-Leste. The footage of the massacre went viral and the pressure on the world powers to cut support to Indonesia grew bigger. But it wasn't until the end of President Suharto's rule in 1998 that Indonesia decided to offer Timor-Leste a special autonomy. Soon, it was announced that the Timorese population would be allowed to choose between independence or autonomy within Indonesia in a referendum.

The public vote was held on August 30, 1999 and 78.5% voted for independence, but the results were not accepted by the pro-integration groups which subsequently caused violence and bloodshed in the country. Finally, at the end of October 1999, the UN established a transitional administration (United Nations Transitional Administration in East Timor – UNTAET) that was supposed to administer the country for the next two years. On May 20, 2002, control was handed over to the first independent government of Timor-Leste and independence was proclaimed. However, the country was in a state of desolation. Most of the infrastructure, including roads and schools, was destroyed and around 80-90% of the population had been displaced from their homes. There were concerns this could have an impact on the language ecology and that the population might shift to a common lingua franca (Hajek 2000b:408). Fortunately, these concerns did not materialize.

Language planning for Timor-Leste started before independence. In 1996, the conference *It's time to lead the way!* was held in Melbourne, discussing the possible future official language. Two criteria were considered: the countries where the Timorese fled to due to the occupation and the languages that the diaspora spoke in these countries. Based on these facts, the three proposed languages were Portuguese, English and Indonesian. However, in the debate that followed, many pros and cons were raised. For one, Portuguese had never been spoken by the majority of the population and only 5% of the Timorese spoke it (and not very well). It would be a nostalgic choice of the old generation and it would represent Portuguese cultural neo-colonialism. Also, Portuguese was, according to many, difficult to learn and there were not enough teachers of

Portuguese. At the same time, the new generation spoke *Bahasa Indonesia* while Portuguese was a foreign language to them. Lastly, English, the universal language, guaranteed economic and technological development (Brito & Corte-Real 2006:124-128).

Despite all the cons, Portuguese language has had a long history in Timor and was the language of the Timorese resistance and the pro-independence movement CNRT (*Concelho Nacional de Resistência Timorense* – Timorese National Resistance Council), led by Xanana Gusmão. In 1998, CNRT adopted a document that would serve as the basis for the future constitution. It was agreed that Portuguese would be the official language with Tetun being a national language. A year later, CNRT held a conference in Melbourne where it was decided that Portuguese and Tetun would replace Indonesian as the language of instruction in schools within a 10-year frame (Hajek 2000b:408). And, shortly before the actual referendum, Xanana Gusmão delivered an independence speech calling for the following principles: "Bearing in mind our history, present reality and the economics and culture of the region surrounding our country, we must develop our Tetum language, generalize and perfect people's command of the Portuguese language and maintain the study of Indonesian language." (cited in Hull 1999a:1). In that same year, Timorese and non-Timorese linguists gathered at a conference in Sydney[27] discussing the roles of foreign and native languages and issued several recommendations: to reinstate Portuguese as an official language; to elevate Tetun to a co-official status; to use the variety of Tetun Prasa for standardizing the Official Tetun; to promote other native languages as a national patrimony and to recognize the country's multilingualism as the nation's most important human resource; to combat illiteracy through mother-tongue-based education; to preserve Malay/Indonesian with no official status; and to not let English threaten the linguistic integrity of Timor-Leste (Hull 1999a:2-4). Most of these recommendations are considered in the current Timorese Constitution of 2002:

---

[27] *East Timor towards Self-Determination: The Social and Cultural Questions,* organized by the UWS Macarthur, Faculty of Education and Languages on July 15-16, 1999.

"**Section 13**

(Official languages and national languages)

1. Tetun and Portuguese shall be the official languages in the Democratic Republic of East Timor.

2. Tetun and the other national languages shall be valued and developed by the State.

Section 159

(Working Languages)

Indonesian and English shall be working languages within civil service side by side with official languages as long as deemed necessary." (Democratic Republic of East Timor 2002)

However, various concerns about national and social identity, as well as its relationship with language, have been voiced. As Taylor-Leech (2008:154) put it:

> "The post-independent era has added new layers to social identity. As a result of this history, the relationship between language and identity in East Timor is complex and hotly contested. Among the many challenges facing this recently independent state is the problem of reconstructing an identity that fully reflects the multilingual character of the nation."

First, there were concerns about the viability of Portuguese as the official language, especially when co-existing with Tetun. There was a whole generation that had no access to this language due to the Indonesian occupation and there were no immediate resources to introduce Portuguese as the language of instruction from a very early age. In 2000, Hajek (2000a:226) admitted: "The future of Portuguese in East Timor is now guaranteed, although it remains to be seen to what extent Portuguese will re-establish itself in East Timorese society. Presently, according to recent press reports, it is estimated that only 10% of the population is proficient in Portuguese." In his research, Leach (2003:148) showed that young people strongly supported Tetun as the language of national identity, mostly because, growing up in Indonesian times, they had no connection to the Portuguese language. The 2004 National Census showed that 94% of the 18-25-

year age group were proficient in Tetun, while at the same time 81% claimed (self-reported) proficiency also in Indonesian, 51% in Portuguese and 42% in English. Although the real numbers are probably much lower, it is interesting to see how big a competition Indonesian and English are to Portuguese among the young generation. On the other hand, Portuguese has never had a negative connotation as a colonial language. The Portuguese ruled the local population on Timorese rules, not replacing the local rulers but trying to pacify and influence them, rather than oppressing them (Nicolai 2004:42). That is also one of the reasons why they were able to remain in Timor for five centuries (Hull 1999b:66).

Fifteen years later, the linguistic landscape is not much clearer. Tetun Prasa has irreversibly established itself as the lingua franca of the whole country and is now spoken even in the furthest ends of the island, such as in the Oecussi enclave in the west and in the Fataluku-speaking area in the east. It is also the language of instruction, together with Portuguese. Although primary school enrolment is 95% and the youth literacy rate is 82% (The World Bank 2015), many children still struggle to stay in school. Students drop out before completing primary education or often repeat grades. This is due to many factors, such as long distance from school, parental concerns about the safety of female students, financial problems, but also the inability to follow the language of instruction. Many children only acquire Tetun when they start school and find it difficult to read and write in a language they do not understand. In 2011, the Council of Ministers approved a new plan for teaching Tetun and Portuguese as subjects in primary education, with Tetun having a priority in the first three years (Taylor-Leach & Caet 2012:296). In the same year, a new Strategic Development Plan (Timor-Leste Strategic Development Plan 2011-2030) was introduced focusing also on disadvantaged children in classrooms:

> "Commission has initiated studies on mother tongue-based multilingual education for Timor-Leste. These studies aim to ensure that children are not disadvantaged and that all have equal access to an education, providing a smooth initial transition to the acquisition of Timor-Leste's official languages." (Government of Timor-Leste 2011:25).

Since then, the pilot project of the mother tongue-based multilingual education has been implemented in various primary and pre-primary schools across the country,

although it has been criticized by the experts as well as by the public, due to various problems like schools catering for kids from various ethnolinguistic groups with different mother tongues or the common, but unsubstantiated, belief that it undermines and slows down the learning of the official languages.

The Portuguese government has invested heavily in the revival of Portuguese through various programs. It has established a network of so-called *escolas de referência* (now renamed *Centros de Aprendizagem e Formação Escolar*, or CAFE) led by Portuguese teachers and with Portuguese as the language of instruction from Grade 1. To hold a job in the government or state administration, employees are expected to be proficient in Portuguese, but this rule is not always enforced. Many university students also do not speak Portuguese, although it is officially the main language of instruction in tertiary education. University professors are still heard giving lectures in Indonesian or Tetun.

English was not widely known before the arrival of the peacekeeping forces, aid agencies and various NGOs in 1999. The international language of this community is English and Timorese staff are expected to speak it. Young people are keener to learn English than Portuguese, because, especially in the capital, English skills mean more employment opportunities and higher salaries. The Indonesian language also cannot be ignored as a language that children pick up at a very early age from television, and as a language that is typologically close to Tetun. Timorese often go to study at Indonesian universities and many have family members living there as well.

According to the 2004 census, these were the literacy and capability rates in the four main languages:

Language	Literacy (ability to speak, read and write)	Capability to either speak, read or write or a combination of these
Tetun	42%	86%
Portuguese	12%	37%
Indonesian	39%	59%
English	5%	21%

Table 4 Adult literacy and capability rates in Tetun, Portuguese, Indonesian and English according to the 2004 census (DNE 2006:135-138)

Six years later, another census was conducted and the figures for adult literacy in all four languages were much higher than in 2004:

17. Adult literacy rate: The percentage of persons 15 and over who are literate (can speak, read and write) in Tetun, Portuguese, Bahasa Indonesia and English Language

Sexes	Total	URBAN	RURAL
Speak, read, and write in Tetun			
Both Sexes	56.1	80.9	44.6
Male	61.3	83.4	50.3
Female	50.9	78.0	39.1
Speak, read, and write in Portuguese			
Both Sexes	25.2	40.1	18.3
Male	29.0	42.8	22.1
Female	21.4	37.0	14.6
Speak, read, and write in Bahasa Indonesia			
Both Sexes	45.3	74.1	31.8
Male	50.0	77.2	36.5
Female	40.5	70.6	27.3
Speak, read, and write in English			
Both Sexes	14.6	24.7	7.6
Male	17.0	32.8	9.1
Female	12.3	26.2	6.2

Table 5 Adult literacy rate in Tetun, Portuguese, Indonesian and English according to the 2010 census (NSD and UNPF 2011:xxii)

The last census took place in 2015 and the latest figures revealed that both adult literacy in Tetun, Portuguese, Indonesian and English and capability to either speak, read or write (or a combination of these) in all four languages increased in urban but also in rural environment:

Sexes	*Total*	*URBAN*	*RURAL*	*Total*	*URBAN*	*RURAL*
	Literacy in Tetun			*Capability to either speak, read or write or a combination of these in Tetun*		
Both sexes	*62%*	*83%*	*53%*	*92%*	*97%*	*89%*
Male	*65%*	*85%*	*56%*	*93%*	*98%*	*91%*
Female	*60%*	*83%*	*50%*	*91%*	*97%*	*88%*
	Literacy in Portuguese			*Capability to either speak, read or write or a combination of these in Portuguese*		
Both sexes	*31%*	*44%*	*25%*	*61%*	*77%*	*54%*
Male	*33%*	*45%*	*27%*	*64%*	*79%*	*57%*

Female	*29%*	*42%*	*23%*	*58%*	*75%*	*50%*
	Literacy in Indonesian			*Capability to either speak, read or write or a combination of these in Indonesian*		
Both sexes	*37%*	*61%*	*26%*	*62%*	*82%*	*54%*
Male	*39%*	*62%*	*29%*	*65%*	*83%*	*57%*
Female	*34%*	*59%*	*24%*	*59%*	*80%*	*50%*
	Literacy in English			*Capability to either speak, read or write or a combination of these in English*		
Both sexes	*16%*	*28%*	*10%*	*39%*	*60%*	*30%*
Male	*17%*	*30%*	*11%*	*41%*	*61%*	*32%*
Female	*14%*	*29%*	*9%*	*37%*	*58%*	*29%*

Table 6 Adult literacy and capability rates in Tetun, Portuguese, Indonesian and English according to the 2015 census (DNE)

According to the 2010 census, the literacy rate increased in all languages: Tetun (42% vs. 56%), Portuguese (12% to 25%), Indonesian (39% vs. 45%) and English (5% to 15%); and it kept rising in the next five years, except for Indonesian[28] (see Table 7). These numbers might not be fully representative because they resulted from self-assessment, but we can see from the developments in the last 10 years that the literacy rate in Tetun and Portuguese is going to keep rising, since school enrolment is ever increasing.

	Tetun	**Portuguese**	**Indonesian**	**English**
2004	42%	12%	39%	5%
2010	56%	25%	45%	15%
2015	62%	31%	37%	16%

Table 7 Literacy rates in Tetun, Portuguese, Indonesian and English

Tetun has a strong position in both urban and rural areas and will continue to do so in the future. Portuguese has been part of the national curriculum since the independence and the number of speakers is ever increasing, but it is difficult to predict the attitudes of the young generation towards it in the upcoming years. Many consider

---

[28] Indonesian is no longer taught is schools and the only domain where it is still widely used is media (TV and news articles). The big difference between the rural and urban population may be caused by the fact that satellite TV is not widely available outside the cities.

Portuguese a form of neo-colonialism and do not see a purpose in learning it, as they take more pride in Tetun as their official and national language. For now, it is clear that the Indonesian language has managed to preserve its position as the second most spoken official/working language although the number of speakers is decreasing. And as a result of globalization, English will play a more important role, especially in the urban environments.

There is no doubt that multilingualism is a natural phenomenon in Timor-Leste. Indeed, to be able to fully function in Dili, one needs to be able to speak at least four languages. Carneiro (2010:10) summarized his own multilingual experience from the Timorese capital:

> "Ao percorrer a cidade, a diversidade de línguas utilizadas não só nas placas e sinalizações, mas também nos diversos contextos de interação surpreende ainda mais: pessoas falando em tétum nas ruas, nas feiras e nas casas; professores portugueses e brasileiros ensinando e interagindo em língua portuguesa nas universidades e em cursos de formação de professores; trabalhadores internacionais dos mais diversos países conversando em inglês nos restaurantes, nas agências internacionais e nas sedes de ONGs; comerciantes de diferentes nacionalidades, mas principalmente indonésios e chineses, utilizando o indonésio e o inglês, dentre os quais ainda se vêem alguns que utilizam o hakka ou o yue, línguas vindas do sul da China que estão presentes no país desde tempos remotos."[29]

---

[29] "Roaming the city, the diversity of languages used not only on signs but also in various contexts of interactions is even more surprising: people speak Tetun on the streets, in the markets and at home; Portuguese and Brazilian professors teach and interact in Portuguese at universities and in teacher training programs; international workers from various countries have conversations in English in restaurants, international agencies and NGO offices; businessmen of different nationalities, but especially Indonesian and Chinese, use Indonesian and English, but one can find among them some, who speak Hakka or Yue, languages originating in South China that have been present in the country for a long time." (my translation)

2.10   Official orthography

Apart from the significant interspeaker variation, another issue I had to face was orthographic. The Tetun language does not have a long literary tradition. The first written materials date back to the late 19th century, and the language itself was elevated to the official status only in the post-colonial times. During the Portuguese colonial period, works in Tetun were written by Portuguese nationals who often applied the Portuguese orthography rules to the local languages. Portuguese borrowings were largely written in their original form while native words were often transcribed using the Portuguese orthography (including <ç>, acute, grave and circumflex accent):

(1)   *hôto* 'finish' (Silva 1889:5) = T. *hotu*

    cátac 'say' (Silva 1889:15) = T. *katak*

But Silva's (1889) orthography conventions were not altogether conservative. He and his successors introduced various innovations that were later incorporated into the official orthography, e.g. the use of /'/ for a glottal stop, the use of /h/ for a glottal fricative, the use of /k/ in places where Portuguese sequence <qu> would be normally used, etc. (Instituto Nacional de Linguística 2004:7-9).

The first attempt to consolidate Tetun orthography and promote literacy and Tetun as a national language came in 1975 from within the Fretilin party with the publication of *Como vamos alfabetizar o nosso povo Mau Bere de Timor-Leste* (How can we provide literacy to our Mau Bere people of Timor-Leste). Fretilin defended an orthography which would follow the orthographies of other post-colonial countries, whose native languages were freed from inadequate representation based on colonial languages (Hull & Eccles 2005:242). Its main contribution was the elimination of the Portuguese-based orthography and its substitution with more phonemic representations, e.g. the circumfix accent was eliminated, <-ão> was substituted by <-aun>, <ou> [o] was substituted by <o>, and <ge> and <gi> were substituted by <je> and <ji>.

During the Indonesian times, when Portuguese was prohibited and replaced by Tetun as a church language, the Dili diocese published the new translation of the *Ordinário de Missa* (Comissão Litúrgica da Diocese de Díli 1980). It was recognized by the Vatican and served to many as a new model of orthography (Albuquerque 2011:97)

which is still used by the Catholic church in present days. Church Tetun and its corresponding orthography have a long history. In 1898, the Catholic church founded a *colégio* in Soibada and its graduates went to constitute the Timorese elite. Some of the orthographic rules included the use of acute accent only in words with irregular stress (other than penultimate) in native and Portuguese words, e.g. *maibé* 'but', *glória* 'glory', *manán* 'win'; the replacement of final <i> with <e> in Portuguese borrowings, e.g. *agradese* 'thank, appreciate', *padre* 'priest' and others. The fact that Soibada is located in a region where Tetun Terik is spoken left its mark on Church Tetun, which uses <w> to represent the native /w/ (replaced by <b> in Tetun Prasa) (Instituto Nacional de Linguística 2004:12).

Although there was no official body in Timor-Leste to study Tetun during the Indonesian times, the International Academic Committee for the Development of East Timorese Languages (IACDETL) was set up in Australia. The international academy, which was led by the Australian linguist Geoffrey Hull, published the *Standard Tetum-English Dictionary* (1999c) and drafted the *Princípios de Ortografia Tétum: Sistema Fonémico* (Principles of Tetun Orthography: the phonemic system) (INL 2001). Their biggest innovation was the introduction of the graphemes <ll> and <ñ> to represent /ʎ/ and /ɲ/, respectively. The reason why the former (and Portuguese-based) <lh> and <nh> were replaced was the fact that, in Tetun, /h/ is aspirated ([h]) and is thus represented by the grapheme <h>, e.g. *bainhira* 'when'. Since the idea was to introduce an orthography that could be later applied to other native languages of Timor which often contain a consonantal sequence <lh> and <nh>, there was a need to find a different representation of the Portuguese palatal sounds. Some suggested the Indonesian graphemes <ly> and <ny> but this was unacceptable also due to political reasons. Thus, graphemes derived from the romance tradition were adopted: <ll> which was used to represent /ʎ/ in medieval Portuguese and <ñ> which is still used in Galician. Other innovations included: the elimination of silent consonants in Portuguese borrowings (*otél* 'hotel' from Port. *hotel*, *asaun* 'act' from Port. *acção*), the introduction of acute accent in stressed long vowels in Portuguese borrowings stressed on the final syllable (e.g. *pás* 'peace' from Port. *paz*), and the differentiation of syllables with double and single vowels (*haree* 'see' and *hare* 'unhusked rice') (Instituto Nacional de Linguística 2004:13-15).

After the independence, the newly-established Instituto Nacional de Linguística (INL – National Institute of Linguistics) published two important works: *Matadalan Ortográfiku ba Tetun Nasionál* (Orthographic Guide for National Tetun, 2002a) and *Hakerek Tetun Tuir Banati: kursu ortografia padronizada nian* (Writing Tetun based on the Model: course of standardized orthography, 2002b). The official Tetun orthography, developed by the National Institute of Linguistics, was standardized by Government Decree No. 1/2004 of 14 April 2004 *The Standard Orthography of Tetun Language*. According to Article 2:

> "1. The orthography of Official Tetun is the phonemic system, as developed and administered by the National Institute of Linguistics on the basis of scientific criteria.
>
> 2. Official Tetun is the variety of Tetun that has asserted itself as the official and national language, which is a modern literary form of the country's most common vernacular based on Tetun-Praça.
>
> 3. This option shall not prejudice the varieties of Tetun used exclusively in certain regions, which the State is preserving and promoting as national languages." (Democratic Republic of East Timor 2004)

The official orthography was disputed by many, especially by the Tetun department of the Dili Institute of Technology led by the Australian linguist Catharina Williams-van Klinken, who is one of the authors of Tetun Prasa grammar (Williams-van Klinken et al. 2002a). The grammar was published before the official standardization and the authors introduced their own rules of representation: instead of using the recommended <ll> and <ñ>, they decided to follow the Portuguese spelling <lh> and <nh> while admitting that the Indonesian alternative <ly> and <ny> would be more convenient for speakers who are unaware of Portuguese spelling conventions. The authors also decided to mark stressed vowels by an acute accent instead of double vowels (e.g. *hát* vs. *haat* 'four') (Williams-van Klinken et al. 2002a:6-7). Another issue was the representation of uvular trill [ʀ] and glottal stop [ʔ]. According to Williams-van Klinken (2007), these sounds are not realized by the majority of speakers, which was documented with a series of short tests for university students. The author concluded that the students showed a very low awareness of which words should and should not be pronounced with

[ʀ] and [ʔ] and she argued for their exclusion from the official orthography. Or, if they were to be preserved for historical and nationalistic reasons, then there must be bigger pressure on students memorizing which Tetun words contain these sounds.

Although there is now an official standardized orthography in place, it is still not widely used and we can see it enforced only in governmental institutions. This proved to be a considerable problem when transcribing and analysing the data in my thesis. It was crucial to adopt the official orthography to consolidate my written corpus (newspaper articles from Sapo, Suara Timor Lorosae and Timor-Leste Government website) and my oral corpus. For Tetun Terik, I applied the same orthographic rules that were developed for Tetun Prasa. As for Tetun Fehan, I transcribed the examples from van Klinken (1999) with their original spelling. Some examples of Tetun Prasa were cited in their original form and in this case, they contain a note 'original spelling'. In case examples from other works also contained glosses and English translations, these were cited *ipsis verbis*.

3. THEORY AND METHODOLOGY

3.1 Theoretical framework

Before proceeding to the description of Tetun Prasa, one fundamental question needs to be answered: What kind of (contact) language is Tetun Prasa? Based on the considerable differences between Tetun Prasa and Tetun Terik on the level of morphology, syntax, phonology and lexicon, and because of the near-affixless character, it has been often labelled as a creole (Grimes et al. 1997, Hull 1999c, Avram 2008), extended pidgin (Avram 2008) or a hybrid language (Hull 1999c) and many authors have commented on the 'simplified' aspect of Tetun Prasa (Thomaz 1981, Hajek 2006b, McWhorter 2007, McWhorter 2011 among others). Sá (1961:xxv) noticed that Tetun Prasa has been departing more and more from its original form and merging with Portuguese, which could lead to a new Timorese creole language.

3.1.1 Theory of creoles, pidgins and mixed languages

Contact between two or more languages is natural and often inevitable but not all types of language contact lead to a change. There are many variables that have influence on its degree, like the nature of the contact or its duration. The most common origin of this contact is when one group moves into another group's territory; the movement can be peaceful but often it turns into a forcible occupation (Thomason 2001:17).

The reasons for language change can be external or internal; the external change is induced by contact with a different language (Hickey 2010:7). According to Thomason (2010:32), this type of contact "is a source of linguistic change if it is less likely that a particular change would have happened outside a specific contact situation." This contact situation can be either stable and permanent, or unstable and short-lasting, but both settings are subject to change based on various social (but not linguistic) factors, e.g. urbanization, number of speakers, institutional support etc. (Thomason 2001:21-22).

Another important thing is the socioeconomic and political relationship between the two groups: the dominant and the subordinate one. They can be in four different positions: indigenous superordinate, migrant superordinate, indigenous subordinate and migrant subordinate. Based on the position, a group is either more or less likely to shift to a dominant/subordinate language, but not all contact situations show this clear asymmetrical relationship (Thomason 2001:23). In Timor, the Portuguese were the migrant dominant group, yet the indigenous group did not shift to their language. As described in Chapter 2, this had to do with the limited number of Portuguese nationals (or native speakers) in Timor and the lack of effort to impose their language on the whole population, although the contact was relatively stable and lasted for four centuries, until other social changes caused the end of it.

There are two main types of linguistic interference, or contact-induced change: *borrowing* and *interference through shift*. Borrowing is:

> "(...) the incorporation of foreign features into a group's native language by speakers of that language: the native language is maintained but is changed by the addition of the incorporated features. Invariably, in a borrowing situation the first foreign elements to enter the borrowing language are words. (...) If there is strong long-term cultural pressure from source-language speakers on the borrowing-language speaker group, then structural features may be borrowed as well – phonological, phonetic and syntactic elements, and even (though more rarely) features of the inflectional morphology. Although lexical borrowing frequently takes place without widespread bilingualism, extensive structural borrowing, as has often been pointed out, apparently requires extensive (though not universal) bilingualism among borrowing-language speakers over a considerable period of time." (Thomason & Kaufmann 1991:37).

The crucial point here is that "in borrowing, the interference features are introduced into the receiving language by people who speak it fluently." (Thomason 2001:68). As presented above, there has never been extensive bilingualism (in terms of Tetun, Portuguese and Indonesian) in Timor-Leste, which means that the language component most affected by this contact is lexicon (Chapter 8). Other components which were affected to lesser extent were phonology, which shows a lot of variation (Chapter

7); morphology, both derivational and inflectional (Chapter 5) and syntax, which was enriched with numerous structures (Chapter 6). Thomason & Kaufmann (1991:38) believed that if there is phonological interference, there will be a similar degree of syntactic interference, while morphological interference lags behind. Indeed, morphological borrowings are very scarce in Tetun Prasa while the language's phonological inventory has been largely expanded.

Interference through shift "results from imperfect group learning during a process of language shift. That is, in this kind of interference a group of speakers shifting to a target language fails to learn the target language (TL) perfectly." (Thomason & Kaufmann 1991:38-39). Unlike in the case of the borrowing, the first elements that undergo language change are phonology and syntax and the whole process can take as little as one generation (Thomason & Kaufmann 1991:39-41).

The two types of interference are not mutually exclusive. Just like Tetun Prasa has been going through language change via borrowing, imperfect L2 learning of Portuguese caused that the Portuguese spoken in Timor has been, to some effect, affected by interference through shift (see Batoréo 2009, Albuquerque 2014, Afonso & Goglia 2015).

But what happens when this language shift from native languages is abrupt? The result is extreme language mixing and phenomena like pidgins, creoles and bilingual mixed languages emerge. The difference between pidgins/creole and bilingual mixed languages is similar to the difference between borrowing and interference by shift: pidgins/creoles are similar to shift-induced interference and develop in social context with no or little bilingualism/multilingualism while bilingual mixed languages are similar to borrowing since imperfect learning plays no major role in their development (Thomason 2001:157-158).

In the next section, I will discuss bilingual mixed languages and pidgins/creoles in more detail while arguing why Tetun Prasa should not be considered one of them.

3.1.1.1 Mixed languages

Until the 1990's, all the cases resembling a possible mixed language were labelled either as codeswitching, adstrate influence or borrowing (Meakins 2013:159). They were first brought to attention in 1988 by Thomason and Kaufman: "[T]here are indeed mixed languages, and they include pidgins and creoles but are not confined to them; mixed languages do not fit within the genetic model and therefore cannot be classified genetically at all (…)" (Thomason & Kaufman 1991:3) because "by definition they are unrelated genetically to the source(s) of any of their multiple components" (Thomason & Kaufman 1991:11). Bakker (2000:29) agrees that in the case of mixed languages a genetic classification is not possible and one can talk about language mixing "[o]nly if the grammatical system and the basic lexicon of a language are of a different origin, OR if both of these components are roughly equally from different language sources." Winford (2003:175) explained one of the reasons why mixed languages were the last ones to be described and studied:

> "(...) classifications of languages based on sociohistorical criteria do not match up exactly with those based on purely structural criteria. This reflects the fact that intertwined languages do not all conform to some ideal prototype, but rather constitute a varied assortment of outcomes, with different histories and structural characteristics." (Winford 2003:175).

Indeed, the degree of mixing can vary from predominantly lexical mixing (as in languages like Anglo-Romani or Media Lengua) to significant structural mixing (as in Michif or Gurindji Kriol) (Meakins 2013:164-165). Bakker (2003:122-125) calls languages that show a split between lexicon and grammar *G-L mixed languages* and the ones displaying structural mixing of verbal and nominal systems *V-N mixed languages*. What is, however, intriguing is that it is almost impossible to predict what sociohistorical background gives rise to either the G-L or V-N groups of languages (Meakins 2013:179). Many linguists have proposed their own theories of the formation of mixed languages, taking into account the relationship between the ancestral and the introduced language, and the direction of the language shift. These theories can be divided into two groups:

unidirectional (borrowing, codeswitching, relexification, paralexification[30], and language repertoire approach[31]) and fusional (language intertwining theory, language competition approaches and centre of gravity hypothesis). While unidirectional approaches suggest that the language shift occurred from the ancestral language towards the introduced language (semi-shift or reversal of the shift), the fusional approaches assume that a mixed language is the result of two languages merging. Most of the G-L mixed languages would be classified as the result of a shift by degree or a reversal of shift.

According to Meakins (2013:183), the shift usually stops part-way for two reasons: either the speakers do not have full access to the introduced language or the ancestral language is a marker of their social identity: "[T]he mixed language serves as an expression of an altered identity, be it new, or differing significantly from an older identity." (Meakins 2013:181). However, the speakers of a new mixed language do not necessarily need to constitute a new ethnic group. Although the ancestral and introduced languages may co-exist in the same environment, the fact that the new language is a native/first language of a certain community whose speakers are not fluent in either of these languages gives it certain autonomy.

According to Meakins (2013:199), language stability, independent development of the source and mixed language, and the presence of structural features from both input languages are the three criteria of mixed language autonomy. As for language stability, Thomason (2003:24) asserts that a language is stable when it is spoken outside of the bilingual context in which it arose, when there is a high degree of consistency within the use of lexicon and grammar and when children are learning it as their mother tongue (Meakins 2013:200-203). An independent development of the source and mixed language means, in this case, that a change in one of the source languages is not necessarily reflected in the mixed language and vice versa. Also, a form in the mixed language can develop a different function or can be distributed differently (Meakins 2013:203-205).

---

[30] According to Mous (2001:113), paralexification is a process "by which parallel word forms for one and the same lexical entry exist while sharing meaning and morphological characteristics." Mous further adds that it differs from intertwining in that paralexification does not always result in a mixed language and it is also not enough to account for all changes in a (mixed) language.
[31] In this approach, Matras (2009) considers a bilingual/multilingual speaker as a possessor of certain lexical repertoires that are used in specific social and communication domains. Gradually, this bilingual/multilingual speaker learns to select the appropriate repertoire depending on social and communication settings.

Based on these criteria and considering the social context and the structure of the language, it is rather unlikely that Tetun Prasa is a mixed language in a traditional sense. Although the number of lexical and structural borrowings is considerably high, it does not reach the level of mixing required for a mixed language to arise, since there is no clear split, either between lexicon and grammar or verbal and nominal systems, as is documented in Chapter 5 (Morphology) and Chapter 6 (Syntax).

3.1.1.2 Pidgins and creoles

According to Thomason (2001:158), while mixed languages are akin to borrowings, pidgins and creoles are related to shift-induced interference, which also means that these languages developed in social situations where bilingualism/multilingualism was not widespread and imperfect learning played a role[32].

Pidgin, in the narrowest sense, is a language that resulted from an extended contact between communities of people that had no language in common. None of these groups learns the pidgin as their native language and usually use it only in certain domains, such as trade. To make it easier for these groups to understand each other, they simplify their languages by dropping inflections, reducing the vocabulary but at the same time extending the meaning of certain words. The difference between a pidgin and a jargon is that a pidgin is more stable and although variation still exists, it has certain norms regarding pronunciation and grammar (Holm 2004:5). However, the grammar (unlike vocabulary) does not come from one of these languages but is somehow a 'crosslanguage compromise' influenced by universals of second-language learning (Thomason 2001:159).

A creole, on the other hand, is spoken as a native language by a whole speech community (usually spoken in all domains of life) and often has a pidgin or a jargon in its ancestry. However, many creoles never went through the stage of pidginization, but

---

[32] On the other hand, McWhorter (2005:253) disagrees that the proficiency and bi-/multilingualism play an important role in the development of mixed languages. According to the author, the decisive factor is the number of native languages in contact: when there is only one native language, a mixed language arises; where there is more than one, a creole develops.

arose gradually or, very often, abruptly[33]. Thomason and Kaufman (1991:147) referred to this abrupt creolization also as a shift without normal transmission. This accelerated change was already noted by van Name (1869-70):

> "(...) [t]he changes which [creoles] have passed through are not essentially different in kind, and hardly greater in extent than those, for instance, which separate the French from the Latin, but from the greater violence of the forces at work they have been far more rapid ... here two or three generations have sufficed for a complete transformation." (van Name 1869-70:123 cited in Holm 2004:25).

Creole languages arose in a specific sociolinguistic context, especially in slavery, where more than two languages were in contact. The process of creolization (or nativization) is opposite to that of pidginization. Phonological rules (e.g. assimilation) and vocabulary are expanded, since the language is spoken in all domains of life, and the grammar is reorganized, with new complex verbal system and phrase-level structures arising (Holm 2004:6-7). The language (usually, but not necessarily, of European origin, a language with higher prestige) that supplied most of the lexicon is called a superstrate language or a lexifier language, while the languages that supplied the grammatical element are called substrate languages (languages of lower prestige, usually non-European). In general, we can say that creoles are vernacular languages that developed mostly throughout the 17th and the 18th century in former European colonies as a result of a language contact between communities that lacked a common language.

There are many crucial details about the origin and development of Tetun that are still missing and it is not clear if Tetun Prasa ever underwent any stage of pidginization. However, Tetun Prasa has often been referred to as a pidgin or a creole language. Grimes et al. (1997:52), used the term "creole" in connection with Tetun Prasa as a language that has largely shifted from its original source. Thomaz (2002:103) described it as a Portuguese-influenced simplified form of Tetun whose characteristics are close to those of creoles and compared it to Língua Geral in Brazil, which was not a creolized form of

---

[33] Abrupt creolization is a process "when the emerging contact language at one becomes the primary language of the community and is learned as a first language (though not necessarily as their only first language) by any children born into the new multilingual community. That contact language therefore expands rapidly into a creole rather than stabilizing as a functionally and linguistically restricted pidgin, though its formative period, before it crystalizes as a language, corresponds to what is generally called a prepidgin stage." (Thomason & Kaufman 1991:150).

Portuguese but rather a language of the Tupi-Guarani family enriched with Portuguese loanwords (Thomaz 2002:69). Hull (1999c:ix) referred to Tetun Prasa as "a fully creolized form" of Tetun and at the same time as "a hybrid language, basically Austronesian, but with a heavy Portuguese superstratum" and compared it to English after the Norman Conquest, which showed the massive French influence it had undergone. Despite the indisputable value of the contribution that Geoffrey Hull has made to the study of Tetun, it must be recognized that he does not use the terms defining creoles (e.g. "superstrate") the way they are currently used by creolists in evaluating the status of Tetun Prasa. The analogy that Thomaz has drawn between Tetun Prasa and Brazil's Língua Geral strikes me as being much more accurate.

So what claims can be made about the sociolinguistic setting in which Tetun Prasa emerged? Taking into account all the different languages that were spoken and used in Timor throughout the centuries, one might wonder why creole languages developed in other former Portuguese colonies in Asia (like Malacca, Macau, but also in Jakarta) and none developed in Timor-Leste. Bidau Creole Portuguese (see section 4.3), which is now extinct, was spoken in the Bidau district of Dili, but it is believed that this restructured variety did not originate there but was brought to Timor when the families of soldiers and officials came to live in Dili after the capital was moved in 1769 (Marcos 1995, Baxter 1996).

As discussed in the chapters on phonology, morphology, syntax and lexicon, both the structural and sociohistorical facts support the hypothesis that Tetun Prasa is not a creole language but the result of a specific kind of language contact that has led to a specific kind of a contact language. When we look at the lexicon, it is clear that Tetun Prasa cannot be considered a creole language because the language lacks the most salient characteristic of a creole: most of its vocabulary is still native and not derived from a superstrate language. Portuguese could well be considered one of its lexical source languages, but not its superstrate as this term is understood in creole linguistics, i.e. the language providing the overwhelming majority of a creole's vocabulary. From the sociolinguistic point of view, creole languages were the result of a communication need between communities that lacked a common language. But this was not the case of Tetun Prasa. In Timor, there was no need to create a new language for the Portuguese and the Timorese to be able to communicate while doing business. They already had one – Malay

– which was the main trade language of the region in those times (Thomaz 1974:253). And later in the 18th century, when the Portuguese settled in Timor, Tetun was already the vehicular language among local peoples (Thomaz 1974:280).

The fact that Tetun Prasa has been claimed to be a creole may have to do with the fact that creole language studies have become a much more prominent part of linguistics since the 1970s, as the study of Tetun was developing. However, the definition of a creole has since evolved and is not understood strictly only in sociolinguistic terms but also in synchronic terms. According to Singler (2008:333), creoles do not form a tightly-knit group and they can't be distinguished from non-creoles based on some rigorously defined linguistic criteria. The author further believed that principles that govern the development of creole languages are the same as principles that govern all languages at their birth, and quoted Muysken (1988:300): "The very notion of a 'creole' language from the linguistic point of view tends to disappear if one looks closely; what we have is just a language." It is also questionable what the motivation of shifting speakers was. Some authors (e.g. Baker 1995) believed that the speakers of substrate languages did not aim to learn the European lexifier as their target language and that their only motivation to create a new language was the necessity of communication[34]. On the other hand, authors like Mufwene (2001) argued that the motivation of speakers played no significant role in shaping the newly emerged language (Singler 2008:342).

In recent years, the traditional universalist[35], superstratist and substratist theories have been refuted by most authors and the attention shifted to the generativists' view of a language. Lefebvre (1998:6) summarized the problems with the universalist approach as follows:

> "The universalist approach does not account for the fact that creole lexicons manifest the properties of their source languages in the way they do (...). Furthermore, creole languages are not uniform; like other natural languages, they

---

[34] Already Greenfield (1830:50) concluded that the reason why the Africans in Suriname failed to learn English (and instead Sranan arose) was not because of the inferior ability to learn the language but rather because of the lack of motivation to acquire it.

[35] Universalist theory was formulated by Coelho (1881: 67-69, translated in Holm 2004:27): "The Romance and creole dialects, Indo-Portuguese and all the similar formations represent the first stage or stages in the acquisition of a foreign language by a people that speaks or spoke another (...) They owe their origin to the operation of psychological laws that everywhere are the same, and not to the influence of the former languages of the peoples among whom these dialects are found."

manifest language-specific features (...). The universalist approach does not account for the variation that exists between creoles". (Lefebvre 1998:6).

Mufwene (1996), one the other hand, criticized the superstratist approach: "One of the problems with the superstrate hypothesis is the absence of any explanation for why creoles lexified by European languages do not correspond to any particular dialect of their lexifiers." (Mufwene 1996:166). Hall (1955) cautioned against the substratist approach:

> "In summary, anyone with some knowledge of pidgin and creole languages cannot deny the existence of the influence of the substrata. (...) Each presumed case for substrate influence should be judged independently on its own merits; thus we shall avoid exaggerations of both those who see signs of substrate everywhere and those who deny it completely." (Hall 1955:9, translated in Holm 2004:43).

In any case, more and more linguists are starting to look at differences between creoles rather than similarities, since the variation between creole languages is as great as between any world languages (Singler 2008:345).

3.1.2   Koineization (dialect mixing)

Another type of contact-induced change, but which does not happen between two or more languages but rather between dialects of the same language, is koineization. According to Kerswill (2002:669), it is quite rapid and possibly dramatic change in which new varieties of a language emerge as a result of contact between mutually intelligible varieties of the same language. Usually it occurs in places to which people have migrated from different parts of the country where a single language is spoken. Siegel (2001:175) defined a koine[36] as follows:

> "A koine is a stabilized contact variety which results from the mixing and subsequent levelling of features of varieties which are similar enough to be

---

[36] 'Koine' comes from Greek *koinē* 'common'. Originally, it referred to a particular variety of the Greek language that became a lingua franca in the eastern Mediterranean during the Hellenistic and Roman periods but is now used as a generic term for other language varieties. The original Koine was based on one regional variety but received influence from many others. It became the lingua franca and was spoken mostly as a second language, although eventually it acquired native speakers (Siegel 1985:358-359).

mutually intelligible, such as regional or social dialects. This occurs in the context of increased interaction or integration among speakers of these varieties."

There are two main types of koines: regional koine and immigrant koine. Regional koine is a result of dialects of the same language being in contact. It does not replace the contributing dialects but remains in the area where the contributing dialects are spoken. On the other hand, immigrant koine is a dialect mixing that occurs outside of the region where the dialects originated, i.e. a new dialect develops in a new settlement and eventually replaces the regional dialects of the original migrants but at the same time does not influence the dialects in their place of origin (Kerswill 2002:671). The difference between koineization and dialect levelling is the fact that in dialect levelling, the dialects affect each other but a new dialect does not arise (Siegel 1985:365). On the contrary, there is usually a decrease in the number of dialects spoken in a certain dialect area (Kerswill 2002:671).

The process of koineization has, according to Siegel (1985:373-374), four stages. The author calls the first one 'prekoine', which is an unstabilized form at the beginning of the developmental continuum. The second stage is the result of stabilization when a new dialect compromise emerges, although often with a reduced morphological complexity. This 'stabilized koine' can become a language of other groups as well as a literary language or a standard language of a country. This extension in use goes hand in hand with linguistic expansion and results in 'expanded koine'. When a koine becomes a first language of a community of speakers, the author talks about a 'nativized koine'. At this stage, besides further linguistic expansion, new innovations which cannot be traced back to the original dialects are also introduced in this variety. Siegel (1985:374) also believed that the nativization could occur after any of the first three stages, not necessarily only after standardization, as is illustrated by Figure 5:

```
      Type 1                      Type 2                      Type 3

     prekoine                    prekoine                    prekoine

        |                           |                           |
        |                      stabilized koine           stabilized koine
        |                           |                           |
        |                           |                      expanded koine
        |                           |                           |
        ↓                           ↓                           ↓
  nativized koine             nativized koine             nativized koine


  (Fiji Hindustani)          (Guyanese Bhojpuri)           (Greek Koine)
```

Figure 5 Developmental stages of various nativized koines (Siegel 1985:375)

Trudgill (1998:198) argued koineization happens in three different stages which correspond to three different, but successive, generations of speakers. The first generation of adult speakers of different regional and social varieties initiates a contact in a new environment which involves some accommodation[37] and dialect levelling. The second generation of speakers has still many linguistic models to choose from and shows great inter- and intra-speaker variation. They are free to select forms from different dialects and subsequently develop new forms and combinations. The stabilization only happens in the third generation when a new variety emerges as a result of focusing. However, there are three main factors that influence focusing and long-term accommodation: the kind and level of social integration (the process of koineization is faster in a socially homogenous community), children's access to peer groups (children need to interact with their peers to be able to establish norms since a stable adult model is missing), and degree of differences between the input varieties which determines the amount of accommodation speakers need to engage in (Kerswill 2002:695).

Koineization shares similarities with other contact-induced changes. Siegel (1985:376) compared the process to pidginization, in which also different varieties are in contact resulting in a new variety that has features of the varieties in contact but is reduced and simplified. However, in case of koineization, the varieties in contact are dialects rather than typologically distinct languages and the process in slower and more gradual.

---

[37] Accommodation theory presupposes that speakers converge when they want to show positive attitudes (gain somebody's approval, show solidarity) and diverge when they do not. Accommodation can be mutual or one-sided (Kerswill 2002:680).

There is also no 'target variety' that the speakers are aiming for. At the later stage, when expansion of function and form occurs, the process of nativization in creolization is analogous with koineization after the initial stage.

Based on the definitions of different types of contact languages/dialects, I believe it would be more fitting to consider the possibility of Tetun Prasa being a koine with heavy Portuguese lexical influence, especially when compared with the more conservative varieties of Tetun Terik and Tetun Fehan. There is hardly any information about the pre-colonial spread of Tetun from its original area but we can assume that the reduction in certain forms (subject marking, possessives, plural marking) occurred already during the initial stage of prekoine. When the Portuguese moved the capital to Dili, people from all across the island migrated there bringing their local L1 and L2 varieties of the 'vehicular Tetun' with them and, at the same time, speakers of the Mambae language (originally spoken in Dili) either shifted to Tetun Prasa or adopted it as L1/L2 while preserving Mambae as their mother tongue[38]. A new immigrant koine started to develop, eventually becoming a mother tongue of children born in Dili, a former Mambae-speaking region, who started to introduce new forms and innovation into this new variety, Tetun Prasa. Although many generations have been growing up with Tetun Prasa as their L1, a lot of variation still remains.

3.1.3 Codeswitching

In order to better understand the contact-induced change, I looked at several mechanisms that play a role in it, with one of the major ones being codeswitching. According to Myers-Scotton (2006:161), codeswitching is "the use of two language varieties in the same conversation. It can occur between speakers, or between sentences in the same speaker's turn, or within a sentence." There are various approaches to codeswitching that try to explain how and why it happens but they all have one thing in common. They assume that "speakers send social messages by switching from one dialect or one language to another, sometimes within the same conversation. All speakers select

---

[38] The same could be said about speakers of other local languages who preserved their mother tongues and, at the same time, adopted Tetun as their L1 or L2.

their speech code on the basis of calculations that they make, even though most of these are unconscious (Myers-Scotton 2006:173)."

Codeswitching can have various forms. It can be inter-sentential or intra-sentential, i.e. we can observe switching that includes full sentences, or just single words occurring within a clause. Myers-Scotton (2006:241) is especially interested in the intra-clausal codeswitching, since "it is only within the clause that the language varieties involved in codeswitching are in contact." Intra-clausal codeswitching can be insertional or alternational. In insertional codeswitching, the grammar of one language is more dominant than the other language, i.e. the elements of the more dominant language are inserted in the morphosyntactic frame of the receiving language. In alternational codeswitching, there is alternation of structure from different languages and a new grammar emerges that is a combination of two languages (Meakins 2013:190-191).

Myers-Scotton further distinguishes between classic codeswitching and composite codeswitching, which is a combination of codeswitching and convergence. Classic codeswitching is defined as codeswitching that "includes elements from two (or more) language varieties in the same clause, but only one of these varieties is the source of the morphosyntactic frame for the clause (Myers-Scotton 2006:241)." However, for a speaker to be able to engage in codeswitching, he or she needs to be proficient enough in the language that provides the morphosyntactic structure. In our case, that language would be Tetun. Moreover, the speaker also needs to have proficiency in the other language that is involved in codeswitching, although they do not need to be bilingual. In our case, this would be Portuguese.

The question is: what is the relationship between code-switching and contact-induced change? Do all code-switches turn into loanwords? Or is it the only way for words to be incorporated into a language? Thomason (2001:132-134) believes that that the boundary is very fuzzy and they both lie on the opposite ends of a continuum. The difference is that "code-switched elements are not integrated into the receiving language's structure, whereas borrowed elements are nativized – adapted to the structure of the receiving language" (Thomason 2001:134). More discussion on lexical borrowings versus code-switches can be found in Chapter 8.

3.1.4 Second-language acquisition

In the discussion above, I have often mentioned a 'native speaker' and its role in the development of a language. However, a definition of a native speaker is often elusive, as commented on by Davies (2004:431), who quoted Ferguson (1983:vii): "Linguists (...) have long given a special place to the native speaker as the only true and reliable source of language data." A person, or rather a child, can be a native speaker of more languages, but the acquisition process needs to start early and before puberty. After that, it is very difficult, but not impossible to become a native speaker of a certain language. Native speakers also have intuitions about their idiolectal grammar and about the features of the standard grammar which are different from their idiolectal grammar (Davies 2004:433-435).

For non-native speakers of a certain language, it is still possible to learn it as their second (or third, or fourth) language. The acquisition process depends on three different factors: cognitive (intelligence, language aptitude, and memory, etc.), affective (motivation, personality, willingness to communicate, etc.) and social (Ellis 2004:530). Clearly, second-language acquisition (SLA) is a psycholinguistic process but it is undeniable that social context and social factors also play a role in it, as shown in Figure 6.
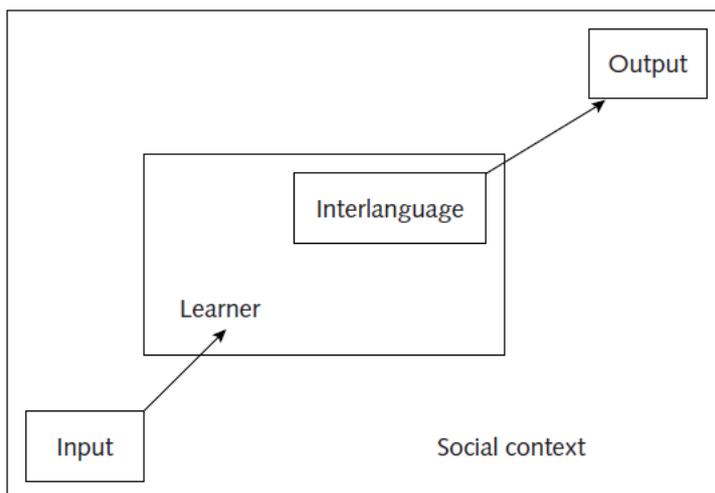
Figure 6 A basic model of language learning (Barkhuizen 2004:556)

The four main social factors that can lead to different levels of L2 proficiency are age, sex, social class, and ethnic identity. The age, at which speakers start learning the second language, can influence their fluency. For example, speakers that start the learning process after the onset of the puberty are more unlikely to have natural native-like accent, and those who start after the age of 15 are less likely to acquire such grammatical ability as those who start earlier[39]. At the same time, adolescents are faster learners than younger and middle-aged learners (Ellis 1994:201). However, these arguments have been challenged by Marinova-Todd, Marshall and Snow (2000:10-11) who believed that older learners have the ability to reach native-like L2 proficiency, too, and argued that there was no 'critical period' for SLA. The authors further claimed that there were 'myriad factors' involved in successful SLA and the most notable among them was the environment in which the language was acquired (e.g. living in the environment where the L2 is standard) and other social, psychological and educational factors (Marinova-Todd, Marshall and Snow 2000:24-25:28).

Sex, or gender, has been proven to also play a role in native-speaker speech as well as SLA. According to Ellis (1994:2002), when it comes to linguistic changes, women are more sensitive to new forms but at the same time are more likely to reject them when they notice change. This would suggest that women are better at SLA because they are open to new linguistic forms in L2 and are more eager to get rid of their interlanguage that deviates from the standard norm.

Another factor is the social class or social status which takes into account income, level of education and occupation. Based on these elements, the members of various social classes have different world and life experiences, which subsequently influence their SLA. In general, the studies have shown that middle-class children do better than low-income working-class children when formal language learning is concerned. However, when it comes to communicative language learning, the social status proved to have no effect (Ellis 1994:206).

---

[39] The importance of social factors was already noted in the 19th century. Van Name (1869-70:124) argued that one of the reason for the development of creoles in the former European colonies was the mature age of the slaves that were taken there. The author suggested that when they arrived in the colony, their vocal organs were not so flexible anymore.

The last factor mentioned by Ellis (1994:207-209) is the ethnic identity, which can take three forms: normative, socio-psychological and socio-structural. The normative view is based on the fact that the bigger the distance between the native language and L2 cultures, the more difficult it is for learners to acquire L2 and achieve high levels of proficiency. The socio-psychological view takes into account the attitude the learners have towards L2. When they have a positive attitude towards the L2 culture and also towards their own native language, they may become 'balanced bilinguals'. However, if their attitude towards their mother tongue is negative, they substitute their L1 with L2, which is referred to as 'subtractive bilingualism'. The socio-structural view deals with the role that ethnic identity plays in the interactions between members of different ethnic groups.

In my analysis, I did not make a distinction between L1 and L2 speakers of Tetun Prasa since all my L2 speakers possessed a native-like proficiency in Tetun Prasa, regardless of their age, sex, social class or ethnic identity. However, if I had to choose one social factor that played a very important role in SLA, it would be ethnic identity. All the speakers showed positive attitudes towards Tetun Prasa and many of them, who have acquired it as L2, now use it as L1 (for practical reasons, rather than attitude-based). This brings us to another concept that is also pervasive in Timor-Leste, namely multilingualism.

3.1.5   Multilingualism

Multilingualism, or proficiency in more than one language, is found at societal as well as at individual level. At the societal level, it is possible to distinguish between 'official' and 'de facto' multilingualism. Some countries have several official languages, but their citizens often speak just one (e.g. Canada or Switzerland), depending on their geographical location. However, many other countries are 'de facto' multilingual, since multiple languages are used in different domains, just like in Timor-Leste. Societal multilingualism is based on various factors, e.g. international migration, colonialism, international borders, *Sprachinslen*, and spread of international languages (Clyne 1998:301-302).

This stratification, or domain-dependent language use, is called diglossia (or polyglossia). The domains rank from highly valued to less valued: the formal domains (media, public administration, Church) prefer the official language, informal domains (casual conversation) use national/local languages. Usually, the official language is of international prestige and is the language of the local elite and/or the Church, which puts the speakers of local languages at a disadvantage (Schiffman 1998). In Timor-Leste, polyglossia exists on two levels: certain languages are used in certain domains, but at the same time, different registers of the same language – Tetun Prasa – are confined to different domains. Portuguese is used in administration, in media and is one of the languages of instruction at schools while Tetun Prasa is confined to all these but also to the informal domain, although its registers differ. The high register of Tetun Prasa is used in the same domains as Portuguese, while low register is used in casual, informal speech (see section 1.3).

Multilingualism has always been the norm in Timor-Leste. With two official languages, two working languages and a couple of dozens of native languages, it is not surprising that people's attitudes towards bilingualism and multilingualism are mostly positive.

3.2 Methodology

In this work, I relied on my own data (written and oral corpus) as well as published work on Tetun (grammars, dictionaries, article). One of the biggest challenges I had to face was the selection of my speakers. The Timorese society is extremely diverse and multilingual and there are many variables influencing the speech of individuals. This inter-speaker variation has been already discussed in Williams-van Klinken et al. (2002a:3), in which the authors agreed that the unusually high levels of inter-speaker variation posed a significant problem to the description of Tetun Prasa. This variation is found on all levels – grammatical, lexical and phonological – and therefore different studies may yield different findings, as none of the previous works dealt with a large population. My results also sometimes contradict previous studies, which can be due to

the composition of the focus group and/or the time that has passed since these studies were published.

I conducted two research trips to Timor-Leste in 2014 and 2015. The objective was to collect recordings of natural speech from speakers of different varieties of Tetun and from different social backgrounds. My oral corpus consists of three different varieties: Tetun Prasa, Tetun Suai and Tetun Viqueque. Recordings of Tetun Prasa were collected in 2014 and 2015 in Dili, while Tetun Suai data come from 2014 and were collected in the town of Suai and neighbouring villages. In 2015, I visited another Tetun Terik speaking area – Viqueque – to consider possible variation between Tetun from Suai and Tetun from Viqueque. The data were collected in the town of Viqueque as well as in the village of Caju-Laran.

In total, I collected roughly 4 hours of recordings from 13 speakers in Dili, 15 speakers in Suai and 7 speakers in Viqueque. I led or participated in 12 recording sessions (6 in Dili, 4 in Suai and 2 in Viqueque) and used a digital voice recorder and written questionnaires. The number of speakers varies across the Tetun varieties, as some of the recordings were deemed unnatural (e.g. my assistants correcting the speakers) and were excluded. I also excluded data on Tetun Prasa collected outside of Dili to narrow down the disparity of my focus group. Ideally, I planned to find and record at least 10 speakers per each variety. In the case of Tetun Viqueque, the speakers were fewer but the total amount of time recorded was comparable to the other two varieties. The speakers were selected either by myself or by my assistents from the local office of Ministry of Education (*señór* Rui Amaral in Suai and *señór* Emídio Amaral in Viqueque). The aim was to make the focus group as respresentable of Timorese society as possible but after realizing the amount of variation found in Timor-Leste, I focused on selecting my spekaers based on the most important social variables: sex, age, level of education and proficiency in Portuguese. As a result, the selection was, to certain extent, subjective. I understand that it is difficult to draw any general conclusions from such a small array of speakers but since I am especially interested in language contact and variation, the volume of oral data was sufficient to do qualitative research. In the case of phonology and lexicon, I decided to attempt quantitative analysis to see whether any generalisations could be made.

All speakers were informed about the objectives of my research and were asked to fill in a questionnaire and sign a consent agreement. Speakers who were illiterate were either helped by me, or the assistants who helped me to approach the speakers. All the speakers participated voluntarily and were not paid. They were either asked to carry out a free dialogue between two or more people, or engage in free narration, or were interviewed by the author or the assistant (semi-structured interviews). In general, the free-dialogue technique was used during my first trip to Timor-Leste when my proficiency in Tetun Prasa was limited and I wanted to avoid influencing the speakers with my L2 variety. During my second trip to Timor-Leste, I conducted the interviews in Tetun Prasa myself, using semi-structured interviews while having had the questions proofread and corrected by a native speaker. Free dialogue or free narration were used when collecting data in Tetun Terik-speaking areas without me interfering in the process. When the dialogue was led by my assistants, their speech was also recorded and analysed. Although one can argue that the use of three different techniques: free dialogue, free narration and semi-structured interviews might have influenced the collected data, due to my initially limited proficiency in Tetun Prasa, no proficiency in Tetun Terik and lack of resources, these strategies seemed to be the most feasible options at the time of collection. Elicitation, another technique to gather information from people, was only used in case of structures that were mentioned in grammars/dictionaries of Tetun Prasa but did not occur in my written and/or oral corpus. My main informants in this case were: prof. Luís Costa (see 4.2.3.11), a native Tetun Terik speaker, my colleagues from National Institute of Linguistics in Dili and my friends who represented the younger generation (less than 30 years old).

The speakers were also informed about the possible topics they could discuss, usually ranging from family background, everyday life, history, current issues and traditional stories. At the very beginning of each interview, the speakers were asked to talk about themselves, their background, family and work. After that, they were free to develop their narration in whichever direction. Whenever they came to a halt, more questions were made concerning the current topic or a new topic/question was suggested. Again, one can argue that the difference in topics could have lead to a different set of data, a fact that I had to acknowledge after I had conducted the initial quantitative analysis at the level of phonology (e.g. see 7.5.2.6 and 7.6).

Being aware of the high level of variation, I decided to collect as much sociolinguistic information about the speakers' background as possible, although in the end, I decided to mainly focus on specific variables like age, sex, level of education, mother tongue, and proficiency in Portuguese (see Annex 1 for Questionnaire). This was mostly due to the fact that the information the speakers provided (about the languages they speak and use and about the languages of instruction) were subjective and I did not check their accuracy (see Annex 2 for the list of speakers and their sociolinguistic profile).

All the recordings were transcribed either by a native speaker (*señór* José Vicente) or by myself with the help of another native speaker (prof. Luís Costa). Although the transcribing done by myself was more laborious and time-consuming, I don't think it has affected the quality of the transcriptions, considering that they were all double-checked by prof. Costa. All transcriptions were done in ELAN, as shown in Image 1. I opted for this tool as it was recommended to me by my collegues from the CIDLES research centre working in the area of descriptive linguistics. Apart from transcriptions, ELAN also allows to add translation, a task that I abandoned due to the lack of time. However, all examples included in this thesis also come with an English translation. Aditionally, Annex 5 contains verbatim transcriptions of Tetun Prasa and Tetun Terik recordings.
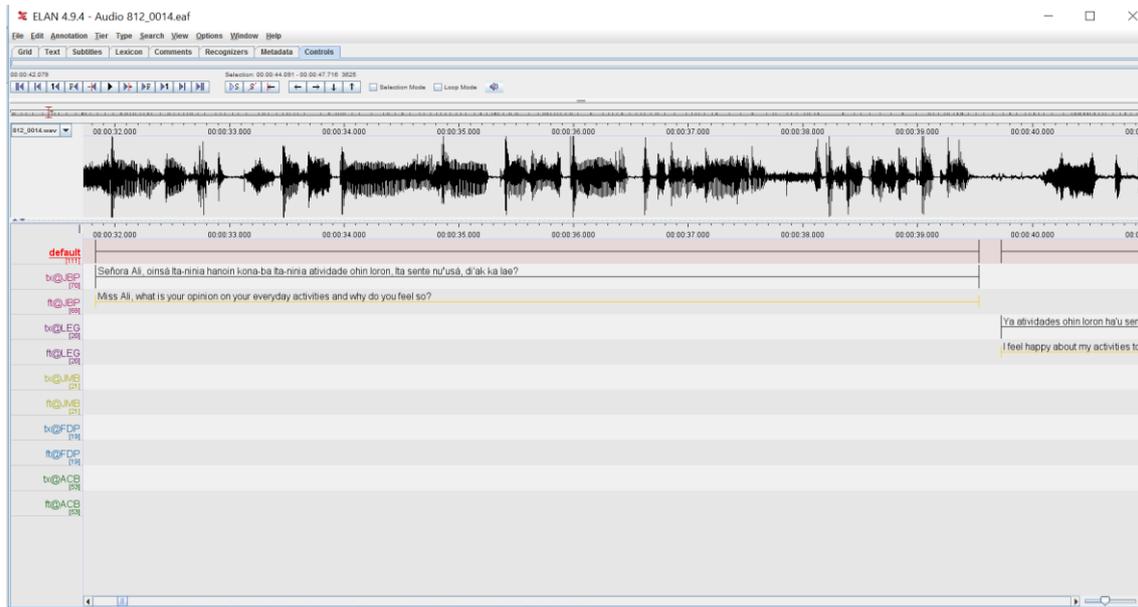
Image 1 Screenshot of ELAN

The data from ELAN were then exported as text files for every single speaker. If a speaker participated in two recordings, two different files were created. To make the file names easily recognizable and to keep the speakers anonymous, I established the following naming conventions:

file.number.Tetun.district_speaker.initials_age_sex_level.of.education_mother.tongue[40]

For example, *0006TD_BE_30_M_Uni_TP_Mak* refers to a recording no. 0006 in Tetun Prasa (TD = Tetun from the district of Dili) and to the speaker BE who is a 30-year-old male with a higher education and whose mother tongues are Tetun Prasa and Makasae. For the full set of abbreviation, see List of abbreviations, section *Recordings* (see also Annex 2 for the list of speakers and their full sociolinguistic profile).

Some native speakers of Tetun, when asked about their mother tongue, referred to the variety they spoke (Prasa or Terik) only as *Tetun* and I respected their answers as they gave me a bit of an insight into how speakers of different varieties view Tetun.

Besides oral recordings, I found it important to look at written sources as well, since the difference between speech and written language can appear striking, especially on the level of lexicon and syntax, where borrowings can make up to 60 %. Tetun represented in media is highly formal, regardless of topic, since many of the newspaper articles are directly translated from Portuguese. To build a written corpus, I decided to rely on newspaper articles published online between 2013 and 2016. I collected articles by *Suara Timor Lorosae* (STL) daily from July 2013 and March 2016 and by *Sapo* daily from March 2016. Unfortunately, I was unable to access the archives of the *Timor Post* daily. In addition to these, I collected news articles published by the Government of Timor-Leste on their website in 2010. All the articles were edited to follow the official orthography.

To analyse both corpora, I used AntConc and Excel. AntConc was crucial especially for morphosyntactic analysis, as it allows to search for concordances and is ideal for qualitative analysis (see Image 2). I also used AntConc for my phonological and

---

[40] In the questionnaire, informants were asked to specify their mother tongue. Many of them are bilingual as their parents come from different linguistic areas or the speakers grew up in a multilingual area. I did not test their proficiency in their mother tongue so in the context of this work, 'mother tongue' is a language the speakers grew up speaking at home.

lexical analysis in order to create word lists containing certain phonemes and borrowings, respectively.
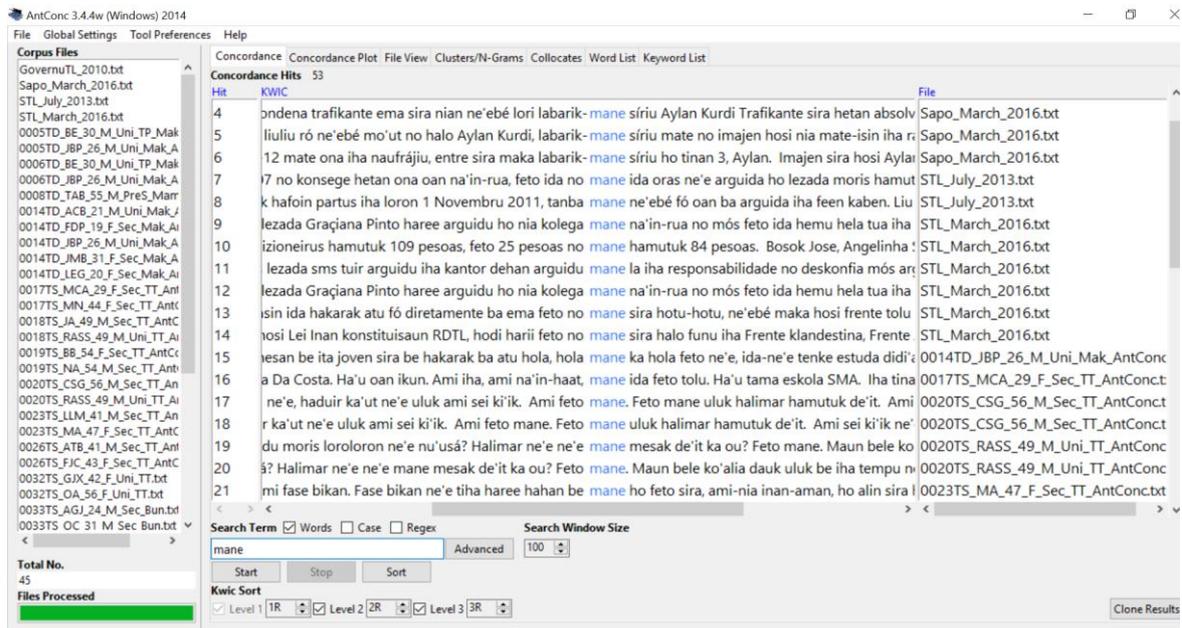
Image 2 Screenshot of AntConc

I found Excel to be the best tool for the qualitative and quantitative analyses required. I was able to pair my linguistic data (see Image 3) with the sociolinguistic variables (see Image 4) to create statistical analysis for borrowed phonemes and words (see Image 5).
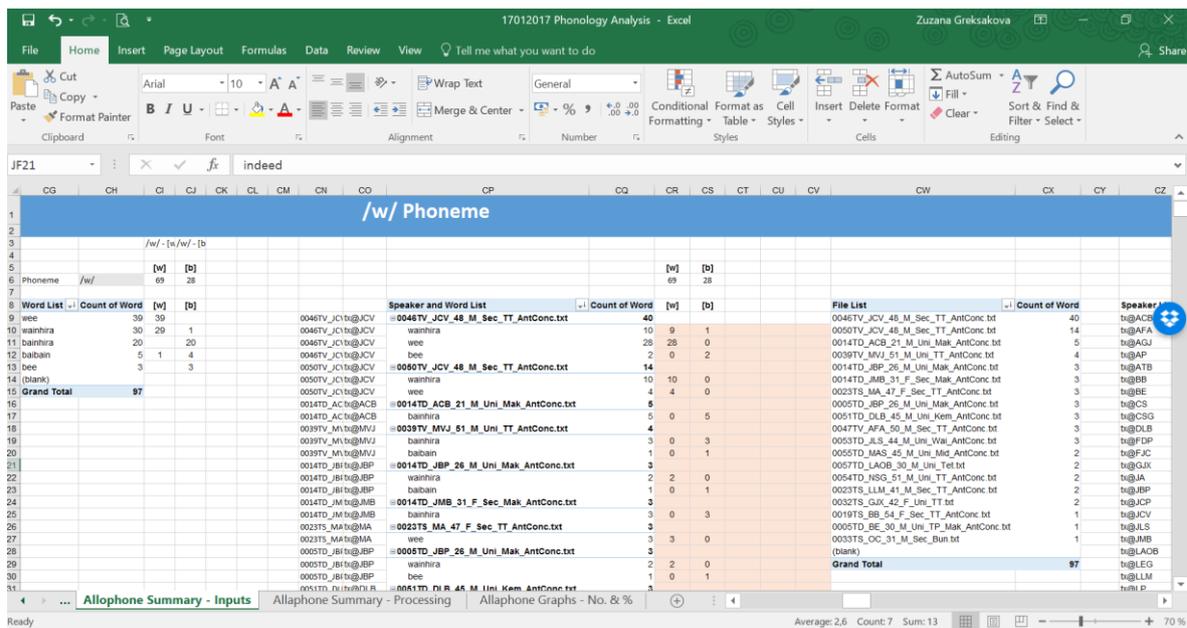
Image 3 Screenshot of the phonological analysis in Excel
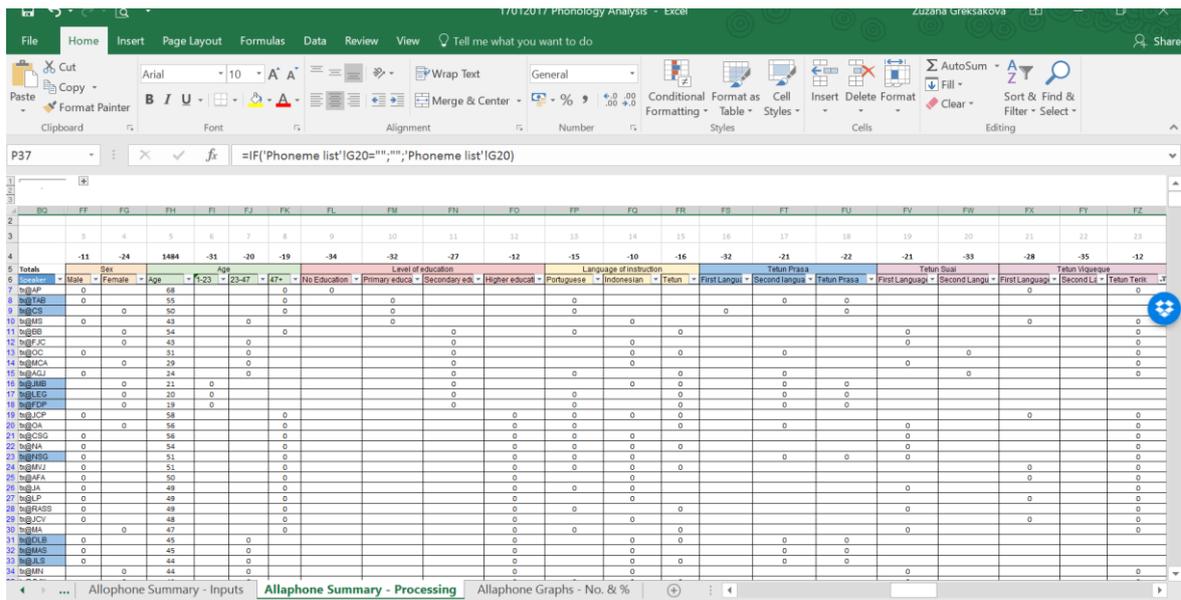
Image 4 Screenshot of sociolinguistic variables in Excel
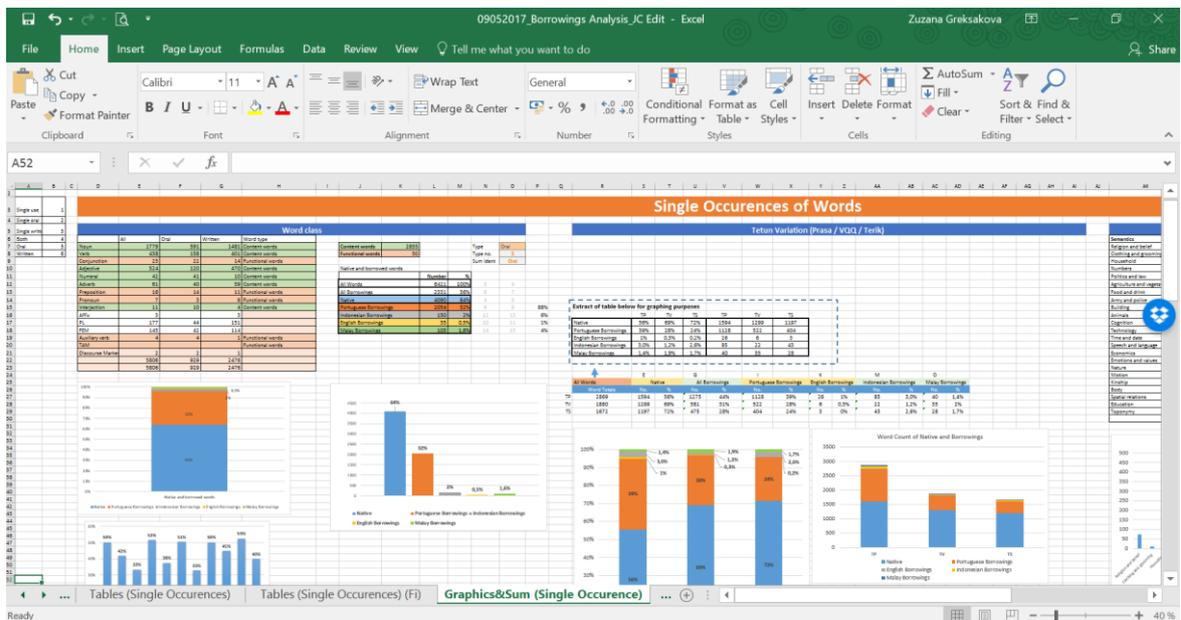
Image 5 Screenshot of lexical analysis in Excel

Despite the limited set of data, I also adopted the variationist approach and used the multivariational regression analysis to understand the 'relative' influence of independent variables (e.g. Tetun variety, age, sex, Portuguese proficiency) on the amount of borrowings Tetun speakers use (dependent variable). This analysis was also carried out in Excel which allows for statistical (Image 6) and graphical (Image 7) representation of the data. For more details on regession analysis and the results, see section 8.6.2 and for the list of independent variables see Table 75.

Tetun	No. Speakers	Level of education				Language of instruction		
		No Education	Primary	Secondary	Higher Education	Portuguese	Indonesian	Tetun
Tetun (All)	35	0,60%	1,61%	0,45%	0,01%	2,91%	0,08%	7,73%
Tetun Prasa	13	Insufficient data	9,37%	26,05%	4,62%	0,01%	0,56%	12,45%
Tetun Terik	22	0,04%	1,32%	0,68%	0,02%	4,65%	0,23%	23,57%
Tetun Suai	15	Insufficient data	Insufficient data	0,55%	0,55%	0,03%	0,22%	14,39%
Tetun Viqueque	7	0,00%	3,75%	Insufficient data	2,23%	60,80%	0,00%	60,80%

Tetun	No. Speakers	Sex		Age	Age groups		
		Male	Female	Age	1-23	23-47	47+
Tetun (All)	35	0,02%	0,02%	2,05%	0,19%	4,57%	5,76%
Tetun Prasa	13	28,82%	28,82%	19,29%	26,20%	8,48%	4,67%
Tetun Terik	22	3,48%	3,48%	0,83%	Insufficient data	3,72%	3,72%
Tetun Suai	15	4,32%	4,32%	2,30%	Insufficient data	8,56%	8,56%
Tetun Viqueque	7	Insufficient data	Insufficient data	2,94%	Insufficient data	3,75%	3,75%

Tetun	No. Speakers	Proficiency in Portuguese		
		Speaks	Understands	No understanding
Tetun (All)	35	0,85%	1,97%	8,08%
Tetun Prasa	13	54,63%	31,05%	6,90%
Tetun Terik	22	4,58%	0,07%	4,60%
Tetun Suai	15	2,94%	0,18%	2,91%
Tetun Viqueque	7	7,83%	Insufficient data	7,83%

Image 6 Statistical representation of the R-square values for independent variables
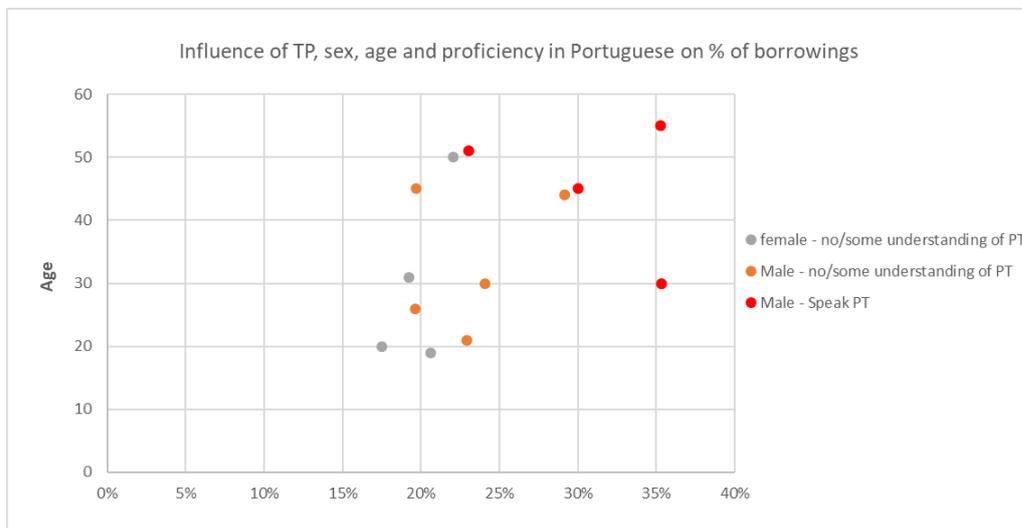
Image 7 Graphical representation of regression analysis with 4 independent variables

94

4. REVIEW OF THE LITERATURE

4.1 Introduction

This chapter will cover two interrelated lines of development of Tetun linguistics: the literature on Tetun Prasa and the works on Bidau Creole Portuguese. I will review relevant literature on the languages of Timor-Leste, particularly Tetun, with the first mentions of Timor dating back to as early as the 13th century (section 4.2.1) and the Portuguese presence in this region from the 16th century (section 4.2.2). I will further describe the contributions of Portuguese missionaries and others dating back to the 19th century (section 4.2.3). The last section 4.3 is dedicated to the discussion on Bidau Creole Portuguese.

4.2 The literature on Tetun

The literature on Tetun before the 17th century is very limited. This is due to various facts. Firstly, the Portuguese came to Insular Southeast Asia[41] by sea to trade and used Malay as a lingua franca (Thomaz 2002:86). Secondly, Tetun, like the other native languages of Timor-Leste, was only oral until writing was introduced in Timor by the Portuguese in the 16th century. Thirdly, the Portuguese only really settled down in Timor in the 18th century when the capital was moved to Dili.

4.2.1   Chinese and Malay accounts of Timor (13th – 17th century)

The proto-historic period can be dated from the 6th to the 16th century A.D. Large quantities of Chinese porcelain in Timor suggest that direct relations with China date back to the Tang dynasty (618 to 907 A.D.) (Thomaz 1974:244). However, Chinese accounts of Timor island are found no earlier than the 13th century when the geographer Zhao Rugua published a descriptive account of several exotic countries, including Timor,

---

[41] Insular Southeast Asia is a geopolitical term describing the former Dutch and Portuguese colonies in the East Indies.

called *Records of the Various Foreign Nations,* although he himself never visited these places. In his book, Zhao said: "They do not have writing or calculation. (…) Local products are sandalwood, cloves, nutmegs, decorated bamboo mats, cloth, iron, swords, various vessels, and other things." (Eccles 2004:180-181 citing Zhao 1930:28).

Further Chinese explorations of the coast of Timor were interrupted during the Ming dynasty (1368 to 1644 A.D.). In 1435, the Chinese kingdom had to deal with invading nomads in the north of China and thus maritime expeditions were suspended. This did not, however, stop the Chinese from trading in Southeast Asia (Thomaz 1974:245). An account of these clandestine travels can be found in Zhang Xie's *Investigations into the Eastern and Western Oceans* (1617):

> "The fields are fertile, and grain is abundant. All along the mountains sandalwood grows, which is cut down for fuel. (…) Men and women cut their hair and wear short skirts. (….) They do not have surnames, nor do they reckon the years, nor do they have writing. They record things with pieces of stone, and for example, they indicate one thousand pieces of stone by a knot on a cord." (Eccles 2004:182-183 citing Zhang 1937:54-55).

It is possible to identify the ports and settlements where Chinese did most of their commerce. They were located mainly in the west of the island and on the north-western shore, as is evident from Map 4.
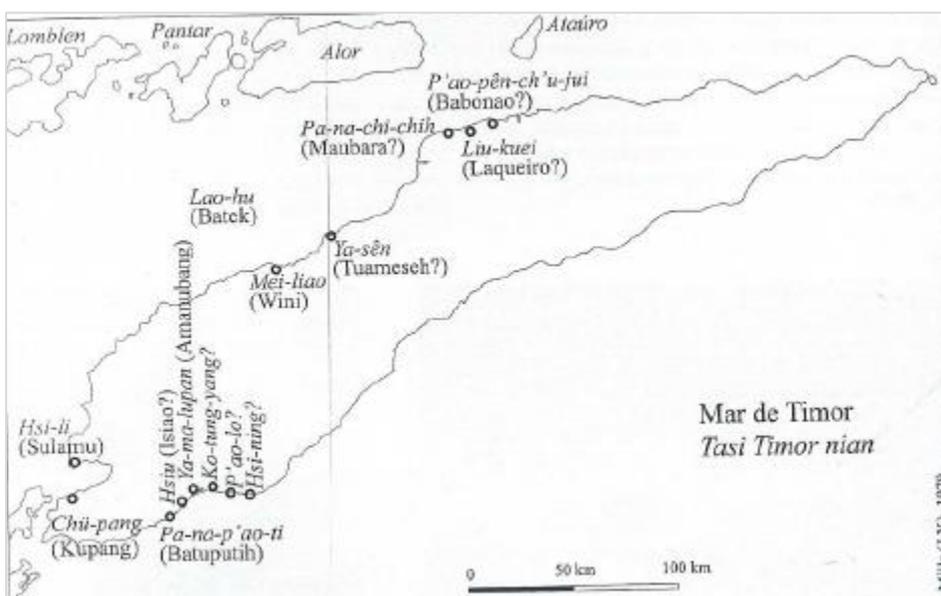
Map 4 The ports of Chinese trade in Timor in the 15th century (Durand 2010:38)

As we can see, the descriptions of the island and of the people that live there are very similar. None of the Chinese authors mentioned the language that they used while trading with local people or the language that the local people used among themselves. However, it is clear that they did not use any writing which supports the assumption that writing was introduced in Timor only with the arrival of the Portuguese in the 16th century.

The earliest Malay accounts of Timor date back to the 14th century. There is a mention of Timor, among other islands, in the famous epic poem Nagara-Kertagama written by the Buddhist monk Rakawi, Prapanca of the Majapahit Kingdom, as a eulogy to the Javanese king Hayam Wuruk. The author lists islands (with Timor being one of them) 'east from the Javanese country', that were 'mindful' (Pigeaud 1960:17), which, according to Pigeaud (1962:35) meant "a relationship of submissiveness". However, I was unable to find any records of the Malays possibly bringing writing to Timor.

4.2.2   Portuguese presence in Insulindia (16th – 18th century)

Only one year after the Portuguese seized the strategic base of Malacca [Malay: Melaka] (see Rio Malaca on Map 5 below) in 1511, the first expedition was sent to explore the islands to the east. The proof that they indeed reached Timor can be found in the maps of a Portuguese cartographer, Francisco Rodrigues, the pilot of a ship whose captain was Simão Afonso Bisagudo. The destination of their voyage was the "Spice Islands" (Moluccas or Maluku Islands in Malay) but they went by way of Timor – "a ilha de Timor onde nasce o sândalo"[42] (Durand 2010:52). Timor became a regular stop for Portuguese ships heading to the Maluku Islands as documented in numerous historical records. What is lacking is a sufficiently detailed account of the ethnolinguistic history of the Timorese people. As Morais (1944:81) put it: "Sólor e Timor foram objecto do maior desinterêsse por parte dos cronistas, o que aliás é perfeitamente explicável, dado a sua mínima importância, no decorrer dos acontecimentos no Oriente."[43] The Dutch and Portuguese coming to Timor were interested only in the abundant reserves of sandalwood,

---

[42] "the Island of Timor where sandalwood grows" (my translation)
[43] "Solor and Timor were not an object of major interest on the part of chroniclers, which is, in fact, perfectly understandable, considering their marginal importance in the history of the Orient." (my translation)

wax, and slaves[44] and since they were able to communicate with the locals in simplified versions of Portuguese and Malay, they either did not pay any attention to local languages (Hull 1998a:6) or their accounts were very brief, like the one written by Duarte Barbosa in 1516: "Indo mais ao diante, deixando estas ilhas de Jaoa Maior e Menor, ao mar delas estão muitas outras, grandes e pequenas, povoadas de gentios, e mouros alguns, entre as quais está uma que chamam Timor, que tem rei e língua sobre si."[45] (Barbosa 1992:172).

The situation started to change after various religious orders began their attempts to convert local people to Christianity. One of the first observations of a local language whose identity is unknown came from a Jesuit priest in Goa, Baltasar Dias, in a letter sent to the Provincial Superior of India in 1559: "A lingoa desta gente dizem ser muito curta, conforme em algumas cousas com a malaia."[46] (Sá 1955:345).
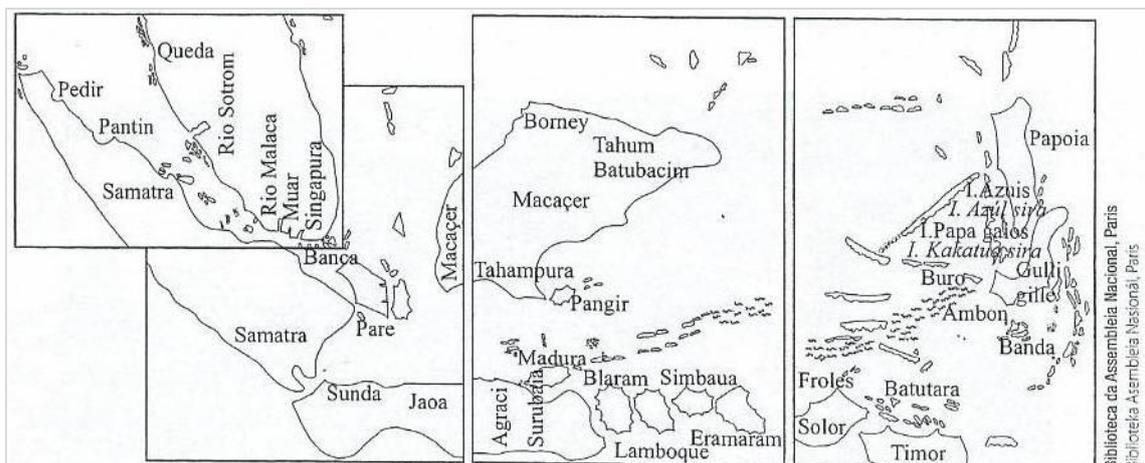
Map 5 Southeast Asia around 1512 according to Francisco Rodrigues (Durand 2010:51)

One of the first references to Tetun can be found in a document dating back to 1624 or 1625: "Nesta ilha há duas lingoas somente, distintas huma da outra que chamão Vaiquenos e Bellos."[47] (Sá 1956:492). The anonymous author was clearly referring to the

---

[44] Although detailed description of the slave trade in Timor is difficult to find, various authors acknowledged slaves as one the main trading commodities. Sá (1961) described the Timorese society as consisting of slaves (*ata*), the people (*ema wai-wai*), chiefs (*dato*), older men acting as advisors (*katuas*) and traditional Timorese rulers (*liurai*). Villers (1985:582:596) noted that the slave trade was very brisk and that slaves were exported all over the archipelago, although the Portuguese never played an important role. On the other hand, by the end of the 17th century, the trade slave, among others, had become Macau's main source of revenue.

[45] "Going further on, leaving these islands of Java Major and Minor behind, there are many others in its sea, small and big ones, inhabited by heathens and some Moors, and among them all there is one called Timor that has a king and its own language." (my translation)

[46] "They say the language of these people is very laconic, in some ways similar to Malay." (my translation)

[47] "In this Island, there are only two languages, distinct from each other, that are called Vaiqueno and Bello." (my translation)

Baikeno language of the Atoni people living in West Timor and the Timor-Leste enclave Oecussi, and to that of the Belu people in the central and eastern part of the southern coast who used Tetun as their mother tongue (Thomaz 1974:217). There is another reference to four local languages from the same period (1635) made by Father António da Encarnação: "A chamada Thimor, tem quatro linguas differentes em si (…)"[48] (Sá 1955:308). Unfortunately, the author did not identify them.

Besides the Portuguese, there were also other Europeans interested in the languages of Timor. The first word list of Tetun was written in 1772 by a Frenchman F. E. de Rosily although the manuscript of his glossary was never published. Thomaz (1982:107-108) wrote a short review of the *Dicctionaire François e Timorien*. The dictionary contains 417 entries from various languages. Thomaz managed to identify words from Makasae, Galolen, Waima'a, Tetun (see Map 6 below), Malay and Portuguese[49]. He was also surprised how little attention Rosily paid to the two vehicular languages, Tetun and Malay. This could be partly explained by the fact that Rosily mostly stayed in small ports between Laivai and Manatuto in the eastern part of the island where Tetun was little used (Thomaz 1982:107-108).
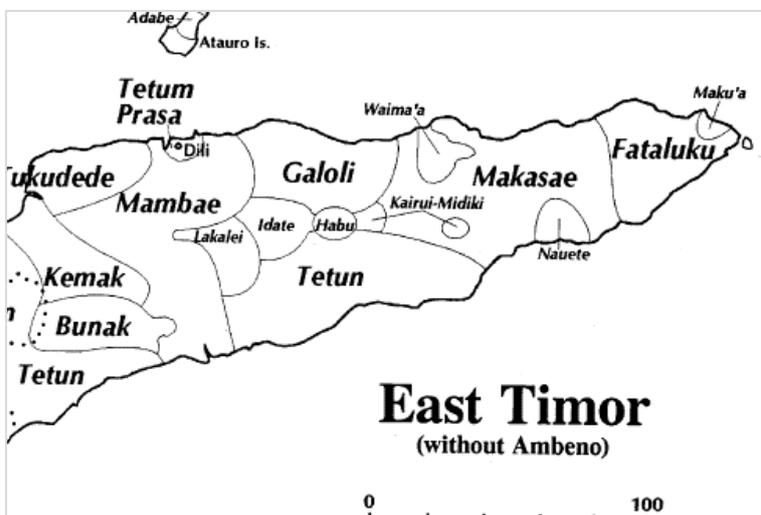
Map 6 The languages of the Timor, excluding Baikeno spoken in the Oecussi enclave (Grimes et al. 1997:38)

---

[48] "The one [island] called Timor has four different languages." (my translation) There is a comment in a footnote made by Sá saying: "O número dos dialectos em Timor, ainda hoje está por definir, ao certo, mas contam-se bem mais de seis." (Sá 1955:308) ["The exact number of dialects in Timor is, even nowadays, only to be set, but the figure is surely more than six." (my translation)]

[49] Rosily transcribed the Portuguese words phonetically: *cobi* 'couve' (cabbage), *roumaon* 'romã' (pomegranate), *carnorou* 'carneiro' (ram), *alfasi* 'alface' (lettuce).

4.2.3 The contributions of Portuguese missionaries and others (19th – 21st century)

Although the Portuguese first reached Timor at the beginning of the 16th century, it was not until 1636 that they actually settled the island. Around 1550, their main base was Solor, from where the Dominican friars regularly carried out trade for sandalwood with four Timorese kingdoms: Mena, Matomea, Camenaça and Servião. When the Dutch withdrew from the area around Solor, the Portuguese established permanent settlements in Timor, mainly around Kupang and Oecussi (Durand 2010:57). Soon they managed to convert various kings and queens to Catholicism and when they defeated the kingdom of Waiwiku-Wehale in 1642, the Portuguese and their faith gained a more prominent role and a Vicar General of Timor was named in 1646. The government was officially transferred from Larantuka (on the island of Flores) to Lifau (in the Oecussi enclave) in 1702 but after the endless rivalry between the Portuguese and the Topasses, the Portuguese governor was forced to leave Lifau and establish a new capital in Dili in 1769 (Baxter 1990:2). For a more detailed discussion see Chapter 2.

4.2.3.1 Affonso de Castro (1867)

There is a short account of the languages spoken in Timor-Leste by Affonso de Castro (1867), a former governor of this territory appointed by the Portuguese Court between 1859 and 1863. Castro mentioned four languages – Tetun, Galolen, Baikeno and Calado[50]. These are his impressions of the local dialects, as he referred to them:

> "São mui pobres todos estes dialectos e nenhum d'elles tem grammatica nem escripta. Linguas inteiramente selvagens, mui asperas e nada parecidas com a suave lingua malaia, que tem seus poetas e bons prosadores. O *Teto* usa muitas palavras portuguezas, que os naturaes de Timor adoptaram, à falta de termos para

---

[50] Castro (1867) probably mistook *Calado* for Mambae which is spoken in the mountainous region around Dili. People living in this central area go by the name of *kaladi* and are usually speakers of Mambae (Luís Costa, p.c.). *Kaladi* is nowadays also used for people from the western part of Timor-Leste and can have a derogatory connotation. According to folk etymology, some people believe that *kaladi* is derived from Portuguese *calado* 'taciturn, calm' since people from the west of Timor-Leste are thought to be less hot-blooded than the *firaku* people from the eastern end of the island. However, it is more plausible that *caladi* is derived from the Malay word for 'taro' *keladi* which is grown in this area.

exprimir objectos que lhes eram desconhecidos antes da conquista."[51] (Castro 1867:328).

As we will see below, this used to be the general view of the local languages and Tetun. This was also possibly the first mention of the Portuguese borrowings in Tetun. But what is more interesting is Castro's reference to the Alfur language:

"Tendo sido a ilha de Timor povoada por emigrados das Molucas e de Borneo, é natural que a lingua *Alfur* fosse fallada pelos primeiros timores n'aquellas remotas eras; mas com o andar dos tempos soffreu taes alterações, que hoje nada se parece com aquella, e não deve admirar tal diferença se notarmos que o Alfur tambem não tem grammatica, nem escripta."[52] (Castro 1867:328).

Castro was wrong in assuming that the Timorese people came from Borneo, since there is no evidence of Bornean influence[53] in Timorese languages but he might have been right about the Alfur language coming from Maluku[54]. However, it is not clear which language the author was referring to, so we can only hypothesize that it was one of the four Papuan languages spoken on the island nowadays, since it is widely accepted that these languages were spoken in Timor before the arrival of Austronesians.

---

[51] "All of these dialects are very poor, none of them has grammar or writing. The languages are entirely savage, very vulgar and nothing like the delicate Malay language with all its poets and good prose writers. Tetun uses many Portuguese words adopted by the natives of Timor due to the lack of terms to describe objects that had been unknown to them before the conquest." (my translation)

[52] "Since the island of Timor had been populated by emigrants from the Moluccas and Borneo, it is natural that the *Alfur* language was spoken by the early Timorese people in those ancient days; but in the course of time it suffered such alternations that nowadays it is not similar to it at all. We should not be surprised by such a difference, though, if we realize that Alfur does not have any grammar or writing either." (my translation)

[53] Historical linguists agree that Malay language originated in Western Borneo. However, the languages of Borneo belong to two distinct groups of Austronesian languages – Bornean languages and Malayic languages (a branch of Sumba-Sulawesi group), none of them containing languages of Timor-Leste.

[54] It was previously thought that *alfuro* is derived from a Portuguese word *forrar* 'to manumit' or *alforria* 'manumission' as the term was used to denote the animist communities in eastern Indonesia. These communities usually lived further inland in the Moluccas and Papua and it was believed that these people constituted the aboriginal inhabitants who lived in this area prior to the Papuan and Austronesian migration waves (Ballard 2008:198). However, according to the Dutch National Archives and AJ van der Aa's 1939 Toponymic Dictionary, "the inhabitants of the Moluccas called themselves *haraforas*, translating Indonesian *anak anak gunung* as 'children of the mountains'."

4.2.3.2 José dos Santos Vaquinhas (1883-1888)

Even more information about the linguistic situation in Timor is available in the letters from José dos Santos Vaquinhas to both Hugo Schuchardt and the Geographic Society of Lisbon *(Sociedade de Geographia de Lisboa)*. Major Vaquinhas, who spent around 18 years on the island and served as the interim governor of Timor between 1880 and 1881, sent a series of three letters to the Society in 1883 with references to Batavia Creole Portuguese, Tetun and Portuguese spoken in Timor. As for Batavia Creole Portuguese (see also section 4.3), he asserted that there was indeed a neighbourhood in today Jakarta called *Jambata-dua* (literally 'two bridges') inhabited by Portuguese descendants who were Christians and spoke Portuguese, "ainda que muito estropiado"[55] (Vaquinhas 1883:277). The author further recalled that ten years earlier there used to be a priest giving sermons in Portuguese, however, at the time Vaquinhas wrote the letters, the mass was conducted in Malay and Dutch. The presence of the Portuguese not only in Batavia but in the whole Indonesian archipelago had influence on Malay language as well as other local languages spoken on the numerous islands and lacking many words and phrases. Vaquinhas noticed the same 'lexical poverty' in Tetun quoting a quartet of a Timorese song:

> "A presente quadra é composta do idioma malaio e portuguez como era o gusto d'aquella epoca, cuja explicação facilmente se comprehenderá, sabendo-se que tendo os timorenses, para exprimirem as suas idéas e pensamentos, apenas um dialecto pobrissimo em vocabulos, se serviam do malaio por ser mais rico e melodioso."[56] (Vaquinhas 1883:280).

> *Rajá Bellus, raja provincial*
> *D. Paulo de Laculó*
> *Suda-denga conselho de padre*
> *Tingar contra o governo*[57]

---

[55] "although very garbled" (my translation)

[56] "The present quatrain is composed of Malay and Portuguese language as was common in that period, and the explanation for that can be easily understood, since we know that the Timorese only possessed a dialect with a very poor vocabulary to express their ideas and thoughts and thus they made use of Malay, which was richer and more melodic." (my translation)

[57] "The kings of Bellus and the province / and D. Paulo de Laculó / heard the advice of the priests / and rebelled against the government." (my translation) The underlined words come from Malay, the rest comes from Portuguese.

4.2.3.3 Aparício da Silva (1885, 1889)

The first public schools were not opened until a century later when the first religious and language studies began to be published. One of the pioneers was a Jesuit priest Sebastião Maria Aparício da Silva, a missionary stationed in Timor between 1877 and 1910. In 1885, he compiled a catechism in Tetun (*Catecismo da Doutrina Christã em Tétum*), the first known publication in Tetun, followed by a Portuguese-Tetun dictionary (*Diccionario de Portuguez-Tétum*) in 1889, both published in Macau. As the author said in the preface to the catechism:

> "Parecerá arrojo da minha parte esta empreza, por nada haver escripto em qualquer dos muitos dialectos que n'aquella ilha há, como V. Ex.ª R. ᵐᵃ muito bem sabe, que me podesse servir de guia para aprender bem a lingua em que escrevi este catecismo, accrescendo a grande difficuldade de explicar algumas verdades da nossa Santa Religião, pela deficiencia de termos para exprimir ideias abstractas."[58] (Silva 1885:III-IV).

It took him more than eight years to finish the catechism and another four to compile the dictionary. In the preface to the dictionary, the author noted that Tetun was the "dialecto o mais generalizado na ilha de Timor, pelo menos na parte que pertence a Portugal"[59] (Silva 1889:I). Interestingly, Silva regretted choosing the dialect of Dili for his dictionary:

> "Dois principaes noto eu, o primeiro dos quaes eu podia ter remediado, mas já era tarde quando o quiz fazer, e é ter formulado as phrases e exemplos em *tétum* segundo a construcção de Dilli, devendo ser segundo *a do interior por ser a mais correcta* (my emphasis ZG). (…) Nos reinos do interior, onde se falla melhor, de reino para reino ha alguma differença, apezar de não ser tão grande como a que se nota entre o de Dilli e o do interior da ilha, sendo o modo de construir outro, principalmente quanto a uma especie de conjugação de verbos [see section 5.2.1],

---

[58] "It might seem bold of me to use it, since nothing has been written in any of the many dialects that are spoken on the island, as your Excellency knows very well, that might have served me as a guide to learn well the language in which I wrote this catechism, making it even more difficult to explain some of the truths of our Holy Religion due to the lack of words that would explain abstract ideas." (my translation)
[59] "the most wide-spread dialect on the island of Timor, at least in the part that belongs to Portugal". (my translation)

que só no interior há, e quanto ao uso dos complementos restrictivos [see section 6.3.2], como se notará nas regras gramaticaes que precederão o diccionario."[60] (Silva 1889:iii-iv).

There are, indeed, a couple of pages dedicated to grammatical rules and it is believed that Silva also published the Tetun grammar separately. Unfortunately, it has been lost (Hull 1998a:8).

4.2.3.4 Works by other authors (1900-1916)

Since not all Catholic missionaries served in Tetun-speaking regions, it is understandable that grammars, dictionaries and translations in other local languages emerged, too. In 1900, a Catholic priest, Manuel Maria Alves da Silva, published a Galolen grammar (*Noções da Grammatica Galoli: dialecto de Timor*), followed by a catechism in 1903 (*Catecismo da doutrina christã em portuguez e galoli*) and a dictionary in 1905 (*Diccionario portuguez-galoli*). According to Hull (1998a:8), another Catholic priest, Manuel Martins Pereira, who served in Timor between 1899 and 1904, compiled a short Tokodede-Portuguese vocabulary that has remained unpublished. Another set of dictionaries that never came to be published were compiled by the priest Manuel Calisto Duarte Neto who was stationed in Timor between 1896 and 1902. According to Fernandes (1931:124), there were three dictionaries ready to be published: a dictionary of Waima'a, a dictionary of Makasae and a dictionary of Baikeno. Correia (1934:357) also knew about Neto working on these dictionaries but to his knowledge, there were only two dictionaries: one of Waima'a and Makasae and one of Baikeno.

Dutch Jesuit priests in West Timor were also studying local languages. Alfons Mathijsen served as a missionary in the interior of the island where Tetun was spoken. He compiled a Tetun-Dutch dictionary (*Tettum-Hollandsche woordenlijst met beknopte*

---

[60] "I observed two main [errors]: the first one I could have avoided but it was already too late – I wrote all phrases and examples in Tetun according to the Dili variety instead of writing them *in the more correct variety spoken in the interior* (my emphasis ZG). (…) In the kingdoms in the interior, where they speak better [Tetun], the differences vary from one kingdom to another, but they are still not as great as the ones between Dili and the interior of the island, for example in the way they construct the language, especially as to the conjugation of verbs that is found only in the interior and as to restrictive complements that is evidenced by the chapter on grammatical rules preceding the dictionary." (my translation)

spraakkunst) that was published by Batavia Society for Arts in 1906, also with a concise grammar (Steenbrink 2007:156).

However, one of the most renowned works was written by a Portuguese lieutenant-colonel Raphael das Dores who published the first Tetun-Portuguese dictionary (*Diccionario teto-português*, 1907), continuing and complementing the work of Father Silva. The author also added some grammatical notes, explaining the phonology and morphology of the language and a short glossary. According to Gonçalves Vianna, the author of the preface, Dores compared the dialect of Tetun with coastal Malay, paying attention to the morphological, syntactic and lexical similarities. The latter were either cognates or borrowings that "por transmissão directa passou do malaio ao tétum, conservando quasi intactas as suas feições originaes"[61] (Dores 1907:ix). Dores also noticed many Portuguese borrowings that were introduced in Tetun by missionaries but he did not list these in his dictionary since he was against the introduction of foreign words into an indigenous language and preferred to use native periphrases. He blamed missionaries for using Portuguese words while giving their sermons in Tetun. Greetings are a good example. Although there are native words for 'good' (*di'ak*), 'morning' (*loron*) and 'night' (*kalan*), locals adopted *bondia* and *bonoite*[62] which, according to Dores (1907:7): "tem feito quebrar a cabeça aos estudiosos que de paises longiquos se dedicam ao estudo circumspecto das linguas crioulas"[63]. Dores was also one of the first to suggest calling Tetun (as well as other local languages) a language and not a dialect[64] and he also refused to use the Portuguese name 'Tétum' to denote it. Looking at the grammatical notes, we can assume that Dores chose as a model the dialect of Dili:

> "Alem d'esta razão, que é muito importante, temos que *na capital, onde se falla o teto talvez mais correctamente* (my emphasis ZG), não existe tal conjugação [de verbos], como tambem não existe na lingua malaia, muito mais conhecida e

---

[61] "passed by direct transmission from Malay to Tetun, preserving almost intact their original form." (my translation)

[62] The reason why Portuguese greetings were adopted might have to do with the fact that Timorese usually greet each other with a question, for example, *Bá ne'ebé?* 'Where are you going?'.

[63] "has been a puzzle for experts from distant countries who dedicate themselves to studies of creole languages." (my translation)

[64] "É, pois, d'essa linguagem ou dialecto que vou tratar, começando por chamar-lhe lingua, resolução que submetto á conspicua apreciação dos mestres em philologia." (Dores 1907:1) "I will, indeed, deal with this tongue or dialect, and I will start by calling it a language, a decision that I subject to the respectable judgement of masters in philology." (my translation)

estudada, portanto não parece racional que ella exista no interior (…)."[65] (Dores 1907:18).

Another missionary, Manuel Mendes Larangeira, compiled a Tetun spelling book (*Cartilha-Tetun*, originally 1916, reprinted in 1932) to teach local children read in Tetun and eventually also in Portuguese. He chose to do it via Tetun since it was generally understood everywhere, even though it had undergone adulterations due to the influence of other languages. Since he spent a lot of time in the southern region of Timor, he decided to use the Tetun he learned there, although he was not sure whether it was the purest and the most traditional variety. The author further believed that Tetun originated from Malay, having later turned into an independent language with a lot of variation in phonology and morphosyntax due to the lack of a written form (Larangeira 1932:iv-v).

4.2.3.5 Works from 1920 to 1940

Apart from the grammars, dictionaries and religious works then being written in and about local Timorese languages, ethnographic and linguistic works started to emerge, too. It appears that the main concern was to determine how many languages there were in Timor, what their origin was and if they were linked to some specific ethnic group.

In 1920, Leite de Magalhães, an infantry lieutenant whose mission in Timor was in the field of agricultural development, wrote an ethnographic paper on Timor. The author (1920:50-51) divided the 25 languages of Timor-Leste into eight distinctive groups that, according to him, also represented peoples of different origin established on the island. He believed that Bataks from Sumatra were the first colonizers of Timor and argued that the second group, where the Tetun language belongs, was of Sumatran origin, too, although the influence of Sulawesi languages (especially from around the city of Manado in the north of the island) was much more evident (Magalhães 1920:57). According to the author, the Malay influence in Timor was overestimated: first of all, in the author's view, Malay was either a sister or a daughter language of Batak, modified by

---

[65] "Apart from this reason, which is very important, we see that *in the capital, where they probably speak more correct Tetun* (my emphasis ZG), this kind of conjugation [of verbs] does not exist, as it does not exist in the Malay language and thus it does not seem rational to exist in the interior." (my translation)

Sanskrit and Arab influence. Secondly, there was no evidence of Islam among the population, which was unlikely in places where Malay had established itself (Magalhães 1920:54-55). Although it was true that ships from all over Insular Southeast Asia came to trade in Timor, it has been widely accepted that all Austronesian languages of Timor-Leste belong to Central-Eastern Malayo-Polynesian branch, unlike languages of Sumatra (where Batak belongs), which are part of the Sunda-Sulawesi branch (Wouk & Ross 2002). Also, the Malay influence in these languages is undeniable since Malay has been the lingua franca in this region for centuries.

A Portuguese commander in Timor, Júlio Garcez de Lencastre (1929:82) suggested that Tetun, the most wide-spread language on the islands, contained a large number of Malay words and a few Portuguese borrowings. On a few pages, the author also presented grammatical and pronunciation rules of Tetun together with a short Portuguese-Tetun vocabulary.

In 1929, a German scholar, Hermann Fiedler, published a book about the island of Timor (*Die Insel Timor*, 1929), in which he dedicated a short section to the languages spoken there. As a linguist, he was able to recognize sound changes in different languages spoken on the island but he was not able to tell how many there were and assumed that apart from Marai and Kemak, all other languages spoken in the Portuguese part could be just as well dialects of Tetun (Fiedler 1929:82). He also noticed that Tetun changed more than other Timorese languages (such as Dawan, Helong, Rote) and did not use subject marking on verbs (Fiedler 1929:81).

Father Abílio José Fernandes (1931:19), based in Timor between 1914 and 1938, referred back to the work of Teofilo Duarte (1930)[66], wondering which twenty languages the author had in mind. According to Fernandes, there were only fifteen languages. In 1937, Fernandes published a short Tetun textbook (*Método prático para aprender o tétum*), divided into three parts: rudiments of grammar with examples, often contrasting Tetun Prasa with Tetun Terik; a list of the most common words; and a short Portuguese-Tetun vocabulary. The author based his work on the Dili dialect:

---

[66] The author commented on the wide variety of languages spoken in Timor-Leste, suggesting that some were incomprehensible to people of other ethnic groups, e.g. a dialect of Lautem. He added that there were more than 20 languages: "Dialectos contam-se mais de 20 (…)." Duarte (1930:123).

"Procurei usar de termos geralmente conhecidos e adoptei a linguagem usada na costa Norte porque, a-pesar-de de menos correcta, é compreendida em tôda a parte, o que não sucede com a linguagem das regiões do interior, onde o tétum é língua própria. (…) A linguagem usada nos reinos de Samoro, Barique, Alas, Lacluta e Luca é mal compreendida pelos habitantes de Dilly e das regiões de dialecto diferente, ainda que falem suficientemente o tétum de Dilly."[67] (Fernandes 1937:5-8).

A difficult task was undertaken by two missionaries, Manuel Mendes Larangeira, mentioned in 4.2.3.4, and Manuel Patrício Mendes, who were appointed in 1915 by their superior to collaborate on a Tetun-Portuguese dictionary. It took them several years to compile and edit the dictionary and another few years to finally publish it in Macau in 1935. The dictionary (*Dicionário tétum-português,* 1935) contained about 8000 entries from Tetun Terik, Tetun Prasa and Western Tetun varieties, and orthographic and phonetic rules of Tetun based on *Cartilha-Tétum* (1916), see above (section 4.2.3.4). In the preface to the dictionary, Mendes refused to choose one correct form of Tetun, arguing that: "esta língua varia muito de região para região e que não há bases nenhumas que nos autorizem a considerar mais pura e legítima uma palavra usada num sítio do que a sua correspondente usada noutro onde também se fale tétum"[68] (Mendes 1935:ii). Mendes also noticed the intensive language contact Tetun has been in, saying that it was inevitable that neighbouring languages influenced each other, "[m]uito menos o tétum que não tem clássicos nem bases seguras de investigação"[69] (Mendes 1935:iii).

Another two missionaries started working on dictionaries of local languages, but their efforts were interrupted by World War 2. Father António Grebaldo da Conceição Fernandes, stationed within the Alas mission between 1924 and 1957, is believed to have worked on quadrilingual dictionary of Portuguese, Tetun, Bunak and Kemak. Father

---

[67] "I tried to use generally known terms and thus I adopted the dialect spoken on the northern coast, which might be less correct but is understood everywhere. That is not true about the dialects in the interior regions where the language is originally from. (…) The dialect used in the kingdoms of Samoro, Bariqu, Alas, Lakluta and Luka is hardly comprehensible to the inhabitants of Dili and of the regions with different dialects, although they speak enough of Tetun Dili." (my translation)

[68] "this language varies greatly from one region to another and there are no bases that would allow us to consider a word used in one area purer and more legitimate than the same word used in other area where Tetun is spoken." (my translation)

[69] "especially Tetun that has no classics nor reliable bases for investigation." (my translation)

Porfírio Campos, who worked as a professor for the Dili parish in 1931, started working on a universal dictionary of the Timorese languages (*Dicionário ideológico universal das línguas de Timor*). Unfortunately, neither of these dictionaries came to be published and as far as we know, both manuscripts were lost (Hull 1998a:11).

Álvaro Eugénio de Neves Fontoura (1940), who served as a governor of Timor-Leste between 1937 and 1940, came up with an idea to compile an album of pictures and maps that would document the types of people living in the area and the languages they speak. The album consisted of 552 photographs and a map of dialects spoken in Portuguese Timor divided by administrative posts. According to Fontoura, there were 29 distinctive linguistic groups.

4.2.3.6 Works from 1941 to 1960

José S. Martinho (1943:267) indicated 16 languages in Timor-Leste with some others having been possibly already extinct. The author distinguished between the Tetun spoken on the South coast (which is, as he says, purer and more resonant) and the Tetun spoken in Dili:

> "Fala-se dentro dos limites da antiga cidade o *tétun*, mas com certa incorrecção, porque não é esta a região originária dêsse dialecto. Bastariam os portugueses, chinas e árabes, que substituem por têrmos seus, principalmente os primeiros, as palavras que faltam no tétun, e os próprios indígenas admitem, para que êsse dialecto se transformasse, ali, numa mistura de línguas e numa algaraviada confusa, embora tendo aquêle por base."[70] (Martinho 1943:271).

Indeed, Tetun Prasa contains a large amount of Portuguese borrowings, however, Chinese and Arabic influence might be a bit overrated. Thomaz (1974:227-228) asserted that the Arab community was quite small and used Malay as their everyday language. Arabic was used only as a liturgical language since most of the Arab people could not

---

[70] "Tetun is spoken within the limits of the old city, but with certain incorrectness, because this is not the original region of this dialect. Due to the Portuguese, Chinese and Arab people – but mostly the first ones – who introduced their own terms for words that Tetun had lacked, this dialect has transformed into a mixture of languages, into a confused language, although having [Tetun] as a base." (my translation)

speak it. Most of the Arabic words (the majority of them being concrete substantives) were imported into Tetun via Malay. Nor did the Chinese language have any impact on Tetun. The Chinese lived in a closed community and communicated with the locals in Tetun (Thomaz 1974:249). Martinho also remarked on the simplification of Portuguese, saying: "O indígena está fazendo progressos, simplificando e adaptando o português às regras dos seus dialectos."[71] (Martinho 1943:271). At the same time, he tried to find a connection between Bornean languages and languages of Timor-Leste. He based his assumption on similar elements in toponyms (e.g. the name of the highest peak of Borneo, Mt. Kinabalu, or Cina Balu, meaning a *Chinese widow* has the same cognate in Tetun – Sina-Bálu). The author suggested that the substitution of /k/ by /s/ is very common in Timorese languages of Austronesian origin. However, according to Capell (1944c:26), the sound change from /k/ in the Original Indonesian, as he calls it, to /s/ never occurred in Tetun – /k/ was either preserved or omitted.

In 1943, Alberto Osório de Castro described how difficult it was to find an interpreter in Timor-Leste due to so many mutually incomprehensible languages. The author listed more than twenty of them, some of which are now considered full languages, some just dialects, like Tetun with its dialects Tetun Terik and incorrectly attributed Kemak, which is not a dialect of Tetun but rather an autonomous language spoken in the northwest of Timor-Leste, closely related to Mambae and Tokodede.

Mendes Corrêa (1944) in his chapter on languages and races of Timor tried to summarize the previous proposals of Silva (1900), Magalhães (1920), Fernandes (1931), Fontoura (1940), Martinho (1943), and Castro (1943) and remarked on how great the divergences in the number of languages in these selected works were. Mendes Corrêa provided a map of linguistic areas (see Map 7 below) based on the studies of Magalhães (1920), Martinho (1943), and Fontoura (1940) and tried to regroup the languages and respective anthropological types of Portuguese Timor based on these works.

---

[71] "The natives are making progress, simplifying and adapting Portuguese to the rules of their own dialects." (my translation)
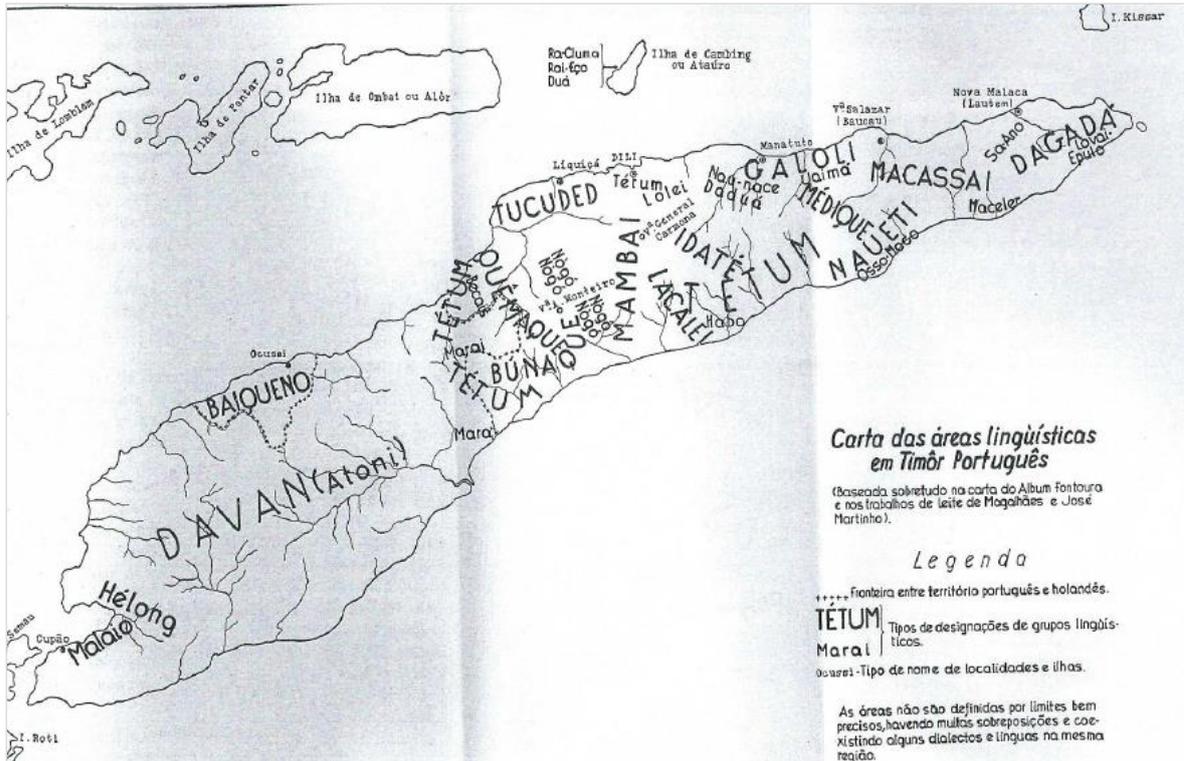
Map 7 Linguistic areas in Portuguese Timor (Corrêa 1944:111)

The first one to recognize the existence of Papuan languages in Timor-Leste was the Australian linguist Arthur Capell. His work on peoples and languages of Timor (1944) was published in three parts: Part I (1944a) gives a general outline of the history and the anthropology of Timor-Leste, Part II (1944b) is dedicated to the study of the languages of Timor, especially the Papuan languages (or non-Indonesian languages, as Capell calls them), and Part III (1944c) deals with Austronesian languages (or Indonesian, according to Capell) in both parts of the island of Timor. Part I describes the people of Timor and their way of life and was based on the studies of Martinho (1943), mentioned above. In Part II, the author gave a grammar and vocabulary outline of Papuan languages (some of them, however, incorrectly identified as Papuan). Regarding Tetun, Capell noted that due to its presence in areas where it was not indigenous, it had been influenced by other languages, Austronesian as well as Papuan. He also referred to the Tetun spoken in the north as Tetun Terik, which is somewhat confusing: "The northern Tetun, known as Tetum terik, is recognized even by its own speakers to be somewhat corrupt, and here the southern has been as far as possible taken as standard." (Capell 1944b:313). In Part III, Capell stated that all Austronesian languages of Timor could be traced back to a common origin, which he called Original Indonesian. He further suggested that the Austronesian

languages of Timor, together with the languages of the islands Roti, Wetar, Leti, and Kisar, formed a regional subgroup. In the subchapter on sound laws and vocabulary, the author presented a table of consonant changes in Tetun, among other languages (Capell 1944c:25-27). Based on these sound changes, Capell suggested that there must have been "a double invasion of the island by Indonesian – or Melanesian – speakers, one from the east and one from the west. It might be thought that the eastern group would have come by way of Ambon, Seran and the lesser islands of Moluccas." (Capell 1944c:27).

Father Ernesto Domingues in his paper *Línguas de Timor* (1947) asserted that there were not more than fifteen local languages and summarized all the works written on these languages up to that date, including the contributions of Raphael das Dores, Manuel Mendes Laranjeira, Manuel Patrício Mendes, Manuel Fernandes Ferreira, and Abílio José Fernandes. He also remarked on three different ways of spelling the name of the Timorese lingua franca: *tétum* (used by Aparício da Silva), *têto* (used by Raphael das Dores) and *tétun* (used by Laranjeira and Patrício Mendes) (Domingues 1947:147). He himself used the Portuguese spelling (*tétum*).

Artur Basílio de Sá, the envoy for the Mission in Macau and Timor and a lecturer in Linguistics and Tetun at the Department of Oriental Languages of the then Superior Institute of Social Sciences and Oversees Politics[72] wrote a paper on Timorese phonetics (*Notas sobre Linguística Timorense: sistema de representação fonética*, 1952a) to promote the methodological studies of the overseas languages in the Portuguese colonies. As a main problem he considered the need to elaborate a system for phonetic transcription. Basílio de Sá drafted a classification of Timorese phonemes and summarized them in a table (Sá 1952a:20). The author also took part in the language (*língua*) versus dialect (*dialecto*) discussion. He decided to call the languages of Timor-Leste *falares*, because, as he explained, the boundary between languages and dialects was not clear. Portuguese was considered a language since it constituted an independent system that a certain group of people used to express their ideas, but it could also be considered, together with Spanish and French, a dialect of Latin because it derived from it. Following this reasoning, Tetun could be classified as a language for its relative

---

[72] Currently known as *Instituto Superior de Ciências Sociais e Políticas* (Higher Institute of Social and Political Sciences) integrated in the University of Lisbon.

autonomy but also as a dialect since it derived from a proto-Austronesian language. Tetun Terik and other varieties of Tetun, however, did not pass beyond the boundaries of a regional subdialect (Sá 1952a:22-23). In his book *Timor* (1952b), the author distinguished two subdialects of Tetun (that is, as the author suggests, spoken also outside of Timor on the islands of Flores and Ende) – *Tetun popular* (popular Tetun) and *Tetun clássico* (classical Tetun). According to the author, popular Tetun was used in speech and classical Tetun was used by the *katuas* (older people) in regions like Viqueque, Lacluta and Somoro. The difference was, however, so big that a person who spoke only popular Tetun was not able to understand classical Tetun which had some features of grammar and literary vocabulary that the colloquial Tetun had dropped (Sá 1952b:37-38).

Hélio A. Esteves Felgas dedicated one of the chapters of his book *Timor Português* (1956) to the languages and dialects of Timor-Leste. He expressed his concern about the fact that Tetun as a lingua franca threatened the existence of minor languages. The author assumed that, in that period, Tetun was spoken by 220,000 Timorese – more or less a half of the population – plus many more who could understand it. The expansion of the language was caused by the constant relocation of construction workers to Dili and the organizing of local markets. Felgas (1956:173) believed that the minor languages would eventually disappear.

A Dutch priest Wilco Wortelboer (1955), who served in central Timor, wrote an essay on the language and culture of the Belu people in West Timor, whom he divided into three groups: the Belu people who came from the villages of Suai and Kamanasa in Portuguese Timor, the Belu people who came from the mountains and spoke Tetun Foho (*foho* meaning 'mountain') and the original Belu people from the Wehali kingdom. The Kamanassa-Suai people arrived in West Timor around 1911-1912 and established their own villages with the same respective names. According to Wortelboer, they spoke the same dialect of Tetun as was then spoken in the Portuguese part, which was very similar to Tetun Foho spoken by the Belu people from the mountains (Wortelboer 1955:174-175).

4.2.3.7 Works from 1961 to 1985

In 1964, the textbook by Abílio José Fernandes (section 4.2.3.5) was reprinted in an abridged form. According to Hull (1998a:15), it was intended for Portuguese soldiers in order to be able "to communicate and fraternize with the natives of the colony in their 'common' dialect so that the European speakers could in turn spread knowledge of the 'national' language by introducing as many Portuguese words as possible into their improvised Tetum," contributing thus to its 'creolization'. This assimilatory policy was made explicit in the 'Scope of this work' (*Finalidade deste trabalho*) printed at the end of the handbook:

> "Na convivência futura, o soldado tem por dever, progressivamente, ir substituindo o tétum por português, primeiro nas palavras de mais utilização, depois em pequenas expressões, até a conversação corrente."[73] (quoted in Hull 1998a:16).

The first Timorese linguist to study (and teach) Tetun was Frederico José Hopffer Rêgo. In his paper on the linguistic situation in Timor-Leste (1968), the author summarized and commented the work of Capell, mentioned above (section 4.2.3.6). He suggested that apart from the two migration waves mentioned in Capell (1944c:27), there might have been a third and a more recent invasion of people speaking the Galolen language. In the section on Tetun (Rêgo 1968:66-67), he looked at the definition of Tetun Terik by Artur Basílio de Sá (see section 4.2.3.6) and Arthur Capell. Sá (1961) considered Tetun Terik the variety spoken in West Timor and along the Timor-Leste border in Batugadé and Balibó. The variety spoken in the south of Timor-Leste was called Tetun Los and the author considered it "teto puro, teto clássico"[74], in opposition to Tetun Prasa – "o teto vulgar, língua franca do Timor Português (…) teto popular, simplificado, invadido por termos e expressões portuguesas, reduzido às suas regras rudimentares (…) a afastar-se cada vez mais das suas origens e a fundir-se com o português num falar típico,

---

[73] "In the future companionship, the soldier should, progressively, substitute Tetun with Portuguese, starting with the most frequent words, then continuing with short expressions until a coherent conversation." (my translation)

[74] "pure Tetun, classical Tetun" (my translation)

que poderá vir a ser o crioulo de Timor."[75] (Sá 1961:xxv). Rêgo doubted both of them, since during his stay in Timor-Leste, he heard people from the south of the country referring to their language as Tetun Terik. However, he admitted that this variety could as well be called Tetun Los, since *los* means 'pure, correct, direct' (Rêgo 1968:67). Very interesting is a reference to a language called Belu that is very similar to Tetun. Some recent authors (see section 4.2.3.10 on Williams-van Klinken and Hajek) consider it a dialect of Tetun spoken in Belu Regency (in West Timor) bordering with Timor-Leste.

Portuguese historian Artur Teodoro de Matos (1974) wrote a book on Portuguese Timor from the arrival of Portuguese in 1515 until the transfer of the capital to Dili in 1769. He supported the hypothesis that the diffusion of Tetun in this period was quite restricted, since most of the vocabulary found in the Portuguese documents from this period was not borrowed from Tetun, but Malay, the main trade language used along the Timorese coast (Matos 1974:21).

The most important contribution to Timorese linguistics was made by Portuguese historian and linguist Luís Filipe Thomaz. In 1974, one year before Indonesian occupation of Timor-Leste, he published a series of articles about Timor in the journal of the University of Lisbon *Portugaliae Historica* under "Timor – Notas historico-linguísticas" (Timor – Historical and linguistic notes, 1974). He dealt with the origin of Timorese population and the migration waves (*O primitivo povoamento do Mundo Oceânico e as origens da população Timorense*), with the classification of the Timorese languages within the Austronesian language family and description of the most salient grammatical features of Tetun (*A posição das línguas de Timor no grupo malaio-polinésico*), with the use of Tetun as a lingua franca of Timor-Leste (*O uso do Tétum como língua veicular em Timor Oriental*), with the Indo-Javanese influence in Timor (*A indianização do sueste asiático e a influência indo-javanesa em Timor*), with the Islamic and Malay influence in Timor (*A hegemonia do reino islâmico de Malaca e a influência malaia em Timor*), with the Chinese presence in Timor (*Os chineses em Timor*), with the reasons why Portuguese never became the lingua franca of the island (*Os portugueses em Timor e o português de*

---

[75] "vernacular Tetun, the lingua franca of the Portuguese Timor (…) popular Tetun, simplified and invaded with Portuguese terms and expressions, reduced to basic rules (…) shifting away, more and more, from its origin and merging with Portuguese into a typical language that could turn into the creole of Timor." (my translation)

Timor), with the now extinct Bidau Creole Portuguese (*Dialectos crioulos em Timor: O «português de Bidau» e o crioulo macaísta*), with the influence of Portuguese language on local languages, especially Tetun Prasa (*A influência do português nas línguas de Timor*), and with the linguistic situation in Timor-Leste in 1974 (*A actual situação linguística do Timor português*). Thomaz distinguished between two varieties of Tetun – the classical Tetun that he called Tetun Terik or Tetun Los, depending on the region, and the popular Tetun used as a lingua franca called Tetun Dili or Tetun Prasa. As discussed in section 2.2, the author presented his own hypothesis as to how Tetun Prasa diffused as a lingua franca.

Due to the long-lasting presence of the Portuguese and the deep transformations they caused in Timorese society, local languages borrowed a large number of Portuguese words. The influence is most notable in Tetun Prasa, which adopted not only nouns, but also prepositions, conjunctions and suffixes such as *-dór*. Even Tetun Terik contains words of Portuguese origin (Thomaz 1974:283-289). Apart from Portuguese, the author also remarked on the high number of Malay borrowings in Tetun Prasa listing more than eighty of them (Thomaz 1974:231-241). It was, however, interesting, that the influence of Tetun on Portuguese was very limited. This was due to the fact that until the end of the 19th century, the Portuguese mostly stayed in the coastal areas where Malay, the trade language, was spoken. When they finally penetrated the interior, their language already contained most of the names of local plants, animals and products (Thomaz 1974:254).

Thomaz also studied how and why Tetun Prasa became dominant in the Dili area, originally a Mambae-speaking region:

"A sua substituição [do mambae] pelo tétum deu-se provavelmente a partir de 1769, com a instalação da capital em Díli, que provocou o afluxo à cidade de pessoas oriundas de diversas regiões da ilha. Pertencentes a diferentes grupos etno-linguísticos tiveram, para se poderem entender, de recorrer ao tétum, *língua franca* de toda a metade oriental da ilha. Trata-se de um tétum muito mesclado de português, conhecido por *tétum praça*, sensivelmente diferente da língua falada na costa Sul e sobretudo da língua clássica, usada pelos *lia-nain* ou oradores tradicionais. Parece, portanto, que se fala tétum em Díli pela mesma razão por que

se fala malaio e não sundanês em Jacarta, mais inglês que malaio na Singapura, e alto-alemão em vez de baixo-alemão em Berlim."[76] (Thomaz 1977:548).

In 1981, Thomaz published another paper on Tetun Prasa (*A formação do tétum-praça, língua veicular de Timor-Leste*) trying to document its formation and its contact with the Malay and Portuguese language. According to the author, this contact led to the fact that Tetun Prasa now exists in a "less pure, simplified and more Portuguese-influenced form"[77] (Thomaz 1981:1). Thomaz assigned to Tetun Prasa an integrating role since it was accepted as a superior variety by the Timorese ethnic groups, although its use in eastern regions like Lautém and the Oecussi enclave was still very limited (Thomaz 1985:325-326).

Thomaz wrote most of his papers on Tetun and Timorese linguistics situation right before the Indonesian occupation (1975-1999). In 2002, he revised and published many of them in the book *Babel Loro'Sae: O Problema Linguístico de Timor-Leste* (Babel Loro'Sae: linguistic issues in Timor-Leste, 2002). Here Thomaz suggested that the fact that the Tetun spoken in Dili was less pure and less grammatically rich, with the inflection of verbs completely lost, gave it a creole character:

> "(…) não, naturalmente, um crioulo de português mesclado de tétum, mas um *crioulo de tétum* (my emphasis ZG), mesclado de português. Assemelha-se assim à língua geral do Brasil. (…) A língua geral não era, como os crioulos de Cabo Verde ou da Guiné, uma forma crioulizada do português, mas antes um tupi-guarani pintalgado de lusismos, e nisso se assemelha ao tétum-praça."[78] (Thomaz 2002:68-69).

---

[76] "The substitution of Mambae by Tetun probably started in 1769 when the capital was transferred to Dili, which resulted in an influx of people from different regions. Belonging to different ethnolinguistic groups, they had to resort to using Tetun, the *lingua franca* of the eastern half of the island, in order to make themselves understood. It was a Tetun mixed with Portuguese, known as Tetun Prasa, significantly different from the language spoken in the South, especially from the classical language used by *lia-nain* (storyteller) and traditional orators. It seems that the reason why Tetun is spoken in Dili is the same as why Malay and not Sundanese language is spoken in Jakarta, why more English than Malay is spoken in Singapore and why High German instead of Low German is spoken in Berlin." (my translation)

[77] In the revised 2002 version, Thomaz (2002:104) reformulates this sentence: "forma simplificada, de características que o aproximam dos crioulos, mais influenciada pelo português" ["simplified form, more influenced by Portuguese and whose characteristics are closer to those of creoles"].

[78] "(…) naturally, not a Portuguese creole modified by Tetun, but a *Tetun creole* (my emphasis ZG) modified by Portuguese. It is similar to *língua geral* in Brazil. (…) Língua geral was not a creolized form

One year after Indonesia proclaimed Timor-Leste its 27th province, the Indonesian linguists P. J. Serantes and I. H. Doko wrote and published a course book (1976a), a short dictionary (1976b) and a short conversation guide to Tetun (1976c). In the introduction, they referred to Timor-Leste as *Tim-Tim*, an abbreviation of *Timor Timur* (lit. Timor East) and to Portuguese Timor as *Timport*. The course book *Pelajaran bahasa Indonesia – Tetun untuk rakyat Timor-Timur* (Course of Indonesian language – Tetun for people of Timor-Leste, 1976a) and the dictionary *Kamus kecil Indonesian – Tetun Belu – Tetun Dili* (A short dictionary Indonesian – Tetun Belu – Tetun Dili, 1976b) were intended mainly for Timorese people but could also serve for Indonesian teachers who were sent to Timor-Leste to teach Indonesian, or for any Indonesian person who got a position or job there. The authors also pointed out that Tetun had been influenced by Portuguese and words found in Tetun Belu had been replaced by Portuguese borrowings in Tetun, although the native words were still understood (Serantes & Doko 1976a:4). Thomaz (1980), in his review of these three works, pointed out that it was clear that the authors were speakers of Tetun Belu (or Tetun Atambua, as he calls it) and although, in the case of the dictionary, they tried to overcome this problem by carrying out a trilingual work, in the other works most of the dialogues are in Tetun Belu, which might be incomprehensible to the speakers of Eastern Tetun, the actual users of these books. Thomaz was suspicious of the fact that the Tetun Dili informants were actually Portuguese speakers since in the dictionary more than a third of the words are of Portuguese origin and many of them are seldom used is spoken Eastern Tetun (Thomaz 1980:350). The authors also tried to eliminate phonemes borrowed from Portuguese which sometimes caused Portuguese loanwords in Tetun to be unrecognizable from the original, e.g. *knovtmenti* (possibly from Port. *conveniente* 'convenient'), *resetu* (possibly from Port. *receio* 'concern')*, basyu* (possibly from Port. *bacio* 'potty). On the other hand, phonemes foreign to Tetun but borrowed from Indonesian were preserved (Thomaz 1980:351). Although Thomaz was critical about the linguistic value of these works, he also acknowledged that the real purpose of these books was not a linguistic study but essentially a language teaching book.

In 1984, an Australian linguist, Cliff Morris, published the first Tetun-English dictionary. He explained why the Tetun language was called Tetun and why it was natural

---

of Portuguese, like those in Cape Verde and Guinea-Bissau, but rather a Tupi-Guarani [language] enriched with Portuguese loanwords and in this way, it is similar to Tetun Prasa." (my translation)

to use this original spelling: "The wise old men 'katuas' tell us that the people lived on the plains 'tetu', and therefore as the people were of the plains 'tetun'." (Morris 1984:x). His dictionary was based on the variety referred to as 'Tetun-Los', which was spoken in the southern central regions of Timor-Leste around Soibada and Samoro, and was also regarded as a standard variety of Tetun. The author also listed other dialects that are discussed in more detail in section 1.3. However, Tetun-Dili was said to be "the dialect taught to the Portuguese and other people needing a common language for commerce" (Morris 1984:x). In 1992, Morris published another version of his 'Tetun-Los' dictionary, *A Traveller's Dictionary in Tetun-English and English-Tetun*.

During the Indonesian occupation, two more dictionaries were published. The first one was written by Franciskus Monteiro (1985) and described Tetun spoken in West Timor (*Kamus Tetun – Indonesia*), the other one, by Domingos M. Dores Soares, the Dili *bupati*, was trilingual, representing words in Tetun, Indonesian and Portuguese (*Kamus Bahasa Tetun – Bahasa Indonesia – Bahasa Portugis*).

4.2.3.8 Works from 1986 to 1999

Most of the Indonesian studies on Tetun were concerned with its western variety and the research of Eastern Tetun started only in the early 1990s. One of the first Timorese linguists, Benjamim de Araújo e Corte-Real, who went on to become the rector of the National University in Timor-Leste (UNTL) and is now the director of the National Institute of Linguistics (INL), wrote his bachelor's thesis in phonology, in which he contrasted Tetun and English consonants and focused on their phonological features. For his study, he decided to choose the Tetun spoken only in the eastern part of the island and the modern version of it, which has suffered influence from Portuguese but which is at the same time understood and recognized as a native Timorese language and referred to simply as Tetun (Corte-Real 1990:13). To better understand the whole consonant inventory in Tetun, the author distinguished between two groups of speakers: non-Portuguese-influenced and Portuguese-influenced. The alteration in pronunciation between these two groups mostly concerns the Portuguese borrowings. The author further dealt with the phonetic distribution of the consonants and consonant clusters. He

concluded that the phonemic inventory of Tetun is more limited in comparison with English, however there is also a significant difference in terms of consonant inventory between the non-Portuguese-influenced and Portuguese-influenced speakers (Corte-Real 1990:91).

Three groups of researchers from Dili appointed by the Ministry of Education and Culture of Indonesia were part of the 1985 team working on a project about the Indonesian language and culture and the province of Timor-Leste. As a result, Saliwangi et al. published a book on the morphological system of verbs in Tetun Prasa (*Sistem Morfologi Kata Kerja (Verba) Bahasa Tetun*, 1991). Although the authors claimed their work was based on Tetun Prasa, they listed features (e.g. affixes *ma-, na-,* etc.) that are not found in this variety. They were followed by Soedjiatno et al. who composed a book on the morphological system of the closed classes in the Tetun spoken in the district of Dili (*Sistem Morfologi Kata Tugas Bahasa Tetum*, 1992). The book dealt with monomorphemic and polymorphemic (affixed, reduplicated and compounded) structures. The third book in the series was *Morfo-sintaktis Bahasa Tetum* (Morphosyntax of Tetun, 1993) by Taryono et al. This book, also based on Tetun Prasa, dealt with morphological and morphophonological processes in Tetun, as well as syntactic constructions and processes.

In 1996, Anton Berkanis wrote a Master's thesis on the phonology of Tetun Prasa. Unlike other authors working on Tetun Prasa who agreed on a more limited number of consonant and vowel segments, Berkanis listed 28 of them – 22 consonant and 6 vowel segments. These sounds included phonemes usually not found in Tetun Prasa like /ʔ/, /w/, /ʤ/, and /ŋ/. Especially the latter two indicated strong Indonesian influence. The author did not mention the palatalization found in Portuguese loanwords, or Portuguese influence in general. This might be, however, due to the fact that the thesis was defended at an Indonesian university during the Indonesian occupation of Timor-Leste.

According to Grimes et al. (1997), there are two kinds of Tetun spoken in Timor-Leste: Tetun Prasa and Tetun, comprising of three dialects – Tetun Loos (spoken in the south of Timor-Leste around Viqueque), Tetun Foho (spoken in the northern part of central Timor) and Tetun Fehan (spoken in the southern part of central Timor), as shown on Map 8. The authors described Tetun Prasa as a 'Tetun-based creole with heavy

Portuguese and Mambae influence'. Since Tetun Prasa was considered a creole, it was treated as a separate language, not a variety of Tetun. Among its distinctive features the authors mentioned the loss of subject markers on verbs, the use of periphrastic rather than morphological constructions (e.g. causatives) and different formation of possession and negatives (Grimes et al. 1997:52).

Map 8 varieties of Tetun spoken in Timor (Grimes et al. 1997:51)

4.2.3.9 Geoffrey Hull and Lance Eccles (1993 – present)

The Australian linguist Geoffrey Hull is one of the best-known specialists in the languages of Timor-Leste. He worked for many years at the National Institute of Linguistics (INL) at UNTL and is the author of innumerous books and articles. Hull has also advocated for the adoption of Tetun and Portuguese as the official languages of Timor-Leste and argued against English and Indonesian for historical reasons (Hull 1999a, 1999b, 2000c).

As a director of the project *Linguistic survey of East Timor* at the University of Western Sydney and later as a researcher and director of publications at the INL (2001 – 2007), Hull worked in the area of language planning and lexicography. His studies of Tetun began through the contact with the local Timorese community in Australia. With the help of his informants, he compiled an English-medium course book on Tetun Prasa (1993) and *Tetum language manual for East Timor* (1999d). He also published a Tetun-

English dictionary[79] (1999c) and *Tetum reference grammar* (2001, with Lance Eccles), republished in the Portuguese language (*Gramática da Língua Tétum*, 2005). While compiling the reference grammar, Hull and Eccles encountered a couple of problems. First of all, they identified six different variants of Tetun: an acrolectal Dili variety, an acrolectal non-Dili variety, a mesolectal variety, a basilectal variety, rural dialects and the liturgical Tetun. This resulted in a very complex chapter on phonology and orthography. The authors referred to Tetun as a creolized language that lost all its morphology, retaining only a few fossilized elements in word derivations. However, they believed that in the future, these 'dead' elements could be revitalized (Eccles 2001:38-39).

During his research on Timor-Leste linguistics, Hull proposed various hypotheses about the origin and classification of Tetun. In 1993, he put Tetun, together with other Austronesian languages of Timor and neighbouring islands, into a Central Malayo-Polynesian language group and he referred to Tetun Prasa as "a mixed language with a Tetum base and a thick Portuguese coating. Portuguese loanwords, syntax and loan translations greatly outnumbered the older Malay layer of borrowings, so that today it is impossible to express oneself in Tetum-Praça without using Portuguese forms." (Hull 1993:viii). In 1998b, the author published an article on the basic lexical affinities of Timor's Austronesian languages, in which he explored the possibility of Tetun being a descendent of Old Butonese[80] spoken in the southeast region of the island of Sulawesi (also known as Celebes) called Buton, and thus belonging to the Western Malayo-Polynesian group of languages. Considering the analytical and isolating character of Tetun together with extensive grammatical simplifications, the author claimed that "in a

---

[79] The *Standard Tetum – English dictionary* (1999c) is a dictionary of the Eastern Tetun. Hull refers to 'Eastern Tetum' as "a hybrid language, basically Austronesian, but with a heavy Portuguese superstratum" and to Tetun Prasa, which he considers a dialect of Eastern Tetun, "a fully creolized form" (Hull 1999c:ix). However, in his other works (Hull 1998a), 'Eastern Tetum' is a variety that refers to Tetun Terik while Tetun Prasa, another variety, is referred to as Díli Tetum. The Tetun that he helped to promote to the official language was based on Tetun Prasa and termed 'National Tetum' or 'Official Tetum'.

[80] Hull (1998b:152-153) suggested that Old Butonese (Proto-Butonic) spread from south-eastern Sulawesi in two waves. The first one originated in the Muna-Buton region and gave rise to the Old Austromunic languages of Timor-Leste: Mambae, Kemak, Tokodede and Idalaka spoken in Central Timor. The second wave originated in the Tukang Besi archipelago and gave rise to the Old Austrofabronic languages spoken in both parts of Timor: Helong, Dawan, Tetun, Galolen, Habun and Kawaimina. In 2001, Hull corrected his previous assumption that Old Austromunic and Old Austrofabronic come from two different Butonic dialects; the differences were caused by hybridization in Timor after the first migration wave. He suggested that Tokodede, Kemak and Mambae were 'offshoots' of Idalaka that was now grouped with the Fabronic languages. Hull thus decided to rename the Old Austromunic languages to Ramaelic since they were spoken around the Mount Ramelau (Hull 2001a:99).

past age Celebic speech was introduced to the region in a grossly simplified variety or, (as is more likely) underwent various degrees of pidginization after implementing itself there." (Hull 1998b:137). However, there are also other elements in Tetun that cannot be traced to the Celebic languages or the Papuan ones spoken in Timor-Leste. Hull suggested that these are the result of a second Austronesian migration wave coming from Ambon island in the Central Moluccas. This led to pidginization and nowadays "[m]odern Timoric languages display all of the classic creole features: little or no inflectional affixation, semantically transparent and potentially optional derivational affixes, the presence of serial verbs and 'exceed' comparative constructions[81]." (Hull 1998b:165). The timeframe suggested by Hull is as follows: Butonic migration – not much older than the 11th century A.D., Ambonic migration – in the 13th century A.D., further pidginization due to the presence of Ambonese Malay – in the 15th century A.D. (see Map 9).
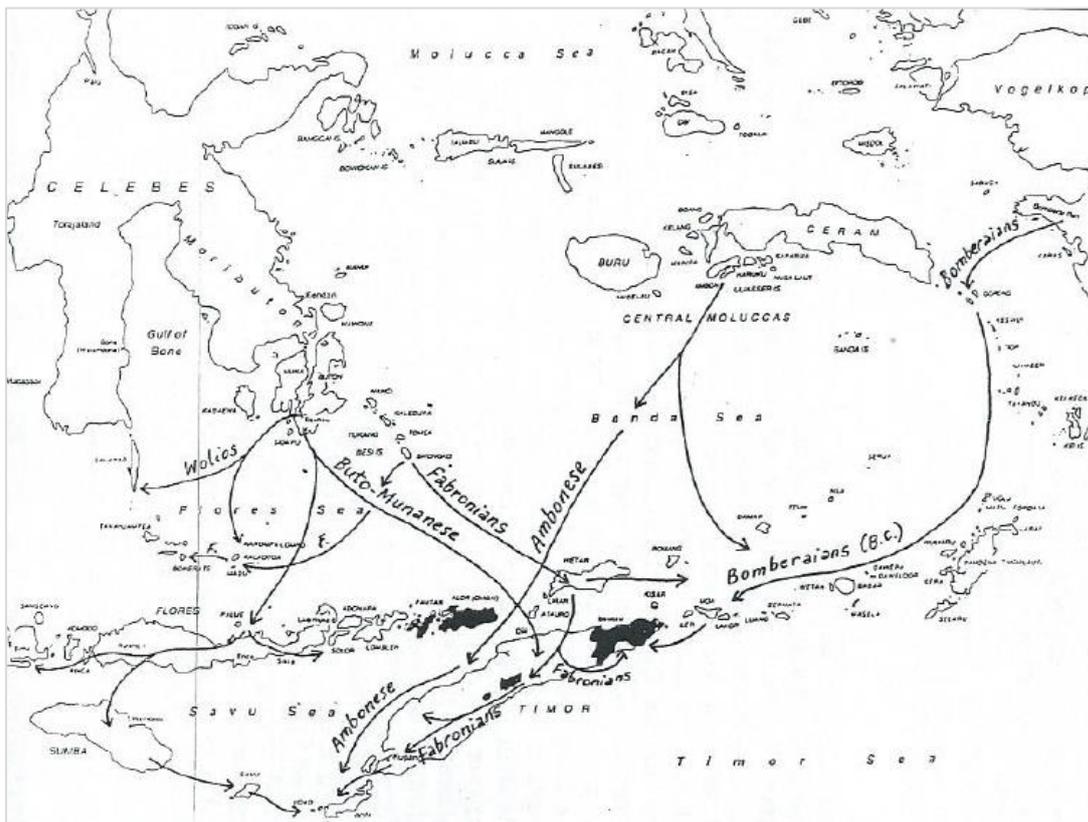
Map 9 Migration waves according to Hull (1998b:151)

Another stage of creolization started with the domination of Belu people who spoke Tetun. Hull referred to this variety of Tetun as 'recreolized' but after having been

---

[81] It should be noted that both serial verbs and 'exceed' comparative structures are also found in non-creoles (Holm 2004:206-210).

established in the capital, it became 'a pure creole' while Tetun Terik was still going through a stage of an 'advanced creolization' (Hull 2000a:58). To further support his hypothesis about the origin of Tetun, Hull looked at the historical phonology of the language. He found that "the sound system of Tetum is fully integrated into that of the Timoric group and presented no peculiarities that could not be explained within the context of a Celebic affiliation" (Hull 2000b:158). However, unlike Proto-Timoric and Proto-Moributonic which have a consonant system of 28 phonemes, Tetun has only 13 consonants and Tetun Prasa, due to the Mambae influence, even fewer (11). The poverty of consonants was, according to the author, reminiscent of Oceanic languages "which belies the Celebic and Hesperonesian (Western Malayo-Polynesian) filiation of Tetun" (Hull 2000b:189).

In 2001, Hull (2001a) looked at the eleven Austronesian and the four Papuan languages of Timor-Leste that form a linguistic area or a *Sprachbund*. The author tried to find an explanation for the simplified structure of these Timoric languages. As the main cause, he considered the two migration waves that originated in Central Moluccas, more precisely the influence of Old Ambonese and Malay, and he reconstructed the past phases of Timoric languages: Old Timorese, the agglutinative language originating in Old Butonese; Early Middle Timorese, morphologically and structurally restructured language due to the influence of Old Ambonese; Late Middle Timorese, a creolized and a relexified language due to the Ambonese Malay influence; and Modern Timorese, the modern Timoric languages with Malay and Portuguese superstrata (Hull 2001a:101). However, Hull's classification of Timorese languages as Western Malayo-Polynesian has never been accepted by other Austronesianists and none of the terms coined by Hull (e.g. Fabronic, Ramaelic, etc.) is used nowadays (see section 4.2.3.11 on John McWhorter).

In 2005, Hull looked more closely at the Malay element in Tetun. He found it difficult to recognize the Malayisms in Tetun since, as he believed, Malay and Tetun belong to the same Western Malayo-Polynesian branch but also because the original Butonic vocabulary underwent relexification and expansion due to Malay influence. To determine which words were actually Malayisms, Hull proposed three criteria: that these words are not found in Moributonia region in south-eastern Celebes, that they have synonyms of Celebic origin in Tetun and that their phonology indicates relatively recent

introduction (Hull 2005a:89). Altogether, he was able to identify 450 Malay loanwords in Tetun (see section 8.3). There were also instances of doublets of native and borrowed words, which could vary only internally. The traditional ones were used by basilectal speakers, and the assimilated ones by mesolectal and acrolectal speakers (Hull 2005a:90). Interestingly, also many Portuguese borrowings entered Tetun via Malay since Malay started to borrow from Portuguese after the Portuguese capture of Malacca in 1511. Since the contact between Portuguese and Tetun really started only in 1769 when the capital was transferred to Dili, most of the earlier Portuguese loanwords entered Tetun via Malay, e.g. *kreda* 'church' (from Malay *gereja*); *dinela* 'window' (from Malay *jendéla*) (Hull 2005a:117).

Hull, in a personal communication to Lance Eccles, also suggested that three TAM markers in Tetun – *atu*, *tiha* and *sei* – appear to have come from Malay. According to Eccles, if this was true, then it would be clear that Tetun was a mixed language since it was very unusual for a language to borrow such items unless in intimately close contact situations (Eccles 1999:35). However, in 2001 Hull published a morphological overview of the Timoric *Sprachbund* which offered insights about the verbal markers in other Austronesian languages spoken in Timor. According to his data, many Timorese languages derived or borrowed verbal markers from Malay rather than 'Old Timorese' which suggests that it is not possible to classify Tetun as a mixed language based just on this criterion.

Hull also contributed to other works, for example, the Malay-Tetun dictionary (with Pollard 2005), a dictionary of Official Tetun published by INL (Correia et al. 2005) as well as Tetun grammar course for professors, translators and tertiary students (with Correia 2005). Additionally, he has also worked on other languages of Timor-Leste, like Mambae (2003a), Galolen (2003c), Makasae (2005b), Baikeno (2001b, 2003b) and Fataluku (2005c).

4.2.3.10 Catharina (Williams)-van Klinken and John Hajek (1999 – present)

Catharina Williams-van Klinken became interested in Timor studies during her doctoral studies at Australian National University, where she wrote her dissertation on the grammar of the Fehan dialect of Tetun spoken in southern West Timor. The Fehan dialect is one of the sub-dialects of South Tetun, the other being Tetun Suai/Tetun Kamanasa spoken along the border of West and East Timor (van Klinken 1999:3). The basis for the Fehan grammar was the everyday speech, although van Klinken recognized three different registers: poetry, the noble register used on formal and ritual occasions, and the sea taboo register used on fishing trips which differed only in special vocabulary (van Klinken 1999:7-9). She also studied registers of Eastern Tetun and compared three high registers used in formal settings: the traditional ritual register[82] of Tetun Terik, the church register – a legacy of the Catholic Soibada mission founded in 1898 – which prefers Tetun Terik forms to Portuguese ones[83], and the press register of Tetun Prasa that began developing after 1999 when Tetun started to be used in the media. The press register can sometimes consist of as much as 50% of Portuguese loans, some of which are not well-known to the Timorese, and the pronunciation of these words is also strongly Portuguese-influenced (Williams-van Klinken 2002:5).

In 2002, Williams-van Klinken moved to Dili to work as a language director for the Peace Corps. In the same year, she co-authored, together with John Hajek and Rachel Nordlinger, a grammar of Tetun Prasa (2002a), which she referred to as Tetun Dili. Since there was no official orthography for Tetun at the time of the writing, the authors adopted a phonemic representation. The focus of the grammar was on spoken language, with comparisons with the written and liturgical registers. The grammar was published in an abridged edition in LINCOM's descriptive grammar series (2002b).

---

[82] This register is used by ritual specialists and incorporates vocabulary that is not used in daily speech and is hardly understood by locals. The author highlights one of the special features of this register – poetic parallelism: "The essence of the ritual register is pairs of poetically parallel lines (or parts of lines), in which the second is an echo of the first, except for the replacement of one or more words or expressions by related ones." (Williams-van Klinken 2002:2).

[83] However, the grammar does not follow Tetun Terik in all aspects, e.g. Church Tetun does not use subject markers on verbs, like Tetun Prasa. The reason why the church register does not follow the ritual one is the fact that Catholic priests did not perform rituals (Williams-van Klinken 2002:4).

Williams-van Klinken also co-authored some other papers with John Hajek, an Australian linguist who specializes in Romance linguistics and phonetics and phonology, and who has published various papers on language ecology and language contact in Timor-Leste. In a paper on the language history of Timor-Leste (2000a), Hajek pointed out the reasons why the Portuguese language had such little negative impact on the linguistic heterogeneity of the eastern part of the island, comparing it to a significantly smaller number of languages in West Timor. It was "[t]he lack of Portuguese control coupled with Portugal's lack of interest in East Timor until the 20th century" (Hajek 2000a:216) that prevented the indigenous linguistic ecology from being affected by Portuguese. Local people traditionally spoke many languages and Portuguese did not cause any disruption; on the contrary, it was rather another contribution to the multilingual repertoire of the speakers[84] (Hajek 2000b:401). Malay also almost disappeared after 1870 and its functions "were taken over by creolized Tetum-Dili and by Portuguese" (Hajek 2000a:219). Hajek wrote this paper just months after the Independence referendum and noticed that right after the vote, the situation became critical due to the Indonesian and anti-independence militants' violence and around 80-90% of East Timorese were displaced from their homes. The author suggested that this dislocation led to establishing new mixed communities whose speakers were most likely going to shift to a common language – Tetun (Hajek 2000a:224). However, there has been no study of the linguistic outcome yet.

Williams-van Klinken and Hajek looked at various morphosyntactic features of Tetun Prasa, like the Portuguese suffix -*dór*, serial verb constructions and patterns of address. In 2003, Hajek and Williams-van Klinken (2003) studied the loans at the level of derivational morphology in Tetun Prasa and discovered only one: the still productive Portuguese agentive suffix -*dór*[85] (see section 5.3.4.1). A similar productive suffix was not found in any other Austronesian language influenced by Portuguese or Spanish. This suffix coexists in Tetun Prasa with a native Tetun root -*tén* 'excrement' (see section

---

[84] The only irreversible change caused by the Portuguese was the replacement of Mambae, spoken in Dili, by Tetun. However, Mambae is still spoken just outside the capital (Hajek, 2000b:402).

[85] In 2009, Williams-van Klinken and Hajek (2009a) suggested that *door* (using their own spelling) was phonologically not an affix but a separate word, since it has the full stress and that syntactically and semantically these derivations were not phrases but compounds.

5.3.4.2) and a native Tetun substantive *na'in* 'owner, master'[86] (see section 5.3.4.3). The difference between these three forms is only on the semantic level: *-tén* conveys a pejorative meaning, *-dór* can also be negative[87], unlike *nain,* which describes somebody who is skillful in something (Hajek & Williams-van Klinken 2003:58-59).

The authors further explored the phenomenon typical for many Asian and African languages – serial verb constructions (see section 6.2.4). Williams-van Klinken, Hajek and Nordlinger published some preliminary remarks on serial verbs in Tetun (2001) before including a full chapter on serial verbs and causative construction in their grammar of Tetun Prasa (2002a). However, Hajek (2006a:239-240) noticed that these verb sequences have started to disappear from Tetun Prasa:

> "Serial verb constructions (SVCs) are a feature of Tetun Dili. However, the language is unusual in that it shows strong evidence of being subject to an ongoing process of substantial deserialization. A series of independent processes, such as grammaticalization, lexicalization, and contact with Portuguese (a clearly non-serializing language) have conspired to significantly reduce the frequency and range of SVC types in T[etun] D[ili] when compared to more conservative varieties of Tetun, such as Tetun Fehan spoken in West Timor (cf. van Klinken 1999)."

But other types of words have undergone changes, too. Williams-van Klinken (2010a:182-183) noticed the rise of the passive voice in Tetun Prasa, a language which did not possess, like other central-eastern Austronesian languages, any passive constructions before. Nowadays one can find passive constructions [undergoer + transitive verb + *hosi/husi* 'from' + actor] or Portuguese passive verb forms (see section 6.2.5).

Another interesting area explored by Williams-van Klinken and Hajek was the language contact and convergence. Hajek (2006b:163-164) compared Tetun Prasa with

---

[86] In Tetun Terik there is a circumfix *mak-(-k)*, see section 5.3.4.4.
[87] Looking at the grammars of Tetun, Hull and Eccles (2005:8) also attributed negative/pejorative meaning to some words formed with *-dór.* However, Williams-van Klinken at al. (2002a:20) defined the derived actor noun as "a person who habitually does X" without commenting on the possible negative meaning. Luís Costa (p.c.) also did not agree with the "negative/pejorative" meaning of the suffix. For a more detailed discussion see section 5.3.4.5.

Portuguese, and then with Tetun Fehan and Mambae to identify the differences in their patterns of influence: the influence of Portuguese "involves a newer process of grammatical reinforcement and complexification (rather than the often presumed pidginization), alongside an earlier longer-term pattern of local convergence and simplification involving Mambae and other languages in the area." Hajek tried to establish a cline of increasing grammatical simplification, ranging from Baikeno, as the least simplified, through Tetun Fehan and Tetun Prasa, to Mambae, the most simplified and thus the most innovative. He believed that the convergence between Mambae and Tetun Prasa (and other Timor-Leste languages) had to do with the long-term historical contact between Austronesian and non-Austronesian language since they had many features in common, e.g. a predominantly isolating nature (Hajek 2006b:176-177). But Hajek was not completely right when comparing more simplified and isolating features of Tetun Prasa to those of Tetun Fehan. For example, he pointed out the use of reduplication to form plurals in Tetun Fehan as a feature that this language 'retains'. However, reduplication in noun pluralization does not really exist in Tetun Terik[88] (Avram 2008:435), the source language from which Tetun Prasa derived and this feature of Tetun Fehan was clearly an influence of the Malay or Indonesian language. The same applies to the tonic/reduced forms of personal pronouns existing in Tetun Fehan but not in Tetun Prasa (Hajek 2006b:175). Again, reduced forms are not found in Tetun Terik either.

Williams-van Klinken (2010b) looked at the two varieties of Tetun (Tetun Prasa and Tetun Terik) to see if these could actually be considered two different languages. She compared all the linguistic levels: vocabulary, style, phonology, morphology, NPs, numerals, prepositions, conjunctions, complementizers, and clauses. Based on these comparisons, she defined the most likely sources of the changes in Tetun Prasa as extensive loans; simplification; removal of exceptions; calquing on Mambae and Portuguese; and grammaticalization and reanalysis of grammatical morphemes as part of the root. The most deep-rooted differences could be attributed to the Mambae influence, while the Portuguese influence was more common in formal speech and writing, which

---

[88] It is unclear whether noun pluralization by reduplication ever existed in Tetun Terik. Sá (1961:199) gives examples like *ema liurai-liurai* 'all the liurais' and *ema ata-ata* 'all the slaves', but when Esperança (2001:70) elicited these examples with Tetun Prasa speakers, he came to a conclusion that this method of pluralization is deemed incorrect by native speakers. For a more detailed discussion see section 6.1.2.4.

was documented in Williams-van Klinken and Hajek (2009b). According to their analysis of selected newspapers, 30-40% of word tokens are Portuguese loans; within open class words it is around 75%. Statistically, there were on average two words per five paragraphs that people under 30 with a tertiary education did not understand. Most of these loanwords and other innovations have not yet entered the spoken language, which is not untypical of contact situations. There are cases of other languages (e.g. Japanese) in which the change came through the written language (Thomason & Kaufman 1991:66).

4.2.3.11   Works from 2000 to present

Important work was carried out by the Timor/Asia Documentation Centre at Lisbon University led by Professor Artur Marcos until 1998. In 2000, the centre published a Tetun-Portuguese dictionary compiled by Luís Costa, a Timorese professor and a native Tetun Terik speaker. The dictionary, with more than 9000 entries was based on two previous Tetun-Portuguese dictionaries by Raphael das Dores (1907) and Manuel Patrício Mendes (1935). As there was no official orthography at that time, Costa presented his own standardization based on the phonetic realization of Tetun. In 2001, Costa published a Tetun conversation guide that tried to find a balance between a vocabulary of Tetun only and one full of Portuguese loans. His latest contribution to the Timorese linguistics in a Tetun grammar *Língua Tétum – Contributos para uma Gramática* (2015) which is based on the Tetun that is, linguistically, close to the Church Tetun.

João Paulo Esperança is one of the few Portuguese linguists working on Tetun. In 2001, the author published a collection of articles about Timorese linguistics, *Estudos de linguística timorense*. In the article *Inventário de fonemas do tétum, um olhar sobre a sílaba e algumas questões sobre reduplicação* (2001), Esperança looked at the phonetic inventory of Tetun and discussed native and borrowed phonemes based on their position in a syllable. He also asserted that in Tetun Prasa, only partial reduplication existed. He drew on the examples of full reduplication from Sá (1961) and Hull (1993) and used elicitation to determine whether they could and would be used by Tetun Prasa speakers. The author asserted that the form of a partial reduplicated word was primarily subjected to the phonological characteristics of a root (Esperança 2001:69).

However, in his paper on reduplication and compounding in Tetun Prasa, a Romanian linguist, Andrei A. Avram (2008) acknowledged that full reduplication was a productive process in this variety although to a lesser extent than in other Austronesian languages. In the case of Tetun Prasa, it was restricted to four syntactic categories: nouns, adjectives, adverbs and cardinal numbers. Unlike Austronesian languages, especially Malay/Indonesian, there was no reduplication of verbs, not even partial. It was found in Tetun Terik, but it was very rare. The author also believed that phonological constraints were the deciding factor in total reduplication: only bases which were at least disyllabic could be fully reduplicated. As for meaning, neither Tetun Prasa nor Tetun Terik expressed diversity or plurality by reduplication. Avram also compared reduplication in Tetun Prasa to that of creoles and pidgins, following the proposal of Bakker (1994) that reduplication was a productive process in creoles and expanded pidgins but not in other pidgins. Based on this reasoning and the fact that Tetun Prasa had native speakers, the author asserted that "Tetun Dili seems to qualify for a double status, of a creole, for a minority of speakers, but an expanded pidgin for the majority of its speakers" (Avram 2008:437). However, unlike Tetun Prasa, verbal reduplication was attested in various Pacific pidgins and creoles. For a more detailed discussion on reduplication in Tetun Prasa see section 5.5.

In 2007, an Indonesian linguist, Yohanes Manhitu working on Tetun and other Timorese languages, published a bilingual Indonesian-Tetun dictionary (*Kamus Indonesia-Tetun Tetun-Indonesia*, 2007). Manhitu lived in the Oecussi enclave in the 1980s where he learned Tetun Prasa and Dawan, and then worked as an interpreter for the office of the United Nations High Commissioner for Refugees in Kupang. Since there was no Indonesian-Tetun Prasa dictionary, he decided to compile one with the support of INL and Geoffrey Hull, following the official orthography. As for the classification of Tetun, he followed the work of Morris (see section 4.2.3.7), stating that his dictionary was based on the Tetun Prasa dialect and was destined especially for Indonesian speakers and general public who wished to learn the official language of Timor-Leste.

John McWhorter, an American linguist specializing in language contact studies, researched Malay and other Austronesian languages spoken in Eastern Indonesia. He noticed that languages in central Flores and in the eastern part of Timor were nearly or completely isolating unlike other languages from the Central Malayo-Polynesian branch.

According to McWhorter, the differences between Tetun Terik and Tetun Prasa could not be attributed only to the contact with Portuguese; there must have been another explanation to the strikingly simpler grammar of Tetun Prasa, considering that Tetun Terik itself was already simplified. He did not believe that this was due to normal language change, because, compared to related languages, Tetun Prasa "stands out among its relatives in terms of overspecification, complexity and irregularity (McWhorter 2007:246)." According to him, the difference between Tetun Terik and Tetun Prasa was clearly caused by non-native acquisition. McWhorter followed the hypothesis of Hull (1998b, 2001a) that it was the Ambonese influence (13th century) that left its imprint and an unusual grammatical profile on Tetun Prasa. This brought him to the conclusion that the grammar of Tetun Prasa was very peculiar even in the cross-linguistic sense: "The vast majority of the world's languages have either inflectional morphology or tones that are either lexically or morphosyntactically contrastive. Only a few languages that are not creoles lack both." (McWhorter 2007:249).

However, in 2011, McWhorter (2011) contested the hypothesis of Hull suggesting that Ambonese influence was just an adstrate matter because it did not explain why the languages in central Flores were even more analytical then the languages in Timor (McWhorter 2011:236). He also did not agree with the Hull's classification of Timorese languages as Western Malayo-Polynesian, although looking at the lexicon, the lexical items in these languages were phonetically similar to their cognates in Tukang Besi, a Western Malayo-Polynesian language spoken in south-eastern Sulawesi. McWhorter proposed that there was a second Austronesian migration from Sulawesi and that "at some point much later than 3000 years ago, speakers of Tukang Besi and/or related languages migrated to Flores and Timor in large numbers and acquired the local languages incompletely amidst language shift." (McWhorter 2011:245). It was this incomplete acquisition that resulted in relexification and grammatical simplification. However, there was no grammatical mixture as in the case of creoles, which lead the author to call these languages nonhybrid conventionalized second-language varieties (NCSLs). This also refuted the idea that the loss of affixation was caused by the fact that Tetun Prasa has been used as a lingua franca. There are other languages spoken in Timor-Leste like Tokodede, Mambae or Waima'a that have very little affixation and were never lingua francas (McWhorter 2011:233).

Davi Albuquerque, a Brazilian linguist specializing in ecolinguistics and language contact between Timorese languages and the Portuguese spoken in Timor-Leste, wrote his master thesis (2011) on the first grammar sketch of Tetun Prasa that had been produced in Portuguese (excluding the 2005 translation of Hull and Eccles's grammar). In his second chapter, Albuquerque focused on the influence of Portuguese on not only Tetun but also on other Timorese languages and proposed the following diagram:

Macau CP   =>   Tetun Prasa   =>   other native languages

Malacca CP   Timor-Leste Portuguese

                 Bidau CP
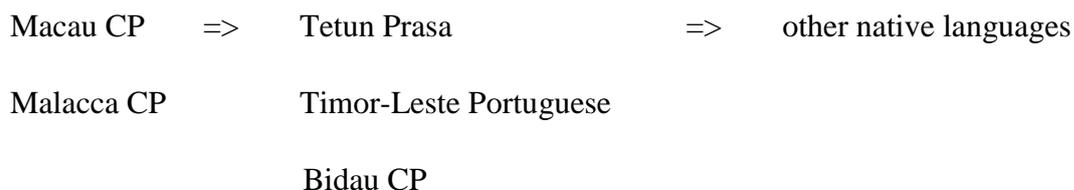
Considering the limited number of native European Portuguese speakers, the influence on Tetun and Timor-Leste Portuguese occurred via Macau Creole Portuguese and Malacca Creole Portuguese. Later on, the native Timorese languages would go to Tetun to borrow the necessary lexicon (Albuquerque 2011:30).

Looking at the works of Silva (1889) and Dores (1907), Albuquerque concluded that items like conjunctions and prepositions were borrowed directly from standard Portuguese since these loans were not attested in older works. These new Portuguese borrowings, especially the scientific terms, were sometimes very similar to the Indonesian ones (derived from Dutch and ultimately very often also from Latin), which created confusion regarding their pronunciation and orthography, e.g. Port. *definição* 'definition', TP *definisaun*, Ind. *definisi* (Albuquerque 2011:36). Albuquerque divided the influence of different 'varieties' of Portuguese into five periods. The first one was the period of the Malacca Creole Portuguese which was the lingua franca in this region from the 16th to the 18th century. After the transfer of the capital to Dili in 1769, which was administrated by the diocese in Macau, many speakers of Macau Creole Portuguese came to the new capital. This period of Macau Creole Portuguese influence lasted until the 19th century. The third period of the influence of the European Portuguese started with the foundation of the Soibada mission in 1898 and lasted until the Indonesian invasion in 1975. Between 1975 and 1999, during the Indonesian occupation, the use of Portuguese was prohibited. It was spoken only in the rural and mountainous regions by Timorese resistance groups. The fifth period started with the independence of Timor-Leste in 2002 and the fact that

Portuguese was proclaimed a co-official language together with Tetun (Albuquerque 2011:73).

In the chapter on phonology, Albuquerque looked at the foreign phonemes incorporated into Tetun Prasa like /p/, /v/, /z/ or /g/ and others that are only found in loanwords like /ɲ/, /ʎ/, /ʒ/ and /ʃ/. The author believed that these palatal consonants were only pronounced by older and educated speakers and the same applied to nasalization (Albuquerque 2011:85-89). Since the stress in Tetun Prasa is fixed on the penultimate syllable, speakers with little knowledge of Portuguese encounter problems pronouncing Portuguese borrowings that kept their original stress pattern (Albuquerque 2011:92).

Albuquerque also commented on the three different orthography proposals. He strongly disagreed with the official orthography based on the proposal of *Instituto Nacional de Linguística* for two main reasons: it makes use of inadequate graphemes like <ll> to represent [ʎ] or <ñ> to represent [ɲ] and it is very artificial for both L1 and L2 speakers of Tetun Prasa. The proposal of the Catholic Church follows the tradition of the Soibada mission and uses the translation of the New Testament as a base for their orthography. Another, more simplified proposal was carried out by Williams-van Klinken (see section 4.2.3.10) based on her linguistic studies (Albuquerque 2011:96-98).

As for the morphology of Tetun Prasa and its lack of flexion, Albuquerque followed the theory of irregular linguistic transmission of Lucchesi (1994) and the theory of non-native acquisition of McWhorter (2007), but he also worked with the hypothesis of Ansaldo and Nordhoff (2009) according to which grammatical simplicity and complexity is not linked to the age of a language, but to the linguistic environment[89]. Albuquerque argued that Tetun Prasa has developed a 'new' flectional morphology that has been going through a stage of grammaticalization due to extensive contact with several Portuguese varieties from the 16ᵗʰ until the 19ᵗʰ century (Albuquerque 2011:99-100). This influence could, for example, be seen in serial verbs. He suggested that there was a clear Portuguese origin behind the serial verb constructions in Tetun Prasa, e.g. *baa*

---

[89] This means "that in contact environments speakers derive their new grammars by a process of competition and selection of existing features. Therefore, in a morphologically rich environment, morphology will emerge. In a typologically isolating environment, this would obviously not be the case. Just as simple does not imply 'creolization' (Ansaldo and Matthews 2001; Sampson 2006), 'complex' does not seem to imply old age." (Ansaldo & Nordhoff 2009:360).

dansa 'go dancing' (from Port. *ir dançar*), *foo han* 'give food, feed' (from Port. *dar de comer*) although he was aware of the fact that these kinds of constructions were productive even before the rise of Tetun Prasa (Albuquerque 2011:127).

4.3   Literature on Timor-Leste Portuguese and Bidau Creole Portuguese

Many authors, linguists or not, commented on the presence of the Portuguese and its creolized form Bidau Creole Portuguese in Timor. Some described it as a 'corrupted Portuguese', some mistook it for other varieties of Asian Creole Portuguese, and some recognized it as a local variety of restructured Portuguese. One of the first people who thought a special variety of creole Portuguese might have developed in Timor was Hugo Schuchardt, a German linguist at the University of Graz in Austria. He maintained correspondence with various Portuguese nationals stationed on the island, asking them for more detailed accounts of the linguistic situation in Timor, and its capital Dili.

The first to comment on Portuguese spoken in Timor was in 1867 Affonso de Castro (see section 4.2.3.1):

> "Têem os timores tão grande facilidade para aprender o malaio, quanta difficuldade para o portuguez, que rarissimos fallam correctamente, e que uma grande parte dos chefes e dos habitantes de Dilly estropiam, fazendo um crioulo, que nos custa quasi tanto a comprehender, como os dialectos timores."[90] (Castro 1867:328).

However, it is disputable whether what Castro (1867) described as a creole was actually a creole rather than an L2 variety of Portuguese or a set of local interlanguages. As Baxter and Cardoso (2017:267) put it: "In this particular case, the description appears to imply that what was classified as "crioulo" was synchronically produced by a certain difficulty in the acquisition of Portuguese as an L2. Therefore, Castro is probably not referring to a variety that would be used by its speakers as L1 at all."

---

[90] "The Timorese have as much ease in learning Malay as they have difficulty in learning Portuguese, which very few speak correctly, and which a majority of the chiefs and inhabitants of Dili mutilate, producing a creole that is almost as difficult for us to understand as the Timorese languages." (Baxter & Cardoso 2017:266)

Another reference to Portuguese spoken in Timor, and especially in Dili, was made by a Portuguese agronomist Tancredo de Casal Ribeiro (1882 in Cardoso 2016b) in a letter to Hugo Schuchardt. Although he himself denied the existence of a creolized variety of Portuguese in Timor, he noted that there was a small part of the Timorese population living in Dili that spoke:

> "(...) a 'corrupted' form of L2 Portuguese highly influenced by their L1 and that, in the hinterland, only a few members of the local nobility spoke Portuguese at all. This form of Portuguese spoken in Dili was known locally as 'lingua da praça', the word *praça* referring not only to a marketplace but also, in the Luso-Asian context, to an important town." (Baxter & Cardoso 2017:268).

Casal Ribeiro (1882) listed a couple of examples which might point to the direction of a creole, although many of the phenomena he described can be explained on the basis of substrate influence. Baxter and Cardoso (2017:272), taking in consideration the possible substrate influence of Tetun and Malay, believed that certain features, like the TAM markers *já* 'PFV' and *há de* 'FUT', 3S verb forms and the negator '*nunca'*, found also in other SE Asian varieties of creoles or L2 varieties of Portuguese, could be an evidence of a creole present in Dili.

There is a very intriguing account of languages spoken in Dili by João Gomes Ferreira, the Vicar General of Timor who wrote a letter to Hugo Schuchardt in 1885. Schuchardt was interested in the sort of Portuguese that was spoken in Dili but the Vicar General asserted that Portuguese was spoken as L1 only by the Portuguese from Europe who lived there and that the vernacular of the capital was actually Tetun: "Na praça de Dilli falla-se a lingua = *Tetum* =, que é para assim dizer a lingua official de Timor inteiro, pois que em todos os reinos se encontra alguem que a saiba"[91] (João Gomes Ferreira in Sousa 2013:1). However, the Vicar General did notice there was some sort of "portuguez corrompido, a que se não pode chamar lingua da praça"[92] (João Gomes Ferreira in Sousa 2013:1-2) spoken by many indigenous people. According to Baxter (1990:6), the Vicar General failed to notice the existence of a creole Portuguese, now called Bidau Creole

---

[91] "In the town of Dili, the Tetum language is spoken and is, so to speak, the official language of the whole Timor, since in all kingdoms one finds someone who knows it." (Baxter 1990:5)

[92] "corrupt Portuguese, which cannot be called the language of the town" (Baxter 1990:5)

Portuguese (BCP) – a well-known phenomenon among speakers of a superstrate language who have difficulty identifying pidgins and creoles based on the same language in diglossic and multilingual situations:

"Por estes poucos exemplos que ahi ficam (see example (2), e que V.Exc.[ia] poderá examinar comparando o portuguez corrompido com o *tetum*, vê-se claramente que esse modo de fallar é mais ou menos a traducção á letra do mesmo tetum. E em todo o caso não constitue uma lingua – Em Macau dá-se uma cousa inteiramente differente: os Macaistas fallam o portuguez corrompido, mas é essa a lingua d'elles. Nao têem outra."[93] (João Gomes Ferreira in Sousa 2013:4).

Here are some of the examples of *portuguez corrompido* excerpted from the letter of the Vicar General:

(2)   a. (original spelling, João Gomes Ferreira in Sousa 2013:2-3)

Portuguese:	*É do Governo*
Tetun:	*Estado ni nia*
Portuguez corrompido:	*Estado sua.*
English:	It's the government's.

b. Portuguese:	*O pôrco está gôrdo.*
Tetun:	*Fahi bócur.*
Portuguez corrompido:	*Pôrco tem gôrdo.*
English:	The pig is fat.

c. Portuguese:	*Que estás a fazer? Nada.*
Tetun:	*Ó hallo sá ida? Lai.*
Portuguez corrompido:	*Tu faze que cousa? Não.*
English:	What are you doing? Nothing.

Vaquinhas (see section 4.2.3.2), in the series of letters he sent to Schuchardt, besides mentioning Batavia Creole Portuguese, also answered some questions regarding the Portuguese spoken in Timor. Vaquinhas (1884) confirmed the existence of Batavia Creole Portuguese, to which he referred as *portuguez corrupto* 'corrupted Portuguese'.

---

[93] "From these few examples shown here you will be able to compare the corrupt Portuguese with that of Tetum. It can be seen clearly that this way of speaking is more or less a direct translation of Tetum. And in any case, it does not constitute a language. In Macao, an entirely different thing occurs. The Macanese speak corrupt Portuguese but it is their language. They have no other." (Baxter 1990:5)

He found the same 'corrupted Portuguese' in some places in Flores and neighbouring islands. In his letter from 1885, he said he would provide Schuchardt with some notes on the creole Portuguese spoken in Dili, which was, according to him, the same as the one spoken in Larantuka. Unfortunately, the annex of this letter has not been found, but a part of it is repeated in his other letter to Schuchardt from 1886. Vaquinhas (1886) believed that the reason why other people denied the existence of these varieties spoken in Batavia and the Lesser Sunda islands is because they failed to learn the local languages, travel into the interior of the islands or find a reliable interpreter. This might be the reason why some authors, like Casal Ribeiro (1882) and Ferreira (1885) did not recognize a Portuguese creole or a restructured variety of Portuguese spoken in Timor, unlike Vaquinhas who spent sixteen years living in Oceania and learned many local languages. Although not a trained linguist, I believe that the linguistic data Vaquinhas provided can be trusted.

In his doctoral dissertation on Portuguese dialectology, Leite de Vasconcelos (1901) included a chapter on the Portuguese of Timor. He got his information from the Portuguese lieutenant-colonel Raphael das Dores (see section 4.2.3.4) who had visited Timor four times. In his first letter to Leite de Vasconcelos in 1901, Dores (1901) denied the existence of a Portuguese creole in Timor completely, saying: "Em Timor os indígenas que fallam portuguez, fallam-o exactamente como nós, não havendo patuá, como em Macau e nas outras colónias."[94] (Dores 1901). In his second letter, Dores (1901) acknowledged the presence of Macau Creole Portuguese, saying it had been brought over by Macanese who had settled down in Timor:

> "Algumas raparigas, vindas do interior para criadas, e servindo em casa de pessoas de Macau residentes em Timor, aprendem palavras do *crioulo macaísta*, mas tanto estas raparigas, como as próprias pessoas de Macau, com o tempo chegam a fallar o português como nós, o que eu observei, e mesmo se nota em Macau nas senhoras que regressam de Timor."[95] (Vasconcelos 1901:151-152).

---

[94] "In Timor, the natives who speak Portuguese speak it exactly like us, there not being any *patois*, as in Macau and in the other colonies." (Baxter & Cardoso 2017:285)

[95] "Some girls, coming from the interior to work as maids in the houses of people from Macau living in Timor, learn words from Macau Creole Portuguese, but these girls as well as the people from Macau eventually start speaking Portuguese like we do – this is what I have observed, and the same can be said about ladies who return to Macau from Timor." (my translation)

It is possible that Dores mistook Bidau Creole Portuguese for Macau Creole Portuguese, which is not surprising due to the close proximity between these two languages as well as Malacca Creole Portuguese and Tugu Creole Portuguese, all influenced by Malay (Thomaz 1985:332-334).

The fact, that some authors in the late 19th century failed to notice Bidau Creole Portuguese (Dores) and some have mistaken it for L2 Portuguese (Casal Ribeiro and Ferreira) was discussed by Baxter and Cardoso (2017:288). The authors listed as one of the main reasons the fact that the contact situation at the end of the century involved a whole range of L2 contact Portuguese and possibly some creole languages, especially when we consider that Bidau and the speakers of Bidau Creole Portuguese could be described as a melting pot. Also, many of these authors were not trained linguists and there might have been a lot of variation in the understanding of a creole language.

Alberto Osório de Castro (see section 4.2.3.6) also noticed a variety of creole Portuguese spoken in Bidau by the families of soldiers and officials. Here is a short example:

"- Vósse bai na ôndi? – Eu bai na riba. – Vósse bai faze cuza, bê? – Eu leva êsté cànico, vai sólè águ. – Vosse tira águ, faze cuza, bê? – Eu tira águ, bai cozinha arroze, ântis meu marido bên, cómi."[96] (Castro 1943:56).

The author said it would be interesting to compare the Creole Portuguese of Bidau spoken by the second generation of families from Sikka (Flores) and the Sikkanese language spoken on Flores (Castro 1943:57).

Luís Filipe Thomaz (see section 4.2.3.7) was the first historian to admit the possible presence and influence of creole Portuguese in Timor. According to him, Portuguese spread in this region three ways: through commerce (from 1515), missionary work (from 1590), and political domination (from 1702). The Portuguese used in commerce was, however, not the official and literary Portuguese the missionaries and Portuguese administration used: "não repugna admitir que se tratasse já de um crioulo,

---

[96] "- Where are you going? – I'm going up there. – What are you going to do there? – I'm taking this jug to draw some water. – What are you going to do with that water? – I'm going to get water to cook rice before my husband comes back to eat." (my translation)

gramaticalmente simplificado, mais que do português oficial e literário"[97] (Thomaz 1985:316). Thomaz (1974) also commented on the Bidau Creole Portuguese that became extinct in the 1960's, which he attributed to the fact that the Portuguese maritime network disappeared.

In 1990, an Australian linguist, Alan Baxter, wrote the most exhaustive paper on Bidau Creole Portuguese till now, *Notes on the Creole Portuguese of Bidau, East Timor*. As the basis for his research, he used the few existing data available: the letter from the Vicar General of Timor (see above) and the tape recordings collected between 1953 and 1954 by the Anthropological Mission of Timor led by António de Almeida (Baxter 1990:6). The author compared the phonological and morphosyntactic features of Bidau Creole Portuguese to those of other Southeast Asian varieties of Creole Portuguese like Malacca, Macau and Tugu. Baxter (1990) concluded that Bidau Creole Portuguese is closely related to Malacca Creole Portuguese and Macau Creole Portuguese, although it was very difficult to say which one it resembled the most since in the period when Bidau Creole Portuguese was formed, these two creoles were very similar. However, there are some phonological and syntactic features that set Bidau Creole Portuguese apart, which might have been caused by its contact with Tetun (Baxter 1990:28-29).

Based on the published works on Bidau Creole Portuguese, Artur Marcos (1995:123) concluded that no creole language ever originated in Timor-Leste but was brought there by immigrants: in the case of Bidau Creole Portuguese, by the Larantuqueiros coming from Flores via Lifau. Bidau Creole Portuguese can be indeed considered an extension of the variety of Portuguese spoken in Larantuka by the people coming from Malacca and Makassar after both of the cities were conquered by the Dutch (Alan Baxter, p. c.).

Grimes et al. (1997:52) also identified a Timor Pidgin (also referred to as Timor Creole Portuguese) spoken around Bidau in Dili and Lifau. It had two dialects: *Português de Bidau* and *Macaísta*. Not many details were given about the number of speakers or the language features. The authors only commented on the fact that similar varieties of creole Portuguese were spoken in Jakarta as well as in Flores, Solor, Sumatra, Kalimantan,

---

[97] "we cannot refuse to accept that it was already a creole, grammatically simplified, rather than the official and literary Portuguese." (my translation)

Sulawesi and Maluku (Grimes et al. 1997:47). As for the Malay language, the authors also identify a variety spoken around Timor-Leste called Dili Malay (Grimes et al. 1997:10).

Hajek (2000) (see section 4.2.3.10) also noticed two other varieties of Portuguese: Macau Creole Portuguese that was spoken in Dili in the 19th century by people coming from Macau, and Bidau Creole Portuguese, which became extinct in the 1960s. He believed that the speakers of these two varieties slowly shifted to standard European Portuguese.

João Paulo Esperança (2001) (see section 4.2.3.11), in his paper on Bidau Creole Portuguese, tried to show the influence of this language on Tetun Prasa which was, according to the author, significant. However, Esperança did not list a single example in which the influence of Bidau Creole Portuguese would be obvious – he himself said that the morphosyntactic influences he listed were just possibilities (Esperança 2001:25). As for the vocabulary, Esperança followed the proposal of Baxter (1990) that Bidau Creole Portuguese was a variety of Southeast Asian Creole Portuguese (SEACP) and tried to verify whether the loans that Thomaz (1974, 1981, 2002) considered borrowings from Portuguese and Malay were actually found in the SEACP.

The lastest contribution to the discussion on Bidau Creole Portuguese has been done by Baxter and Cardoso (2017) who revisited Baxter's article on Creole Portuguese of Bidau (1990) and looked at various and until now unpublished archival epistolary sources – letters sent to Hugo Schuchardt and José Leite de Vasconcelos. Because of the limited linguistic data and a lot of language variation, it is often difficult to identify and reconstruct the languages spoken in a certain linguistic space. In order to interpret these data, the authors looks at specific linguistic features of Portuguese and Creole Portuguese in Timor: those that were common for SEACP, those that were specific for Timor area and those that resemble L2 acquisition, to confirm the presence of Portuguese-based creoles and L2 variety of Portuguese spoken in Timor.

5.   MORPHOLOGY

5.1 Introduction

The structure of Tetun Prasa reveals a lot about the sociolinguistic history of the language and the partial restructuring that Tetun Terik underwent. In this chapter, I will discuss the morphological features in Tetun Prasa that differ from those in Tetun Terik as evidence of the contact with various Austronesian and European languages. Compared to other languages from the Central Malayo-Polynesian branch of Austronesian languages, Tetun Prasa has a strikingly analytical and isolating nature (Hull 1998b, McWhorter 2011). It does not show rich inflectional and derivational morphology and the few inflectional affixes that can be found in Tetun Prasa survive in a fossilized form. However, they are still actively used in various dialects of Tetun Terik (section 5.2). As for derivational morphology, there are several verbal prefixes and some nominalization suffixes (section 5.3). Other forms of word-formation include productive compounding (section 5.4) and reduplication with its various semantic properties (section 5.5).

5.2 Inflectional morphology

In this section, I discuss inflectional morphology, focusing mainly on subject marking, a feature still preserved in Tetun Terik but absent in Tetun Prasa (section 5.2.1). Other examples of inflectional morphology are described in section 5.2.2, followed by a discussion on Tetun Prasa's analyticity (section 5.2.3).

5.2.1   Subject marking

The loss of inflectional morphology in Tetun Prasa is a clear sign of language contact and second language acquisition, considering that inflection is still present in the more conservative variety of Tetun Terik. Aparício da Silva (1889:iii-iv) himself noticed the differences in inflectional morphology between the Tetun spoken in Dili and that spoken in the interior. Although he opted to describe Tetun Prasa, he later realized that

due to the higher complexity of the Tetun from the interior, this should be considered the 'more correct' one.

Silva (1889:18) described the inflection of Tetun Terik verbs that start with a vowel (except for *i-*) and *h-* (except for *hi-*). These change to *c-* in the 1st person singular, to *m-* in the 2nd person singular, to *n-* in the 3rd person singular and to *r-* in the 3rd person plural. As an example, Silva (1889:20-21) conjugated the verb *haruka* 'send' in Tetun Prasa and Tetun Terik.[98] These inflections are not confined to verbs only, but are applied to other parts of speech as well, e.g. the preposition *ho*[99] 'with' or the adverb *hamutuk* 'together'. What Silva (1889) and Dores (1907) really observed was a subject marking on verbs that is still used in some dialects of Tetun Terik (see Table 8).[100]

Van Klinken (1999:172-175), studying Tetun Fehan, noticed that all *h-*initial verbs (transitive and intransitive) in this dialect take subject-marking with the prefix replacing the initial consonant *h-*. In the case of other consonants, subject marking is used only in the first person singular (*ha'u k-*). This, apparently, has to do with the fact that all consonant clusters in Tetun Fehan begin with *k-* (van Engelenhoven & Williams-van Klinken 2005:735). In the case of serial verbs, only the first verb in a verb sequence takes subject-marking. Vowel-initial verbs have no subject marking.

The marking is obligatory on *h-*initial verbs in written Tetun Fehan though often omitted in speech. A decisive factor favouring subject marking is the presence of an overt personal pronoun in the clause, although there are cases of subject marking with covert pronouns, too. Interestingly, subject marking is not restricted to nouns, but can be found also with predicative adjectives, e.g. *di'ak* 'well', as shown in (3), and prepositional verbs, as van Klinken (1999) refers to them, e.g. *hó* 'with, accompany', as shown in (4).

---

[98] Raphael das Dores (1907:17-18), basing his work on that of Silva (1889), also noticed that there is no conjugation of verbs in the Tetun spoken in Dili, but he attested that in the interior, the verb *haruka* 'send' changes to *karuka, naruka, raruka*. He was able to observe this conjugation only with this particular verb and commented on the fact that it is not found in Tetun Prasa.

[99] In the works of Silva (1889) and Dores (1907), *ho* (also spelled *hó*) is considered a preposition, however, in Tetun Fehan, it is still used as a verb meaning 'accompany' (van Klinken 1999:302). It is possible that it was originally a verb also in Tetun spoken in the eastern part of the island and that later it has grammaticalized into a preposition.

[100] There is no subject marking for the first and second person plural. It is believed that these markers were lost in the diachronic process, since they existed in Proto-Central Malayo-Polynesian (PCMP) language (van Klinken 1999:173) and still exist in many Austronesian languages (Himmelmann 2005:150). Blust (1993:269) lists the following subject markers for PCMP: 1S **ku-*, 2S **mu-*, 3S **na-*, 1PI **ta-*, 1PE **ma-*, 2P **mi-,* 3P **da-*.

(3)   Tetun Fehan (van Klinken 1999:176):

    ha'u   ***k**di'ak*   *basuk*
    1S   1S.good   very

    'I'm very well.'

(4)   Tetun Fehan (van Klinken 1999:272):

    ha   ***k**bá*   ***k**ó*   *feto*   *sia*
    1S   1S.go   1S.with   girl   PL

    'I go with the girls.'

The use of subject marking varies by region. Hull and Eccles (2005) documented this phenomenon in Tetun Terik (spoken in the district of Viqueque) and in Tetun Belu (spoken in the districts of Balibó and Suai). According to their research, subject marking in Tetun Terik is used only with *h-* and vowel-initial verbs with the subject marker replacing the initial *h-*. In the case of Tetun Belu, it is the consonant-initial verbs (except for those starting with *h-*) that take the subject marking, which is added to the verb without any change to the original form of the word.

Based on the available literature, I summarized the data collected by Hull and Eccles (2005) and van Klinken (1999) about subject marking in different varieties of Tetun, as documented by Table 8.

halo 'do'/ *bá* 'go'	Tetun Prasa	Tetun Terik (Hull & Eccles 2005)	Tetun Belu (Hull & Eccles 2005)	Tetun Fehan (van Klinken 1999)
1S *ha'u*	*halo/bá*	***k**alo/bá*	*halo/**k**bá*	***k**alo/**k**bá*
2S *ó*	*halo/bá*	***m**alo/bá*	*halo/**m**bá*	***m**alo/bá*
3S *nia*	*halo/bá*	***n**alo/bá*	*halo/**n**bá*	***n**alo/bá*
1PI *ita*	*halo/bá*	*halo/bá*	*halo/bá*	*halo/bá*
2PE *ami*	*halo/bá*	*halo/bá*	*halo/bá*	*halo/bá*
2P *imi*	*halo/bá*	*halo/bá*	*halo/bá*	*halo/bá*
3P *sira*	*halo/bá*	***r**alo/bá*	*halo/**n**bá*	***r**alo/bá*

Table 8 Subject marking on verbs in Tetun Prasa, Tetun Terik (Hull & Eccles 2005), Tetun Belu (Hull & Eccles 2005) and Tetun Fehan (van Klinken 1999)

However, my data contradict the distinction suggested by Hull and Eccles (2005). My data from Suai (where so called 'Tetun Belu' is spoken) show that subject marking is still used in everyday speech and it is used almost exclusively with *h-* initial verbs, as in (5). I was unable to document subject marking used with C-initial words in this variety.

(5)   Tetun Suai: *hanorin* 'to teach'

*Agora ó   **m**anorin   iha-ne'ebé?*
now   2S   2S.teach   LOC-where
'Where do you teach now?' (0017TS_MN_44_F_Sec_TT)

*Ha'u **k**anorin   iha   EBC   Suai Loro,   **k**anorin   kelas*
1S   1S.teach   LOC   EBC   Suai Loro,   1S.teach   class

empat.
five

'I teach at the EBC Suai Loro, I teach the 5th year.'
(0017TS_MCA_29_F_Sec_TT)

At the same time, I found examples of *h*-initial but also vowel-initial words, as in (6), that take subject marking in Tetun from Viqueque ('Tetun Terik' according to Hull and Eccles 2005). I encountered one example where the subject prefix is used with a C-initial word, as in (7), and, as in the case of Tetun Fehan, in Tetun Terik from Viqueque this prefix was also *k-* '1S'.

(6)   Tetun Viqueque: *oho* 'kill'

[…] *entaun ha'u mós   la,   la   **k**oho animál*
        so   1S   also   NEG   NEG   1S.kill animal

lekirauk   ho   ida-ne'e.
monkey   with   this.one

'[…] so I did not kill the monkey.' (0049TV_MS_43_M_Mid_TT)

(7)   Tetun Viqueque: *rona* 'hear' and *husi* 'from'

[...]	*ha'u*	*krona*	*ba,*	*krona*	*kusi*	*lia*
	1S	1S.hear	to	1S.hear	1S.from	voice

ida-ne'e [...]
this.one

'[...] I heard this voice [...]' (0047TV_AFA_50_M_Sec_TT)

Interestingly, subject marking is never used with Portuguese borrowings, the only exceptions being Portuguese borrowings which have been fully adopted in Tetun Prasa and are used with the native causative prefix *ha-*, such as *hasai* 'take out, remove'.

I analysed my Tetun Terik oral corpus to see how frequent subject marking is in certain contexts. I only considered *h*-initial verbs, excluding *oho*[101] 'kill' and *rona*[102] 'listen' although these two verbs occurred with subject markers, too. Also, I only considered cases in which there was an overt personal pronoun. I created two lists: one that contained only sequences of *personal pronouns + h-verb* and one that contained sequences of *personal pronoun + another verb/element + h-verb* within the same clause. I was interested in finding answers to the following questions:

- Is an overt personal pronoun required for subject marking?
- Can any other element stand between a personal pronoun and an *h*-verb (within the same clause)?
- Can subject marking occur after noun phrases?
- Can subject marking occur in a subordinate clause cross-referencing to the subject of the main clause?

In Tetun Fehan, the subject does not need to be specified, as in (8), since the subject marking shows the person and number of the implied subject (van Klinken

---

[101] According to Hull (1993:247), verbs beginning in a vowel get inflected in Tetun Terik. Van Klinken (1999:172) found only one such verb and so did I: *oho* 'kill'. The author believed that the reason why *oho* is inflected for person is because in Tetun Fehan, the same verb is *h*-initial: *ho'o* 'kill'.

[102] *Rona* is the only C-initial verb in my corpus that received subject marking. Just like in Tetun Fehan, it was only inflected for 1S (*krona*), as /kr/ is a frequent consonant cluster in Tetun, unlike other combinations.

1999:174). In case it is specified, the subject can be either a personal pronoun, as in (9), or a noun phrase, as in (10).

(8)   Tetun Fehan: *haré* 'see' (van Klinken 1999:174)

   Nák   *"Ó,*   ***k**aré*   *ti'an"*.
   3S.say   Oh   1S.see   already.

   '(He) said "Oh, (I) have seen (it)".'

(9)   Tetun Fehan: *hatene* 'know' (van Klinken 1999:174)

   Ó   ***m**atene*   *ká*   *lale?*
   2S   2S.know   or   no

   'Did you know (him) or not?'

(10)   Tetun Fehan: *hobun* 'watch' (van Klinken 1999:174)

   Ibu   ***n**obun*   *dansa*.
   mother   3S.watch   dance

   'You (mother) will watch the dancing.'

The same applies to Tetun Terik in which the subject, either a personal pronoun or a noun phrase, can be expressed, as illustrated by (11) and (12), or it can be omitted, as illustrated by (13).

(11)   Tetun Terik: *hadomi* 'love'

   Tanba   *sira*   ***r**adomi*   *na'i-lulik*   *ne'e*.
   because   3P   3P.love   priest   this

   'Because they love the priest.' (0050TV_JCV_48_M_Sec_TT)

(12)   Tetun Terik: *hatudu* 'show'

   Depois   *malae*   *ne'e*   ***n**atudu*   *tiha*   *nia*
   after   foreigner   this   3S.show   PFV   3S

   uman   *sira*   *ne'e*   […]
   house.PL   PL   this

   'Then the foreigner showed (them) his houses [...]'
   (0046TV_JCV_48_M_Sec_TT)

(13)   Tetun Terik: *haree* 'see', *hadeer* 'wake up'

>   ***N**aree ba   laran, ba   na'i-lulik   ne'e   toba   la*
>   3S.see to   inside to   priest   this   sleep   NEG
>
>   ***n**adeer*
>   wake.up
>
>   '(He) looked inside, the priest was sleeping and would not wake up.'
>   (0050TV_JCV_48_M_Sec_TT)

My corpus also shows sequences of verbs (and other word classes) following a subject. In Tetun Fehan, all *h*-initial verbs in this sequence can take subject marking. As for C-initial verb sequences, only the first verb is marked (van Klinken 1999:174-175). The same applies to Tetun Terik:

(14)   Tetun Terik: *hodi* 'in order to', *haree* 'see'

>   [...]   *lees   tiha   didin,   **n**odi   **n**aree   tama   ba   laran.*
>   tear   PFV   wall   3S.in.order.to   3S.see   enter   to   inside
>
>   '[…] (he) tore down a wall in order to look inside.'
>   (0050TV_JCV_48_M_Sec_TT)

As mentioned above, subject marking on *h*-initial verbs seems to be obligatory in written Tetun Fehan, although van Klinken (1999:175) noticed that 5% of these verbs were uninflected in her corpus. Inflection of C-initial verbs is much less frequent and mostly occurs with an overt subject pronoun within a clause. In my corpus of Tetun Terik, these numbers were much lower. In the case of sequences of a personal pronoun immediately followed by an *h*-initial verb, only 34% of these verbs occurred with subject marking, as shown in Table 9:

Total no. of *personal pronouns* + *h-verb*	*personal pronouns* + *h-verb* **with** subject marking	*personal pronouns* + *h-verb* **without** subject marking
95 occurrences	32 occurrences (34%)	63 occurrences (66%)

Table 9 Subject marking with personal pronouns immediately followed by an h-verb

The numbers were similar for sequences of *personal pronoun + other verb/element + h-verb* within the same clause, as shown in Table 10:

Total no. of *personal pronoun + another verb/element + h-verb*	*personal pronoun + another verb/element + h-verb* **with** subject marking	*personal pronoun + another verb/element + h-verb* **without** subject marking
53 occurrences	17 occurrences (32%)	36 occurrences (68%)

Table 10 Subject marking with personal pronoun + other verb/element + h-verb within the same clause

Often, these elements standing between a personal pronoun and the inflected verb were the negator *la*, as in (15), directional verbs *may* 'come' and *bá* 'go', as in (16), SVC elements, as in (17), preposed TAM markers *sei* 'FUT', *atu* 'IRR', *nafatin* 'continue', etc. and others.

(15)   Tetun Terik: *hateten* 'tell'

Lian	*seluk*	*ha'u*	*la*	***k**ateten*	*tanba*	*la'ós*
language	other	1S	NEG	1S.tell	because	NEG

ha'u	*lian.*
1S	language

'I do not speak other languages because they are not my languages.' (0047TV_AFA_50_M_Sec_TT)

(16)   Tetun Terik: *halo* 'make, do'

[…]	*nia*	*bá*	*ona*	***n**alo*	*tatakak*
	3S	go	ANT	3S.make	sling

'[...] he went to make a sling (to catch the monkey)' (0049TV_MS_43_M_Mid_TT)

(17)   Tetun Terik: *fó-hatene* 'inform', lit. 'give.know'

No	*ó*	*bá,*	*ó*	*tenke*	*fó-**m**atene*	*ba*	*ema*	*sira.*
and	2S	go	2S	must	give-2S.know	to	person	PL

'And go and let people know.' (0046TV_JCV_48_M_Sec_TT)

Apart from these two most frequent types of subject marking, there are other, more marginal types (53 occurrences in total). I found cases in which subject marking was used following a NP within the same clause, as shown in (18), without an expressed subject, as shown above in (13), or with a preposition *husi* 'from' and a conjunction *hodi* 'in order to', as shown in (19). I did not find examples of subject marking in subordinate clauses cross-referencing to the subject in a main clause.

(18)   Tetun Terik: *hariis* 'bathe'

Maibé malae-feto   ida   no   mane ida   sei
but   foreigner-woman   one   and   man   one   FUT

***r**ariis iha   wee   laran.*
3P.bathe LOC water   inside

'But the two foreigners, a woman and a man, will bathe in the water.' (0046TV_JCV_48_M_Sec_TT)

(19)   Tetun Terik: *husi* 'from'

*Ha'u   **k**usi   Suai Loro.*
1S   1S.from   Suai Loro

'I am from Suai Loro.' (0017TS_MCA_29_F_Sec_TT)

I was also interested in finding out whether younger speakers (14 -18 years old) of Tetun Terik still use subject marking. Although I found some instances of it, I am not sure whether these data could be deemed 'natural'. Speakers were instructed by my assistant to speak Tetun Terik, which might have interfered with their spontaneity. They were often heard correcting themselves while using subject marking, as if they had to concentrate on using it properly. Nonetheless, it can be concluded that young speakers of Tetun Terik do know how to use subject marking but further study with more natural data is needed to determine whether it is used in their everyday speech.

I also looked at the frequency of TAM markers used with verbs marked for subject but I have found only 6 occurrences:

(20)   Tetun Terik: *hateten* 'tell'

Ne'e-duni	*ha'u*	*ohin*	**k**ateten	*liu*	*tiha*	*ba*
so	1S	today	1S.tell	more	PFV	to

istória	*kona-ba*	*ha'u-nia*	*moris.*
history	about	1S-POSS	life.

'So I have said enough today about my personal history.' (0049TV_MS_43_M_Mid_TT)

Although most of the narratives refer to the past, the speakers often leave out the aspect markers *tiha* (PFV), *ona* (ANT) and *tiha ona* (PRF), which is a common strategy in the spoken discourse. See section 6.2.3 for further discussion on TAM markers.

Based on my data and previous studies, it can be concluded that subject marking is not used in Tetun Prasa anymore. However, it still has its place in Tetun Terik and other Austronesian languages spoken in Timor-Leste[103]: Galolen marks vowel-initial verbs for subject in all persons (Hull 2003c:22), Baikeno uses personal prefixes and circumfixes with all persons and with all verbs (Hull 2003b:29) and Mambae only uses personal prefixes on vowel-initial verbs in the third person singular (3S) (Hull 2003a:25).

5.2.2   Other examples of inflectional morphology

Among other cases of inflectional morphology found in Tetun Terik but not preserved in Tetun Prasa, or preserved only in a fossilized form, one can find the plural suffixes *-r/-n* (see section 6.1.2.5) and the possessive suffixes (genitive clitics) *-n/-r* (see section 6.1.3.3). At the same time, many Portuguese loanwords in Tetun Prasa are inflected for gender, e.g. *viziñu* (from Port. *vizinho* 'neighbour') and *viziña* (from Port. *vizinha* 'female neighbour') (see section 6.1.4) and number (see section 6.1.2.3) and these inflections can be found also in Tetun Terik speakers. However, because the use of these inflections often depend on the context, they are discussed in Chapter 6 and dealt with at the phrasal level.

---

[103] So-called Papuan languages Makasae and Fataluku do not present subject marking on verbs.

5.2.3 Discussion

It is not common for Austronesian languages to lack any kind of inflectional morphology as is the case of Tetun Prasa (and other languages from the Flores-Timor region). As McWhorter (2007:249) noted: "The vast majority of the world's languages have either inflectional morphology or tones that are either lexically or morphosyntactically contrastive. Only a few languages that are not creoles lack both." McWhorter (2007) believed this was due to non-native acquisition of Tetun Prasa. Williams-van Klinken (2010b), on the other hand, considered the most probable source the neighbouring languages and simplification.

This possibility was further discussed by Albuquerque (2011:100), based on the work of Ansaldo and Nordhoff (2009) that goes against the hypothesis of McWhorter (2005:10, 42-43) according to which 'new languages' like pidgins and creoles are less complex than other natural languages because they have existed for a shorter period of time. The authors based their analysis on Sri Lankan Malay that could be considered 'a young language' but shows signs of 'age' like inflectional and derivational morphology or agglutinative structure. Although only around 300 years old, Sri Lankan Malay has inherited these features via close contact with its adstrate languages Sinhala and Tamil that have rich inflectional morphology[104]. This led the authors to conclude that "in a morphologically rich environment, morphology will emerge. In a typologically isolating environment, this would obviously not be the case." (Ansaldo & Nordhoff 2009:360).

How does this apply to Tetun Prasa? Tetun Prasa has been in close contact with Portuguese, which has a rich verbal inflectional morphology but, at the same time, also with Mambae, spoken just outside of Dili, which has a limited amount of inflectional affixation. Albuquerque (2011:100) asserted that Tetun Prasa has in fact been developing its own inflectional morphology that is currently going through the stage of grammaticalization due to the co-existence with Portuguese and English. The author argued that Tetun Prasa developed new forms of 'functional transference', including

---

[104] Although McWhorter (2005:317) assumed that some grammatical features, like inflectional affixation, arise only over time and are thus not found in pidgins and creoles, he also acknowledged cases in which a creole can acquire inflectional affixes: "(…) such creoles result only from contexts in which the languages in contact happen to be closely related genetically, or in which the creole develops in close contact with an inflected older language." The latter is, for example, the case of Sri Lanka Creole Portuguese.

gender marking (*mane* 'MASC' and *feto* 'FEM'), definite marker *ida* 'one', TAM markers and directional verbs *ba* 'go' and *mai* 'come', although it is hard to imagine how these could be considered 'inflectional morphology'. This would also mean that these features (with the same functions) did not exist in the conservative variety of Tetun Terik and only became grammaticalized in Tetun Prasa after the contact with restructured varieties of Malay and Portuguese. This hypothesis seems very implausible considering that they not only exist in Tetun Terik, but they are also found in Tetun Fehan, which has been in limited contact with Tetun Prasa. The same strategies can be also found in other Timorese languages so I do not see how these could be contact-induced innovations in Tetun Prasa.

5.3 Derivational morphology

Tetun Prasa, as mentioned above, does not possess much derivational affixation, which might be caused by the fact that the word classes in this language are relatively fuzzy and conversion (or zero derivation) is thus one of the productive methods of forming new words (see section 5.3.8). Also, the result of the derivational affixation is not always clear in terms of word classes. Another problem arises when it comes to the distinction between a bound morpheme (an affix) and a free morpheme (a lexeme). Authors of the grammars and descriptions of Tetun Prasa (Williams-van Klinken et al. 2002a, Hull & Eccles 2005, Albuquerque 2011) do not always agree on the level of grammaticalization of certain morphemes, e.g. the agentive suffix/lexeme *na'in*, an issue discussed in the following subsections.

In this section, I discuss derivational morphology, analysing various productive and fossilized verbal, nominal, adjectival and other affixes found in Tetun Prasa, Tetun Terik and Tetun Fehan, as well as zero derivation.

5.3.1 Derivation of causative verbs: *ha-*[105]

The derivation of causative verbs is, according to Himmelmann (2005:170), "probably the most widely attested productive derivation in western Austronesian languages". This prefix derives causative transitive verbs from intransitive verbs, as in (21), adjectives, as in (22), and a small number of nouns, as in (23).

(21)   Tetun Prasa: *hasai* 'take out' from *sai* 'exit' (from Port. *sair*)

PNTL	*Hasai*	*Tan*	*umakain*	*Haat*	*hosi*	*otél*
PNTL	remove	more	household	four	from	hotel

Rezende
Rezende

'PNTL removed four families from Rezende hotel.' (STL_July_2013)

(22)   Tetun Prasa: *hamanas* 'heat up' from *manas* 'hot'

Ó	*lalika*	*hamanas*	*karik*	*di'ak*
2S	need.not	heat.up	maybe	good.

'You do not need to heat it up, if it is still good.'
(0014TD_JBP_26_M_Uni_Mak)

(23)   Tetun Prasa: *hadomi* 'love' from *domin* 'love'

[…]	*tenke*	*perdua*	*no*	*hadomi*	*malu*	[…]
	must	forgive	and	love	RECIP	

'(They) need to forgive and love each other […]' (STL_March_2016)

It is also possible to turn some transitive verbs into causative ones. In this case, the resulting verb has a slightly different or figurative meaning, as shown in (24) (Hull & Eccles 2005:86). The prefix *ha-* can be also applied to a base derived from Portuguese, as shown in (25) and Indonesian, as shown in (26). Historically, *ha-* is derived from a Proto-Austronesian prefix **pa-* and is still productive in Tetun Prasa.

---

[105] This prefix can trigger certain morphophonemic processes applied to the root. If the base is vowel-initial, *hah-* is used instead. Along with *ha-*, Morris (1984:xviii) listed another causative prefix to verbs *hak-*. However, van Klinken (1999:66) listed four different uses of *hak-* but none of them was that of causation. According to Hull and Eccles (2005:90), *hak-* (sometimes *ham-* before /l/, /r/ and /s/) is used to derive dynamic verbs from adjectives and nouns.

(24)   (Hull & Eccles 2005:86)

    hadulas 'surround; go around' from *dulas* 'spin, twist'

(25)   *hapara* 'put a stop to' from *para* 'stop' (from Port. *parar*)

    haforsa 'strengthen' from *forsa* 'strength' (from Port. *força*)

    hamenus 'reduce' from *menus* 'less' (from Port. *menos*)

(26)   *hadame* 'reconcile' from *dame* 'peace' (from Indo. *dame*)

There are also verbs that start with *ha-* but do not have causative meaning. In this case, in order to give them causative meaning, other constructions are possible, e.g. causative *fó* 'give' or serialization with *halo* 'make, do' (see section 6.2.4).

5.3.2   Derivation of intransitive verbs: *nak-*

The prefix *nak-* derives intransitive verbs from transitive verbs. If the verb base starts with *k-* and sometimes with *l-*, the prefix changes to *nam-*. Causative verbs that are already derived with *ha-* lose this prefix and replace it with *nam-* (Hull & Eccles 2005:90). In Tetun Terik, the prefix to derive intransitive verbs is *hak-* with a fossilized form *nak-* (3S subject marking inflection) used for verbs with non-human subjects (Williams-van Klinken et al. 2002a:19):

(27)   Tetun Terik:

    a. *naksobu* 'come apart' from *sobu* 'take to pieces'

Normál	*Koligasaun*	*CNRT-PD*	*Naksobu*
normal	coalition	CNRT-PD	come apart

    'It's normal that the CNRT-PD fell apart.' (STL_March_2016)

    b. *nakfera* 'be split' from *fera* 'split'

Ne'e	*tiha*	*sala*	*balu*	*mós*	*aat*	*tiha,*	*nakfera*
this	PFV	room	some	also	bad	PFV	be.split

kala	*anin*	*ne'e.*
maybe	wind	this

155

'Some classrooms are also in a bad state, they were shattered by wind.'
(0017TS_MCA_29_F_Sec_TT)

I have found only a few examples of *nak-* prefixation in my oral and written corpus. As for loanwords, this type of derivation has not been documented with Portuguese bases, except for *naksai* 'break out, emerge', derived from *hasai* 'take out, remove' (from Port. *sair* 'exit').

5.3.3 Other verbal affixes

Hull and Eccles (2005:88-89) listed a verbal prefix *na-* (*nah-* with a vowel-initial base) that derives inchoative verbs from adjectives, as in (28a), and nouns, as in (28b). These intransitive verbs express a beginning of a process or a natural state. This prefix is no longer productive in Tetun Prasa, in which a periphrastic construction with *sai* 'become' is used.

(28)   (Hull & Eccles 2005:88)

      a.   *namanas*[106] 'become/grow hot' from *manas* 'hot'

      b.   *nabee(n)* 'melt, liquefy' from *been* 'liquid'

This category also includes animal sounds and natural sounds:

(29)   (Hull & Eccles 2005:89)

      a.   *nameek* 'to bleat'

      b.   *nameo* 'to mew'

According to van Klinken (1999:69), 'make sound' verbs in Tetun Fehan are derived by applying a prefix *ha-* or *hak-* to the sound, as shown in (30a). In some cases, as in (30b), a suffix *-k* is added to the verb.

(30)   (van Klinken 1999:70)

      a.   *hakmé* 'to bleat'

---

[106] *Namanas* has also other meaning, that of 'busily, in the thick of' (Williams-van Klinken 2011), 'eager, eagerly' (Hull 1999:241).

   b. *hahoe(k)* 'to bellow'

Other uses of the *ha-* or *hak-* prefixes include: derivation of intransitive verbs from transitive ones, as mentioned above; derivation of verbs from other verbs without a change in their transitivity but with a similar or more restricted meaning; and derivation of reciprocal action verbs with a plural subject[107] (van Klinken 1999:66-69).

5.3.4 Derivation of actor nouns

To derive actor nouns, Tetun Prasa uses several strategies, such as the agentive suffixes *-dór*, *-teen*, and *-na'in*. There is one more strategy used exclusively in Tetun Terik, which is the agentive circumfix *mak-(-k)*.

5.3.4.1 Agentive suffix *-dór*

One of the few fully productive affixes[108] in Tetun Prasa is the agentive suffix of Portuguese origin *-dór* which is used in Tetun to derive agentive nouns from verbs. Interestingly, no other Austronesian language which has been in contact with Portuguese or Castilian (e.g. Tagalog and Chamorro) uses this or any other Romance suffix in a productive way (Hajek & Williams-van Klinken 2003). It is the reflex of Latin *-tore* and has been documented in Tetun Prasa as early as the 19th century. In his dictionary, Silva (1889:3:84) recognized it as a particle that was introduced from Portuguese as an imitation of words ending in *-dor*, and that it was used to express a certain habit to work/behave, for example *húcic dór* 'hunter' (original spelling, from Port. *caçador*), although he also listed the native expressions for 'hunter': *ma húcic, ema coáin, ema ba húcic, húcic ná'in* (original spelling).

---

[107] In this case, a circumfix *hak-k* is added to the base (van Klinken 1999:69). The same circumfix is listed in Hull and Eccles (2005:90-91) but with a different function: it derives intensive verbs, i.e. it gives intensive or figurative meaning to simple or causative verbs (e.g. *haktuir* 'to imitate, to recount' from *tuir* 'to follow').

[108] Hajek and Williams-van Klinken (2003) argue that *-dór* is the only productive agentive suffix in Tetun Prasa, coexisting with a base *-teen* and a noun *na'in*. In this thesis, I consider both *-teen* and *-na'in* derivational suffixes as I believe they have grammaticalized into affixes, losing their original meaning of 'faeces' and 'master', respectively.

Since then, the suffix has been used not only in Portuguese loanwords, as in (31), but also with Tetun roots, as in (32).[109] Native roots are used to derive agentive nouns with a habitual semantic feature or agentive nouns, sometimes with an adjectival meaning, that can carry a pejorative meaning (Hull & Eccles 2005:8, Albuquerque 2011:114), e.g. *handór* 'glutton', *hemudór* 'drunkard.' According to Williams-van Klinken (2011), the suffix has a neutral connotation, for example, *handór* does describe a person who eats a lot, but it is not necessarily a vice.

(31)   *administradór* 'administrator' (from Port. *administrador*)

         oradór 'speaker' (from Port. *orador*)

(32)   a.   *sa'edór* 'climber' from *sa'e* 'climb, ascend'

                sa'edór kuda 'horse rider', *sa'edór ai* 'tree climber'
                (0044_JCP_58_M_Sec_TT)

         b.   *ohodór* 'murderer' from *oho* 'kill' (0049TV_MS_43_M_Mid_TT)

The two examples in (32) come from my oral corpus and were recorded in a Tetun Terik-speaking area around Viqueque which can be seen as a proof that the borrowed suffix has been fully integrated into the native morphology of Tetun.

Some words containing this agentive suffix may resemble a Portuguese loan but further analysis shows that they were derived from a Tetun verbal base:

(33)   (Hajek & Williams-van Klinken 2003:61)

         pintadór 'painter' from *pinta* 'to paint' (from Port. *pintar*) in contrast with Port. *pintor* 'painter'

Some have a different form and/or meaning from their Portuguese counterpart:

(34)   *pasiardór* 'gadabout' from *pasiar* 'go for a stroll' (from Port. *passear* 'go for a stroll' and *passeador* 'stroller')

---

[109] Hajek and Williams-van Klinken (2003:60) also listed Malay roots, although they are relatively uncommon. As for gender agreement, only Portuguese loanwords can take a feminine form, e.g. *administradór* – *administradora* 'administrator – female administrator' (Williams-van Klinken et al. 2002a:20). However, in Tetun Prasa, the more common and accepted feminine form is *administradór feto* (female administrator, literally 'administrator.woman').

Some are Tetun Prasa neologisms that do not exist in Portuguese, e.g. *nervozudór* 'nervous person, irritated' from *nervozu* 'nervous, irritated'. And in some case, borrowed words were assimilated to fit the morphological structure of Tetun Prasa, e.g. *mandór* 'foreman' (from Indo. *mandur* 'foreman').

The same suffix can be also found in words that do not denote a person, but a thing or an instrument, as illustrated by (35). However, in this case, one cannot consider *-dór* a productive derivational morpheme, since it is only used with Portuguese loanwords. The reason for this possibly lies in the fact that the instrumental function of this suffix became productive in Portuguese only much later, probably as late as the 19th century[110], according to Albuquerque (2011:114). The same suffix can be found attached to verbs to form adjectives, as shown in (36); again, in this case, this derivation is only found in Portuguese loans.

(35)   *jeradór*   'generator'   (GovernuTL_2010)

       komputadór   'computer'   (0014TD_JMB_31_F_Sec_Mak)

(36)   *unifikadór*   'unifying'   (GovernuTL_2010)

       konservadór 'conservative' (Sapo_March_2016)

5.3.4.2 Agentive suffix *-teen*

Another agentive suffix used with verbal, adjectival and nominal bases is *-teen*[111], as shown in (37). According to the official orthography, it is used with a hyphen. The resultant word has a pejorative meaning, describing a person who habitually does the activity expressed by the root, or its negative moral qualities (Williams-van Klinken et al. 2002a:20-21; Hull & Eccles 2005:9). The negative connotation might have to do with the

---

[110] In archaic Portuguese, *-dor* had only agentive or experiencer meaning. The meaning never extended to non-human agents in this period. But with the spread of electricity, the suffix gained more productivity and was applied also to instruments that could function without any human interaction (Costa & Coelho 2013:70).

[111] According to Williams-van Klinken et al. (2002a:20-21), the result of this derivation is an adjective, although in English it is often translated as a noun. This might have led Hull and Eccles (2005:9) to list this suffix under the section on actor nouns. On the other hand, Silva (1889:9) asserted that many of these derived adjectives had a *-oso* ending in Portuguese, which is typical for the class of adjectives: *bosok-teen* (Port. *mentiroso*) 'liar', *baruk-teen* (Port. *preguiçoso*) 'lazybones'.

etymology of the noun *teen* 'faeces'. The suffix can be also applied to roots of Portuguese origin, e.g. *kafé-teen* 'coffee addict'.

(37) Tetun Prasa: *na'ok-teen* 'thief' from *na'ok* 'to steal'

Lezadu	*Paul Saldanha*	*mai*	*hasai*	*liafuan*
victim	Paul Saldanha	come	release	word

tolok	*no*	*dehan*	*na'ok-teen*	*ne'e*	*maromak*	*lalika*
insult	and	say	thief	this	god	must.not

tama	*uma.*
enter	house

'The victim Paul Saldanha swore and said that that God [statue of Saint Mary] should not enter the thief's house.' (STL_March_2016)

5.3.4.3 Agentive suffix *-na'in*

The third agentive suffix used in Tetun is *-na'in*. Originally, *na'in* meant 'owner, master' and the meaning listed in Silva's dictionary (1889:3) is that of Port. *senhor* 'mister; owner, master'. However, its meaning has been extended[112] and currently is also used to derive actor nouns. Again, there is no consensus on the nature of this suffix and the result of its derivation. Williams-van Klinken at al. (2002a:21) consider *na'in* a noun that follows a verb and forms a nominal compound that has the meaning of the "one who is or does x". Hull and Eccles (2005:8-9) and Albuquerque (2011:115) consider *-na'in* a suffix that can be attached to a verb (38a) or a noun (38b). When affixed to a noun, the suffix often retains its original meaning of 'master'[113]. The suffix can be attached to both native and borrowed bases, as shown in (38) and (39), respectively.

(38) Tetun Prasa:

a. *kaben-na'in* 'married person' from *kaben* 'to marry'

Ha'u	*ema*	*kaben-na'in.*
1S	person	marry-master

'I'm married' (0051TD_DLB_45_M_Uni_Kem)

---

[112] *Na'in* is also a classifier for humans (see section 6.1.5.).
[113] According to Hull and Eccles (2005:9), *-na'in* can also mean 'author'.

b. *lia-na'in* 'traditional orator' from *lia* 'news'

Iha	*ema*	*sira-nia*	*liurai*	*iha,*	*sira-nia*	*xefe,*
LOC	person	3P-POSS	liurai	LOC	3P-POSS	chief

sira-nia	*kbahen,*	*lia-na'in*	*sira,*	*kompletu.*
3P-POSS	elderly	news-master	PL	complete

'There were people, their liurais, their chiefs, their elderly, the traditional orators, everybody.' (0050TV_JCV_48_M_Sec_TT)

(39)   Tetun Prasa: *lei-na'in* 'lawmaker' from *lei* 'law' (from Port. *lei* 'law')

Nia	*hatutan*	*iha*	*governu*	*nia*	*laran*	*ne'e*
3S	extend	LOC	government	POSS	inside	this

ema	*lei-na'in*	*no*	*asesór*	*mós*	*barak* […]
person	law-master	and	advisor	also	many

'He added that there were also many lawmakers and advisors in the government.' (STL_March_2016)

5.3.4.4 Agentive circumfix *mak-(-k)*

Apart from the three agentive suffixes mentioned above, there is a circumfix *mak-(-k)*[114] which is still productive in Tetun Terik. It is used to derive actor nouns from native bases:

(40)   Tetun Terik: *maksa'ek* 'climber' from *sa'e* 'rise, climb, ride'

Ema	*ne'e*	*ami*	*dehan*	*sa'edór*	*maibé*	*Tetun*	*nasionál*
person	this	1PE	say	climber	but	Tetun	national

ninian	*dehan*	*katak*	*ema*	*maksa'ek.*
3S-POSS	say	COMP	person	climber

'We call this person a climber [sa'edór], but in national Tetun they say a climber [ema maksa'ek].' (0044TV_JCP_58_M_Sec_TT)

---

[114] If a word starts with a vowel, the prefix changes to *mah-*; if it starts with *h-*, the prefix changes into *ma-*. Also, if the base ends in a consonant, the suffix is omitted (Hull & Eccles 2005:8).

I elicited examples of such derivation with speakers of Tetun Prasa and although they were able to understand some of them, they claimed they were only used in Tetun Terik.

According to van Klinken (1999:70) a similar prefix/circumfix *mak-(-n)* and *ma-(-n)*[115] is used in Tetun Fehan to derive words describing somebody who has performed the action of the verb, or one who habitually performs it or intends to do so. Van Klinken (1999:77) also believed that his actor prefix/circumfix and the elative clause marker *mak* have a lot in common, since they are both translatable as 'who, which' and both function as noun modifiers.

5.3.4.5 Summary

I summarized the different ways of deriving of agentive nouns in Tetun Prasa and Tetun Terik/Terik Fehan in Table 11, considering their productivity and semantic value. Although the Portuguese suffix *-dór* has been competing with other three native affixes, it is clear that it has been fully adopted into Tetun morphology, evidenced by the fact that it is productive with native roots too in the more conservative dialects of Tetun.

Affix	Tetun Terik/Fehan	Tetun Prasa
-dór	marginally productive, neutral	productive, neutral
-teen	productive, pejorative	productive, pejorative
-na'in	productive, neutral	productive, neutral
mak-(-k)	productive	marginally productive

Table 11 Summary of agentive affixes in Tetun

---

[115] *Ma-* is applied to *h-* initial roots and stems, *mak-* is applied to disyllabic bases beginning with any other consonant. As for the suffixes, there is dialectal variation and no apparent logic behind their use. However, they are applied only if the base is vowel-final. No suffix is applied if the word has a following object NP (van Klinken 1999:73-74).

5.3.5 Other nominal suffixes and circumfixes

Hull and Eccles (2005) listed two more affixes that are further found only in van Klinken's grammar of the Fehan dialect (1999), which would suggest that they are not productive in Tetun Prasa anymore. However, they exist in this variety in a fossilized form. The first one is the suffix *-n* used to derive deverbal nouns that are a product or result of an action:

(41)   (Hull & Eccles 2005:7)

   a. susun 'breast' from *susu* 'suckle'

   b. *futun* 'bundle' from *futu* 'bind'

Both Hull and Eccles (2005:6) and van Klinken (1999:82) believed that many instances of this suffix were historically genitive markers. This especially applies to body parts (like *susun* 'breast'), which belong to the group of inalienable nouns. These are often inherently possessed and many of them have preserved a final *-n*, a fossilized genitive clitic (see section 6.1.3).

The second one is the circumfix *ka*[116]*-(-k)* used to derive either concrete nouns from adjectives, as in (42a), and dynamic verbs, as in (42b), or nouns with a metaphoric meaning from a nominal base, as in (42c) (Hull & Eccles 2005:7). However, in Tetun Fehan, the same circumfix derives an instrument or undergoer.

(42)   (Hull & Eccles 2005:7)

   a.   *kamedak* 'stain' from *medak* 'dirty'

   b.   *kafotik* 'rise (of land)' from *foti* 'to raise'

   c.   *kabeen* 'saliva' from *been* 'liquid'

I was also told that *kafotik* 'rise' was an expression used in Tetun Suai (Luís Costa, p.c.) but not in Tetun Prasa. Both *kamedak* 'stain' and *kabuén* 'saliva' (original spelling) are listed in Costa's dictionary (2000). However, I was unable to find any examples of this kind of derivation in my corpus and my informants either didn't understand the words

---

[116] If a word starts with a vowel, the prefix changes to *kah-*.

presented in (42) or guessed their meaning. It can be thus assumed that although this affix is recognized by some speakers, it is not productive in Tetun Prasa anymore.

Two Portuguese derivational suffixes made their way into Tetun Prasa, too, although I have found them attached only to Portuguese bases so their actual productivity with native bases is disputable: *eis-* 'ex-' (from Port. *ex-*), used with both animate, as in (43), and inanimate, as in (44) bases, and *vise-* 'vice' (from Port. *vice-*), used with animate bases only, as in (45).

| (43) | *eis-kombatentes* | 'former fighters' | (from Port. *ex-combatentes*) |
| | *eis-prezidente* | 'former president' | (from Port. *ex-presidente*) |

| (44) | *eis-otél* | 'former hotel' | (from Port. *ex-hotel*) |

| (45) | *vise-ministru* | 'vice-minister' | (from. Port. *vice-ministro*) |
| | *vise-prezidente* | 'vice-president' | (from Port. *vice-presidente*) |

5.3.6   Derivation of adjectives

Derivation as a word-formation strategy is very marginal when it comes to adjectives. Although there are several derivational affixes mentioned in the literature on Tetun, my corpus shows that none of them seem to be productive in Tetun Prasa. Even in Tetun Fehan, which makes greater use of derivational morphology than Tetun Prasa, most of these affixes are unproductive (van Klinken 1999:85).

5.3.6.1 Suffix *-k*

The suffix *-k* is used to derive adjectives from nominal and intransitive verbal bases. These adjectives can be used both attributively and predicatively. According to Hull and Eccles (2005:147) this suffix is not productive in Tetun Prasa anymore, but it is still used in the literary language. Williams-van Klinken et al. (2002a) did not list this suffix in their grammar of Tetun Prasa and van Klinken (1999:85) also deemed it unproductive in Tetun Fehan. On the other hand, Albuquerque (2011:113) believed this

suffix is a part of the Tetun Prasa morphology and that it is used to derive adjectives from nouns. He gave examples as *tasik* 'marine' from *tasi* 'sea' and *motak* 'fluvial' from *mota* 'river'. However, neither of these words are found in the dictionary of Tetun Prasa by Williams-van Klinken (2015) and in the dictionary of Tetun by Costa (2000) which also contains a lot of Tetun Terik words. At the same time, I have not found examples of this derivational suffix in my corpus and my informants did not recognize the meaning of words listed in Albuquerque (2011).

5.3.6.2 Suffix *-n*

This suffix derives adjectives whose form is equivalent to that of the past participle of a transitive verb in English and Portuguese (Hull & Eccles 2005:148; van Klinken 1999:91):

(46)   *sonan* 'fried' from *sona* 'fry'

         tunun 'baked' from *tunu* 'bake'

Again, this suffix is believed not to be productive in Tetun Prasa anymore as I failed to find any examples in my corpus and my informants did not understand the derived forms mentioned in (46), opting for a basic form of a verb instead (*na'an sona* lit. 'meat fry'). However, Albuquerque (2011:113) mentioned it as one of the affixes still used to derive adjectives from verbs and Luís Costa (p.c.) also believes that these derived forms are still used.

The use of the derived adjectives is syntactically restricted. They cannot be modified by the irrealis aspect markers *atu* and the perfective aspect marker *tiha*. In this case, the basic word is used (Hull & Eccles 2005:149):

(47)   (all examples Hull & Eccles 2005:149-150)

         a.   *Asu-oan   ne'e   boot   tiha.*
               dog.young   this   big   PFV

               'The puppy is already grown up.'

b.   *Asu   ne'e   **kore<u>n</u>** tiha.
        dog   this   free   PFV

        'The dog got loose.'

c.   *Asu*   *ne'e*   ***kore*** *tiha.*
        dog   this   free   PFV

        'The dog got loose.'

This could have to do with the fact that *tiha* is rarely used with adjectives. However, it is grammatically possible to give the sentence a past meaning using the perfect aspect:

(48)   Asu   ne'e   ***kore<u>n</u>*** tiha ona.
         dog   this   free   PRF

    'The dog got loose.'

The same syntactic restrictions of this suffix apply to the suffix *-k*.

5.3.6.3 Circumfix *k(a)-(-k)*

The last of the affixes used to derive adjectives from other adjectives is the circumfix *k(a)-(-k)*. The meaning of the resultant word is slightly different, but is in the same semantic field, as in (49). This circumfix is not productive in Tetun Prasa but can be still found in literary language (Hull & Eccles 2005:147-148). In Tetun Fehan, it has much wider use besides just deriving adjectives (van Klinken 1999:86).

(49)   (Hull & Eccles 2005:148)

    a. *kabeer* 'smooth, tidy' from *beek* 'diluted' (the final consonant can be irregular)

    b. *kabeik* 'nauseous, nauseated' from *beik* 'stupid'

5.3.7 Derivation of adverbs: *-mente*

Hull and Eccles (2005:176) suggested there was an adverbial suffix of Portuguese origin *-mente*, but the analysis of my corpus revealed that it cannot be considered

productive in Tetun Prasa since it is only attached to Portuguese roots, or rather, it is only found in adverbs that have been borrowed from Portuguese, as shown below:

(50)   a. *presizamente* (from Port. *precisamente* 'exactly, precisely')

           b. *prinsipalmente* (from Port. *principalmente* 'especially')

5.3.8   Zero derivation

Zero derivation (change in word class without a change in form) is used in Tetun Prasa to derive verbs from nouns, as shown in (51), as well as to derive abstract nouns from verbs and adjectives. This word-formation process has become even more common with the adoption of more and more Portuguese (and other) loanwords (see section 8.6.4).

(51)   a.   *eskola* n. 'school' (from Port. *escola* 'school')

           b.   *eskola* v. 'to attend school'

There is, however, some inter-speaker variation. According to Williams-van Klinken et al. (2002a:18), conservative speakers prefer to use the prefix *ha-*, besides deriving causative verbs from intransitive verbs (see section 5.3.1), in order to derive transitive verbs from a nominal base, as in (52a), while younger speakers prefer zero derivation, as in (52b). Unfortunately, I was unable to find instances of this kind of zero derivation in my corpus.

(52)   *kail* 'fishing hook'

           a. *hakail*   v. 'to fish with a line'

           b. *kail*   v. 'to fish with a line'

This 'affix dropping' is also very common in colloquial/informal Indonesian, in which the intransitive prefixes *ber-* and *meN-* tend to be dropped in the most frequently used verbs (Ewing 2005:251), as in (53), where the verb *bertanya* 'ask' occurs without the prefix *ber-*. Unfortunately, there has been no work on the regularity of this process in Tetun Prasa yet.

(53)   Indonesian (Ewing 2005:251):

Kita   *bisa*   *langsung*   *tanya*   *ya?*
1PI   can   direct   ask   yes

'We can ask directly, can't we?'

5.3.9 Discussion

As we can see from the examples above, derivational morphology is much more limited in Tetun Prasa than it is in more conservative dialects of Tetun Terik and Tetun Fehan. With the influence of Portuguese and Indonesian, borrowings started to be preferred relative to affixation and many of my younger informants do not even recognize some of the derived forms. On the other hand, some derivational affixes are still widely used in Tetun Prasa, e.g. *ha-* to derive causative verbs or *-teen* to derive actor nouns. The only real borrowed affix is *-dór*, which has been attested already in the early works on Tetun and has since spread across all dialects of Tetun.

5.4 Compounding

According to Payne (2010:92-93), a compound "is a word that is formed from two or more different words" while fulfilling formal and semantic criteria, like the stress pattern of a single word; unusual word order; morphophonemic processes characteristic of single words; morphology specific to compounds; and more specific or entirely different meaning of a compound than the combined meanings of the separate words. This means that the compound has just one primary stress and neither of the members can be modified on their own.

Compounding as a word-formation process is especially dominant in isolating languages. Although new compounds keep emerging using Tetun Terik words, few people understand them. It is more frequent that these compounds are borrowed or calqued from Portuguese (Albuquerque 2011:119). But there is a considerably large group of nominal and verbal compounds that are of native origin.

Williams-van Klinken et al. (2002a:23-24) distinguished four types of nominal compounds in Tetun Prasa:

- possessor-head constructions that derive names for many body parts[117] (*liman-laran* 'palm', literally 'hand-inside'), animal products (*susubeen* 'milk', literally 'breast-juice') and plant parts (*ai-kulit* 'bark', literally 'plant-skin'), plus many other compounds;

- activity-place compounds referring to places where specific activities are performed, such as *servisu-fatin* 'workplace', literally 'work-place' or *vota-fatin* 'electoral station', literally 'vote-place';

- head-modifier compounds with noun, adjective or verb modifiers, such as *oan-feto* 'daughter', literally 'child-woman';

- and generic-specific compounds consisting of two nouns, with the first noun being more generic, often a nominal classifier, such as *ai-dila* 'papaya', literally 'tree-papaya'.

Some of these nominal compounds, namely the possessor-head type, the head-modifier type and the generic-specific type can be truncated. For the first member of the compound to be truncated, it has to be monosyllabic or disyllabic, for example *bairua* 'the day after tomorrow', literally 'day-two' (Avram 2008:442-443). However, all the truncated examples listed in Avram (2008) exist also in their full form (e.g. *bainrua* 'the day after tomorrow').

Hull and Eccles (2005:9-11) listed seven possible combinations of word classes to form nominal compounds, although some of them can also function as adjectives[118]. Albuquerque (2011:120) followed the classification described in Booij (2007:79) and divided compounds into:

- endocentric compounds with a head (e.g. *bibi-malae* 'sheep', literally 'goat-foreigner');

- exocentric compounds without a head (e.g. *fororai* 'python, wood snake', literally 'mountain-earth');

---

[117] All the compounds have a possessive suffix *–n* in the first element, unless they end in a consonant.
[118] These are: noun + noun + possessive suffix *–n*, noun + qualifying noun, noun + quantifiable noun, noun + adjective, noun + verb, verb + noun (complement/object), and verb + possessive noun.

- dvanda or copulative compounds[119] in which the two members are in a relation of coordination (e.g. *inan-aman* 'parents', literally 'mother-father');
- bahuvrihi compounds that denote a "person who is in possession of the entity mentioned by the compound" (Booij 2007:80) (e.g. *isin-manas* 'fever', literally 'body-hot').

This last category is often referred to as 'body-good expressions' and denotes "[c]haracter, emotions, health and physical attributes of people" (Williams-van Klinken et al. 2002a:57). Morphologically, these expressions are, in general, made up of a noun – a body part (most commonly *laran* 'inside', *isin* 'body', *ain* 'leg', *matan* 'eye' and *oin* 'face') and an adjective:

(54) Tetun Prasa: *laran-susar* 'grieving, very sad' from *laran* 'inside' and *susar* 'difficult'

País	*ida*	*riku,*	*ho*	*laran-susar,*	*sei*	*ho*	*kiak*
country	one	rich	with	grieving	still	with	poor

barak.
many

'A rich country, but grieving and still very poor.' (GovernuTL_2010)

The authors treat them separately from compounds for a number of reasons: although these expressions behave as compounds in some contexts, as in (55), in other contexts they act as regular predicative sentences, as in (56). But in both cases, they have the same meaning.

(55) (Avram 2008:440)

Ha'u   raan-nakali.
1S   blood-boiling

'I am angry.'

---

[119] Williams-van Klinken et al. (2002a:24) refer to them as 'semantically coordinate compounds'.

(56)   (Avram 2008:440)

Ha'u-nia   laran   nakali.
1S-POSS   inside   boiling

'I'm outraged' (lit. 'My inside is boiling')

Williams-van Klinken et al. (2002a:58) called this behaviour a 'split personality' and added that only a minority of these expressions are, based on morphosyntactic, phonological and semantic criteria, true compounds (for example *isin-rua* 'pregnant', lit. 'body-two').

When it comes to negation, there is a lot of interspeaker variation and the negator *la* can occur either before or within the body-good expression (Avram 2008:441). Since the negator *la* is used to negate verbs and adjectives, the preferred position is between the noun and the adjective:

(57)   (Avram 2008:441)

Ha'u   isin   la   di'ak
1S   body   NEG   good

'I'm not well.'

This looks like a predicative sentence, unless the negator and the adjective are analysed as one lexicalized item (Loch & Tschanz 2005:40):

(58)   *Ha'u   isin   ladi'ak.*
        1S   body   bad

'I'm unwell.'

Some, however, use the negator before the noun of the body-good expression and treat it as a complex adjective (Williams-van Klinken et al. 2002a:59) rather than a noun with a modifier:

(59)   (Avram 2008:441)

Ha'u   la   isin-di'ak.
1S   NEG   body-good

'I'm not well'

Avram (2008:441) also pointed out the adjectival status of the negated structure. This construction resembles the clausal negation in Portuguese or English and was deemed 'unusual' by native speakers of Tetun Prasa.

Another possibility is to employ the contrastive negation marker *la'ós* (Williams-van Klinken et al. 2002a:59), which can be used with any word class except for dynamic verbs:

(60)   (Williams-van Klinken et al. 2002a:59)

    Nia   *la'ós*   *isin-di'ak.*
    3S   NEG   body-good

    'He is not well.'

An interesting, and rather frequent phenomenon in Tetun Prasa is the creation of so-called loanblends that can be characterized as "hybrid borrowings which consist of partly borrowed material and partly native material", while the structural properties are also borrowed (Haspelmath 2009:39):

(61)   a. *uma-andár* 'multi-storey building' (from T. *uma* 'house' and Port. *andar* 'storey')

    b. *karau-vaka* 'ox, cow' (from T. *karau* 'buffalo' and Port. *vaca* 'cow')

They can be divided into several semantic categories, like plants, animals, kin terms, household, etc.:

(62)   a. *ai-farina* 'cassava' (from T. *ai* 'tree' and Port. *farinha* 'flour')

    b. *liis bombai* 'yellow onion' (from T. *liis* 'garlic, onion' and *Bombai*, probably a calque from Indo. *bawang bombai* 'yellow onion')

    c. *aman-sarani* 'godfather' (from T. *aman* 'father' and M. *sarani* 'Christian')

    d. *mina-azeite* 'olive oil' (from T. *mina* 'oil' and Port. *azeite* 'olive oil')

    e. *sapatu-talin* 'shoelace' (from Port. *sapato* 'shoe'[120] T. *talin* 'string, rope')

    f. *andár-leten* 'top floor' from (Port. *andar* 'storey' and T. *leten* 'top')

---

[120] Possibly borrowed via M. *sepatu* 'shoe' and later adjusted to the original Portuguese form *sapatu*.

Some borrowings which are compounds in Tetun Prasa are not necessarily compounds in the source language:

(63)　a. *armáriu-livro* 'bookshelf' (from Port. *armário de/para livros* lit. 'cupboard for books')

　　　b. *lian-ensinu* 'language of instruction' (from *lian* 'language' and Port. *ensino* 'education')

　　　c. *mentál-fraku* 'self-conscious' (from Port. *mental* 'mental' and Port. *fraco* 'weak')

Another category resembling compounds are serial verbs. These will be discussed in section 6.2.4.

5.4.1 Discussion

To summarize, we can divide Tetun Prasa compounds into three categories:

- native compounds consisting of two native words, which are the most numerous group and are often less semantically transparent, especially the so-called 'body-good expressions';
- Portuguese borrowings, which often are not compounds in the source language, are preferred to Tetun Terik borrowings;
- hybrid borrowings consisting of one native and one borrowed element.

It is expected that in the next few years, the second category of the Portuguese borrowings will grow the most.

5.5 Reduplication

Reduplication is a morphological process which "involves copying a set amount of phonological material from a base form (root or stem) and fusing it with that base to form a stem onto which other morphemes may then be added" (Velupillai 2012:101). There are two types of reduplication: a full base reduplication (simple repetition of the base) and a partial reduplication that can follow one of the four patterns: C-reduplication, Ca-reduplication, CV-reduplication or monosyllabic reduplication, and CV(C)CV-reduplication or disyllabic reduplication (Himmelmann 2005:121-123). Reduplication

can also be either simple (a mere repetition of a given material from the base) or complex (involving material from the base which undergoes slight alteration). It is also possible to find examples of discontinuous reduplication in which other material may be inserted between the base and the reduplicant (Velupillai 2012:101).

Reduplication is a very prominent and productive morphological process in many Austronesian languages. However, in Tetun Prasa, reduplication is not such a widespread morphological operation as it is in other languages. First of all, it is restricted to four word classes: nouns, adjectives, adverbs and numerals, and has a limited range of semantic properties that include habituality, intensification, distributivity, and manner adverbs. There are only a few examples of pluralization by reduplication in Tetun Prasa (see section 6.1.2.4) but this can be assigned to the influence of Malay/Indonesian.

Also, unlike other Austronesian languages, there is no reduplication of verbs, not even partial. It is found in Tetun Terik (*book* 'to mix' – *bobook* 'to mix many times/for a long time' (Costa 2000:58), but it is still very rare (Avram 2008:435). Reduplication of verbs is a productive strategy in Tetun Fehan, in which the reduplicated verb carries the meaning of "action done aimlessly, without reason, or heedless of the prescribed rules" (van Klinken 1999:93). A similar function of verb reduplication can be found in Indonesian, in which the reduplicated verb can carry the meaning of "action done in a casual or leisurely way" (Sneddon et al. 2010:23).

Both Esperança (2001:85) and Avram (2008:430) followed theories that consider reduplication a special form of affixation in which a partial or full copy of the base is attached to it as a prefix or a suffix (Booij 2007:35). One of the characteristics that reduplication shares with affixation is the fact that no stress shift occurs. In Tetun Prasa, there is only a prefixal reduplication. However, the final form of a reduplicated word depends on various phonological constraints and especially on the phonological characteristics of the root.

First of all, full reduplication is possible only with bases (native or borrowed) that are at least disyllabic. Bases that have more than two syllables, can be only fully reduplicated (Avram 2008:430-431). As for the partial reduplication, according to Avram (2008:432), it occurs with monosyllabic and disyllabic bases. However, Albuquerque

(2011:117) did not find any reduplicated monosyllabic bases in his corpus.[121] Further, both vowel- and consonant-initial bases can be reduplicated. In the case of disyllabic bases, the reduplicant is either monosyllabic or disyllabic.

The main semantic properties of reduplication in Tetun Prasa include habitual, intensification and distributivity, as will be shown in the following sections. I will also discuss the possibility of reduplicating borrowed bases.

5.5.1   Habituality

Reduplication of numerals and nouns referring to units of time derives temporal adverbs with the meaning of "each/every + time unit". This reduplication can be full or partial, depending on the phonological constraints mentioned above. Most of the adverbs preserve the original meaning of the base (e.g. *lororoloron* 'every day' from *loron* 'day'), but there are cases in which the reduplicated adverbs assume a more general meaning (e.g. *baibain* 'usual; usually' from *bain* 'day'), as shown below:

(64)   Tetun Prasa:

    […]   *sira*   *ko'alia*   *baibain*   *lororoloron*   *nia.*
           3P   speak   RDP.day   RDP.day   POSS

    '[…] they usually speak (Tetun) every day.'
    (0051TD_DLB_45_M_Uni_Kem)

This reduplication strategy can be also used with Portuguese bases, as in (65a), but is often substituted with a Portuguese expression '*kada* + time unit', as in (65b). There are also examples of *kada* + reduplication, as in (66).

---

[121] This might have to do with the unclear boundaries of syllabic segmentation. Originally, there were no diphthongs in Tetun and two neighbouring vowels were considered to belong to two different syllables. For example, Avram (2008:431) treated *rua* 'two' as a monosyllabic word, while Albuquerque treated it as a disyllabic. For more discussion on syllable structure of Tetun, see sections 7.5.1 and 7.5.3.2.

(65) a. *oraoras* 'from time to time' (from Port. *hora* 'hour')

  semana-semana 'weekly' (from Port. *semana* 'week')

  b. *kada Segunda* (from Port. *cada segunda* 'every Monday', instead of *Segunda-Segunda* 'every Monday')

(66) Tetun Terik:

Ne'ebé	*kada*	*tinan-tinan*	*sai*	*hanesan*	*ne'e*	[…]
REL	every	RDP.year	COP	like	this	

'Every year, (he) became like this.' (0046TV_JCV_48_M_Sec_TT)

5.5.2 Intensification

Another function of reduplication in Tetun Prasa is that of intensification. In this case, adverbs, as in (67), non-numeral quantifiers, as in (68), and adjectives can be reduplicated (Hull & Eccles 2005:152; Avram 2008:429; Albuquerque 2011:117).

(67) Tetun Prasa:

[…]	*ha'u*	*mós*	*kontente*	*tebetebes*	*tanba*	*haree*
	1S	also	happy	RDP.really	because	see

ha'u-nia	*kolega*	*universidade*	*sira*	[…]
1S-POSS	colleague	university	PL	

'[…] I'm also extremely happy because I got to see my university colleagues [...].' (0014TD_ACB_21_M_Uni_Mak)

(68) Tetun Prasa:

[…]	*ema*	*barak*	*ona,*	*ema*	*barbarak*	*ona*	*aprende*
	person	many	ANT	person	RDP.many	ANT	learn

iha	*lian*	*malaiu* […]
LOC	language	Malay

'Many people, very many people studied in Malay.'
(0053TD_JLS_44_M_Uni_Wai)

Sometimes, these adjectives can have two different meanings: that of intensification and at the same time that of a manner adverb, for example *momoos* 'very clean' or 'clearly' from *moos* 'clean' (Hull 1999:228; Costa 2000:248). Portuguese bases

have not been attested with this type of reduplication, nor does this intensification strategy exist in Indonesian.

5.5.3 Distributivity

Reduplicated numerals and numeral quantifiers carry a distributive meaning. In the case of the numeral *ida* 'one', the meaning of *ida-ida* is 'one at a time; individually; each, every (one)', as shown in (69), and that of *ida-idak* is 'each'. In the case of numerals higher than one, e.g. *rua* 'two', the meaning of *ru(a)-rua* can be 'in pairs/in a group of two; two at a time; two by two'. Reduplication of non-numeral quantifiers is also possible, as shown in (70). The same strategy exists in Indonesian, e.g. *satu-satu* 'one at a time', but has not been attested with Portuguese bases.

(69) Tetun Prasa:

Nia	*mós*	*husu*	*atu*	*ida-ida*	*halo*	*nia*	*aktividade*
3S	also	ask	IRR	RDP.one	do	3S	activity

hanesan	*baibain*
like	RDP.day

'He also asked that everyone does their activities as usual.' (STL_March_2016)

(70) Tetun Prasa:

Só	*ke*	*seidauk*	*hotu-hotu*	*iha*	*livru*	*hosi*
only	REL	not.yet	RDP.all	EXIST	book	from

dialetu	*ida-idak*	*maibé*	*parte*	*balu*	*iha*	*ona*	*liu*
dialect	RDP.one	but	part	some	EXIST	ANT	more

ne'e.
this

'However, there are not books in every dialect yet but there are already some.' (0051TD_DLB_45_M_Uni_Kem)

5.5.4   Manner adverbs

Reduplication of adjectives to form manner adverbs is, together with the habituality, the most productive mechanism of reduplication in Tetun Prasa. Examples include *di'diak* 'well' from *di'ak* 'good', or *lailais* 'quickly' from *lais* 'quick'. However, there is a lot of ambiguity and limited transparency when it comes to the meaning of the resultant word. Interestingly, *neineik* 'slow; slowly', although reduplicated, also acts as an adjective with its root's (*neik*) meaning unclear, as in (71). The double reduplicated form *neineik-neineik* 'gradually' gained a different meaning, as in (72).

(71)   Tetun Prasa:

Ha'u	*sei*	*estuda*	*neineik*	*atu*	*nune'e*	*ha'u*	*bele*
1S	FUT	study	slowly	IRR	so	1S	can

aprende.
learn.

'I will keep studying, slowly, so I can learn (foreign languages).'
(0014TD_LEG_20_F_Sec_Mak)

(72)   Tetun Prasa:

No	*neineik-neineik,*	*neineik-neineik*	*ita*	*bele*
and	RDP.slowly	RDP.slowly	1PI	can

muda	*ita-nia*	*moris*.
change	1PI-POSS	life

'And gradually, gradually we can change our lives.'
(0014TD_JBP_26_M_Uni_Mak)

Similar ambiguity exists with *loloos*. While the adjective *loos* generally means 'right, true, correct', as in (73), its reduplicated form *loloos* can act as a synonym of the adjective, as in (74), or act as an adverb 'truly, properly, exactly', as in (75).

178

(73)   Tetun Prasa:

No   ha'u   bele   dehan   katak, ita   iha   dalan   loos,   ita
and   1S   can   say   COMP   1PI   LOC   road   correct 1PI

iha   ona   kapasidade   ruma.
have   ANT   ability   some

'And I can tell that we are on the right path, that we had the abilities.'
(GovernuTL_2010)

(74)   Tetun Prasa:

Portavós   la   hateten   númeru   loloos hosi
spokesperson   NEG   know   number   correct of

pasajeiru   sira   ne'ebé maka   iha   "ferry"   laran   […]
passenger   PL   REL   FOC   LOC   ferry   inside

'The spokesperson does not know the real number of people who were
on the ferry.' (Sapo_March_2016)

(75)   Tetun Prasa:

[…]   ne'e   dunik presiza   fronteira   para   hatene
        this   indeed need   border   to   know

loloos   rikusoin   ne'e   Austrália   nia   ka
RDP.correct   wealth   this   Australia   POSS   or

Timor-Leste
Timor-Leste

'[We] need to know exactly whether the wealth belongs to Australia or
Timor-Leste.' (STL_March_2016)

Just like with other types of reduplication, except for habituality, Portuguese bases
have not been attested.

5.5.5 Discussion

Some linguists used the fact that reduplication is a productive derivational strategy in Tetun Prasa to support their claims about Tetun Prasa being a creole. According to Bakker (1994:43), reduplication is very common, but not universal, in creoles, but it is almost non-existent in pidgins (even when the contributing languages are rich in reduplication), except for extended pidgins. So Avram (2008:437-439) looked at the reduplication in pidgins and creoles, especially in those with Austronesian substrate and/or Portuguese superstrate and, although recognizing there has been some dispute about the status of Tetun, suggested that Tetun Prasa might have a double status of a creole (for a minority of speakers) and an extended pidgin (for the majority of speakers), an assertion I disagree with (see section 3.1.1.2). Another feature in common are the phonological characteristics of reduplication and of reduplicants: one can find monosyllabic and disyllabic reduplicants in both types of languages, as well as full and partial reduplication is possible in either of them.

Avram (2008) further affirmed that all the syntactic categories that serve for reduplication in Tetun Prasa are also used in other pidgins and creoles, with the exception of the verbal reduplication, which has not been attested in Tetun Prasa. I compared reduplication strategies in Tetun Prasa and Malacca Creole Portuguese and indeed they match. However, I did the same exercise for Tetun Fehan and the reduplication strategies in this variety are exactly the same as in Malacca Creole Portuguese, as is illustrated by Table 12.

We can thus conclude that Bakker and Avram are right to claim that reduplication is a common strategy in creoles, just like in Malacca Creole Portuguese. However, these characteristics are found throughout many Austronesian languages that are, by no means, creoles or pidgins. Reduplication is found in all Tetun varieties but also in Indonesian, which uses a full range of reduplication, as shown in Table 12.

	Tetun Prasa	Tetun Fehan (van Klinken 1999)	Indonesian (Sneddon et al. 2010)	Malacca CP (Alan Baxter, p.c.)
habituality (e.g. *semana* 'week' → *semana-semana* 'weekly')	yes	**yes** (e.g. *loron* 'day' → *loro-loron* 'daily')	**no**	**yes** (*niora-niora* 'often')
intensification (e.g. *moos* 'clean' → *momoos* 'very clean')	yes	**yes, but limited** (*mesan* 'alone' → *mesa-mesan* 'all alone, really alone')	**no**	**yes, but limited** (e.g. *limpu-limpu* 'very clean')
distributivity (e.g. *ida* 'one' → *ida-ida* 'one at a time')	yes	**yes** (e.g. *lima* 'five' → *lima-lima* 'five at a time')	**yes** (e.g. *satu* 'one' → *satu-satu* 'one at a time')	**yes** (e.g. *uma* 'one' → *uma-uma* 'one at a time')
manner adverbs (e.g. *di'ak* 'good' → *didi'ak* 'well')	yes	**yes** (e.g. *di'ak* → *di'a-di'ak* 'well')	**yes** (e.g. *cepat* 'quick' → *cepat-cepat* 'quickly')	**yes** (e.g. *bagar* 'slow' → *bagabagar* 'slowly')
RDP.verbs	no	**yes** – 'do heedlessly' (e.g. *fasi* 'wash' → *fasi-fasi* 'wash heedlessly')	**yes** (e.g. *berjalan* 'walk' → *berjalan-jalan* 'walk about')	**yes** – repetition and duration (*rema-remá* 'row and row')
pluralization	limited/ fossilized	**yes** – repetition of a head noun or reduplication of an adjective (e.g. *hudi* 'banana' → *hudi hudi* '(any) types of bananas')	**yes** (e.g. *rumah* 'house' → *rumah-rumah* 'houses')	**yes, but limited** (e.g. *krensa-krensa* → children)

Table 12 Reduplication strategies in Tetun Prasa, Tetun Fehan, Indonesian and Malacca Creole Portuguese

The fact that reduplication is a widespread phenomenon in Austronesian languages and is not confined to contact languages makes Avram's hypothesis of Tetun Prasa being a creole and/or an extended pidgin less convincing.

5.6 Conclusions

In this chapter, I tried to document the influence of Portuguese and Malay/Indonesian on the inflectional and derivation morphology of Tetun Prasa. Based on the limited occurrences of derivational affixes in my corpus it can be concluded that, besides a few exceptions like the agentive suffixes, Tetun Prasa does not make much use of derivational morphology. And as we will see in Chapter 8, the most common strategy of vocabulary expansion in Tetun Prasa is lexical borrowing. Inflectional morphology is also very limited and the subject marking is only found in more conservative varieties of Tetun. But even in Tetun Terik spoken in Suai and Viqueque, subject marking is often dropped. The younger generation of speakers is aware of it, but most likely does not use it on a daily basis, although a more detailed study is needed. Compounding and reduplication continue to be productive, although the use of borrowed elements is limited, except for the class of so-called loanblends.

The lack of richer inflectional and derivational morphology in Tetun Prasa made various authors compare this variety to creoles and pidgins. However, as detailed in this chapter, there are many Austronesian languages that are rather analytical and isolating so the sole fact that Tetun Prasa has limited inflectional and derivational morphology is not enough to claim it is a creole.

6. SYNTAX

In this chapter, I will discuss the syntax of noun phrases and verb phrases, as well as complex sentences. As special emphasis is given to language contact, my aim is not to offer a complex description of Tetun Prasa grammar but rather focus only on syntactic features in this variety that have been in some way influenced by Portuguese, Indonesian or other languages. To understand the full extent of this influence, I will draw comparisons with other varieties of Tetun (Tetun Terik and Tetun Fehan) and several Portuguese- and Malay-based creoles spoken in the Southeast, East and South Asia.

6.1 The noun phrase

On the level of a noun phrase, the most striking differences between Tetun Prasa and other varieties of Tetun can be found in terms of pluralization strategies (section 6.1.2), possession (section 6.1.3), and gender marking and agreement (section 6.1.4). Tetun Prasa also differs from other varieties in the number of classifiers it employs (section 6.1.5). When it comes to numerals and quantifiers, Portuguese and Indonesian borrowings are often preferred to the native ones (section 6.1.6).

6.1.1   The unmarked noun

Most nouns in Tetun Prasa are unmarked. As I presented in Chapter 5, there is very little inflectional morphology in Tetun Prasa, and nouns marked for number (section 6.1.2) and/or gender (section 6.1.4) are mostly Portuguese borrowings. Some Tetun nouns are also marked for inalienability, which means they are inherently possessed (section 6.1.3).

6.1.2 Plural forms

There are five different mechanisms of forming the plural in Tetun. These include:

- the unmarked plural (6.1.2.1);
- the postposed plural marker *sira* 'they' (6.1.2.2);
- the Portuguese plural marker *-s/-es* (6.1.2.3);
- reduplication (as found in Malay/Indonesian) (6.1.2.4);
- the Tetun Terik plural suffix *-r/-n* (6.1.2.5).

The first linguists to observe different strategies of forming the plural in Tetun Prasa were Aparício da Silva (1889:4) (see section 4.2.3.3) and Raphael das Dores (1907:10) (see section 4.2.3.4).

Silva (1889:4) noticed that there was no fixed rule to form plurals. In some regions, he observed the suffix *-r* or *-n* being added to nouns ending in a vowel, e.g. *sala* ''sin' – *salan* 'sins', *ema* 'person' – *emar* 'people'. Other words that ended with an *-n* would change the last consonant to an *-r* in plural, e.g. *oin* 'face' – *oir* 'faces'. Another strategy was to reduplicate the same word, either fully (*mane* 'man' – *mane mane* 'men') or partially (*fuan* 'heart' – *fufuan* 'hearts'), or to postpose a plural particle *sira* (*mane sira* 'men').

Dores (1907:10), basing his work on Silva (1889), noticed the very same plural strategies: the reduplication characteristic of Malay language, either full (e.g. *feto – feto-feto* 'woman – women'), or partial (e.g. *fúan – fufúan* 'heart – hearts'), the plural marker *sira* (e.g. *mâne – mâne sira* 'man – men') and the plural suffixes *-r* and *-n* used in some regions (e.g. *ria – rían* 'cousin – cousins', *ulún – ulúr* 'chief – chiefs').

To understand which pluralization mechanisms are used in which contexts and varieties, I will address each one of them separately.

6.1.2.1 Unmarked plural

In modern Tetun Prasa, no plural markers are used with indefinite nouns since their plurality is usually clear from the context, as shown in (76) (Hull 1993:31; Hull &

Eccles 2005:13; Costa 2015:53). To differentiate indefinite nouns expressing singularity, these are often marked with *ida* 'one; a', as shown in (77).

(76) Tetun Prasa:

Atu	*kontrola*	*sidadaun*	*estranjeiru*	*ne'ebé*	*tama*
in.order.to	control	citizen	foreign	REL	enter

to'o	*aldeia*	*no*	*suku*	*iha*	*territóriu*	*Timor-Leste* […]
reach	village	and	suku	LOC	territory	Timor-Leste

In order to control (the number of) foreign citizens that enter villages and sukus in the Timor-Leste territory [...] (STL_March_2016)

(77) Tetun Prasa:

[…]	*iha*	*ha'u-nia*	*subdistritu*	*iha*	*suku*	*ida*	*sei*
	LOC	1S-POSS	subdistrict	EXIST	suku	one	still

ko'alia	*ne'e*	*ne'e.*
speak	this	this

'[...] in my subdistrict there is a suku that still speak this (language).' (0055TD_MAS_45_M_Uni_Mid)

6.1.2.2 *sira* 'they'

In most of the cases, definite nouns take a plural particle *sira*, as shown in (78), which is identical to the third person plural pronoun meaning 'they'. According to Williams-van Klinken et al. (2002a:30) this plural marker is not restricted to human beings although it can be omitted in some contexts in rural Tetun and Tetun Prasa when referring to animals and things (Hull 1993:31). This mechanism of forming plurals is common in Austronesian languages (Himmelmann 2005:174) and it is found in various pidgins and creoles around the world like Tok Pisin, Portuguese-based creoles in Gulf of Guinea or creoles spoken in the Caribbean and northern South America (Maurer et. al. 2013). A similar strategy with a postposed particle is used also in other native languages of Timor, for example Baikeno *sin* 'they', as in (79), however, the meaning of these particles (or suffixes, as in Fataluku) is not always equivalent to the personal pronoun

'they'. In Mambae, the plural particle *sêr*, as in (80), is used instead of the third person plural pronoun *rom* 'they'.

(78)   Tetun Prasa:

ema	*estudante*	*tékniku*	*agrikultura*	*sira*
person	student	technical	agriculture	PL

'technical students of agriculture' (0005TD_JBP_26_M_Uni_Mak)

(79)   Baikeno (Hull 2003b:11):

fafi 'pig' – *fafi sin* 'pigs'
ume 'house' – *ume sin* 'houses'

(80)   Mambae (Hull 2003a:11):

mu 'banana' – *mu sêr* 'bananas'
hin 'woman' – *hin sêr* 'women'

6.1.2.3 Portuguese plural marker *-(e)s*

Tetun Prasa contains a large number of Portuguese loanwords, especially abstract nouns and technical terms. There are some nouns that were borrowed into Tetun Prasa already with their plural markers, e.g. *Estadus Unidus* 'United States' (from Port. '*Estados Unidos*') but most of them can take either the native plural marker *sira* or the Portuguese suffix. This depends on the literacy level of the speaker and the situation (Albuquerque 2011:103). The Portuguese plural marker with Portuguese loanwords, as illustrated by (81) and (82), is usually used by educated people with some knowledge of Portuguese as well as by the media (newspapers and television) in which these borrowings often make up over 60% of the words types (see section 8.6.7). However, speakers with limited knowledge of Portuguese often combine the Portuguese plural with the Tetun plural marker (83). This way of marking plurals is disapproved of by many and is not found in the literary language (Hull & Eccles 2005:14-15; Williams-van Klinken et al. 2002a:31). Although the Portuguese suffix is still very productive, there are some words that exist with the plural marker in a fossilized form conveying singular meaning, for example *oras* (from Port. *horas*) 'hour', *uvas* (from Port. *uvas*) 'grape'.

(81)   Tetun Prasa:

[…]   *ita-nia*   *atividades*   *loroloron*   […]
       2S.HON-POSS   activity.PL   RDP-day

'[…] your everyday activities […]' (0014TD_JBP_26_M_Uni_Mak)

(82)   Tetun Prasa:

[…]   *Austrália*   *mós*   *simu*   *refujiadus*   *Timor nian*   […]
       Australia   also   accept   refugee.PL   Timor POSS

'[…] Australia also accepted Timorese refugees […]'
(0008TD_TAB_55_M_PreS_Mam)

(83)   Tetun Prasa:

[…]   *ami*   *hakbesik-an*   *ba*   *ema doadores*   *sira*   […]
       1PE   approach   to   person donor.PL   PL

'[…] we approached the donors […].' (0006TD_JBP_26_M_Uni_Mak)

In my corpus, I have found only few instances of a Portuguese plural marker in Tetun Terik:

(84)   Tetun Terik from Suai:

Ha'u   *alunus*   *iha*   *tolunulu-resin hitu.*
1S   pupil.PL   EXIST   thirty-excess   seven

'I have thirty-seven pupils.' (0017TS_MCA_29_F_Sec_TT)

In five cases, the plural nouns were followed by a plural particle *sira*:

(85)   Tetun Terik from Viqueque:

[...]   *ha'u*   *fó*   *fiar*   *ba*   *ha'u-nia*   *primus*   *sira*
         1S   give   trust   to   1S-POSS   cousin.PL   PL

ne'ebé la'ós   *servisu*   *nu'udar*   *funsionáriu*   [...]
REL   NEG   work   as   civil.servant

'[...] I have faith in my cousins who do not work as civil servants [...]'
(0046TV_JCV_48_M_Sec_TT)

187

It occurs only with Portuguese borrowings and mostly with animate nouns, as in (86a), with a few examples of inanimate nouns, as in (86b). Except for *alunus* 'pupils' (11 occurrences in total in my corpus), other words occurred only once.

(86)   a.   *alunus* 'pupils' (11 occurrences in my corpus)

                profesores 'professors' (1 occurrence)

                primus 'cousins' (1 occurrence)

      b.   *berlindus* 'marbles' (1 occurrence)

                livrus 'books' (1 occurrence)

                dúvidas 'doubts' (1 occurrence)

                ekipamentus 'equipment' (1 occurrence)

6.1.2.4 Reduplication

The reduplication of nouns in Tetun can have various functions, one of them being to mark plurality. This mechanism of forming plurals is very common in western Austronesian languages like Malay and Indonesian where the "noun is usually reduplicated unless it is clear from the context whether one or more than one is referred to and then only if this is important to what the speaker wishes to convey" (Sneddon et al. 2010:20). In Tetun Prasa, this process is no longer productive, although some reduplicated pronouns exist in a fossilized form, e.g. *seluk – seseluk* ('other – others'). However, Costa (2015:53) noticed that in Tetun spoken in Dili and as a vehicular language, pluralization through partial reduplication can be found in limited number of cases: *ai-fuan* 'fruit' – *ai-fufuan* 'fruits'.

It is, however, not surprising that Tetun Prasa adopted this feature in the past; pluralization of nouns via reduplication can be also found in Malayo-Portuguese, particularly in Batavia Creole Portuguese (Maurer 2011:22-23), Malacca Creole Portuguese (Baxter 1988:102), in Singapore Bazaar Malay and Ambon Malay where Malay is one of the substrate languages (Holm 1989:290-291). It is also used in Macau Creole Portuguese (Batalha 1959:186). We can assume that the use of reduplication to mark plural nouns was used more widely in the past when the influence of Malay and

Indonesian on Tetun was greater. This is supported by the fact that in earlier texts in Tetun, like in the one from Sá (1961), there are examples of pluralizing reduplication:

(87)   *reinu-reinu*   lit. 'kingdom-kingdom' - 'many kingdoms'

       kuda mutin-mutin   lit. 'horse white-white' - 'all the white horses'[122]

However, when Esperança (2001:70) elicited these examples with his Tetun Prasa informants, they did not recognize it as a plural mechanism in Tetun Prasa anymore. His informant confirmed that the correct way to express 'all + the + noun' would be 'noun + *sira + hotu*', e.g. *reinu sira hotu* or *kuda mutin sira hotu*.

The same also goes for Tetun Terik, in which this strategy does not occur either. However, we can find examples of plural reduplication in Tetun Fehan (van Klinken 1999:95), which, again, is most probably caused by the influence of Malay/Indonesian. Interestingly, reduplication as a plural strategy is also possible in other languages spoken in Timor. Hull (2003a:11) asserted that generic plural can be expressed by reduplication in Mambae: *man-man nor hin-hin* 'men and women'. Another exception, where reduplicative plural still exists, alongside a periphrastic form with *sia* '3PL', are the languages spoken on Atauro (Hull 2001a:110).

6.1.2.5 Plural in Tetun Terik

Apart from the plural particle *sira*, there are two native suffixes used with Tetun Terik nouns and pronouns to mark plurality: *-r* and *-n* (e.g. *ema – emar* 'person - people', *belu – belun* 'friend - friends'). This type of plural marking is not productive in Tetun Prasa anymore, although one can find it in a fossilized form with either plural (e.g. *emar* 'people') or singular meaning (e.g. *belun* 'friend'). It the latter case, it is often used with a human classifier *na'in* or a plural particle *sira* to express the plural meaning 'friends' (88) (Hull 1993:31-32; Hull & Eccles 2005:14):

---

[122] The very same strategy can be found in the Indonesian language. According to Sneddon et al. (2010:22), "[r]eduplication of an adjective usually occurs when the noun it describes is plural; reduplication indicates that the characteristic indicated by the adjective applies to all the objects."

(88)   Tetun Prasa:

Maibé mós   hanesan   belun sira.
but   also   like   friend PL

'But they are also friends.' (Sapo_March_2016)

According to Costa (2015:53), most of the Tetun Prasa speakers are unaware of these plural forms although they can be found in song lyrics. It is also important to bear in mind that these suffixes are not found in all subvarieties of Tetun Terik. Van Klinken (1999:134), who studied the Tetun Fehan variety, argued that although these suffixes exist in this subvariety, they have different functions. The suffix *-n* is a genitive marker and *emar* is an alternative form of *ema* which can also mean 'person'.

In my Tetun Terik corpus, I have found *emar* only used as a synonym to *ema*, as shown in (89). The data show that this plural suffix is very little, if at all, productive in both Tetun Terik areas that I studied. I was unable to document cases of this kind of plural marking, but noticed that *sira*, or alternatively *sia*, is widely used throughout the whole area.

(89)   Tetun Terik:

Wainhira   to'o   mai   iha-ne'e   ida   na'i-lulik   ne'e
when   arrive come here   one   priest   this

emar ida   prepara   ai-han, fase nia   ropa [...]
person one   prepare   food   wash 3S   clothes

'When the priest arrived here, a person (Gáspar) prepared food for him, washed his clothes [...]' (0050TV_JCV_48_M_Sec_TT)

6.1.3 Possession

In this section, I will discuss possessive constructions in Tetun Prasa, Tetun Terik and Tetun Fehan. I will use terms *possessor*, *possessum* and *possessive marker* for the constituents of a possessive noun phrase. For pronominal possessors, I will use *possessive determiners* to describe pronouns which occur in determiner position and *possessive*

pronouns to describe pronouns which occur in predicative structures. Since the exact status of possessive markers (possessive-marked pronouns, affixes, or function words) is not clear because they differ in form and distribution, I will refer to them as *(possessive) markers*.

Possessive constructions carry the semantic feature of ownership, although the relationship between the possessor and the possessum is not always that of possession. This relationship can be formally expressed in various ways, for example by genitive case markers (e.g. German) or genitive clitics (e.g. English), by a specific word order (e.g. Indonesian) or by possessive markers (e.g. Tetun Prasa).

Austronesian languages are often divided into two typological categories: *symmetrical voice languages* and *preposed possessor languages*[123]. The latter one comprises languages in which the possessor precedes the possessum, although this is not the only possible order but rather the most common or unmarked (Himmelmann 2005:112-113). Capell (1944c:31) delimitates this group geographically as languages spoken "in eastern Indonesia, east of a line known as Brandes' Line, which runs west of Moluccas, but east of the Philippines and north-eastern Celebes and east of Sumbawa". All Austronesian languages of Timor-Leste have a preposed possessor.

One of the main characteristics of these preposed possessor languages is the distinction between alienable and inalienable possession. Semantically, alienable possession can be terminated, which means that the possession is transferrable. On the other hand, inalienable possession is not transferrable and cannot be terminated, as is the case of body parts and kinship relations (Payne 2010:105). Syntactically, alienable possession usually takes the form of *possessor + possessive marker/ligature + possessum* and inalienable possession is realized as *possessor + possessum + possessive enclitic* (possessive marker in the case of Tetun Prasa) (Himmelmann 2005:164).

---

[123] According to Himmelmann (2005:113-114), these two categories also correlate with other characteristics, i.e. the symmetrical voice languages (60% of the non-Oceanic/western Austronesian languages) usually postpose the possessor and conversely the preposed possessor languages (25%) generally do not show any voice alteration or these alterations are asymmetrical. Languages that do not fit in either of these categories are sometimes referred to as transitional languages (15%). There is also at least one language (the Formosan language Pazeh) that has both a symmetrical voice system and a preposed possessor.

6.1.3.1 Possessive constructions in Tetun Prasa

Historically, there was no distinction between alienable and inalienable possession in Old Timorese. It was only with the Old Ambonese influence during the Middle Timorese period that Timorese languages adopted this distinction (Hull 2001a:115). Tetun Prasa, like most of the Timorese languages nowadays, makes a distinction between alienable and inalienable possession, although it is not as clear-cut as in other Tetun dialects. There are two grammatically distinct orders of possessive constructions and there are two factors affecting them: the alienability/inalienability of the possessum and the lexical nature of the possessor (pronominal versus nominal).

In Tetun Prasa, there are two groups of items (body parts and kinship terms)[124] that are inherently possessed and have a final *-n*, for example *alin* 'younger sibling' or *ulun* 'head'[125]. It is believed that this final *-n* is a fossilized form of the Tetun Fehan genitive clitic *-n*[126] that is required in Tetun Fehan with all vowel-final possessors, body parts and kinship terms when they are possessed (van Klinken 1999:146). Historically, Old Timorese had possessive suffixes that could be applied to any noun. However, only one suffix survived in Tetun Prasa, the 3S (and 3P) suffix *-n* which was also extended to certain alienable nouns (Hull 2001a:115-118).

The two possessive strategies in Tetun Prasa are as follows:

1. *possessor + POSS (-(ni)nia) + possessum*
2. *possessum + possessor + POSS (-(ni)nian)*

---

[124] There are two other words that can take on the genitive clitic *-n*: *rai* 'land/country' and *lia* 'language'. When *rai* is possessed, it can mean either 'land' or 'country' but only in the latter case it takes on the genitive clitic, i.e. *ha'u-nia rain* 'my country' compared to *ha'u-nia rai* 'my land'. The same can be applied to *lia* 'language'. When it is possessed, it must be used with a final *-n*, e.g. *ha'u-nia lian* 'my language'; in case it denotes a particular language, e.g. *lia Tetun* 'Tetun language', it is used without the genitive clitic. This distinction is especially made among Tetun Terik speakers (Williams-van Klinken et al. 2002a:36).

[125] According to Williams-van Klinken et al. (2002a:36) the final *-n* is used only with native kinship terms except of two Portuguese borrowings – *tia* 'aunt' and *tiu* 'uncle' – that take the final *-n* when they are possessed. Other borrowings, like *mana* 'older sister', do not take any genitive clitic. However, I have found another example of a Portuguese borrowing in my Tetun Terik corpus which comes with a possessive suffix: *avón* 'grandfather'. The fact that Tetun Terik applies the same suffix to Portuguese loanwords might suggest that the speakers do not perceive them as borrowings anymore.

[126] Historically, the final *-n* was the third person singular possessive suffix *n'a* in Original Indonesian, the common origin of all Austronesian languages in Timor, according to Capell (1944c:32).

Tetun Prasa has one possessive marker that has the form of the 3S personal pronoun. In this section, I will gloss it as 3S.POSS.DET. The marker can vary in its form depending on the order of other constituents in the possessive NP. In general, *nia/ninia* is used in the preposed possessor order, as illustrated by (90), and *nian/ninian* is used in the postposed possessor order, as illustrated by (91).[127] Both of them follow the possessor.

(90) *Inspetór*   *nia*   *hanoin*
      inspector   3S.POSS.DET   opinion

      'inspector's opinion' (GovernuTL_2010)

(91) *baliza marítima*   *tasi*   *nian*
      border maritime   sea   3S.POSS.DET

      'the sea's maritime border' (0008TD_TAB_55_M_PreS_Mam)

The first strategy with a preposed possessor marks a possessive relationship and is used with inalienable and alienable nouns. According to Williams-van Klinken et al. (2002a:34), the category of inalienable possessums includes four semantic relationships:

a. *part-whole relationship*: although there are many examples of this relationship without any possessive marker that are compounds (see section 5.4, e.g. *ai-kulit* 'tree bark', literally 'plant.skin'), some, especially entities including body parts, are used with a possessive marker:

(92) *La*   *loos*   *mak*   *arguidu*   *la*   *kaer*   *lezada*
      NEG   right   FOC   defendant   NEG   grasp   injured

      nia   *ulun*   *riba*   *ba*   *parede*   *maibé* *só*
      3S.POSS.DET head   fling   to   wall   but   only

      tuku   *de'it.*
      punch   only

      'It's not true that the defendant did not grasp the injured female's head and smashed it against the wall but only punched her.'
      (STL_March_2016)

---

[127] According to Williams-van Klinken et al. (2002a:33), the longer forms *ninia* and *ninian* are restricted to formal register.

b. *spatial relationship*: if the possessor is a human or is modified, *nia* is used; if the noun is not modified and does not refer to a human, the possessive marker can be either used or omitted:

(93)   *Saida*   *mak*   *atu*   *halo*   *ho*   *petróleu,*   *ho*   *Austrália*
           what   FOC   IRR   make   with   oil   with   Australia

           no   *ho*   *minerál*   *sira*   *ne'ebé*   *mak*   *iha*
           and   with   mineral   PL   REL   FOC   EXIST

           país   *nia*   *laran.*
           country   3S.POSS.DET   inside

           'What to do with the oil, Australia and the minerals that are found in the country.' (GovernuTL_2010)

c. *kinship terms*: preposed possessors are used with all kinship terms (e.g., mother, father, parents, etc.), as shown in example, the only exception being *família* 'family' which allows both orders:

(94)   *lezada*   *nia*   *inan-aman*
           injured   3S.POSS.DET   mother-father

           'the injured person's parents' (STL_July_2013)

d. *naran* 'name': with this word, preposed possessor order is used in all case:

(95)   *Ida*   *ha'u*   *hatene*   *nia*   *naran*   *maibé*   *ida*   *ha'u*
           one   1S   know   3S.POSS.DET   name   but   one   1S

           la   *hatene.*
           NEG   know

           'I know the name of one of them, but not of the other.'
           (0006TD_BE_30_M_Uni_TP_Mak)

The preposed possessor order is also used, although not always required, with alienable nouns, namely with deverbal nouns when they are possessed, as in (96a), with physical possession, as in (96b), or with a non-pronominal possessor that refers to a particular individual, as in (96c) (Williams-van Klinken et al. 2002a:34-35).

(96) a. *Ita-nia*   *hanoin*
          2S.HON-3S.POSS.DET   thought

          'your thought'

      b. *ó-nia*   *tolun*
          2S-3S.POSS.DET   egg

          'your eggs'

      c. *ema estudante ne'e nia*   *knaar*
          person student   this 3S.POSS.DET   duty

          'the student's duty'

On the other hand, the postposed possessor order marks an associative relationship in which the relationship between the possessor and the possessum is more general. The relationship is not strictly that of a possession and the possessor often does not refer to a particular entity (Williams-van Klinken et al. 2002a:35). This order is also more common in formal registers and literary Tetun when the possessor is a non-human or an abstract thing, as shown in (97), and it is always preferred in titles and similar expressions, as shown in (98) (Hull & Eccles 2005:20).

(97) […] *ami-nia*   *grupu iha sektór agrikultura*
           1PE-3S.POSS.DET   group LOC sector agriculture

      nian
      3S.POSS.DET

      '[…] our group in the agricultural sector.'
      (0006TD_JBP_26_M_Uni_Mak)

(98) […] *iha   universidade UNPAZ   nian* […]
           LOC   university   UNPAZ   3S.POSS.DET

      '[...] at the UNPAZ university [...]' (0014TD_ACB_21_M_Uni_Mak)

However, when the referent is unique, it is possible to omit *nia*, e.g. *prezidente CNRT* 'the president of the CNRT'. The omission is also very common in informal speech and texts (Williams-van Klinken et al. 2002a:35).

In the case of pronominal possessors, the form of the possessive determiners is the same as that of personal pronouns followed by a possessive marker *nia* (Hull & Eccles

2005:20), as shown in Table 13. The third person singular possessive form is shortened from *nia-nia* to *ninia* or just *nia*.

Singular	Plural
1. *ha'u-nia uma* 'my house'	1. *ami-nia uma* 'our (exclusive) house' *ita-nia uma* 'our (inclusive) house'
2. *ó-nia uma* 'your house' *ita-nia uma* 'your (honorable) house'	2. *imi-nia uma* 'your house'
3. *nia/ninia uma* 'his/her/its house'	3. *sira-nia uma* 'their house'

Table 13 Pronominal possession in Tetun Prasa

In general, the longer form *ninia* should not be used with pronominal possessors and can be considered ungrammatical by some speakers. However, in my oral and written corpus of Tetun Prasa, the longer form *ninia*, and even *ninian*, is used quite frequently with pronominal possessors:

(99)   […]   *ita*   *ninia*   *programa*   *ohin*   *loron*
              1PI   3S-3S.POSS.DET   plan   today   day

       nian   […]
       3S.POSS.DET

       '[…] our program today […]' (0006TD_JBP_26_M_Uni_Mak)

In Tetun Terik, three speakers used the longer form *ninia* in preposed possessors and *ninian* in postposed possessors, either nominal or pronominal, on more than one occasion. On the other hand, we often see *ninia* in a postposed position where *ninian* should be used instead. I believe this might be due to the fact that speakers of Tetun Terik often omit the final *-n* (and other sounds) in many words (Hull 1993:249), for example *han* 'to eat' is often pronounced as [há].

Based on the frequency of its occurrence (see Table 14), I believe that *ninia*, when used as a possessive marker after personal pronouns or nouns, should no longer be considered ungrammatical but rather a variant of *nia*. As you can see, this variant is especially common in Tetun Prasa speakers (6 out of 13) with secondary or university education. It is also rather common in two dailies, namely Sapo and Suara Timor Lorosae. Three out of seven Tetun Terik speakers from Viqueque also used this construction. No occurrences of this use were found in Tetun Terik speakers from Suai.

Source/speaker	Total number of occurrences of *ninia*	Number and percentage of occurrences of *ninia* as a possessive marker used after nouns or personal pronouns	
Governo T-L	7	0	0%
Sapo	28	18	64%
Suara Timor Lorosae	63	19	30%
0005TD_JBP_26_M_Uni_Mak	20	17	85%
0006TD_BE_30_M_Uni_TP_Mak	6	2	33%
0008TD_TAB_55_M_PreS_Mam	2	1	50%
0017TS_MN_44_F_Sec_TT	1	0	0%
0026TS_ATB_41_M_Sec_TT	1	0	0%
0033TS_OC_31_M_Sec_Bun	3	0	0%
0039TV_MVJ_51_M_Uni_TT	49	17	35%
0044TV_JCP_58_M_Sec_TT	4	2	50%
0045TV_LP_49_M_Sec_TT	1	0	0%
0050TV_JCV_48_M_Sec_TT	17	10	59%
0051TD_DLB_45_M_Uni_Kem	11	4	36%
0053TD_JLS_44_M_Uni_Wai	3	3	100%
0054TD_NSG_51_M_Uni_TT	10	7	70%
0055TD_MAS_45_M_Uni_Mid	2	0	0%

Table 14 Frequency of use of ninia as a possessive determiner after nouns and personal pronouns

It is interesting to note that a similar construction exists in Portuguese-based creoles, like Malacca Creole Portuguese (*sa/sua*), Batavia Creole Portuguese (*soe/soea*) and Macau Creole Portuguese (*sa/sua*), as illustrated in Table 21. The possessive determiner *sa/sua*, which was originally the 3S possessive pronoun, came to be used as a possessive marker in all circumstances. This is why I believe that *ninia* is on the way to being grammaticalized as a possessive marker, synonymous to *nia*.

According to Hull and Eccles (2005:21), there is one more strategy to form possessive constructions in Tetun Prasa and that is with the prepositions *hosi* 'from' and *ba* 'to', often complemented by *nia(n)*. However, I do not find this strategy being highly productive in Tetun Prasa. It is clearly a calque of the Portuguese possessive strategy formed with *de* 'from, of' and it is mostly found in formal, written registers, as in (100), as well as in speakers proficient in Portuguese, as in (101):

(100) [...] *haktuir* *hosi* *gabinete* *ofisiál hosi* *estatística*
follow from office official from statistics

UE *nian.*
EU 3S.POSS.DET

'[...] followed by the official EU Statistical office.' (Sapo_March_2016)

(101) *Mezmu* *ne'e* *iha* *ke* *tama* *iha* *kurríkulu* *husi*
although this EXIST REL enter LOC curriculum from

eskola sekundária *nian.*
school secondary 3S.POSS.DET

'Despite this, it [mathematics] is part of the secondary school curriculum.' (0057TD_LAOB_30_M_Uni_Tet)

6.1.3.2 Predicative possession

In Tetun Prasa, possessive pronouns take the form of personal pronouns followed by a possessive marker *-nian*, as shown below:

Singular			Plural		
1.	*ha'u-nian*	'mine'	1.	*ami-nian* 'ours (exclusive)' *ita-nian* 'ours (inclusive)'	
2.	*ó-nian* *ita-nian*	'yours' 'yours (honorable)'	2.	*imi-nian*	'yours'
3.	*nian/ninian*	'his/hers/its'	3.	*sira-nian*	'theirs'

Table 15 Predicative possession in Tetun Prasa

Nominal possessors in Tetun Prasa are also followed by the possessive marker *nian* in case they occur in predicative constructions, as shown in (102), or headless noun phrases, as shown in (103) (Williams-van Klinken et al. 2002a:33):

(102) *Livru* *ne'e* *Maria nian.*
book this Maria 3S.POSS.PRON

'This book is Maria's.'

(103) *Maria   nian   iha-ne'ebá.*
       Maria   3S.POSS.PRON   there

       'Maria's is there.'

6.1.3.3 Possessive constructions in Tetun Terik and Tetun Fehan

Unlike in Tetun Prasa, adnominal possessive constructions in Tetun Terik do not require any possessive marker, whether the possessor is nominal or pronominal. The possession can be expressed by a simple juxtaposition 'possessor + possessum', as illustrated by (104) and (105). The same applies to the archaic form of literary Tetun in which *nia/nian* can be omitted, especially before nouns referring to kin terms and body parts (Hull & Eccles 2005:36).

(104) Tetun Terik:

       Ha'u   Ø   ina   sei   moris, ha'u   Ø   ama   sei
       1S   Ø   mother   still live   1S   Ø   father still

       moris.
       live

       'My mother is still alive; my father is still alive.'
       (0017TS_MCA_29_F_Sec_TT)

(105) Tetun Terik:

       Ita   Ø   aman Ø   naran saida?
       2S.HON   Ø   father Ø   name   what

       'What is your father's name?' (0017TS_MN_44_F_Sec_TT)

Various studies (Hull 1993, Hull & Eccles 2005) suggest that possessive determiners in Tetun Terik do not take the possessive marker *-nia*. Indeed, the possessive marker is not required, but my data show that it is relatively widely used in Tetun Terik from both Viqueque, as in (106), and Suai, as in (107).

(106) Tetun Terik from Viqueque:

Ne'e   hanesan   ha'u-nia   istória badak kona-ba
this   like   1S-3S.POSS   history short   about

ha'u-nia   moris ne'e […]
1S-3S.POSS   life   this

'This is my short history of my life […]' (0042TV_MS_43_M_Mid_TT)

(107) Tetun Terik from Suai:

E   ami-nia   mestra mak   agora sei   iha,   naran be
and   1PE-3S.POSS teacher FOC   now   still   EXIST name   umm

Olivia.
Olivia

'And our teacher, who is still here, is called Olivia.'
(0019TS_NA_54_M_Sec_TT)

It seems like the use of the possessive marker is not conditioned by the alienability/inalienability of possessum, since it can be found with both alienable and inalienable possessums in my corpus. It is not unusual to find speakers who switch freely between the two strategies:

(108) Tetun Terik:
Ha'u-nia   naran, ha'u   naran, naran AFA.
1S.-3S.POSS name   1S   name   name AFA.

'My name is... my name is AFA.' (0047TV_AFA_50_M_Sec_TT)

However, in the southern Tetun Terik, it is possible to form possessive determiners and possessive pronouns also with a genitive suffix *-n/-k* attached to a personal pronoun , as illustrated by (109), e.g. *ha'un/ ha'uk* '1S.GEN' (Hull 1993:249, Hull & Eccles 2005:36). The analysis of my data of Tetun Terik from Suai shows that this construction mostly appears in a preposed position, especially with kinship terms, as shown in (110). Possessive pronouns take the same form in this dialect, e.g. *ha'un/ha'uk* 'mine', as shown in (111).

(109)   Tetun Terik from Suai:

*Ita**k**   buat   ida   lakon   iha   to'os.*
1PI.GEN   thing   one   lose   LOC   field

'Our thing got lost in the field' (0020TS_CSG_56_M_Sec_TT)

(110)   Tetun Terik from Suai:

*Ami**k**   ina[128], ami**k**   ama sia,   halo to'os de'it.*
1PE.GEN   mother 1PE.GEN   father PL   make farm only

'Our mothers and our fathers only worked in the field.'
(0023TS_LLM_41_M_Sec_TT)

(111)   Tetun Terik from Suai:

*[...]   iha   Tetun Terik sei   dehan livru   ne'e   ha'u**n**.*
        LOC   Tetun Terik still   say   book   this   1S.GEN

'[..] in Tetun Terik they still say: This book is mine.'
(0054TD_NSG_51_M_Uni_TT)

Although Hull and Eccles did not account for it, I found instances of the genitive suffix also in Tetun Terik spoken in the Viqueque region (112)-(113). However, in this variety, the possessive determiner is always postposed in my corpus, a phenomenon not recorded in Suai.

(112)   Tetun Terik from Viqueque:

*Lian   ida   ha'u**k**   kontinua   nanu'u   ohin.*
language   one   1S.GEN   continue   thus   today

'My language thus continues until today.'
(0047TV_AFA_50_M_Sec_TT)

(113)   Tetun Terik from Viqueque:

[...]   nia   bá   mai   fali   iha   rai   wa'iwa'in,   rai
        3S   go   come   again   LOC   land   common   land

---

[128] The final *-n* is often eliminated in Tetun Terik, even with kinship terms that are, in general, inherently possessed.

*ita**k***   *ne'e.*
1PI.GEN   this

'[…] he returned again to the usual country, to our country.'
(0046TV_JCV_48_M_Sec_TT)

In Tetun Fehan (van Klinken 1999:143), the possessor can either precede or follow the possessum, although the preposed order is more common. Preposed nominal possessors either do not take any possessive markers, as in (114), or are followed by a possessive marker, which has the form of a personal pronoun *ni/nia* for singular nouns and *sia* for plural nouns. This possessive marker can be further marked by another possessive marker *-kan*, as in (115).

(114)   Tetun Fehan (van Klinken 1999:143):

        ó   *fén*
        1S   wife.GEN

        'my wife'

(115)   Tetun Fehan (van Klinken 1999:143):

        ina-ama   *sia-kan*   *sasoin*
        mother-father   POSS.PL-POSS   wealth

        'parent's wealth'

Preposed pronominal possessors are also often marked by optional *-kan*:

(116)   Tetun Fehan (van Klinken 1999:143):

        ha'u-kan   *tais*
        1S-POSS   cloth

        'my cloth'

When the possessor is preposed, a vowel-final possessum or an inalienable possessum in most contexts takes a genitive suffix *-n* when used in singular, as shown above in (114), or *-r* when used with plural possessors, as shown in (117). This is different from Tetun Terik where the genitive suffix is attached to the possessor, not the possessum.

(117) Tetun Fehan (van Klinken 1999:143):

sa   atar
3P   slave.GEN

'their slave'

Alienable possessums never take this marker, even when they are possessed (except for *rai* 'land' and *uma* 'house'). The possessor can be further optionally followed by a possessive marker *ni* or *nia* (for singular nouns) or *sia* (for plural nouns) (van Klinken 1999:142-143). In case of a postposed possessor, a genitive suffix *-k* is employed with pronominal possessors:

(118)   Tetun Fehan (van Klinken 1999:143-144):

asu   ók
dog   2S.GEN

'your dog'

To summarize, let us look at the various possessive strategies in all Tetun varieties with pronominal possessors, as illustrated by Table 16, and with nominal possessors, as illustrated by Table 17.

variety of Tetun	preposed possessive determiner	possessum	postposed possessive determiner
Tetun Prasa	*ha'u-nia*	*uma* 'house'	*ha'u-nian*
Tetun Suai	*ha'u Ø* *ha'u-nia* *ha'uk/ha'un*	*uma* 'house'	
Tetun Viqueque	*ha'u Ø* *ha'u-nia*	*uma* 'house'	*ha'uk*
Tetun Fehan	*ha'u Ø* *ha'u-kan*	*uma* 'house'	*ha'uk*

Table 16 Pronominal possession in Tetun

variety of Tetun	possessive NP with *feto* 'woman'	possessum	possessive NP with *feto* 'woman'
Tetun Prasa	*feto nia*	*uma* 'house'	*feto nian*
Tetun Suai	*feto nia*	*uma* 'house'	*(feto nian)*
Tetun Viqueque	*feto nia*	*uma* 'house'	*(feto nian)*
Tetun Fehan	*feto ni* *feto nia-kan* *feto sia-kan (PL)*	*uma* 'house'	*feto niak*

Table 17 Nominal possession in Tetun

6.1.3.4 Origin and development of possessive constructions in Tetun

As is shown above, there are certain types of possessive constructions and possessive markers that are found predominantly in Tetun Prasa and others that are typical for Tetun Terik and Tetun Fehan. It can thus be hypothesized that these constructions in Tetun Prasa were borrowed or calqued from other languages spoken in the region. Let us look at the possible development of possessive constructions in Tetun Prasa and the possible origin of the possessive marker *nia*.

Different possessive constructions were already documented in the oldest written works. Aparício da Silva (1889:13) described the most common strategy in his Portuguese-Tetun dictionary and according to him, all possessive constructions with a pronominal possessor in Tetun Prasa were formed by adding *nia* to a personal pronoun (in case of 3S, the possessor took the form of *ninia* or just *nia*). In the dialects spoken in the interior of the island, *nia* was often omitted with pronominal possessors, e.g. *ha'u aman* 'my father'. Silva found this strategy irrational "porque em alguns casos, mudando o tom de voz, pode entender-se outra coisa"[129] (Silva 1889:14). As an example, he listed *ó asu* 'your dog', which could also mean 'you are a dog'. Silva described possessive constructions in Tetun Prasa, in which the possessor preceded the possessum (as in *ó-nia kuda* 'your horse'), although he mentioned other possibility that was, according to his own words, not that correct: *kuda ó-nia* 'your horse' (Silva 1889:29).

---

[129] "Because is some cases, if we change the tone of our voice, the meaning can be different." (my translation)

Raphael das Dores (1907:12), also working on Tetun, mentioned two possible ways of forming possessive constructions. The first one used *nia*, identified by Dores as the third person singular 'he/she' (*haú-nia* 'my'). The other one was the suffix *-n/k* used, according to the author, on the southern coast (*haún* 'my'). As for the order of the constituents, Dores (1907:18-19) described only one possible order: that of a preposed possessor (*haú-nia áman* 'my father').

Capell (1944c:32) noted that the possessive marker had the form of the third person singular *nia* and appeared in between two nouns (*katuas nia uma* 'old man's house'). The author believed that this type of construction was common to all Timorese languages, except for Waima'a.

Some decades later, Morris (1984:xv), in his Tetun-English dictionary based on the southern dialect of Tetun, described three possible strategies for possessive constructions with a pronominal possessor: pronoun + *nia* + possessum[130] (*ha'u-nia asu* 'my dog') which is the most common order, pronoun + possessum + *-n/-k* (*ha'u asun* 'my dog') which is used in conversations; and pronoun + *nian/-n/-k* (*ha'u nian/ ha'un/ ha'uk* 'mine') when the possessive pronoun stands alone. In the case of nominal possessor, the noun is followed by *nia* and the possessum (*karau nia dikur* 'buffalo's horn'), although *nia* can be omitted in conversation (*karau dikur* 'buffalo's horn').

According to Hull (2003a:17), the same construction exists in Southern Mambae, in the Daissau-Betano dialect, in which the possessive link-word is *ni*, in both adnominal (*hin ni um* 'the woman's house') and predicative (*kud rai Pedro ni* 'this horse is Pedro's') possessive constructions. In the case of pronominal possessors, this dialect also prefers the compound forms, e.g. *au-ni* '1S-3S.POSS.DET', i.e. 'my' (Hull 2003a:17) which might have been caused by contact with Tetun Prasa. However, in this case, *ni* is not the third person singular personal pronoun (which is *urá* 'he/she/it'), but rather the form of the third person singular possessive determiner (*ni* 'his/her/its') which resembles the Tetun Prasa 3S *nia*.

I have done my own survey among native speakers of Mambae and my three informants offered three different sets of data, as is documented in Table 18. I was unable

---

[130] According to Morris (1984:xv), in the case of the third person singular, *nia nia* is just as correct as *ninia*.

to verify Hull's claims about the compound forms, although one informant from Same (Informant 3, see Table 18 below) used them rather irregularly. This made me believe that Mambae uses, in general, a simple juxtaposition of a personal pronoun and a noun (*au uma* '1S house', i.e. 'my house'), except for the third person singular which uses a possessive determiner *ni* (*ni uma* '3S.POSS.DET house', i.e. 'his/her/its house'). The order of the constituents can be also inverted (*um au* 'house 1S', i.e. 'my house').

Tetun Prasa	Gloss	Informant 1	Informant 2 from Ainaro	Informant 3 from Same
ha'u-nia uma	'my.house'	*au-uma*	*au fada*	*au uma*
ha'u-nia aman	'my.father'	*au-ama*	*au aman*	*au ama*
ninia inan	'his.mother'	*ni-ina*	*ura inan*	*ura <u>ni</u> ina*
ninia uma	'his.house'	*ni-uma*	*ura fada*	*ura <u>ni</u> uma*
sira-nia oan	'their.child'	*rom-ana*	*ro anan*	*ura <u>nia</u> ana*
sira-nia fahi	'their.pig'	*rom-háhi*	*ro haifa*	*ura <u>ni</u> haeh*
uma ha'u-nian	'house.mine'	*um-au*	*fada aun*	*um au <u>ni</u>*
labarik ninian	'child.his'	*anta-ni*	*ankoloban nin*	*ankate ni*

Table 18 Pronominal possession in Mambae

6.1.3.5 Adnominal possession in Timorese languages with nominal possessors

Most of the authors agree on the fact that the possessive marker *nia* was derived from the third person singular pronoun *nia*. Hull and Eccles (2005:20) believed that *nia* originated in Malay *punya* 'to own' before it assimilated to the 3S personal pronoun *nia*. This theory is further discussed in Baxter (1990:14) who compared the possessive structures in Tetun Prasa and in the restructured varieties of Malay and SE Asian Portuguese. He noted that the restructured Bazaar Malay made use of the same type of possessive construction with a preposed possessor followed by a possessive marker *punya* (i.e. *possessor + punya + possessum*, for example *lu punya rumah* 'your house').

But we do not even have to look outside of Timor to find parallel constructions for possessive noun phrases. Austronesian languages like Mambae, Baikeno and Galolen and Papuan languages like Makasae and Fataluku can all form both preposed and postposed possessive constructions when the possessor is a noun. In Makasae and Fataluku, the order depends on the animacity of the possessor: preposed order is used with human possessors, postposed order is used with non-human possessors (Hull 2005b, Hull 2005c). The possessive marker in the preposed constructions usually has the form of the 3S personal pronoun or 3S possessive determiner and slightly differs from the possessive marker in postposed constructions.

- Mambae[131]

Hull (2003a) claims there is a possessive marker *ni* used in the southern Mambae, both with preposed possessors, as in (119), and postposed possessors, as in (120).

(119)   Mambae (Hull 2003a:12):

Hin   *ni*   *um*
woman   POSS   house

'woman's house'

(120)   Mambae (Hull 2003a:12):

Au   *tad*   *gala*   *iskola*   *ni*
1S   know   name   school   POSS

'I know the name of the school.'

- Baikeno[132]

In preposed constructions, a possessive marker *in* '3S' is used, as show in (121). In postposed construction, it is possible to use either a simple juxtaposition, as shown (122), just like in Malay, or a possessive markers *ini* '3S.POSS.PRON', as in (123), and *in kuna* (Hull 2003b:13-14). Baikeno is a dialect of Uab Meto, the most widely spoken language in West Timor, however, we do not find a possessive marker in the West Tetun

---

[131] An Austronesian language spoken outside of Dili and all the way to the southern coast with heavy Tetun influence (Hull 2003a:1).

[132] The most grammatically complex Austronesian language in Timor, spoken in the Oecussi enclave (Hull 2003b:1).

variety. According to Benu (2016), Uab Meto uses a simple juxtaposition with nominal possessors, as shown in (124).

(121) Baikeno (Hull 2003b:13):

Atoni   in   bikase
man   POSS   horse

'the man's horse'

(122) Baikeno (Hull 2003b:13):

Bikase atoni
horse   man

'the man's horse'

(123) Baikeno (Hull 2003b:13):

Au   uhín   nakaf skol   ini
1S   know   name   school POSS

'I know the name of the school'

(124) Uab Meto (Benu 2016):

Na   Lukas   ume
DET   Lukas   house

'Lukas' house'

- Makasae[133]

In preposed constructions, a possessive marker *gi* '3S/3S.POSS.DET' is used, as shown in (125). Postposed constructions are in general used with non-human possessors and a possessive marker *ge'e/gige'e* '3S.POSS.PRON', as shown in (126) (Hull 2005b:12).

---

[133] A non-Austronesian (Papuan) language spoken in the districts of Baucau and Viqueque, strongly influenced by other Austronesian languages, with a fully isolating character (Hull 2005b).

(125) Makasae (Hull 2005b:12):

Filipe gi   ina
Filipe POSS mother

'Filipe's mother'

(126) Makasae (Hull 2005b:12):

Bu'u   Timor ge'e
mountain   Timor POSS

'the mountains of Timor'

- Fataluku[134]

In preposed constructions, a possessive marker *i* '3S.POSS.DET' is used, as shown in (127). Postposed constructions are in general used with non-human possessors and a possessive marker *hini/ihini* '3S.POSS.PRON', as shown in (128) (Hull 2005c:16-17).

(127) Fataluku (Hull 2005c:16):

Tupuru   i   lee
woman   3S.POSS.DET   house

'the woman's house'

(128) Fataluku (Hull 2005c:17):

Ili   ere   Timor ihini
mountain   PL   Timor 3S.POSS.PRON

'the mountains of Timor'

Based on the summary in Table 19, it is clear that all Austronesian and Papuan languages of Timor-Leste mentioned above use possessive markers with nominal possessors to mark both preposed and postposed possession. This comparison is important, because as we will see in the next section 6.1.3.6, this is not the case with pronominal possessors.

---

[134] A non-Austronesian (Papuan) spoken in the easternmost part of Timor-Leste (Hull 2005c).

Language	preposed order	postposed order
Tetun Prasa	Pr + *nia* + Pm	Pm + Pr + *nian*
Tetun Terik	Pr + *nia* + Pm	Pm + Pr + *nian*
Mambae	Pr + *ni* + Pm	Pm + Pr + *ni*
Baikeno	Pr + *in* + Pm	Pm + Pr + *(ini)*
Galolen	Pr + *ni* + Pm	Pm + Pr + *nin*
Makasae	Pr + *gi* + Pm	Pm + Pr + *(gi)ge'e*
Fataluku	Pr + *i* + Pm	Pm + Pr + *i(hini)*

Table 19 Preposed and postposed possession in Timorese languages with nominal possessors

6.1.3.6 Adnominal possession in Timorese languages with pronominal possessors

As noted in Esperança (2001:30), none of the languages of Timor-Leste uses possessive markers in constructions with *preposed pronominal possessors*. Indeed, the languages that employ the same strategy to mark nominal and pronominal possessors form a small minority of the world's languages (Dryer 2007:182). In what follows, I provide a summary of how other Timorese languages form possessive determiners:

- **Tetun Prasa**: different form for personal pronouns and possessive determiners (personal pronoun + *-nia(n)*).

- **Mambae**: in general, the pronominal possessors take the form of personal pronouns, except for 3S (*ni* 'his/her/its').

- **Baikeno**: no possessive marker used with possessive determiners, except for a possibly dialectal variation when possessive determiners take a possessive marker *-kun* which is similar to Tetun Fehan *-kan*., e.g. *au(-kun) ume* '1S-POSS house' (Hull 2003b:19-20).

- **Galolen**: possessive determiners are the same as personal pronouns, except for 1S (Hull 2003c:14), adnominal possession is formed by a simple juxtaposition.

- **Makasae**: Makasae has its own set of possessive pronouns that are affixed to the possessum. The forms of possessive pronouns 2S, 3S, 1PI and 2P are formally identical to personal pronouns, although they can't be stressed, but 1S (*asi-*) and

1PE (*isi-*) have special possessive forms. In case of 3P, the 3P personal pronoun *era* takes on a possessive marker *gi-*, which is the form of a general possessive marker used with nominal possessors (Huber in press).

- **Fataluku**: possessive determiners have a different form from personal pronouns, but they do not take any possessive marker (Hull 2005c:21-22).

- **Bunak**[135]: has different strategies for alienable and inalienable possession. In case of inalienable possession, a possessive prefix is attached to the possessum, e.g. *n-up* '1-tongue', meaning 'my tongue'. In case of alienable possession, a possessive prefix is attached to a particle *e*, which precedes the possessum, e.g. *ni-e zap* '1-POSS dog', meaning 'my dog' (Donohue & Schapper 2008: 322).

As we can see in Table 20, pronominal possessors (possessive determiners) behave differently from nominal possessors. They usually have a form of a personal pronoun or a possessive pronoun but no other language but Tetun Prasa (and to some extent also Tetun Terik) uses a possessive marker to form possessive determiners.

Language	*possessive NP with {possessive determiner}*
Tetun Prasa	{pron + *-nia*} + Pm
	Pm + {pron + *nian*}
Tetun Terik	{pron} + Pm
	{pron + *-k/-n*} + Pm
Mambae	{pron} + Pm
Baikeno	{pron} + Pm
Galolen	{pron} + Pm
Makasae	{POSS pron} + Pm
Fataluku	{POSS pron} + Pm

Table 20 Possessive NP with possessive determiner in Tetun and other Timorese languages

6.1.3.7 Discussion

Based on the evidence, we may ask whether this possessive construction (*pronominal possessor + -nia + possessum*) always existed in Tetun or if it is an innovative form of Tetun Prasa.

---

[135] A Papuan language spoken in the western part of Timor-Leste, especially in the districts of Bobonaro, Ainaro and Cova Lima.

The fact that only Tetun Prasa makes use of a possessive marker with preposed pronominal possessors can be, according to Esperança (2001:33), attributed to the influence of a creole. Indeed, similar constructions are also found in restructured varieties of Malay and Portuguese spoken in the region. According to Paauw (2008:408), one of the strategies to express possession is Ambonese Malay[136] is by using a possessive marker *pung*[137]:

(129)   Ambonese Malay (Paauw 2008:408):

antua   pung   rambut
3S.FORMAL   POSS   hair

'her hair'

This possessive marker can be used both with nominal and pronominal possessors. This would support the hypothesis of Hull and Eccles (2005:20) that the possessive marker *nia* used in Tetun Prasa originated in Malay *punya* and then assimilated to the 3S personal pronoun.

We can also find similar parallels in all Asian varieties of Creole Portuguese (with the exception of Diu and Daman), including Malacca Creole Portuguese, Macau Creole Portuguese (Batalha 1959:188) and Bidau Creole Portuguese. Baxter (1988:91-93) documented the possessive constructions in Malacca Creole Portuguese, including the ones with pronominal possessors, in which two possessive markers are used: *sa*, which follows the possessor (*possessor + sa + possessum*), as shown in (130) and (131), and *di*, which precedes the possessor (*possessum + di + possessor*), as shown in (132). Both markers are derived from Portuguese: *sa* is most likely a contracted form of Port. *sua*

---

[136] The official Malay language uses the *possessum + possessor* order in which the possessor simply follows the possessum, e.g. *rumah Ruslan* 'Ruslan's house' (Collins 1983:29). However, historically, *punya* possession existed in the official Malay, too, but became obsolete and was avoided by the end of the 19th century. Nowadays, constructions with *punya* in Indonesian are considered non-standard (Tjia 2004:54).

[137] Collins (1983:30) further asserted that in Ambonese Malay it is also possible to delete [*puŋ*] which leads to a possessive construction possessor + possessum. This deletion can most likely be linked to the influence of indigenous languages of the Moluccas (Collins 1983:35), which belong to the preposed possessor languages. Tjia (2004:54) believed that this kind of construction is unique to Ambonese Malay and is not found in other Malay varieties. He, however, disagreed with Collins (1983) when it came to the origin of this deletion. According to Tjia (2004:55-56), the deletion of *pung* was not caused directly by the substrate influence but rather by language-internal phonological reduction which eventually led to a complete deletion.

212

'3S.POSS.FEM', i.e. 'her' (Rêgo 1998:66), since other creoles use similar forms[138]. The possessive marker *di* comes from Port. *de* 'of, from'.

(130) Malacca Creole Portuguese (Baxter & Bastos 2012:60):

Singapura   sa   jenti   lo   beng   Malaka
Singapore   POSS   person   FUT   come   Malacca

'People from Singapore will come from Malacca'

(131) Malacca Creole Portuguese (Baxter & Bastos 2012:66):

bos   sa   fila
2S   POSS   daughter

'your daughter'

(132) Malacca Creole Portuguese (Baxter 1988:94):

kaza   di   pedra
house   POSS   stone

'house of stone'

The construction with *di* is clearly of Portuguese origin, but the *sa* strategy can be traced to various sources. Clearly, one of them is the already mentioned Bazaar Malay *punya*, mentioned in section 6.1.3.5. Baxter (1988:91) believed that it could have been Hokkien language from the Sino-Tibetan language family that introduced this form into the restructured variety of Malay spoken in Melaka, although the Hokkien connection seems rather unlikely. Baxter and Bastos (2012:75) advocated for Indo-Portuguese connection since similar structures are found in Malabar (Clements 2000), Korlai or Cannanore (Cardoso 2009:114).

Baxter and Bastos (2012:75) proposed that the reanalysis of the 16th-century Portuguese 3rd person *sua/seu* was based on the third person possessive and association with the genitive suffixes in the substrate languages in Southern India, which was further grammaticalized to different degrees in different varieties of Asian Creole Portuguese, as is evident from Table 21.

---

[138] Baxter (1988:92-93) originally suggested that the marker *sa* came from Port. *vossa/nossa* '2S.HON.FEM/1P.FEM', however, in his joint paper with Bastos (2012), he revised his position.

	Pronominal possessor [[Pronoun] + GEN] + NP		Nominal possessor [[DP] + GEN] + NP	Possessive Pronoun = [[Pronoun] + GEN]	
MALACCA	*sa, sua* (arch.)† R:66:		*sa, sua* R:66.	R:66:	
	1SG: *yo sa*	1PL: *nus sa*		1SG: *yo sa*	1PL: *nus sa*
	2SG: *bos sa*	2PL: *bolotu sa*		2SG: *bos sa*	2PL: *bolotu sa*
	3SG: *eli sa*	3PL: *olotu sa*		3SG: *eli sa*	3PL: *olotu sa*
BATAVIA	S:100:		*soe, soea,* S:122.	S:100–101:	
	2SG *boos soea*	1PL *noos soea, noosoter soea*		1SG *mienja soea*	1PL *noosoter soea*
		2PL *vosoter soea*		2SG *bosée soea*	2PL *vosoter soea*
		3PL *iloter soea*			3PL *iloter soea*
MACAU	F:237–238:		*sa,* B:483, *sua,* B:30.	F:237–238:	
	1SG *iou-sua*††	1PL *nôs-sua*		1SG *iou-sua*	1PL *nôs-sua*
	2SG *vosso-sua, vôs-sua*	2PL *vosôtro-sua*		2SG *vosso-sua, vôs-sua*	2PL *vosôtro-sua*
	3SG *êle-sua*‡	3PL *ilôtro-sua*		3SG *êle-sua*	3PL *ilôtro-sua*

Table 21 Post-nominal genitive markers in East and Southeast Asian Creole Portuguese (Baxter & Bastos 2012:57)

It can be hypothesized that it was after being in contact with the restructured varieties of Malay or Portuguese that Tetun Prasa started to mark possession on personal pronouns indirectly by way of using a possessive marker *nia*. However, we need to take into consideration the fact that other varieties of Tetun and other Timorese languages have a possessive marker, although it is not always used also pronominally. This would mean that the possessive marker *nia* is native of Tetun but its use in pronominal constructions has been copied/calqued from creoles of Malay and/or Portuguese origin.

6.1.4   Gender marking and agreement

In general, nouns are not marked for gender in Tetun Prasa. To refer to humans, animals, and plants, Tetun Prasa uses classifiers for the sex distinction. In the case of humans and certain plants, the classifiers used for female/male distinctions are *feto* 'woman' and *mane* 'man', respectively (*labarik-feto* 'girl', *labarik-mane* 'boy' from *labarik* 'child'). In the case of animals, *aman* 'father' is used with males and *inan* 'mother' is used with females (*bibi-aman* 'billy-goat; ram', *bibi-inan* 'nanny-goat; ewe' from *bibi* 'goat'). Portuguese loanwords depicting humans (especially kinship terms and professions) are borrowed in their original form, e.g. *primu* 'male cousin' (from Port. *primo),* and *prima* 'female cousin' (from Port. *prima*).

When it comes to borrowings, adjectives are, in general, borrowed in the singular masculine form, whether they modify a Tetun noun, as in (133a), or a Portuguese loanword, as in (133b).

(133)   a.   *tulun umanitáriu* 'humanitarian help'

                (from *tulun* 'help' and Port. *humanitário* 'humanitarian.MASC')

       b.   *empreza públiku* 'public company'

                (from Port. *empresa* 'company.FEM' and *público*

                'public.MASC')

However, as Williams-van Klinken et al. (2002a:37) noted, there are two exceptions: *bonitu* 'handsome' (from Port. *bonito*) used for males, and *bonita* 'beautiful' (from Port. *bonita*) for females. In addition to these, I found other examples of gender marking in my corpus: *morna* 'lukewarm.FEM' (from Port. *morno/a*), which was borrowed in the feminine form, possibly because it is mostly used to describe water (Port. *água*.FEM) and drinks (Port. *bebida*.FEM), which are feminine nouns in Portuguese. Other examples would be the adjective *materna* 'mother.FEM', as in *lian materna* 'mother tongue', which is always used in the feminine form, since Port. *língua* 'tongue' is also feminine, and *grávida* 'pregnant' (from Port. *grávida*).

Apart from these exclusively feminine forms, I have found numerous examples of feminine gender agreement, e.g. as the one shown in (134), and a few examples of feminine forms that are used with native Tetun words, as shown in (135).[139]

(134) *nesesidade*   *bázika*   *país*   *nian*
       need(FEM)[140] basic.FEM   country   POSS

       'the country's basic need' (GovernuTL_2010)

(135)   a.   *rai*   *gratuita*
              land   free.FEM   (cf. Port. terra 'land.FEM')

            'free land' (GovernuTL_2010)

       b.   *lian*   *materna*
              language   mother.FEM   (cf. Port. *língua* 'language.FEM')

            'mother tongue' (0055TD_MAS_45_M_Uni_Mid)

---

[139] Some Portuguese masculine nouns were followed by a feminine modifier. They were almost exclusively found in written texts and the most probable cause was just bad editing.
    ex.   *partidu*   *demokrátika*
            party(MASC)   democratic.FEM
            'Democratic Party' (STL_July_2013)
[140] Although Tetun nouns do not encode gender, I will list the gender in parenthesis as it is used in Portuguese.

    c.   *moras*   *kontajioza*
               disease   contagious   (cf. Port. *doença* 'disease.FEM')

               'contagious disease' (Sapo_March_2016)

    d.   *buat   ruma konkreta*
               thing   some concrete
               (cf. Port. *alguma coisa* 'something.FEM')

               'something concrete' (STL_July_2013)

Comparing the written and spoken corpora, the majority of cases of gender agreement were found in my written corpus. This was probably due to the fact, that the media use more technical language and more fixed multiword expressions.

Gender agreement is almost always found in fixed multiword expressions. Ordinal numerals also tend to agree with the head noun, especially when they precede it, i.e. follow the Portuguese word order:

(136) a. *primeiru anu*   'first.MASC year(MASC)'
       (0057TD_LAOB_30_M_Uni_Tet)
      b. *kuarta klase*   'fourth.FEM grade(FEM)'
       (0053TD_JLS_44_M_Uni_Wai)
      c. *terseira filla*   'third.FEM daughter(FEM)'
       (0056TD_CS_50_F_Pri_Tet)

When the ordinal numeral follows the head noun, the gender agreement is not always found:

(137) a. *pergunta primeira* 'first.FEM question(FEM)'
       (0051TD_DLB_45_M_Uni_Kem)

      b. *pergunta segundu* 'second.MAC question(FEM)'
       (0014TD_JBP_26_M_Uni_Mak)

I looked at the distribution of gender agreement with Portuguese borrowings in my corpus and analysed the adjectives based on their syntactic context (see Table 22):

- modifiers to NPs;
- predicative adjectives in predicative constructions;

- adjectives used in relative clauses.

	1 - MASC default	2 - MASC agreement	3 - FEM agreement	4 - FEM non-agreement	5 - gender neutral	6 - with native words
1 - NP + modifier	137	658	420	21	258	75
2 - predicative phrases	14	21	3	0	31	0
3 - REL clause	23	33	2	1	33	10

Table 22 Syntactic contexts and types of gender agreement considered in the analysis

In my analysis, I did not consider two types of loaned adjectives: gender-neutral adjectives (number 5 in Table 22; 322 occurrences) which are invariant and do not encode gender, and those that combined native words (number 6 in Table 22; 85 occurrences). In a limited number of cases (4 occurrences), the adjectives used with native words were used in the feminine form, as shown above in (135). This could be due to the fact that these nouns are feminine in Portuguese. The rest of the occurrences (81 occurrences) had the default, masculine form.

I divided the remaining occurrences into four categories, based on the type of agreement (see Table 23):

- adjectives that would have feminine form in Portuguese but were used in the default, masculine form, in Tetun (MASC default);
- adjectives that would have feminine form in Portuguese and were used in feminine form in Tetun (FEM agreement);
- adjectives that would have masculine form in Portuguese and were used in masculine form in Tetun (MASC agreement);
- adjectives that would have masculine form in Portuguese and were used in feminine form in Tetun (FEM non-agreement).

	NPs with a [FEM] head		NPs with a [MASC] head	
	1 - MASC default	3 - FEM agreement	2 - MASC agreement	4 - FEM non-agreement
1 - NP + modifier	25%	75%	97%	3%
2 - predicative phrases	82%	18%	100%	0%
3 - REL clause	92%	8%	97%	3%

Table 23 Gender agreement in Portuguese borrowings

In the case of modifiers to NPs, these are most likely to agree in gender as many of them are fixed multiword expressions and were borrowed as such:

(138) a. *eskola primária bázika* 'basic.FEM primary.FEM school(FEM)'
(0057TD_LAOB_30_M_Uni_Tet)

b. *ortografia Tetun padronizada*
'Tetun standard.FEM orthography(FEM)'
(0051TD_DLB_45_M_Uni_Kem)

c. *funsionáriu públiku* 'civil.MASC servant(MASC)'
(0051TD_DLB_45_M_Uni_Kem)

d. *língua materna* 'mother.FEM tongue(FEM)'
(0054TD_NSG_51_M_Uni_TT)

However, with other noun phrases, the masculine form of the modifier was preferred:

(139) a. *kestaun polítiku* 'political.MASC question.FEM'
(0051TD_DLB_45_M_Uni_Kem)

b. *nasaun viziñu* 'neighbouring.MASC nation.FEM'
(0008TD_TAB_55_M_PreS_Mam)

When it comes to relative clauses, as in (140), and predicative constructions, as in (141), we can see that (feminine) gender agreement travels only sometimes across the phrase boundary. It can be argued that the bigger the distance between a noun and an adjective, the higher the chance of the masculine default form.

(140) *Nasaun* *sira* *ne'ebé* *riku* *liu* [...]
       nation(FEM) PL   REL   rich.MASC more

'Nations that are richer […]' (Sapo_March_2016)

(141) *Ha'u* *hakarak* *imi* *hotu* *komprende* *katak*
       1S   want   2P   all   understand   COMP

situasaun *ne'e* *perigoza*.
situation(FEM) this dangerous.FEM

'I want all of you to understand that this situation is dangerous.'
(Sapo_March_2016)

The predominance of masculine forms indicates that speakers/writers do not necessarily aim for gender agreement and do not take into consideration the gender of nouns borrowed from Portuguese.

6.1.4.1 Analysis of the oral corpus

To see if proficiency in Portuguese has any influence on gender agreement, I divided my speakers into two groups: speakers who claim to speak Portuguese (with various levels of proficiency) and speakers who do not speak (but can understand some) Portuguese. Just like in other cases, I found significant interspeaker variation.

Speakers proficient in Portuguese would, in general, aim for gender agreement, especially when the adjective immediately follows the noun (see Table 24). Gender agreement was mostly found in borrowed fixed multiword noun expressions, as illustrated by (138) above.

PROFICIENT IN POTUGUESE	NPs with a [FEM] head		NPs with a [MASC] head	
	1 - MASC default	3 - FEM agreement	2 - MASC agreement	4 - FEM non-agreement
1 - NP + modifier	5%	95%	100%	0%
2 - predicative phrases	50%	50%	100%	0%
3 - REL clause	50%	50%	0%	0%

Table 24 Gender agreement in Portuguese borrowing used by speakers proficient in Portuguese

As for the non-Portuguese speakers, I have found instances of gender agreement, too, especially with ordinal numerals and borrowed multiword expressions (see Table 25). This might be caused by the fact, that although these speakers do not possess active knowledge of the language, they are in daily contact with it through media, in which gender agreement is a rather common phenomenon.

NOT PROFICIENT IN POTUGUESE	NPs with a [FEM] head		NPs with a [MASC] head	
	1 - MASC default	3 - FEM agreement	2 - MASC agreement	4 - FEM non-agreement
1 - NP + modifier	33%	67%	89%	11%
2 - predicative phrases	0%	0%	0%	0%
3 - REL clause	0%	0%	0%	0%

Table 25 Gender agreement in Portuguese borrowing used by speakers not proficient in Portuguese

Another variable that proved to be rather insignificant was the level of education. When it comes to "NP + modifier" structures, the numbers were almost the same for university-educated speakers and speakers with primary or secondary education, as shown in Table 26 and in Table 27.

	NPs with a [FEM] head		NPs with a [MASC] head	
PRIMARY & SECONDARY EDUCATION	1 - MASC default	3 - FEM agreement	2 - MASC agreement	4 - FEM non-agreement
1 - NP + modifier	8%	92%	100%	0%
2 - predicative phrases	0%	0%	0%	0%
3 - REL clause	0%	0%	0%	0%

Table 26 Gender agreement with Portuguese borrowings used by speakers with primary and secondary education

	NPs with a [FEM] head		NPs with a [MASC] head	
HIGHER EDUCATION	1 - MASC default	3 - FEM agreement	2 - MASC agreement	4 - FEM non-agreement
1 - NP + modifier	10%	90%	95%	5%
2 - predicative phrases	50%	50%	100%	0%
3 - REL clause	50%	50%	0%	0%

Table 27 Gender agreement with Portuguese borrowings used by speakers with higher education

6.1.4.2 Analysis of the written corpus

The highest number of Portuguese adjectives was found in the written corpus. Again, many of the cases of gender agreement were found in fixed multiword expressions, as shown in (142). However, there were also many examples of gender non-agreement, i.e. adjectives with feminine forms used with Portuguese masculine nouns, as shown in (143). As we can see from Table 28, gender agreement is preferred to default masculine form, especially when an adjective immediately follows a NP.

(142) a. *rekursu*   *umanu*
           resource(MASC)   human.MASC

           'human resources' (GovernuTL_2010)

       b. *estadu*   *demokrátiku*
           country(MASC)   democratic.MASC

           'democratic country' (GovernuTL_2010)

c. *forsa* *armada*
force(FEM) armed.FEM

'armed forces' (Sapo_March_2016)

(143) a. *sidadaun* *australiana*
citizen(MASC) Australian.FEM

'Australian citizen' (Sapo_March_2016)

b. *observedór* *política*
observer(MASC) political.FEM

'political observer' (STL_March_2016)

WRITTEN	NPs with a [FEM] head		NPs with a [MASC] head	
	1 - MASC default	3 - FEM agreement	2 - MASC agreement	4 - FEM non-agreement
1 - NP + modifier	27%	73%	97%	3%
2 - predicative phrases	87%	13%	100%	0%
3 - REL clause	96%	4%	97%	3%

Table 28 Gender agreement with Portuguese borrowings in written corpus

Unlike in the oral corpus, I have also encountered many occurrences of gender agreement in predicative phrases, as in (144), and relative clauses, as in (145), as Table 28 above illustrates. This might have to do with the fact that newspaper articles make better use of complex sentences with multiple subordinate clauses. However, default masculine form is also widely used, as shown in (146).

(144) *Akordu* *ho* *Austrália* *ne'e* *tranzitóriu.*
agreement with Australia this transitory

'The agreement with Australia is transitory.' (Sapo_March_2016)

(145) [...] *atu* *hetan* *sistema* *ida-ne'ebé* *ekilibradu,*
in.order.to get system one-REL equilibratory

eficiente *no* *efikás* […]
efficient and effective

'[...] in order to get a system that is equilibratory, efficient and effective [...]' (GovernuTL_2010)

(146) *Ha'u*   *hanoin rezolusaun*   *ne'e*   *klaru* [...]
       1S   think   resolution   this   clear

'I think this resolution is clear [...]' (STL_July_2013)

6.1.5 Classifiers

Apart from classifiers based on biological gender, Tetun also has a number of sortal classifiers[141] that are used with nouns of high countability. These classifiers occur in NP constructions together with numerals and their main function is to divide the count nouns into semantic classes according to their shape, functions, animacy, etc. (Gil 2013). Their collocation is not always predictable but "there is always some semantic relationship between a sortal classifier and the common noun with the same phonological form", e.g. *tahan* lit. 'leaf' is a classifier for things that have a similar shape as a leaf (van Klinken 1999:54).

Sortal classifiers are very common throughout the whole Southeast Asia (Gil 2013). They are found in Tetun Terik, Tetun Fehan but their use is very limited in present-day Tetun Prasa. In all dialects, classifiers are not obligatory but rather optional although some, like *na'in* 'CLF:human', are considered the norm in the Official Tetun. According to Hull and Eccles (2005:25), only three classifiers survived in modern Tetun Prasa: *na'in*[142] lit. 'owner', the classifier for humans; *kain* lit. 'stem', the classifier for long objects; and *fuan* lit. 'heart', the classifier for roundish objects like fruits and eggs, as shown Table 29.

Williams-van Klinken et al. (2002a:40) mentioned three other classifiers: *tahan* lit. 'leaf' used with thin flat objects; *musan* lit. 'seed' used with very small round objects; and *lolon* lit. 'trunk' used with long cylindrical objects (see Table 29). There is another classifier used in Tetun Fehan which was lost in Tetun Prasa, namely *matan* lit. 'eye',

---

[141] Nouns of low countability (so called uncountable nouns like water, smoke, etc.) are also used with classifiers, but unlike sortal classifiers, mensural classifiers are semantically transparent common nouns, e.g. one glass of water (Gil 2013).
[142] To avoid confusion, *na'in* is, according to the Official orthography, attached to the postposed numeral with a hyphen, unlike the agentive affix *-na'in* which is attached to a preceding noun.

used with two groups of animals: buffalos and pigs. However, for the majority of nouns, there are no classifiers.

Classifier	Applied to	Example	Found in TP?
na'in *'owner'*	humans	*kolega na'in-rua* 'two friends'	yes
kain *'stem'*	long objects	*batar kain-rua* 'two husks of corn'	yes
fuan *'heart'*	roundish objects	*paun fuan-rua* 'two loaves of bread'	yes
tahan *'leaf'*	thin flat objects	*odamatan tahan-rua* 'two doors'	no
musan *'seed'*	very small round objects	*aimoruk musan-rua* 'two tablets'	no
lolon *'trunk'*	long cylindrical objects	*kilat lolon-rua* 'two guns'	no

Table 29 Classifiers in Tetun Prasa and other varieties (examples from Williams-van Klinken 2011 and my own corpus)

The reason why Tetun Prasa dropped most of the sortal classifiers might be the language contact. If we look at creoles and other restructured languages in the APiCS database, we see that this feature is marginal and is only found in four of them, including Sri Lankan Portuguese and Ambonese Malay[143] (Maurer and the APiCS Consortium 2013). The contact with Portuguese, which does not have sortal classifiers, and second-language acquisition by non-native speakers of Tetun Prasa might have caused that the less frequent classifiers became obsolete.

6.1.6   Quantifiers and numerals

Tetun Prasa has a whole range of quantifiers. Most of them are of native origin, like *hotu* 'all', *balu(n)* 'some', *uitoan* 'few' or *hira* 'several'. According to Hull (2005a), two of them are old Malay borrowings: *barak* 'many' from M. *banyak*, as in (147), and *lubun* (or Tetun Terik *lubuk*) from M. *lumpuk* 'group of', always followed by an indefinite

---

[143] The other two are Chinese Pidgin English and Gullah. Rubino (2012), in his paper on a noun classifier in Zamboagueño Chavacano, investigated the functions and semantic scope of 'bilung' which could have been influenced by southern Filipino adstrate languages.

article *ida* 'one', as in (148). However, similar forms are also found in other Austronesian languages, so the Malay connection is questionable.

(147) *problema*   *barak*
       problem   many

       'many problems' (0057TD_LAOB_30_M_Uni_Tet)

(148) *matéria*   *lubun*   *ida*
       subject   group.of   one

       'a group of subjects' (0053TD_JLS_44_M_Uni_Wai)

There is only one Portuguese quantifier: *kada* 'every, each' which is almost exclusively used in written language. It is mostly used with time references, as shown in (149), and has a native counterpart in a reduplicated form of the head noun, as shown in (150), also discussed in section 5.5.1. The use of this Portuguese quantifier is very marginal in spoken language. In my oral corpus, there are only 3 speakers (interestingly, all from Viqueque) who used it, although they often combined it either with a reduplicated head noun, as in (151), or with another quantifier *ida-idak* 'one.RDP', i.e. 'each', as in (152).

(149) *kada*   *fulan*
       every   month

       'every month' (STL_July_2013)

(150) *tinan* 'year' – *tinan-tinan* 'each/every year'

(151) *kada*   *tinan-tinan*
       each   year.RDP

       'each year' (0046TV_JCV_48_M_Sec_TT)

(152) *kada*   *uma-lisan*   *ida-idak*
       each   household   one.RDP

       'every household' (0039TV_MVJ_51_M_Uni_TT)

But when it comes to numerals, Portuguese and Indonesian borrowings are often preferred. Indonesian numerals are used mostly with time, as in (153a), percentages, as in (153b), and grades at school, as in (153c), followed by Indonesian nouns.

(153)   a. *dua menit* 'two minutes'

      jam tujuh tiga puluh '7:30 o'clock'

    b. *duabelas persen* 'twelve percent'

    c. *kelas empat* 'fourth grade'

Portuguese numerals are preferred for larger numbers like years and dates, as in (154), and ordinal numbers, as in (155).

(154) | *Entaun* | *iha* | *setenta* | *i* | *sinku,* | *dia* | *sete* | *de* |
       |----------|-------|-----------|-----|----------|-------|--------|------|
       | So | LOC | seventy | and | five | day | seven | of |

Dezembru	*komesa*	*Indonézia*	*mai*	*ona*	*ho*
December	start	Indonesia	come	ANT	with

aviaun.
plane

    'So in 75', on 7th of December, Indonesia started to arrive with planes.'
    (0039TV_MVJ_51_M_Uni_TT)

(155) | [...] | *primeiru* | *fillu* | *ne'e* | *eskola* | *iha* | *Colégio* | *Ossú,* |
       |-------|------------|---------|--------|----------|-------|-----------|---------|
       | | first.MASC | son | this | school | LOC | Colégio | Ossú |

segundu	*fillu*	*iha*	*Colégio*	*Ossú*	*iha*	*tempu,*
second.MASC	son	LOC	Colégio	Ossú	LOC	times

portugés	*nia*	*tempu.*
Portuguese	POSS	times

    [...] the first son studied in Colégio Ossú, the second son (studied) in Colégio Ossú in the Portuguese times. (0039TV_MVJ_51_M_Uni_TT)

6.2 The verb phrase

In this section, I will focus on new, emerging features, like the copula (6.2.1), passive voice (6.2.5), existential and other structures that allow VS order (6.2.6) as well as functional lexical borrowings like modal auxiliaries (6.2.2), and coordinating (6.3.1) and subordinating conjunctions (6.3.2 and 6.3.4). I will also discuss syntactic features of Tetun Prasa that have limited use is this variety, like serial verbs (6.2.4), but that are present in Tetun Fehan and/or Tetun Terik. As discussed in 5.2, verbs in Tetun Prasa, unlike verbs in Tetun Terik and Tetun Fehan, do not take any subject-marking inflections and there is no inflection to mark tense, person nor number. Instead, TAM markers are used to encode tense, aspect and mood (6.2.3).

6.2.1 Copula 'COP'

Tetun Prasa, as most languages in the Pacific region, uses a zero copula strategy with predicate nominals and attributive clauses. However, "quite often, the distribution of zero copulas within a language family or area is rather unpredictable. Languages which are closely related areally or genetically may differ considerably in the extent to which they allow zero encoding." (Stassen 2013). This is true especially for Tetun Fehan which contains a copula *ní* 'be' and two semi-copulas *tu'an* 'grow' and *dadi* 'become' (van Klinken 1999:182-183):

(156)   Tetun Fehan (van Klinken 1999:182):

Na	*fé-n*	*kbesik*	*á*	***ní***	*ha'u.*
3S	wife-GEN	direct	DEF	COP	1S

'His true wife is me.'

In the case of nonverbal predicates, when the predicate is nonreferential and consists of an indefinite noun, as in (157a), adjective, as in (158a) or an adverb, Tetun Prasa uses simple juxtaposition with no verbal element.

(157) Tetun Prasa:

    a.   *Nia   Ø   juis.*
            3S   Ø   judge.

            'She is a judge.'

    b.   *Nia   maka juis.*
            3S   FOC juis.

            '<u>She</u> is the judge.'

    c.   *Juis   maka nia.*
            judge FOC she

            'The judge is she.'

(158) Tetun Prasa:

    a.   *Nia   kolen.*
            3S   tired

            'He/she is tired.'

    b.   *Nia   maka kolen.*
            3S   FOC tired

            '<u>He/she</u> is tired.'

But when the nonverbal predicate consists of a definite noun, as in (157b), or a personal pronoun, as in (157c), the subject is followed by the definite contrastive marker *mak/maka*[144] 'FOC' (Hull 1993:29). This means that the contrasted constituent is interpreted as exclusive[145]. Dryer (2007:233) called this type of nominal predicates referential because the subject and the predicate refer to the same individual. Also, the difference between the two clauses is that in the second type, the subject and the predicate can be reversed. For this reason, Dryer (2007:233) referred to this type of clause as 'true

---

[144] According to Williams-van Klinken et al. (2002a:69), the longer form is preferred in writing but seldom used in speaking. On the other hand, Hull and Eccles (2005:98) argue that the shorter form *mak* is more common in the current written Tetun although they agree that there is a lot of variation suggesting that some authors prefer to use *maka* before a consonant and *mak* before a vowel. According to my oral corpus, *mak* is by far preferred in Tetun Prasa as well as in Tetun Terik, regardless of the following sound.

[145] The contrastive marker can also be used with adjectives (158b).

equational', since the change of word order only causes a possible difference in topic and focus.

In the case of affirmative equational clauses, it is possible to use the demonstrative pronoun *ne'e* 'this' after the subject that takes the place of a copula (Hull & Eccles 2005:108):

(159)   Tetun Prasa:

Akordu   ho   Austrália   ne'e   tranzitóriu.
agreement   with   Australia   this   transitory

'The agreement with Australia is transitory.' (Sapo_March_2016)

Although Tetun Prasa does not have a proper copula like the Tetun Fehan's *ní*, Williams-van Klinken (2010a) suggested there has been a new copula *hanesan* 'like, as' developing, possibly due to the Portuguese influence, that has become rather common especially in translations and newspapers, as illustrated by (160), (161) and (162). It is nothing unusual that the word serving as a copula is nonverbal and that it has grammaticalized while preserving also its original meaning (Dryer 2007:225-226).

(160)   Tetun Prasa:

korrupsaun   hanesan   krime ne'ebé mak   estraordináriu
corruption   like   crime REL   FOC   extraordinary

'corruption is an extraordinary crime' (STL_March_2016)

(161)   Tetun Prasa:

Atividade   ne'ebé ke   ita   halo   hanesan   ekonomi.
activity   REL   REL   1PI   make   like   economy

'The activity that we do is economy.' (0014TD_ACB_21_M_Uni_Mak)

(162)   Tetun Prasa:

Setór privadu   hanesan   komponente importante   ba
sector private   like   component   important   for

Governu.
government

'Private sector is an important part for the government.'
(GovernuTL_2010)

Another preposition with a similar meaning is *nu'udar* 'like, as', which is, according to Williams-van Klinken (2011) used mostly in formal situations and church language. However, I have found several instances of *nu'udar* in my written corpus, too:

(163) Tetun Prasa:

Ida-ne'e	*nu'udar*	*ezemplu*	*ida*	*klaru*	*hosi*
this.one	as	example	one	clear	from

Governu	*kona-ba*	*étika,*	*responsabilidade*	*no*
government	about	ethics	responsibility	and

transparénsia	*iha*	*governasaun.*
transparency	in	governance

'This is a clear example from the government about the ethics, responsibility and transparency in governance.' (GovernuTL_2010)

To understand how widespread the use of this emerging copula is, I elicited sentences with *hanesan* with the professors of Tetun at National Institute of Linguistics, showing them the examples from Williams-van Klinken (2010a) and my own corpus. Not only did they fail to recognize it as a copula, they found the examples that I presented 'sounding strange.'

6.2.2 Modal auxiliaries

Modal verbs, in general, express modality that can be either epistemic (concerned with degree of certainty of knowledge) or deontic (concerned with moral obligations or permissions) (Noonan 2007:138). They constitute a closed class and in my analysis, I treated them separately as function words. There are two modals in Tetun Prasa: *tenke* 'must' which can express obligation and *bele* 'can' which can express ability, possibility and permissibility.

To express obligation, all varieties of Tetun use the Portuguese loanword *tenke,* including *tenki,* (from Port. *tem que* + verb 'must; has to'), as shown in (164) and (165), or, rarely, *tende* (from Port. *tem de* + verb 'must, has to'), as shown in (166):

(164) Tetun Prasa:

Nia	*tenke*	*esplika*	*ne'e*	*di'ak.*
3S	must	explain	this	good

'He/she has to explain it well.' (0014TD_JBP_26_M_Uni_Mak)

(165) Tetun Terik:

Entaun	*sira*	*tenke*	*oho*	*ema*	*sira*	*ne'e.*
so	3P	must	kill	person	PL	this

'So they had to kill these people.' (0039TV_MVJ_51_M_Uni_TT)

(166) Tetun Terik:

[...]	*tende*	*uza*	*lian*	*ne'e*	[...]
	must	use	language	this	

'[...] (they) must use this language [...]' (0044TV_JCP_58_M_Sec_TT)

Bele 'can' can be used in an epistemic, as in (167), deontic, as in (168), and dynamic, as in (169), sense :

(167) Tetun Prasa:

Ita	*halo*	*ona*	*relatóriu*	*EITI*	*nian*	*no*	*ha'u*	*bele*
1PI	make	ANT	report	EITI	POSS	and	1S	can

dehan	*katak,*	*relatóriu*	*ne'e*	*hatudu*	*katak*	*la*	*lakon*	*dolar*
say	COMP	report	this	show	COMP	NEG	lose	dollar

ida	*iha*	*sistema*	*estrasaun*	*petróleu*	*nian.*
one	LOC	system	extraction	oil	POSS

'We wrote an EITI report and I can say that this report shows that we do not lose a (single) dollar in the oil extraction system.' = 'it is possible for me to say' (GovernuTL_2010)

(168) Tetun Prasa:

Ema   ne'ebé de'it   mak   bele   iha   asesu ba   informasaun
person which only   FOC   can   have   access to   information

ida-ne'e?
this.one

'What people can access this information?' = 'may/are permitted to'
(GovernuTL_2010)

(169) Tetun Prasa:

Entaun   ita   bele   uza,   ita   bele
so   2S.HON   can   use   2S.HON   can

hatene, ita   bele   komprende   didi'ak   […]
know   2S.HON   can   understand   RDP.good

'So you can use (the language), you can understand it, you can
understand it well […]' = 'you are able to'
(0051TD_DLB_45_M_Uni_Kem)

According to Hull (2005a:102), *bele* 'can' is derived from Malay *boleh* 'to be allowed'. Although the same word is used in Indonesian, the author believed that *bele* is indeed of Malay origin, since the meaning of Indo. *boleh* is restricted to 'may' and does not express ability (Hull 2001a:156). However, it is rather unlikely that this modal auxiliary was borrowed from Malay into Tetun as we can find similar Austronesian reflexes in other Timorese languages that were not in contact with Malay.

6.2.3   Verbal markers

Many languages that are analytical or isolating in nature make use of tense-aspect-mood (TAM) markers. These are:

"(...) operations that anchor or ground the information expressed in a clause according to its sequential, temporal, or epistemological orientation. Tense is associated with the sequence of events in real time, aspect with the internal temporal 'structure' of a situation, while 'mode' relates the speaker's attitude

toward the situation of the speaker's commitment to the probability that the situation is true" (Payne 2010:233-234).

Different language families prefer to mark different features. For example, according to Payne (2010:234), Austronesian languages pay less attention to mode or tense but more to aspect, "a device used to grammatically express different views of events in relation to their respective start and end points" (Velupillai 2012:208).

In Tetun Prasa, temporal information is usually specified by various expressions of time (e.g. *aban* 'tomorrow', *iha Segunda* 'on Monday'). But, when it comes to aspect, most verbs do not carry the aspectual meaning that is normally associated with their translation equivalent. As Williams-van Klinken et al. (2002a:73) attested, the sentence *nia di'ak* lit. 'he/she good' can mean both that the referent *is* good or *becomes* good. So in order to resolve this type of ambiguity, Tetun Prasa prefers to mark aspect.

TAM markers in Tetun Prasa are not grammatically obligatory. In natural speech, they are often left out and the meaning of the verb depends generally on the context (Williams-van Klinken et al. 2002a:73-74). There are five main aspects in Tetun Prasa:

- perfective (*tiha* = PFV);
- anterior/inchoative (*ona* =ANT);
- perfect (*tiha ona* = PRF);
- continuous (*hela* = CONT)[146];
- progressive (*daudaun* = PROG).

Apart from the aspectual category, Tetun Prasa also uses TAM markers to express future tense (*sei* = FUT), irrealis mood (*atu* = IRR), and other meanings. All TAM markers in Tetun Prasa are free morphemes derived from adverbs and auxiliary verbs (Hull 2001a:157). And although, originally, they were understood as content items, they have all been going through a process of grammaticalization. Many have already lost their original meaning and have become grammatical words (Eccles 1999:35).

---

[146] The continuous aspect *hela* was termed stative by Eccles (1998:39), indicating that "the participant in the action of the verb remains in a state produced by that action". However, I decided to follow the classification by Williams-van Klinken et al. (2002a:79), who defined *hela* as a continuous aspect that can be used with both stative and non-stative predicates. With the non-stative predicates, *hela* acts in the same way as the progressive aspect *daudaun*.

To give an example of the use of TAM markers in Tetun Prasa, I will discuss the three main aspectual markers *tiha* (PFV), *ona* (ANT) and *tiha ona* (PRF) that are often a source of confusion to non-native speakers (section 6.2.3.1). In section 6.2.3.2, I will explore various strategies of expressing future in Tetun Prasa, using a tense marker *sei* (FUT), a mood marker *atu* (IRR) and a verb *bá* 'go'.

6.2.3.1 *tiha* (PFV), *ona* (ANT) and *tiha ona* (PRF)

The perfective marker *tiha* and the anterior marker *ona* may seem very similar. In fact, some Timorese languages did not develop the distinction between these two markers (Hull 2001a:160). Mathijsen (1906:xi), studying Western Tetun, recognized both *tián/tianá* (in Tetun Prasa *tiha*) and *ona/na* (in Tetun Prasa *ona*) as having the same meaning 'already'. According to van Klinken (1999:238), the marker *ona* went on to develop a different meaning in Tetun Fehan: *onan* 'imminent'. The reason why some dialects do not make a distinction between these two aspects may be the fact that, although they are of different origin, they used to have the same meaning ('already').

The use of 'already' to mark the perfective aspect is very common in many world languages, including creoles. According to Hull (2001a:160-161), the anterior marker *ona* is a reflex of the Old Timoric **pena* 'already'. On the other hand, the author believed that *tiha* developed from Malay *telah* 'already' (*telah* > **tehal* > *tiha*). However, I find the Malay connection unlikely, since a similar marker also exists in other Timorese languages[147]. In Tetun Fehan, the perfective marker is *ti'an/ti'a/ta*. Although, originally, both markers used to have the same meaning, in the process of grammaticalization, they developed into two different aspectual markers in Tetun Prasa: *tiha* indicates a completion of a process while *ona* indicates the beginning of a new situation.

The perfective aspect *tiha* has to do with how we view an event – as a whole but without having knowledge about its internal structure. We see it in its totality, regardless of its durativity or internal structure. However, for an event to be viewed in its totality, it

---

[147] Hull (2001a:162) believed a similar reflex of M. *telah* existed also in other Timorese languages and dialects: *tel* in Aileu dialect of Mambae, *tala/tela/tel* in Southern Mambae, *taho/ta/te* in Isni, a dialect of Idalaka, and *ta* in Lakalei, spoken in the Manufahi district.

usually must come to an end and that is why the perfective aspect carries usually, but not always the past meaning (Velupillai 2012:210-211). According to Timberlake (2007:303), perfective aspect looks at whether events are bounded contextually and it differentiates the not-now from the now. When used with stative verbs or adjectives, the post-verbal marker *tiha* indicates that the subject has entered a certain state, as in (170). When used with non-stative verbs, it indicates that the activity happened and was completed in the past, as in (171). In the case of two grammatically coordinated clauses with the same temporal meaning, the aspect marker is used only after the second clause, as shown in (172) (Hull & Eccles 2005:122). However, in speech, it is also possible to find the marker after each verb phrase:

(170)   Tetun Prasa:

Sira	*hatene*	*tiha,*	*aprende*		*tiha*	*mas*	*depois*	*sira,*	*ikus*
they	know	PFV	learn		PFV	but	then	3P	last

sira	*haluha* […]
3P	forget

'They knew (Official Tetun), they learned (Official Tetun), but in the end, they forgot it […]' (0051TD_DLB_45_M_Uni_Kem)

(171)   Tetun Prasa:

Ha'u	*kose*	*tiha*	*nehan.*
1S	brush	PFV	tooth

'I brushed my teeth.' (0014TD_LEG_20_F_Sec_Mak_)

(172)   Tetun Prasa:

Sira	*na'ok*	*ami-nia*	*fahi,*	*oho*	*tiha.*
3P	steal	1PE-POSS	pig	kill	PFV

'They stole our pig and killed it.' (Hull & Eccles 2005:122)

Tiha is also used in backgrounded clauses and expressions meaning 'after that' (*Hotu tiha…* 'After…') (Williams-van Klinken et al. 2002a:74-75). In this case, *tiha* comes in the first clause. Although the authors claim these expressions are used by Tetun Terik speakers from the south, I have found many examples in Tetun Prasa as well:

(173) Tetun Prasa:

Han   hotu   tiha   ami   bele   deskansa   kalan.
eat   finished   PFV   1PE   can   rest   night

'Having finished eating, we can rest in the evening.'
(0014TD_LEG_20_F_Sec_Mak)

(174) Tetun Prasa:

Hotu   tiha,   matabixu   hotu   tiha,   hamoos
finished   PFV,   breakfast finished   PFV,   clean

uma   laran   no   dasa   fo'er   iha   li'ur.
house inside and   sweep dirt   LOC   outside.

'After I ate breakfast, I cleaned the house and swept the dirt outside.'
(0014TD_JMB_31_F_Sec_Mak)

As stated above, the perfective marker can also be used with a non-past meaning. *Tiha* can be used for commands/requests, as in (175), and when used in a negated clause, *la…tiha* changes its meaning to 'no longer', as in (176).

(175) Tetun Prasa (Williams-van Klinken et al. 2002a:74):

Soe   tiha   de'it!
throw   PFV   only

'Just throw it out!'

(176) Tetun Prasa (Williams-van Klinken et al. 2002a:74):

Nia   la   iha   tiha.
3S   NEG   EXIST PFV

'She is no longer here.'

The use of *ona* might seem very similar to *tiha* because they both mark a certain boundary: in the case of *tiha*, it is the completion of the process, but in the case of *ona*, it is the beginning of a new situation, regardless of the nature of the verb (dynamic or stative) (Eccles 1998:40). *Ona* can have the following meanings: 1. an action that is beginning now or has already begun; 2. a new resulting situation; and 3. a new situation, described by a stative verb, that has come into existence, irrelevant of whether the action takes place in the past, present or future (Eccles 1998:44). However, especially in

235

informal speech, speakers often use *ona* to express past tense in general. Again, this might be because they still recognize that *ona* has its own meaning ('already') (Williams van-Klinken et al. 2002a:75).

The marker *ona* has been termed 'anterior' by Williams van-Klinken et al. (2002a:75) and 'inchoative' by Eccles (1998:44), but also referred to as 'inceptive' (Hull & Eccles 2005:124) and 'ingressive' (Albuquerque 2011:105). In short, we could say that *ona* acts as an anterior marker (ANT) with dynamic verbs and as an inchoative marker (INCH) with stative (and sometimes dynamic) predicates. It is used in post-verbal position, although in transitive clauses it can either precede or follow the object noun phrase[148]:

(177)   Tetun Prasa:

    a.   *Sira   uza   ona   Tetun Prasa.*
           3P   use   ANT   Tetun Prasa

    b.   *Sira   uza   Tetun Prasa   ona.*
           3P   use   Tetun Prasa   ANT.

    'They have been using Tetun Prasa (with the outsiders).'
    (0054TD_NSG_51_M_Uni_TT)

When *ona* is used with non-stative verbs, the meaning of the clause is that of an event that has happened but has an effect on the present, as shown in (178). According to Bybee et al. (1994:61), "the goal of the utterance is not to locate a situation at some definite point in the past, but only to offer it as relevant to the current moment."

(178)   Tetun Prasa:

    Ha'u   hanoin   ha'u   hakerek   ona   disionáriu
    1S   think   1S   write   ANT dictionary

    badak kona-ba   lian   Midiki ho   lian
    short   about   language Midiki and language

    Waima'a
    Waima'a.

---

[148] In Literary Tetun, *ona* always precedes the complement (Hull & Eccles 2005:126).

'I think I wrote a short dictionary of Midiki and Waima'a language (and it's still written).' (0055TD_MAS_45_M_Uni_Mid)

When used with stative verbs and adjectives, the meaning of *ona* is that of a subject entering a certain state:

(179)   Tetun Prasa:

Tuir	*organizasaun*	*ne'e,*	*kontinente*	*amerikanu*	*iha*
follow	organization	this	continent	American	EXIST

ona	*suspeitu*	*kazu*	*vírus*	*Zika* [...]
INCH	suspect	case	virus	Zika

'According to this organization, the American continent has had suspected cases of the Zika virus.' (Sapo_March_2016)

Just like *tiha*, *ona* often occurs in backgrounded conditional and temporal clauses. When negated, the meaning is that of 'no longer, not again' (Williams-van Klinken et al. 2002a:76:

(180)   Tetun Prasa:

Ne'e,	*oras*	*ne'e*	*ema*	*la*	*uza*	*selamat*	*pagi,*
this	hour	this	person	NEG	use	blessed	morning

ema	*la*	*uza*	*ona.*
person	NEG	use	ANT

'Nowadays, people do not use *selamat pagi* ('good morning' in Indonesian), they do not use it anymore.'
(0054TD_NSG_51_M_Uni_TT)

According to Hull and Eccles (2005:124), *ona* can also be used as an inchoative marker indicating that an action is starting or is happening now, as shown in (181). However, when elicited, some speakers understood this sentence as simple past and as synonymous with *Sira oho **tiha** karau ne'e*. ('They killed the buffalo.').

(181)   Tetun Prasa (Hull & Eccles 2005:138):

Sira	*oho*	***ona***	*karau*	*ne'e.*
3P	kill	ANT	buffalo	DET

'They are killing the buffalo.' or 'They will kill the buffalo.'

Ona is further used in commands and invitations:

(182)   Tetun Prasa:

Bá   ona!
go   INCH

'Come on now!'

This use is very similar to Tetun Fehan's marker *onan* 'imminent' (183) which is most likely derived from the same word:

(183)   Tetun Fehan (van Klinken 1999:239):

Bá   m-ika   onan!
go   2S-back   IMM

'Go back now.'

Some markers in Tetun Prasa can cluster together. Both *tiha* and *ona* can cluster with the future marker *sei* 'FUT' (see 6.2.3.2) and *foin* 'just', but the most common combination is *tiha ona* 'PRF':

(184)   Tetun Prasa:

Nune'e,   ita   kria   tiha ona,   ANP   hodi
thus   1PI   create   PRF   ANP   in.order.to

regula   setór   ne'e   […]
regulate   sector   this

'Thus, we have created ANP [and it still exists] to regulate the sector [...].' (GovernuTL_2010)

(185)   Tetun Prasa:

Linguista   sira   iha   INL   hatuur   tiha ona.
linguist   PL   LOC   INL   set   PRF

'The INL linguists have set [the rule].'
(0055TD_MAS_45_M_Uni_Mid)

When the two markers *tiha* and *ona* are combined, they convey the perfect aspect, i.e. "the contextual occasion of a (present) perfect includes the here-and-now of the speech event and extends back, as a continuous interval, to include the actual event

reported by the predicate" (Timberlake 2007:289). In this case, both conditions, the perfective *tiha* and the inchoative *ona,* can be applied. The focus in on both the completion *tiha* and the resulting situation *ona*. In other words, it not only marks the verb as being completed, but also as having produced a situation that continues (Eccles 1998:49). It is, however, rarely used with states.

As my data show, aspect markers are not obligatory in Tetun Prasa as aspect and tense is mostly implicit. From the statistical point of view, the perfect marker *tiha ona* is rather marginal in spoken language while the marker *ona* is more frequent in both oral and written corpus, as Table 30 illustrates:

TAM marker	Oral corpus	Written corpus
tiha (PFV)	153	52
ona (ANT)	173	301
tiha ona (PRF)	18	40

Table 30 Frequency of TAM markers tiha, ona and tiha ona in written and spoken corpus

6.2.3.2 Future

In Tetun Prasa, future can be expressed in three different ways: with future marker *sei* (FUT), irrealis marker *atu* (IRR) and verb *bá* ('go'). In this section, I will refer to all three of them as future markers, although in general, I consider *bá* a lexical word and not a TAM marker.

Originally, there used to be only one future marker in Old Timorese, namely **umba* 'go', which later developed into Tetun Prasa *bá/ba* 'go/to'. As Payne (2010:237) argues, this is not surprising since future markers "often derive historically from free verbs meaning 'want', 'come' or 'go.'" According to Hull (2001a:167-168), it is possible that Tetun Prasa adopted *sei* 'FUT' from Malay *(ma)sih* 'still' and *atu* 'IRR' from Malay *untuk* 'for'. However, Aone van Engelenhoven[149] (p.c.) finds the possible Malay origin hardly imaginable in this case due to the unlikely sound change.

---

[149] Aone van Engelenhoven has dedicated his career to the study of historical linguistics (especially the proto languages of Austronesian and non-Austronesian languages in the SW Moluccas) and descriptive linguistics, publishing grammatical descriptions of minority languages in Indonesia and Timor-Leste.

The future marker *bá* 'go' is, according to some authors, a borderline case. Williams-van Klinken et al. (2002a:82) do not consider *bá* 'go' a future marker but treat constructions like (186) as SVCs because, as the authors say, "these constructions are used only when an actual act of going precedes the activity specified by the following verb", although this definition disagrees with the actual status of SVCs in which two verbs must constitute a single event. In the majority of cases, this construction can be translated in English as '*going to* + verb' or '*go and* +verb'.

(186)   Tetun Prasa (Williams-van Klinken et al. 2002a:82):

   Ha'u   bá   toba.
   1S   go   sleep.

   'I'm going to sleep.'

On the other hand, Eccles (1999:32-33) and Hull and Eccles (2005:137) do not consider similar constructions composed of *bá* 'go' + verb SVCs. They believe that *bá* 'go' developed into a preposition *ba* 'to' which can act as an 'approximative marker' conveying less sense of urgency than *atu* 'IRR' and suggests some physical proximity, as in (187). Historically, the marker developed from the verb OT **umpa* > TP *bá* 'go' into an unstressed preposition *ba* 'to'.

(187)   Tetun Prasa (Eccles 1999:33):

   Aban   ha'u   ba   selu   nia.
   Tomorrow   1S   to   pay   3S

   'I'm going to pay him tomorrow.'

Another way of expressing future, and probably the most prototypical one, is by using a preverbal marker *sei* 'FUT'. Hull and Eccles (2005:133) refer to *sei* as a prospective marker and distinguish two of its functions: durative with the meaning 'still', and future with the meaning of 'FUT'. Sometimes, there might be a certain ambiguity between these two meanings, especially when context is not provided:

(188) Tetun Prasa:

Ha'u   hanoin   sei   iha.
1S   think   FUT/still   EXIST

'I think that there will be.' or 'I think that there still are.'
(0055TD_MAS_45_M_Uni_Mid)

But Tetun is not the only language in which *sei* has two meanings. Parallel structures also exist in Mambae and other Timorese languages. This could be explained by the fact that the future marker possibly developed from 'still' because the verbal process is "still waiting to be accomplished" (Eccles 1999:30).

As Williams-van Klinken et al. (2002a:82) suggested, one way to resolve this ambiguity is to either add a future time phrase (e.g. *aban* 'tomorrow') or the serializing *bá/mai* 'go/come' to reinforce the future meaning of *sei* 'FUT', as in (189). In the case of *sei* 'still', it is possible to add the continuous aspect *hela* 'CONT' or *nafatin* 'continue', as in (190).

(189) Tetun Prasa:

Ne'ebé	*agora*	*seidauk*	*halo*	*hotu*	*mas*
REL	now	not.yet	make	finished	but

iha-ne'e	*knaar*	*INL*	*ne'e*	*para*	*bele*
here	duty	INL	this	so.that	can

aban-bainrua	*sei*	*bele*	*prodús*	*hotu*	*disionáriu*
in.the.future	FUT	can	produce	too	dictionary

monolíngua	*ba*	*dialetu*	*lokál*	*ou*	*lian*	*lokál*	*sira*
monolingual	for	dialect	local	or	language	local	PL

ne'ebé	*mak*	*iha,*	*eziste.*
REL	FOC	EXIST	exist

'It has not been done yet, but it is the INL's duty to also produce, in the future, monolingual dictionaries for local dialects and local languages that exist here.' (0051TD_DLB_45_M_Uni_Kem)

(190) Tetun Prasa:

Ami	*sei*	*buka*	**nafatin**	*solusaun*	*justu*	*sira*	*no*
1PE	still	search	CONT	solution	just	PL	and

pragmátiku	*sira*	*hodi*	*rezolve*	*ita-nia*	*diferensa.*
pragmatic	PL	in.order.to	resolve	1PI	difference

'We are still searching for just and pragmatic solutions to resolve our differences.' (Sapo_March_2016)

The ambiguity is also lost when the phrase is negated. In the case of *sei* 'FUT', the negative marker *la* 'NEG' is postposed, as shown in (191). In the case of *sei* 'still', the negative *seidauk* 'not yet' is used, as shown above in (189).

(191) Tetun Prasa:

Xina	*sei*	*la*	*para*	*atu*	*defende*	*ninia*
China	FUT	NEG	stop	in.order.to	defend	POSS

soberania.
sovereignty

'China will not stop to defend its sovereignty.' (STL_March_2016)

A preposed irrealis marker *atu* (IRR), almost exclusively used with non-stative verbs, indicates a desire, intention, imagination of future event and marks immediate future, i.e. when something is about to happen:

(192) Tetun Prasa:

Instituisaun	*hirak*	*ne'ebé*	*atu*	*kria.*
Institution	certain	REL	IRR	create

'These institutions that are to be created.' (GovernuTL_2010)

(193) Tetun Prasa:

Atu	*bá*	*ne'ebé?*	*Ha'u*	*dehan*	*atu*	*bá*	*Indonézia.*
IRR	go	where	1S	say	IRR	go	Indonesia

'Where are you going? I said I was going to Indonesia.' (0056TD_CS_50_F_Pri_Tet)

Unlike *sei* 'FUT', which marks future certainty, in case of *atu* 'IRR', which marks future possibility, the speaker is unsure whether the event will come about. On the other hand, the difference between *atu* 'IRR' and *bá* 'go' is that the intention in the case of *atu* 'IRR' is stronger than in case of *bá* 'go' (Hull & Eccles 2005:135). *Atu* and *bá* are not syntactically exclusive which gives us another reason not to consider *bá* a TAM marker but rather a lexical word.

The degree of future certainty/intention from the lowest to the highest:

bá ⟹ *atu* ⟹ *sei*

When used in clauses with past time reference, *atu* marks counterfactuality:

(194)   Tetun Prasa (Williams-van Klinken et al. 2002a:83):

Ha'u	***atu***	*bá*	*misa,*	*maibé*	*la*	*biban*	*ona,*	*tanba*
1S	IRR	go	mass	but	NEG	get.to	ANT	because

misa	*tama*	*tiha ona.*
mass	enter	PRF

'I meant to go to the mass, but can't do it anymore because the mass has already started.'

Eccles (1999:32) further states that when the phrase is negated, *la* 'NEG' is preposed: *la atu* + verb. This could, according to the author, mean that *atu* 'IRR' retains something of the quality of a verb, although it cannot stand independent of a following word and therefore is a grammatical word that has only now stopped being a content item. Contrary to Eccles's claims, I have not found an example of a preposed negator. In my corpus, *la* always follows *atu* in all varieties:

(195)   Tetun Terik:

Fó	*tiha*	*ba*	*hodi*	*hatudu*	*katak*	*sira*	***atu***
give	PFV	to	in.order.to	show	COMP	3P	IRR

la	*bele*	*fila*	*ba*	*sira-nia*	*rain.*
NEG	can	return	to	3P-POSS	country

'They gave them (the money) to show that they cannot return to their country.' (0050TV_JCV_48_M_Sec_TT)

Atu is often used in purpose and complement clauses with the meaning of 'to; so that' (196). According to Williams-van Klinken et al. (2002a:83-84), the presence of *atu* in these clauses is often syntactically optional and although it "may be part-way towards becoming a semantically bleached complementizer", it is not at that stage now. It is also not clear what the difference in meaning is in clauses with and without *atu*. In the example below, the phrases (196a) and (196b) have the same meaning. However, some speakers say there is a difference between *Ha'u hakarak bá Japaun* 'I want to go (but it's just my wish)' and *Ha'u hakarak atu bá Japaun* 'I want to go to Japan (and there is a possibility I'll go)' (Williams-van Klinken et al. 2002a:84).

(196) a. *Iha   tempu ne'ebá la   fásil   **atu**   enfrenta,   la*
           LOC   time   that   NEG   easy   to   face   NEG

           *fásil **atu** enfrenta   moris.*
           easy   to   face   life.

           'In those times, it wasn't easy to face life.'
           (0057TD_LAOB_30_M_Uni_Tet)

       b. *Iha   tempu ne'ebá la   fásil   enfrenta,   la   fásil*
           LOC   time   that   NEG   easy   face   NEG   easy

           enfrenta   moris.
           face   life.

           'In those times, it wasn't easy to face life.'

6.2.4   Serial verbs

Another characteristic of languages with little or no verbal morphology are serial verbs. They are frequently found in languages of Southeast Asia, Oceania, New Guinea, West Africa but also in creoles (Aikhenvald 2006:1). A serial verb construction (SVC) can be defined as a sequence of two or more verbs that are not compounded and share the following characteristics: they are part of the same clause; they share the same intonation contour and grammatical features (e.g. negation, aspect and tense); neither of the verbs is

subordinate or modifies the other; and they have at least one shared argument[150]. A SVC has a single illocutionary force and it is interpretable as a single, complex event (Williams-van Klinken et al. 2002a:49).

In Tetun Prasa, there are only SVCs containing two verbs. Hajek (2006a:242), however, also identified some multipart SVCs which "consist of iterations of nested asymmetrical serial verb constructions involving motion-direction, for example [[*monu tun*] *mai*] lit. 'fall.descend.come' meaning 'fall down (this way)'."

Based on the composition, Aikhenvald (2006:3) divided SVCs into symmetrical and asymmetrical constructions. Symmetrical SVCs contain two or more verbs from a semantically and grammatically unrestricted class while one of the verbs in asymmetrical SVCs is from a semantically and grammatically restricted class. Tetun Prasa does not have any symmetrical SVCs, because they were all lost in the process of lexicalization and developed an idiomatic meaning (Aikhenvald 2006:34). As an example, Hajek (2006a:242) lists *hanoin hetan* lit. 'think.find' meaning 'to remember' which appears to be symmetrical but is often subject to ellipsis (just *hanoin*) and many speakers do not understand its meaning. Symmetrical SVCs are neither very prominent in Tetun Fehan, which has more extensive serialization, so Hajek (2006a) decided to treat the few examples in Tetun Prasa as fully lexicalized. Unfortunately, there are no studies of serial verb constructions in Tetun Terik so we can only hypothesize whether the number of SVCs in Tetun Terik is higher than in Tetun Prasa due to the limited contact with Portuguese.

Based on the contiguity, Williams-van Klinken et al. (2001:49-50) divided SVCs in Tetun Prasa into two types:

a. nuclear (i.e. contiguous) SVCs, which are tightly-bound sequences whereby no element can intervene between the two verbs (for example 'motion-direction serialization', as in (197), '*halo* and *fó* causative serialization', as in (198).

---

[150] Some authors tried to extend the definition of serial verb construction also to the SVCs in which the two verbs do not share an argument. Crowley (1987:40) calls this 'ambient serialization'. For the discussion against the extension of the concept of verb serialization in Tetun Prasa, see Williams-van Klinken (2008).

(197) Tetun Prasa:

*Pois, nia **mai tama**, la iha esplikasaun* [...]
thus 3S come enter NEG EXIST explanation

'So he entered, without any explanation […]'
(0057TD_LAOB_30_M_Uni_Tet)

(198) Tetun Prasa:

*Depois ha'u **fó-hatene** ba ha'u-nia ria:*
then 1S give-know to 1S-POSS brother-in-
law

"Markus, di'ak liu ita rua fila ona."
Markus good more 1PI two return INCH

'Then I informed my brother-in-law: "Markus, we better go back."'
(0014TD_ACB_21_M_Uni_Mak)

b. core layer SVCs (i.e. non-contiguous) which allow for each verb to introduce
its own argument or an adverb that can stand between the two verbs (for example 'motion-
action serialization', as in (199), 'motion-direction serialization', as in (200), and
'instrument serialization').

(199) Tetun Prasa:

[...] *tanba ha'u hatene katak kolega sira atu **mai***
because 1S know COMP friend PL IRR come

***halimar** iha-ne'e.*
relax LOC-this

'[…] because I knew that the friends were coming to relax here.'
(0014TD_JMB_31_F_Sec_Mak)

(200) Tetun Prasa:

*Nia bá Manatuto to'o **fila fali mai** uma la*
3S go Manatuto arrive return again come house NEG

fó-hatene ha'u [...]
give-know 1S

'She went to Manatuto, returned back home and didn't tell me […]'
(STL_March_2016)

In case of Tetun Fehan, van Klinken (1999:255) also distinguished between nuclear and core layer serialization. Apart from the parameters presented above for SVCs in Tetun Prasa, there is one distinguishing feature characteristic for Tetun Fehan. In this variety, the first verb in SVCs takes subject marking; the subsequent verbs do not take subject marking in nuclear serialization, as shown in (201), but do in core layer serialization, as shown in (202), if the phonology allows (van Klinken 1999:251):

(201)   Tetun Fehan (van Klinken 1999:257):

Ó   *mai,*   **_má_**   **_sít_**   *nú*   *kain*   *ne'e*
2S   come   2S.eat   be.cut.off   coconut   stalk   this

lain.
FIRST

'You come and chew through this coconut stalk now.'

(202)   Tetun Fehan (van Klinken 1999:260):

Nia   **_simu_**   **_nola_**   *sala*   *ne'e.*
3S   receive   3S.take   fine   this

'He accepted this fine.'

In Tetun Terik, verbs in the second slot can take subject marking also in nuclear serialization:

(203)   Tetun Terik from Viqueque:

No   *ó*   *bá,*   *ó*   *tenke*   **_fó-matene_**   *ba*   *ema*   *sira*
and   2S   go   2S   must   give-2S.know   to   person   PL

ne'ebé   *te'in*   *masin*   *iha*   *tasi* [...]
REL   cook   salt   LOC   sea

'And you go, you must inform the people who are making salt at the sea […]' (0046TV_JCV_48_M_Sec_TT)

When compared with Tetun Fehan, the range, type, and frequency of serial verb constructions in Tetun Prasa is much more restricted. This is due to two ongoing processes

in SVCs, namely deserialization, discussed in section 6.2.4.3, and grammaticalization[151]. But first I will look closer at two types of innovative asymmetrical serialization: facilitative (*fó* + major verb) and causative (*halo* + major verb) that are also discussed, in general, in Baxter (2009:68). Both of them are valency increasing processes.

6.2.4.1 Facilitative serialization *fó* + verb

The facilitative construction using *fó* 'give' is an interesting case of serialization and deserialization at the same time. According to Williams-van Klinken (2010b), the strategy with '*fó* + verb', seems to be rather prominent in Tetun Prasa but rare in Tetun Fehan and Tetun Terik. In this kind of construction, *fó* 'give' occupies the first position and introduces a recipient. In the case of nuclear serialization, the construction with *fó* acts as a compound since it is lexically restricted (different verbs follow different causation strategies)[152] and the two verbs cannot be separated (Williams-van Klinken et al. 2002a:100). Core serialization with a transitive verb (*fó* + O + ba + Vt) is also possible.

This causative serialization changes the valency of the verb in the second slot, so intransitive verbs turn into transitive and transitive verbs into either transitive or ditransitive:

(204)   Vi → Vt: *hariis* 'bathe' → *fó-hariis* lit. 'give.bathe', i.e. 'to bathe sb.'

Vt → Vt: *han* 'eat' →'*fó-han* lit. 'give.eat', i.e. 'to feed sb.'

Vt → Vdt: *empresta* 'borrow' → '*fó-empresta'* 'give.borrow', i.e. 'to lend'[153]

---

[151] The opposite process to the deserialization of symmetrical SVCs, that has happened in Tetun Prasa but also in other varieties, is the grammaticalization of the minor verbs (from limited set of possibilities) in asymmetrical SVCs. Some of the minor verbs that are going through the process of grammaticalization are transitive verbs *lori* and *hodi*, both 'carry, bring, take' used in core layer instrumental SVCs.

[152] Causative constructions in Tetun Prasa can be formed in different ways but usually there is only one possible strategy for each verb. Apart from *fó* serialization, it is possible to express causation using *halo* 'make, do' or a derivational prefix *ha-* (see section 5.3.1). Although it might seem that *ha-* is a grammaticalized form of *halo*, in fact, *ha-* is a reflex of Proto-Austronesian causative prefix *pa-* (Hull 2001a:147, Hajek 2006a:250).

[153] In Portuguese, *emprestar* is always ditransitive and cannot be used in the sense of 'borrow'. There seems to be no single word verb with the meaning of 'borrow'.

In my oral corpus of Tetun Prasa, I have found only few examples of *fó* + native verb serialization (see Table 31, not in bold), which contradicts the assumptions of Williams-van Klinken (2010b) that this construction is very prominent in this variety. On the other hand, it can be confirmed that serialization with *fó* + native verb is also limited in Tetun Terik:

SVC	Translation	Written	TP	TS	TV
fó-sai	announce	51		7 (L2)	
fó-hanoin	instruct, remind	19			
fó-hasoru	greet	1			
fó-fiar	have faith in	1			1
fó-agradese	give thanks	3			
fó-hatene	inform	5	4	1	1
fó-fila fali	return	4			
fó-sala	recompense	7			
fó-han	feed	1		1	2
fó-tuku	punch	1			
fó-fila	give back, return	2			
fó-komenta	comment		1		
fó-tuir	to make follow		1		
fó-volta	do a U-turn	1			

Table 31 Distribution of 'fó' SVCs in the written and oral corpora

Some *fó* constructions have Portuguese roots in the second slot. A few of them are verbal (see Table 31, in bold) but many are nominal and thus cannot be considered SVCs. Still, it is very interesting to notice that Tetun Prasa often prefers to adopt Portuguese nouns rather than verbs and then derive causative verbs by using *fó* + noun, e.g. *apoiu* (n.) 'support' – *fó-apoiu* (v.) 'to support' instead of *apoia* (v.) 'to support' (from Port. *apoiar* 'to support'):

(205)   Tetun Prasa:

[...]	*ami*	*hakbesik*	*an*	*ba*	*ema*	*doadores*	*sira,*
	1PE	approach	self	to	person	donor.PL	PL

kompañia	*balun*	*atu*	*bele*	*fó*	*apoiu*	*ba*
company	some	IRR	can	give	support	to

ami-ninia	*grupu*	*iha*	*sektór*	*agrikultura*	*nian*
1PE-POSS	group	LOC	sector	agriculture	POSS

'We approached the donors and some companies that could support our agricultural group.' (0006TD_JBP_26_M_Uni_Mak)

In some registers, e.g. journalistic and technical texts, '*fó* + noun' constructions, as shown in (206), give way to single word Portuguese borrowings, as documented by (207). However, compared to the oral corpus, SVCs are still very frequent in written texts (see Table 31).

(206)   Tetun Prasa:

[...]   *tenke   fó   kontribuisaun hodi   hametin*
        must   give   contribution   in.order.to   strengthen

liután   relasaun   bilterál.
even.more   relation   bilateral

'[…] must contribute to strengthen even more the bilateral relations.' (Sapo_March_2016)

(207)   Tetun Prasa:

Tenke   kontribui   hodi   promove   solusaun   foun
must   contribute   in.order.to   promote   solution   new

sira",   nia   afirma   ona.
PL   3S   affirm   ANT

'"(We) must contribute to promote new solutions", he affirmed.' (STL_March_2016)

According to Williams-van Klinken (2010b), a possible source of this strategy in Tetun Prasa is Malay or the expansion of the low-frequency Tetun Terik construction. At first glance, the Malay connection might seem valid but Indonesian language does not have serial verbs or periphrastic constructions with 'give'. This type of serialization, in which 'give' occupies the first position and introduces a recipient is only found in a limited number of creoles: Malacca Creole Portuguese, as in (208), Batavia/Tugu Creole Portuguese, as in (209), Macau Creole Portuguese, as in (210), Ambon Malay, as in (211), Indo-Portuguese of Daman and Diu, and three varieties of Chabacano (Maurer and the APiCS Consortium 2013). All of them are spoken in the South, East or Southeast Asia region and some of them (Malacca Creole Portuguese, Batavia Creole Portuguese, Macau

Creole Portuguese and Ambonese Malay) have Malay as one of the contributing languages.

(208) Malacca Creole Portuguese (Baxter 2009:72):

Nu	*lo*	***dá***	***kumí***	*ku*	*olotu*	*aros*
1P	FUT	give	eat	DAT	3P	rice

'We will feed them rice.'

(209) Batavia Creole Portuguese (Maurer 2011:73-75):

Isti	*belu*	***da***	*sabe*	*kung*	*ile*	*ki*	*esta*
DEM	old.man	give	know	OBJ	3S	COMP	DET

teng	*lugar*	[...]
COP	place	

'This old man told him that this was the place (...).'

(210) Macau Creole Portuguese (Barreiros 1943/44:454 cited in Baxter 2009:76):

Mestê	***dá***	***come***	*pâ*	*unga*	*casa*	*intêro.*
must	give	eat	AC	one	house	entire

'Must feed an entire household.'

(211) Ambonese Malay (Adelaar 2005:215):

Kase	***makaŋ***	*ana-ana*	*tu*	*dolo*,	*doŋ*
give	eat	RDP-child	ANAPH	first	3P

su	*lapar*
already	hungry

'Feed the children, they are hungry.'

However, this construction is confined to the 'pidgin-derived Malay varieties' and is not found in Proto-Malayic (Adelaar 2005:12). Unfortunately, Adelaar (2005) did not explain where this construction originated and if it could be explained on the basis of substrate languages or language contact. Baxter (2009:73-74) tried to shed more light on the possible origin of this construction. He noticed that the same structure was used in colloquial vernacular Malay and its varieties but also in Hokkien language which could

have influenced 'Lingua Franca Malay' and, subsequently, Malacca Creole Portuguese. The author further considered the possible influence of Portuguese, in which certain structures with 'give' (*dar a[154]* + verb) have partially similar semantics. This is the more probable explanation for the existence of this causative structure in Indo-Portuguese of Daman and Diu. However, in Portuguese, the choice of the verb in the second slot is restricted.

6.2.4.2 Causative construction *halo* + verb

Another innovative SVC in Tetun Prasa is the periphrastic causative construction with *halo* 'do, make' which allows more syntactic and semantic possibilities (Williams-van Klinken et al. 2002a:98). According to Hajek (2006a:252) this might be due to the influence of Portuguese in which a fuller range of variation of word order can be observed. In the case of the causative construction in European Portuguese, the object can either precede the second verb, as shown in (212a) and (212b), or follow the second verb, as shown in (212c). However, the last option is relatively uncommon.

(212)   Portuguese (Lima-Salles & Pilati 2014:203-205):

a.	*A*	*mãe*	***fez***	*o*
	the.FEM	mother	make.PRF	the.MASC

menino	***dormir****.*
boy	sleep

b.	*A*	*mãe*	***fê-lo***	***dormir****.*
	the.FEM	mother	make.PRF-3S.ACC	sleep

c.	*A*	*mãe*	***fez***	***dormir***
	the.FEM	mother	make.PRF	sleep

o	*menino*.
the.MASC	boy.

'The mother put the boy to sleep.'

---

[154] Baxter (2009:75) suggested that the preposition *a* 'to' could have been either absorbed by the final vowel [a] in rapid speech in certain forms or it could have been omitted by second-language acquisition.

In Tetun Prasa, both orders are possible:

(213)   Tetun Prasa (Williams-van Klinken et al. 2002a:98):

    a.   *Moras*   *AIDS* **halo** *ema*   *barak* **mate**.
            disease   AIDS make person many die

    b.   *Moras*   *AIDS* **halo** **mate** *ema*   *barak.*
            disease   AIDS make die   person many

‘The AIDS illness has caused many people to die.’

The same strategy is used in Malacca Creole Portuguese, as in (214), Macau Creole Portuguese, as in (215), and Batavia/Tugu Creole Portuguese, as in (216), but also in Indo-Portuguese of Daman and Diu, as in (217), and Sri Lanka Creole Portuguese, which would support the Portuguese connection. However, this strategy is even more prominent in Malay, which makes use of verbs *buat* ‘make’ and *bikin* ‘make’ to form causative serialization, as shown in (218). The same structure has been observed also in restructured varieties of Malay (Baxter 2009:70-72). But Malay influence would not explain the presence of this structure in the South Asian Creole Portuguese, where substrate influence can be disqualified since Indian and Sri Lankan languages use suffix to form direct causation. We can also find this type of serialization in certain varieties of Chabacano, which have Spanish superstrate and Tagalog substrate. Most of the structures with *ase* ‘make’ resemble the Spanish ones, but there are examples that do not have a parallel in the Romance language, which could, again, point to the influence of the Austronesian substrate.

(214)   Malacca Creole Portuguese (Baxter 2009:70):

    a.   *Bos*   *já*   **fazé** **kai**   *ku*   *eli*
           2S   PFV   make fall   AC   3S
           ‘You made him fall.’

    b.   *Bos*   *já*   **fazé** *ku*   *eli*   **kai**
           2S   PFV   make AC   3S   fall
           ‘You made him fall.’

(215) Macau Creole Portuguese (Barreiros 1943/44: 33 cited in Baxter 2009:71):

Aquelle	*tentação*	*de*	*animal principiá*	*corê* ...[...]...
DEM	temptation	of	animal begin	run

de	*sorte*	*que*	*já*	***faze***	***ri***	*tudo*	*aquelle*
of	luck	that	PFV	make	laugh	all	DEM

gente	*na*	*Praia*	*Grande.*
people	LOC	Praia	Grande

'The cursed animal began to run ... [...] ... so that it made all the people on the Praia Grande esplanade laugh.'

(216) Tugu Creole Portuguese (Schuchardt 1891:48 cited in Baxter 2009:71):

Mas	*dianti*	*nos*	*machika*	*aka*	*neli*	***fai***	***sai***	*suwa*
more	before	1P	tread	DEM	rice	make	exit	POSS

gaba	*deri*	*suwa*	*pau.*
kernel	from	POSS	stalk

'First, we tread the rice [and] make its kernels come out of its stalks.'

(217) Diu Indo-Portuguese (Cardoso 2009:237):

Yo	***fez***	***fuj-i***	*pə*	*peacock.*
1S	make.PST	flee-INF	DAT	peacock

'I made the peacock run away.'

(218) Malay (both examples Baxter 2009:72):

a.
Aku	***buat***	***jatuh***	*sama*	*dia.*
1S	make	fall	AC	3S

'I made him fall over.'

b.
Aku	***buat***	*sama*	*dia*	***jatuh.***
1S	make	AC	3S	fall

'I made him fall over.'

As we can see from Table 32, there are many varieties of Asian Creole Portuguese and Creole Spanish that express causation with a serial verb construction. Although they

all have a Romance superstrate, there is evidence of significant substrate influence as well. Baxter (2009:86) believed that these structures were present:

> "(...) in the contact environments in earlier stages in these communities on the Portuguese Asian trade networks, originating in the interaction of superstrate, foreign-talk and early contact varieties based on Portuguese that functioned in the Indo-Portuguese contact. Subsequently, with the development of the Portuguese trade network in an easterly direction, and with creolization and stabilization of these contact varieties in different multilingual settings, the SVCs that had parallels in the various local substrates, received reinforcement." (Baxter 2009:86).

It is thus possible that Tetun Prasa adopted these SVCs from the restructured varieties of Malay and Portuguese that were used in the region.

Serial V	Diu	Daman	Ceylon	Malacca	Batavia	Tugu	Macau	Cavite, Ternate	Zamboanga
Causative	X	X	X	X	X	X	X	X	X
Facilitative	X	X	X	X	X	X	X	X	X

Table 32 Causative and facilitative verb serialization in Asian Creole Portuguese and Philippine Creole Spanish (Baxter 2009:86)

6.2.4.3 Deserialization

Tetun Prasa is going through a process of deserialization of SVCs, as pointed out by Hajek (2006a:251). This is exactly the opposite of what has been happening in other geographical areas. According to Aikhenvald (2006:52), verb serialization "as a grammatical feature tends to diffuse. (…) Languages with SVCs tend to form areal clusters." SVCs are indeed an areal feature in Timor-Leste and Eastern Indonesia, and it seems like local varieties of Malay also adopted this structure, like the causative constructions in Melayu Tenggara Jauh (van Engelenhoven 2002:185-186). But language contact can also cause the loss of SVCs. As one of the most significant factors of

deserialization, Hajek (2006a) lists the long-term contact with Portuguese and the large influx of Portuguese borrowings which tend to replace the native SVCs, especially in high registers of Tetun Prasa. So, for example, instead of the SVC *fó-hatene* 'inform' (lit. 'give know'), as shown above in (198), Tetun Prasa adopted a one-word Portuguese borrowing *informa* (from Port. *informar* 'inform'):

(219) Tetun Prasa:

[...]	*Jacinto*	*Rigoberto*	*informa*	*katak,*	*número*
	Jacinto	Rigoberto	inform	COMP	number

dezmobilizasaun	*ba*	*tinan*	*ida-ne'e*	*hamutuk*	*219.*
demobilization	to	year	this.one	together	219

'[…] Jacinto Rigoberto informed (somebody) that the number of demobilizations for this year is 219.' (STL_July_2013)

The second important factor that has led to the deserialization is "the rise of an independent Adjective class, also triggered by the influx of loans" (Hajek 2006a:252). In Tetun Fehan, adjectives are treated as verbs, also because of their tendency to act predicatively. According to van Klinken (1999:52), 60% of adjectives in Tetun Fehan are used predicatively as a subject complement. However, due to the influence of Portuguese and the noun-like nature of Portuguese adjectives, Hajek (2006a:252) does not consider Tetun Prasa causative constructions with adjectives as one of their elements SVCs.

6.2.5   Passive voice

As I have mentioned earlier, Tetun Prasa can be grouped together with other preposed possessor languages which differ from the symmetrical voice languages in the way they express voice. They "either do not show any grammaticized voice alternations at all or the voice alternations are clearly asymmetrical" (Himmelmann 2005:114). Historically, Tetun Prasa, like most of the Austronesian languages in central and eastern Indonesia, did not have a passive voice (Klamer 2002:371). Presently, there are various possibilities to express passive meaning, or rather, the demotion of agent or actor argument in transitive constructions in Tetun Prasa.

In this thesis, I will follow the work of Keenan and Dryer (2007) who divide passive constructions into basic and non-basic. One of the basic passive strategies is to use the basic form[155] of the verb following the object:

(220)   Tetun Prasa (Hull 1993:52):

Ikan	*ida-ne'e*		*fa'an*	*iha-ne'ebé?*
fish	this.one		sell	LOC-where

'Where is this fish sold?'

This strategy might resemble anticausative constructions (or pseudo-reflexives in Portuguese), however, as documented by the example (220), the main verb is, in its basic form, transitive. This strategy also fulfils two other characteristics of 'basic passives': there is no agent phrase and the main verb in its basic form expresses an action in which the subject is the agent and the object is the patient. In this case, the subject of the active voice is simply eliminated and the verb is not distinctively morphologically marked (Keenan & Dryer 2007:328-330). The fact that these passive constructions usually appear without agent phrases is not uncommon. There are languages, for example Latvian, that only allow agentless passives and then there are languages that prefer agentless passives although the agented ones are fully grammatical (Keenan & Dryer 2007:331-332).

A more natural solution to introduce the agent, without actually using the agent phrase, would be to use the emphatic construction with the OSV order:

(221)   Tetun Prasa:

Ikan	*ida-ne'e*	*vendedór*	*ida*	*mak*	*fa'an.*
fish	this.one	seller	one	FOC	sell
O	S				V

'This fish is sold by a seller.'

This strategy is also described in Silva (1889:5) with a parallel construction in Tetun Terik which uses *ha'e* 'FOC':

---

[155] In languages that distinguish between active and passive form of a verb, the active form is used.

(222)   Tetun Terik (original spelling, Silva 1889:5):

    a.   *Maromac*   *nala*   *laléhan*
             God   3S.do heaven
             S   V   O

             'God created heaven.'

    b.   *Laléhan*   *Maromac*   *há'e* *nala*
             heaven   God   FOC   3S.do
             O   S   V

             'Heaven was created by God.'

For passive meaning, it is also possible to omit the agent of the transitive verb and shift the patient before the verb:

(223)   Tetun Prasa (Hull & Eccles 2005: 116-117):

    a.   *Ha'u* *lakon* *karteira.*
             1S   lost   wallet

             'I lost the wallet.'

    b.   *Ha'u-nia*   *karteira*   *lakon.*
             1S-POSS   wallet   be.lost

             'My wallet got lost.'

Tetun Prasa also possesses so called 'strict morphological passive' (Keenan & Dryer 2007:333), using a prefix *nak-* (see section 5.3.2) which turns transitive verbs into intransitive ones with a passive meaning (Hull 1993:183-184):

(224)   Tetun Prasa (Hull 1993:183):

    a.   *Sira* *sobu*   *uma.*
             3P   pull.apart   house

             'They destroyed the house.'

    b.   *Uma* *ne'e* *naksobu.*
             house this   come.apart

             'This house is destroyed.'

However, not all Tetun verbs take the detransitivising prefixes when used with passive meaning. Based on the analysis of my corpus, it seems that this strategy is very little productive in Tetun Prasa. Portuguese loanwords never take the prefix *nak-* but are instead used in the basic active form (Hull 1993:184):

(225) Tetun Prasa (Hull 1993:184):

Osan   hotu   sei   gasta.
money all   FUT   spend.INF

'All the money will be spent.'

Compare with the Portuguese passive construction:

(226) Portuguese:

Todo   o   dinheiro   será   gasto.
all   the   money   be.FUT   spend.PPP

'All the money will be spent.'

Some Portuguese loans are used in constructions with passive meaning with an auxiliary verb *hetan* 'get, find' (Hull 1993:185). This verb of reception is often used in so-called 'periphrastic passive' constructions in which the second element is often not the usual transitive verb, as shown in (227), but, in the majority of cases, its nominalized form, as illustrated by (228) (Keenan & Dryer 2007:337):

(227) Tetun Prasa:

Ministru   Ekonomia   no   Dezenvolvimentu   hetan
minister   economy   and   development   get

simu   hosi   Dr.   Basílio   Horta […]
receive   by   Dr.   Basílio   Horta

'Minister of Economy and Development was received by Dr. Basílio Horta [...]' (GovernuTL_2010)

(228) Tetun Prasa: *rekoñese* (v.) 'recognize, acknowledge'

Ita   hamutuk   besik país   30   mak   luta
1PI   together   close country   30   FOC   fight

hela atu   hetan rekoñesimentu   ba   ita-nia   naran
CONT IRR   get   recognition   to   1PI-POSS   name

di'ak.
good

'We have, together with almost 30 countries, been fighting to be recognized for our good name' (GovernuTL_2010)

When a passive clause includes an agent phrase, Keenan and Dryer (2007:342) talk about so-called 'non-basic passives'. Agent phrases are not an integral part of passive constructions for three reasons: many languages do not permit agent phrases; agent phrases occur also in non-passive constructions; and when they are present in active voice, they take the form of an independently existing oblique NP. Agent phrases are not common in Tetun Prasa, but they are possible. In my example (229), the actor could come at the end, introduced by *hosi* 'by':

(229) Tetun Prasa:

Ikan	*ida-ne'e*	*fa'an*	*hosi*	*vendedór*	*ida*
fish	this.one	sell	by	seller	one

'This fish is sold by a vendor.'

This type of construction was described by Williams-van Klinken (2010a:184), who suggested that it was an 'incipient passive-like construction' found in certain registers of Tetun Prasa. This passive has no morphologically marked verb nor is there any auxiliary, and the actor can be introduced by a preposition *hosi* 'from, by':

(230) Tetun Prasa:

Kongregasaun	*IS-MAIK*	***funda***	*iha*	*Dili*	*tinan*	*1989,*
congregation	IS-MAIK	found	in	Dili	year	1989

hosi	*Irmã*	*Maria Lordes*	*"Mana Lu"* […]
by	Sister	Maria Lordes	"Mana Lu"

'IS-MAIK congregation was founded in Dili in 1989 by Sister Maria Lordes "Mana Lu" […]' (Timor Post 24/08/2015, pg. 2)

This passive-like construction very much resembles the passive construction in Portuguese and it is believed that it has indeed been calqued from Portuguese, as shown in (231). This kind of passive-like constructions in Tetun Prasa can be mostly found in media language and official documents – journalists and authors tend to translate literally

the passive constructions from Portuguese, English, as in (232), but also from Indonesian, as in (232).

(231) Portuguese:

A	*congregação*	*IS-MAIK*	**foi**	**fundada**
the.FEM	congregation	IS-MAIK	be.PST	found.PPP

em	*Dili*	*no*	*ano de 1989*	**pela**	*Irmã*
in	Dili	in.the.MASC	year of 1989	by.the.FEM	Sister

Maria	*Lordes*	*"Mana Lu"* […]
Maria	Lordes	"Mana Lu"

(232) English:

The	*congregation*	*IS-MAK*	**was**	**founded**	*in*
the	congregation	IS-MAK	be.PST	found.PPP	in

Dili	*in*	*1989*	**by**	*Sister*	*Maria Lordes*	*"Mana Lu"* […]
Dili	in	1989	by	Sister	Maria Lordes	"Mana Lu"

(233) Indonesian:

Kongregasi	*IS-MAK*	**_didirikan_**	*di*	*Dili*	*pada*	*tahun*
congregation	IS-MAK	PPP.found	in	Dili	in	year

1989	**oleh**	*Suster*	*Maria Lordes*	*"Mana Lu"* […]
1989	by	Sister	Maria Lordes	"Mana Lu"

Occasionally, the Portuguese past participle is used which presupposes that the speaker is proficient in Portuguese. However, Hull and Eccles (2005:119-120) do not consider these forms real past participles but rather attributive adjectives, as shown in (234), since they are not used in passive constructions with a complement clause:

(234) Tetun Prasa:

Prezidente	*nia*	*biografia*	**_autorizada_**.
president	3S.POSS	biography	authorized

'The president's authorized biography.'

Contrary to Hull and Eccles' (2005) claims, I have found many examples of complement clauses following the past participle in the written language of daily newspapers:

(235)   Tetun Prasa:

*Tinan   rua   hafoin   **nomeadu**   hanesan   xefe   forsa*
year   two   after   name.PPP   like   chief force

defeza   australianu.
defence   Australian

'Two years later he was appointed chief of the Australian Defence Force.' (Sapo_March_2016)

Other passive-like constructions include the continuous marker *hela*, as shown in (236). According to Williams-van Klinken et al. (2002a:80), *hela* can be used to "focus attention on the resulting state of a transitive verb, when the object is fronted and the subject is omitted; this gives a passive-like sentence."

(236)   Tetun Prasa (Williams-van Klinken et al. 2002a:80):

Bero   ida   para   hela   iha-ne'ebá
boat   one   stop   CONT   LOC-there

'A boat stopped/anchored there.'

Passive constructions in general cause valency reduction. However, the example above looks rather like a middle voice. A middle voice construction can be considered "a semantically transitive situation in terms of a process undergone by the PATIENT, rather than as an action carried out by an AGENT (Payne 2010:216)." Middle voice resembles a passive voice but lacks one of its main characteristics: the possibility of expressing the agent, hence the 'passive-like' construction:

(237)   English (Keenan & Dryer 2007:352-353):
   a.   The ship sank.
   b.   *The ship sank by the enemy.

6.2.6 Existential structures

In Tetun Prasa, the prototypical word order is SVO (or OSV in topicalized clauses). However, there are verbs that allow VS order, one of them being *iha* 'EXIST'. *Iha* has also other meanings, that of a transitive verb 'have' and that of a preposition 'LOC', which can cause some ambiguity.

The existential *iha* is equivalent to English 'there is/there are', Portuguese *há/(tem)* and Indonesian *ada/(punya).* Depending on its nature, the argument can follow or precede existential *iha.* Usually, indefinite arguments, as used in (238), and arguments modified by a relative clause[156], as illustrated in (239), follow the verb.

(238) Tetun Prasa:

Liuliu,	*ita*	*iha*	*petróleu,*	*Austrália*	*no*
especially	1PI	have	oil	Australia	and

iha	*indikasaun*	*katak*	*iha*	*minerál*	*oioin.*
EXIST	indication	COMP	EXIST	mineral	various

'Especially, we have oil, Australia, and there is an indication that there are various minerals.' (GovernuTL_2010)

(239) Tetun Prasa:

Iha	*rezolusaun*	*rua*	*mak*	*ko'alia*	*kona-ba*	*loron*
EXIST	resolution	two	FOC	speak	about	day

nasionál	*veteranus* [...]
national	veteran.PL

'There are two resolutions that deal with the National veterans' day […]' (STL_March_2016)

When the argument is definite, the reversed order is preferred, as shown in (240). However, in this case, the more accurate reading of *iha* would be 'be present, be here' (Williams-van Klinken et al. 2002a:55-56):

---

[156] Relative clauses that act as an argument to *iha* 'EXIST' are not introduced by a relativizer, see section 6.3.3.

(240) Tetun Prasa:

[...]	*no*	*momentu*	*ne'e*	*arguida*	*rasik*	*nia*
	and	moment	this	defendant	own	POSS

pasaporte	*la*	*iha.*
passport	NEG	EXIST

'[...] and at that time, the defendant's own passport was not there.'
(STL_March_2016)

Some sentences can have a double reading: existential and possessive. However, when *iha* has a possessive meaning, the most common word order is SVO (Williams-van Klinken et al. 2002a:56):

(241) Tetun Prasa:

Eskola	*folin*	*la*	*iha,*	*osan*	*la*	*iha,*	*buat*	*hotu*
school	cost	NEG	EXIST	money	NEG	EXIST	thing	all

la	*iha*	[...]
NEG	EXIST	[...]

'There is no money for school fees, there is no money, there is nothing.'
or '(They) have no money for school fees, they have no money, they have nothing.' (0053TD_JLS_44_M_Uni_Wai)

But *iha* is not the only verb that allows VS order. Especially in formal language, we find examples of verbs like *mosu* 'appear', as in (242), *akontese* 'happen' (from Port. *acontecer*), as in (244), *falta* 'lack' (from Port. *faltar*), as in (243), *moris* 'live' and *hela* 'remain' that follow the same order[157] (Williams-van Klinken et al. 2002a:57). This might be, again, due to the influence of Portuguese, which allows VS order with verbs like *haver* 'exist', *ser* 'be', *faltar* 'lack', *aparecer* 'appear', etc.

---

[157] In journalistic texts, the verb *dehan* 'say' is used to introduce direct speech. In this case, the subject can either precede or follow the verb. For more discussion see 6.3.2.1.1.

(242) Tetun Prasa:

Iha   2006   mosu   dunik   krize   liuliu   foos   la
in   2006   appear indeed   crisis   especially   rice   NEG

iha.
EXIST

'Indeed, in 2006, a crisis appeared, there was especially no rice.'
(GovernuTL_2010)

(243) Tetun Terik:

Ha'u   nu'udar   profesora   língua   portugesa   ha'u
1S   as   professor   language   Portuguese   1S

iha   obstákulu   boot   ida   tanba   falta   livrus
have   obstacle   big   one   because   lack   book.PL

ba   alunus,   hanesan   falta   materiál,   […]
for   pupil.PL   like   lack   material   […]

'As a Portuguese professor, I face one big obstacle, because there is a lack of books, a lack of materials, […]' (0032TS_OA_56_F_Uni_TT)

(244) Tetun Prasa:

Tanba   akontese   dala   rua   ona   ema   la   haree
because   happen   time   two   ANT   person   NEG   see

ha'u.
1S

'It happened twice to me that the people did not see me.'
(0056TD_CS_50_F_Pri_Tet)

SV order is, in general, preferred when the subject is definite, as shown in (245) (Williams-van Klinken et al. 2002a:57). These verbs also frequently occur in relative clauses modifying the previous NP and the verb *akontese* in particular is often followed by an adverbial phrase, as illustrated by (246).

(245) Tetun Prasa:

Problema	*rai*	*la'ós*	*mosu*	*de'it*	*ba*	*povu*
Problem	land	NEG	appear	only	to	the.people

kbiit-laek	*sira,*	*maibé*	*kazu*	*ne'e*	*akontese*	*mós*	*ba*
power-less	PL	but	case	this	happen	also	to

ema-boot	*sira*	*hanesan*	*eis-membru*	*governu*	*no*
people-big	PL	like	ex-member	government	and

emprezáriu	*sira.*
businessman	PL

'Problems with land do not only appear to powerless people but this case also happens to VIPs, like former MPs or businessmen.' (STL_March_2016)

(246) Tetun Prasa:

Tuir	*Constancio*	*katak*	*problema*	*ne'ebé*	*akontese*
follow	Constancio	COMP	problem	REL	happen

entre	*estudante*	*Timor-oan*	*iha*	*Indonézia* […]
between	student	Timorese	LOC	Indonesia […]

'According to Constancio, the problem that happened between the Timorese students in Indonesia […]' (STL_July_2013)

6.3 Clause structure

Originally, Tetun's complex sentences were mostly dominated by coordination. But as Thomaz (2002:107) put it, with time, the language started to borrow new vocabulary and new syntactic processes: "…o tétum teve de recorrer à imitação dos modelos frásicos do português e ao empréstimo das suas conjunções, através dos quais conseguiu desenvolver a hipotaxe a par da sua parataxe tradicional".[158]

---

[158] "…Tetun had to resort to the imitation of Portuguese phrasal model and borrow conjunctions, which made it possible to develop hypotaxis, existing alongside the traditional parataxis." (my translation)

However, simple juxtaposition is still a common strategy to express not only coordination, as in (247), and repetition, as in (248), but also to link subordinated clauses, like conditional or temporal ones, as in (249).

(247) Tetun Terik:

Ha'u-nia   *inan*   *naran*   *Bui Lera,*   *aman*   *naran*   *Kai Seran.*
1S-POSS   mother name   Bui Leta   father   name   Kai Seran

'My mom is called Bui Lera, my dad is called Kai Seran.'
(0045TV_LP_49_M_Sec_TT)

(248) Tetun Prasa:

[...]   *ami*   *ko'alia*   *ho*   *ami-nia*   *lian*
        1PE   speak   with   1PE-POSS   language

maternal,   *lian*   *inan.*
maternal.FEM   language   mother

'[…] we speak our mother tongue, mother tongue.'
(0055TD_MAS_45_M_Uni_Mid)

(249) Tetun Prasa:

Fa'an *manu-tolun*   *ema*   *dehan* […]
sell   bird-egg   person say   […]

'(When) I was selling eggs, people would say […]'
(0014TD_JBP_26_M_Uni_Mak)

In this section, I will look closely at coordination and its copulative (6.3.1.1) and adversative (6.3.1.2) conjunctions as well as subordination and its complement clauses (6.3.2), relative clauses (6.3.3) and adverbial clauses (6.3.4) while examining the extent of the influence of language contact.

6.3.1 Coordination

This section discusses two types of coordinating conjunctions: copulative ('and') and adversative ('but'), as they occur in my corpus.

6.3.1.1 Copulative conjunctions

There are three main copulative (additive) conjunctions in Tetun Prasa: *no*, *ho* and *i* 'and'. In general, it is said that *ho* coordinates noun phrases, *i* coordinates clauses and sentences and *no* coordinates phrases, clauses, and sentences, but it is mostly used in Church Tetun and formal situations, while being rare in everyday speech (Williams-van Klinken et al. 2002a:104).

In written texts, *ho* is mostly used with the meaning of 'with' and in limited number of cases, as a conjunction connecting two NPs. However, in oral speech, *ho* 'and' is still rather widespread:

(250)   Tetun Prasa:

[...]	*nesesidade*	*hodi*	*hadi'a*	*kualidade*	*relasaun*
	need	to	improve	quality	relation

entre	*estadu*	*ho*	*sidadaun*	[…]
between	state	with	citizen	

'[…] the need to improve the quality of the relation between the state and its citizens […]' (GovernuTL_2010)

The conjunction *i* has a limited use in oral speech but I was also surprised by its low frequency in written texts. It is derived from Portuguese (*e* 'and'), in which this conjunction connects phrases, clauses and sentences.[159] In Tetun, we usually find it sentence-initially, as shown in (251), or connecting the final digits of Portuguese numerals, as illustrated by (252).

(251)   Tetun Prasa:

i	*agora*	*ne'e*	*daudauk*	*ha'u*	*servisu*	*ho*
and	now	this	PROG	1S	work	with

Universidade	*Nasionál*	[…]
university	national	

---

[159] Although I follow the official orthography of Tetun in this thesis, I decided to adopt *i* for the conjunction 'and', since *e/eh* carries the meaning of 'or' in Hull (1999).

'And now I'm working with the National University […]'
(0053TD_JLS_44_M_Uni_Wai)

(252) Tetun Terik:

i	*iha*	*mós*	*kuatrusentus*	*i*	*kuarenta*	*i*	*dois*,
and	EXIST	also	four.hundred	and	forty	and	two

atus	*haat*	*haatnulu-resin-rua*	[...]
hundred	four	forty-excess-two	

'And there are also 442 (*sukus*) […]' (0039TV_MVJ_51_M_Uni_TT)

Contrary to the observations of Williams-van Klinken et al. (2002a:104), *no* is the most common conjunction in my corpus, with almost 1400 occurrences (including 1325 occurrences in the written and oral corpus of Tetun Prasa), possibly because it connects all kinds of phrases, as shown in (253), clauses, as shown in (254), and sentences, as shown in (255). It is used quite extensively in written texts but is also common in Tetun Prasa oral speech.

(253) Tetun Prasa:

Iha	*liuliu*	*Waima'a*	*no*	*Midiki*	*ne'ebé*
EXIST	especially	Waima'a	and	Midiki	REL

ko'alia	*iha*	*zona*	*Vemasse,*	*Baucau*	*nian.*
speak	LOC	area	Vemasse	Baucau	POSS

'In the area around Vemasse and Baucau, especially Waima'a and Midiki are spoken.' (0055TD_MAS_45_M_Uni_Mid)

(254) Tetun Prasa:

[...]	*ita*	*hatene*	*ona*	*ema*	*ne'ebé*	*mak*	*atu*	*akuza*	*no*
	1PI	know	ANT	person	REL	FOC	IRR	accuse	and

ne'ebé	*mak*	*sai*	*sasin*	*ba*	*kazu*	*ne'e.*
REL	FOC	become	witness	to	case	this

'[…] we know the people who are going to be accused and (the people) who are going to become witnesses in this case.' (STL_March_2016)

(255) Tetun Prasa:

[...] *ha'u-nia* *aman* *mai* *husi* *Baucau* *no*
1S-POSS father come from Baucau and

ha'u-nia *inan* *mai* *husi* *Dili.*
1S-POSS mother come from Dili.

'[…] my father comes from Baucau and my mother comes from Dili.'
(0057TD_LAOB_30_M_Uni_Tet)

According to van Klinken (1999:302), it is possible to analyse Tetun Prasa *no* (*nó* in Tetun Fehan) diachronically as a 3S conjugation of Tetun Fehan *hó* 'accompany'. She based her assumptions on the fact that, in older works on Tetun Prasa, *ho* (also written as *hó*) is analysed as a preposition 'with' but also as a conjunction 'and' (Silva 1889:34, Morris 1984)[160], but never as a verb. This could mean, that *ho* was originally used as a verb and later grammaticalized into a preposition 'with' and subsequently into a conjunction 'and' connecting NPs. In Tetun Terik, where subject marking is applied also to prepositions and conjunctions, *ho* acquired the 3S form *no* which, eventually, started to be used with other persons as well. Williams-van Klinken et al. (2002a:104) provided further evidence that *no* was inherited from Tetun Terik.

It is interesting to see how Tetun Terik *no* spread to coordinate all types of constituents in both Tetun Terik and Tetun Prasa, as is clear from Table 33. It is one of the few inherited elements from the conservative variety Tetun Terik that became the norm in Tetun Prasa, as opposed to the Portuguese borrowing *i* that is still confined to coordinate only clauses and sentences.

Conjunction 'and'	Tetun Prasa (Williams-van Klinken et al. 2002a:104)	Tetun Fehan (van Klinken 1999:302)
no	Coordinates all types of constituents, but rare in speech	Coordinates NPs, rarely clauses and predicates
ho	Coordinates NPs	Coordinates human referents
i	Coordinates clauses and sentences	N/A
hodi	N/A	Coordinates non-NPs
mós	Coordinates clauses	Coordinates clauses

Table 33 Conjunction 'and' in Tetun Prasa and Tetun Fehan

---

[160] Dores (1907) analysed *ho* as 'also'. Hull (1993:4) mentioned *ho* and *i*, but not *no*.

6.3.1.2 Adversative conjunctions

To express contrast, Tetun Prasa uses adversative conjunctions *mas* 'but' and *mais* 'but', both borrowed from Portuguese *mas* 'but,' and a native conjunction *maibé* 'but':

(256) Tetun Prasa:

Mas	*ha'u*	*la*	*mate,*	***maibé***	*ha'u*	*reza*	*maka'as.*
but	1S	NEG	die	but	1S	pray	hard

'But I didn't die, but I prayed hard.' (0056TD_CS_50_F_Pri_Tet)

Maibé is generally preferred by Tetun speakers of all varieties (143 occurrences in the oral corpus) and it is almost a norm in journalistic texts. In my written corpus, I have encountered 229 occurrences of *maibé* but only 8 occurrences of *mas/mais* 'but'.

Regarding the origin of the Portuguese borrowing *mas*, it is possible that the source is directly Portuguese or the so-called Southeast Asian Creole Portuguese. We find the same conjunction in Macau Creole Portuguese *mas* and Malacca Creole Portuguese *mas*. Since it is found in both Tetun Prasa (28 occurrences) and Tetun Terik (28 occurrences), it is probable that the conjunction was borrowed during the earlier stages of contact with Portuguese. At the same time, many Tetun Terik speakers opt for *mais* 'but' (11 occurrences), and so do Tetun Prasa speakers (also 11 occurrences). Hull (2005a:102) believed it was borrowed from Kupang Malay *masi* 'but', but this seems unlikely. It is more probable that *mais* is just a variant of *mas.*

6.3.2 Complement clauses

A complement clause is a clause that functions as an argument of the main predicate. According to Payne (2010:314), prototypical finite complements act like independent clauses, sharing these two characteristics with them: they carry their own tense and aspect, and they express their subject directly (subject reference is not restricted to that of the matrix clause). Williams-van Klinken et al. (2002a:106) distinguished two types of complements in Tetun Prasa: sentential complements and reduced complements.

Sentential complements differ from reduced complements in that "the predicate has the same syntactic relation to its subject and its other arguments that it has in syntactic main clauses" (Noonan 2007:59). On the other hand, the reduced complement is defined as "[a]ny complement type that has fewer syntactic and inflectional possibilities than an indicative main clause" (Noonan 2007:83).

To identify complement clauses, many languages use so-called 'complementizers'. However, their use is often optional or contextually determined by pragmatic, not grammatical considerations (Noonan 2007:55). In Tetun Fehan, there are only few complementizers and even these are used rather infrequently. Van Klinken (1999:283) only mentioned three: *hosi* 'about', *batu* 'so that', and possibly *ne'ebé* 'so that', but usually, there is no complementizer. On the other hand, complementizers are frequent in Tetun Prasa, especially in formal discourse and written texts. This has to do with the close contact with Portuguese clause structure since in Portuguese, finite complement clauses are always introduced by a complementizer[161] (Mateus et al. 2003:597).

In this section, I will discuss sentential and reduced complements, focusing on the nature – native or borrowed – of the complementizers they employ.

6.3.2.1.1 Sentential complements

Sentential complements can either have no complementizer or be introduced by *katak* 'that', *para* 'to, so that' (from Port. *para*), and *atu* 'that'. These complementizers introduce complements for verbs of speaking (so-called 'utterance complements'), thinking and knowing, although they are often omitted in everyday speech (Williams-van Klinken et al. 2002a:106).

According to Hull and Eccles (2005:221), other complementizers can be found in colloquial speech: *ke* (derived from Port. *que* 'that'), *ne'ebé, dehan* and several combined forms (e.g. *katak ke* and *katak dehan*), all of which mean 'that' and introduce indirect

---

[161] Non-finite complement clauses in Portuguese are, in general, not introduced by a complementizer (Mateus et al. 2003:621).

speech. I have found no examples of *ke, katak ke,* and *katak dehan* introducing complement clauses in my corpus. I also do not consider *dehan* 'say' a complementizer but rather a verb of speaking that can introduce or follow direct speech, as documented by (263) and (264), respectively, later in this section. In indirect speech in Tetun Prasa, *dehan* 'say' is, in the majority of cases, followed by one of the complementizers, as shown in (257).

Complementizers often historically derive from pronouns, conjunctions, adpositions, case markers, and, rarely, from verbs (Noonan 2007:57). The complementizer *katak* derives from a verb meaning 'say'[162] and is still used with this meaning in Tetun Fehan. In Tetun Prasa (and based on my data, also in Tetun Terik), *katak* has grammaticalized completely as a sentential complementizer, as illustrated by examples (257) and (258), and is no longer used with its former lexical meaning 'say', although it can be used to mean 'signify, mean' (Williams-van Klinken et al. 2002a:106-107):

(257)   Tetun Prasa:

[…]	*dala ruma*	*ema*	*dehan*	***katak***	*ami*	*Timor-oan*	*rasik*
	sometimes	person	say	COMP	1PE	Timorese	self

servisu	*mak*	*la*	*hamutuk*	[…]
work	FOC	NEG	together	

'Sometimes people say, that we, Timorese, do not work together (cooperate).' (0053TD_JLS_44_M_Uni_Wai)

(258)   Tetun Terik:

Tanba,	*ha'u*	*sente*	***katak***	*Tetun*	*ne'e*	*difisil*	*ba*
because	1S	feel	COMP	Tetun	this	difficult	for

ita.
1PI

'Because of that, I feel that Tetun is difficult for us.'
(0026TS_FJC_43_F_Sec_TT)

---

[162] A large number of languages have a complementizer that is quite transparently derived from the verb 'say' (Schachter & Shopen 2007:48).

A similar complementizer to *katak* 'COMP' can be found in the restructured variety of Ambonese Malay which uses optional *kata* 'COMP', derived from the Malay *kata* 'word, say' (Paauw 2013):

(259) Ambonese Malay (van Minde 1997:307):

Dia	*parlente*	***kata***	*de*	*su*	*bayar.*
3S	lie	COMP	3S	PFV	pay

'He lied that he had already paid.'

The use of *katak* and *para* as complementizers seems quite recent as neither of them is described in the works of Silva (1889) and Dores (1907). Both authors only mention the traditional Tetun Terik complementizers *ha'ak*, *na'ak* and *a'ak*. According to Silva (1889:35), *ha'ak/na'ak* have two meanings: that of a verb 'say', as in (260) and that of a conjunction 'that', as in (262), which is, however, not obligatory, as shown in (261). I only found instances of *na'ak* (mostly followed by *ka* 'or') in my Tetun Terik data, noticing that, although it resembles the historical 3S conjugation of *ha'ak*, the complementizer *na'ak* is nowadays used with all persons, as (262) illustrates. In Tetun Fehan, *ha'ak* 'say' is used to introduce direct speech (van Klinken 1999:279).

(260) Tetun Terik:

Entaun	*sira*	*rakrake*	***na'ak****: Oh*	*na'i-lulik*	*mate*
so	3P	3P.scream	say oh	priest	dead

tiha ona	[...]
PRF	

'So they screamed, saying: The priest is dead [...]'
(0050TV_JCV_48_M_Sec_TT)

(261) Tetun Terik (original spelling, Silva 1889:36):

Ámi	*hacárac*	***Ø***	*ó*	*ba*	*cátac*	*hó*	*ó*
1PE	want[163]	Ø	2S	go	tell	with	2S

nia	*álin*	*nia*	*bêlo*	*nia*
3S.POSS	younger.sibling	3S.POSS	friend	3S.POSS

---

[163] In Tetun Prasa, the complementizer used with the verb 'want' would be *para* 'so that'. *Hau hakarak para nia diak* 'I want her to be good.' (Williams-van Klinken et al. 2002a:107).

áman	*nia*	*ata*	*ato*	*mai*	*laláis.*
father	3S.POSS	servant	IRR	come	quickly.

'We want you to go and tell the servant of your youngest brother's friend's father to come fast.'

(262) Tetun Terik:

[...]	*ami*	*sente*	**na'ak**	*ka*	*ami*	*oan*	*ruma*	*atu*	*bá*
	1PE	feel	COMP	or	1PE	child	some	IRR	go

iha	*semináriu*	*karik,*	*bá*	*eskola*	*atu*	*buka*	*matenek*
LOC	seminary	maybe	go	school	IRR	seek	smart

ruma.
some

'[...] we feel that if some of our children went to the seminary or to school, they could get some knowledge.'
(0050TV_JCV_48_M_Sec_TT)

In Tetun Prasa, no complementizer is used in case of direct quotes, although the direct speech can be introduced by *hanesan ne'e* 'like this', *dehan* 'say' or *dehan katak* 'say that'. In speaking, the main clause always precedes the direct quote, as shown in (263), however, in journalistic writing, both orders are possible, as illustrated by (264) (Williams-van Klinken et al. 2002a:107):

(263) Tetun Prasa:

Fa'an	*manu-tolun*	*ema*	*dehan:*	*Oi,*	*ó-nia*	*tolun*
sell	bird-egg	person	say	hey	2S-POSS	egg

hira	*mak*	*ne'e*?
how.many	FOC	this

'When I was selling eggs, people would ask: How many eggs [in this context: balls] do you have?' (0014TD_JBP_26_M_Uni_Mak)

(264) Tetun Prasa:

"*Ami,* *agora,* *sente* *ho* *enerjia* *maka'as* *no*
1PE now feel with energy hard and

komprometidu *atu* *kontinua* *servisu* *maka'as*
committed to continue work hard

ba *futuru* *ida-ne'ebé* *di'ak* *liu",* *dehan* *Ministra.*
for future REL good more say minister.FEM

"'Now we feel full of energy and are committed to continue working hard for a better future", said the Minister.' (GovernuTL_2010)

Another difference between Tetun Prasa and Tetun Fehan is that nouns in Tetun Fehan cannot take complement clauses. In Tetun Prasa, this is, however, possible in writing and with Portuguese nouns, due to the Portuguese influence (Williams-van Klinken 2010b):

(265) Tetun Prasa (Williams-van Klinken 2010b):

direitu [*atu* *rona* *rádio*]
right [COMP hear radio]

'(the) right to listen to radio'

The purposive conjunction *para* 'to, so that', derived from Port. *para* 'to, so that' can also introduce sentential complements. According to Williams-van Klinken et al. (2002a:107), *para* can be used with verbs of wanting (*hakarak* 'want'), as in (266), but I have not found an example of such use in my corpus. The verb *hakarak*, when followed by a sentential complement, can either be followed by *atu*, as shown in (267), or not have a complementizer, as in (268).

(266) Tetun Prasa (Williams-van Klinken et al. 2002a:107):

Ha'u *hakarak* ***para*** *nia* *di'ak.*
1S want COMP 3S good

'I want her to be good.'

276

(267) Tetun Prasa:

Ha'u   *hakarak*   ***atu***   *nia*   *di'ak.*
1S   want   COMP 3S   good

(268) Tetun Prasa:

Ha'u   *hakarak*   *nia*   *di'ak.*
1S   want   her   good

The irrealis marker *atu* can also introduce verb phrase complements, as shown in (269). This construction is more common in writing than in oral speech (Williams-van Klinken et al. 2002a:84)[164].

(269) Tetun Prasa:

Deputadu   *Timor-oan*   *sira*   *husu*   ***atu***   *governadór*
delegate   Timorese   PL   ask   COMP governor

jerál australianu   *halo*   *intervensaun*   *kona-ba*   *tema*
general Australian   make   intervention   about   topic

fronteira nian
border   POSS

'Timorese delegates asked that the Australian Governor General intervenes in the border issue.' (Sapo_March_2016)

6.3.2.1.2   Reduced complements

Tetun Prasa also possesses reduced complements that cannot have independently specified subjects and the time reference is either future or co-temporal with the main predicate (Williams-van Klinken et al. 2002a:108). This type of complement clauses can be introduced by *atu* 'IRR', as in (270), *para* 'to', as in (271), or a combined *para atu* 'to'[165]. Often, there is no complementizer, as illustrated by (272).

---

[164] See the discussion about the irrealis marker *atu* becoming a semantically bleached complementizer in section 6.2.3.2.
[165] I did not find instances of *para* and *para atu* in my corpus.

(270) Tetun Prasa:

Ha'u *hakarak* ***atu*** *bá* *halo* *osan, atu* *simu*
1S want COMP go make money IRR receive

osan *veteranus.*
money veteran.PL

'I want to go and make money, to receive the pension for veterans.'
(0056TD_CS_50_F_Pri_Tet)

(271) Tetun Prasa (Williams-van Klinken et al. 2002a:109):

Sira *obriga nia* *para* *hatudu dalan.*
3P oblige 3S COMP show way

'They obliged him/her to show (them) the way.'

(272) Tetun Prasa:

Ami *mós* *hakarak* *fó* *atensaun* *espesiál* *ba*
1PE also want give attention special to

meiu *teknolójiku* *sira* *foun* *hanesan* *internet* [...]
means technological PL new like internet

'We also want to give special attention to new technological means like
internet [...]' (GovernuTL_2010)

A purposive conjunction *hodi* 'in.order.to' was not mentioned by Williams-van
Klinken et al. (2002a) as a complementizer but I have found a couple of examples with
hodi introducing reduced complements, as shown in (273) and (274). However, these
examples were deemed unnatural by my informants, suggesting a complementizer *atu*
should be used in (273) and no complementizer should be used in (274).

(273) Tetun Prasa:

[...] *sei* *bele* *ajuda Giné-Bisau* ***hodi***
FUT can help Guinea-Bissau in.order.to

dezenvolve *oferta turística.*
develop offer touristic.FEM

'[…] can help Guinea-Bissau to develop tourism offers.'
(Sapo_March_2016)

(274) Tetun Prasa:

Nune'e	*arguidu*	*Manecas*	*tuirmai*	*haruka*
thus	defendant	Manecas	next	send

polísia Rofino	**hodi**	*bolu*	*lezadu*	*ba*	*iha*
police Rofino	in.order.to	call	plaintiff	to	LOC

eskuadra	*PNTL*	*nian.*
station	PNTL	POSS

'So next, the defendant Manecas sent policeman Rofino to summon the plaintiff to the PNTL station.' (STL_March_2016)

As is clear from the discussion above, Tetun Prasa uses a whole range of complementizers to introduce sentential and reduced complements, which are summarized in Table 34. Most of them are of native origin, except for *para*, which was borrowed from Portuguese and has been attested in both Tetun Prasa and Tetun Terik.

Complementizer	*Sentential complements*	*Reduced complements*
katak	yes	no
para	yes	yes
atu	yes	yes
para atu	no	yes
no complementizer	yes	yes
hodi	no	marginal, considered incorrect

Table 34 Complementizers used with reduced and sentential complements

6.3.3 Relative clauses

Relative clauses can be either restrictive or non-restrictive. In this section, I will only discuss restrictive relative clauses. Andrews (2007:206) defined restrictive relative clause (RC) as a "subordinate clause which delimits the reference of an NP by specifying the role of the referent of that NP in the situation described by the RC". The main clause, in which the RC is embedded, is also called a matrix clause and the noun of the NP that is modified by the RC is called a head noun. When analysing a relative clause, we should consider three main parameters (Payne 2010:326):

- the position of the RC with respect to the head noun (prenominal, postnominal, internally headed and headless);
- the mode of expression of the relativized NP (gap strategy, pronoun retention);
- which grammatical relations can be relativized.

The most common type of relative clause in Austronesian SVO languages is a postnominal RC:

(275) Tetun Prasa:

Oké, obrigadu ba		*tempu*	[*ne'ebé*	*fó*	*mai*	*ha'u.*	]
OK thank.you	to	time	[REL	give	to	1S	]

'Ok, thank you for the time [that you gave me]'.
(0014TD_JMB_31_F_Sec_Mak)

In Tetun Prasa, only the subject, as in (276), object, as in (277), and time phrases can be relativized.

(276) Tetun Prasa:

Korrupsaun	*fenómenu*	*ne'ebé iha*	*konsekuénsia*	*negativa*
corruption	phenomenon	REL have	consequence	negative

ba	*vida*	*sosiál*	*no*	*ekonómika* […]
to	life	social	and	economic

'Corruption is a phenomenon that has negative consequences on social and economic life [...]' (GovernuTL_2010)

(277) Tetun Prasa:

[...]	*Primeiru*	*Ministru*	*fó-hanoin*	*ba*	*delegasaun*
	Prime	Minister	give-think	to	delegation

sira	*ne'ebé hola*	*parte*	*iha*	*soru-mutu*	*ne'e*	*katak* […]
PL	REL get	part	LOC	meeting	this	COMP

'[…] the Prime Minister informed the delegations which participated in the meeting that […]' (GovernuTL_2010)

Relative clauses in Tetun Fehan differ from Tetun Prasa in the fact that they can either precede or follow the head of the NP, although the premodifying relative clauses have a very restricted syntax (van Klinken 1999:318) and the head is always *fatik* 'place' (van Klinken 1999:154):

(278)   Tetun Fehan (van Klinken 1999:154):

Rai	*né*	*rai*	*futu*	*manu*	*sia*	*fatik.*
land	this	land	tie	rooster	PL	place

'This area is a cockfighting place.'

To identify a relative clause, many languages use relativizers. Often, they have the same form as complementizers (Payne 2010:332). In Tetun Prasa, some speakers think that the relativizer is required in most contexts, some disagree, but most of them omit the relativizer when the NP is the argument of a preceding phrase with the meaning of 'there is/there are' (*iha* 'EXIST'):

(279)   Tetun Prasa:

Iha	*ema*	*ida-rua*	*mak*	*sei*	*kontinua*	*halo*	*krime*
EXIST	person	one-two	FOC	still	continue	make	crime

iha	*sidade*	*Dili*	*laran*	[...]
LOC	city	Dili	inside	[...]

'There are still several people that continue committing crimes within the city of Dili.' (STL_March_2016)

The most frequent relativizer is the relative particle *ne'ebé* 'REL' (it can mean 'which', 'who', 'whom' and 'that') although there is a lot of variation when it comes to a) the choice of the relativizer, b) the extent to which it can be omitted and c) the acceptability of the determiner *ne'e* 'this' following the relative clause (Williams-van Klinken et al. 2002a:116).

Superficially, the relativizer *ne'ebé* might resemble the English relative pronouns 'who' and 'which'. However, they differ in function from the Tetun relativizer, one reason why Eccles (2001:40) decided not to refer to it as a pronoun[166].

The relativizer *ne'ebé* can be shortened to *be*, as in (280), in both colloquial and literary Tetun. When the head noun is definite, the relativizer can be omitted. When the relativizer is present, the relative clause often ends in *ne'e* 'this', as shown in (281) (Hull 1993:71-72).

(280)   Tetun Prasa:

Liuliu	*ba*	*hanesan*	*be*	*ita*	*joven*	*sira*	*be*
mainly	for	like	umm	1PI	young.person	PL	REL

hakarak	*ba*	*atu*	*hola*	[...]
want	to	IRR	marry	[...]

'Especially for us, young people who are looking to get married [...].' (0014TD_JBP_26_M_Uni_Mak)

(281)   Tetun Prasa:

[...]	*hodi*	*promove*	*dezenvolvimentu*	*sustentavel*
	in.order.to	promote	development	sustainable

hosi	*país*	*sira*	*ne'ebé*	*mak*	*partisipa*	*ne'e.*
from	country	PL	REL	FOC	participate	this.

'[…] to promote the sustainable development in the countries that participate.' (GovernuTL_2010)

The relativizer *ne'ebé* (but never *be* or *ke*) is often followed by *mak* 'FOC', as in

(282), which, in this context, presents the relativized constituent as known information (Williams-van Klinken et al. 2002a:116). However, in this case, *mak* does not serve as a focus marker.

---

[166] For example, the relativizer can be omitted also in cases in which English relativizers can't be omitted. English can only omit the relative pronoun if it is the object of the verb in its clause but Tetun *ne'ebé* is not dependent on the verb in the following clause.

(282) Tetun Prasa:

Programa   **ne'ebé mak**   *prinsipál*   *liu*   *ba*   *ha'u* […]
program   REL   FOC   main   very   to   1S

'The program that is very important to me […]'
(0014TD_JMB_31_F_Sec_Mak)

According to Williams-van Klinken et al. (2002a:116), the use of *mak* in Tetun Prasa could be a relic of its function as a relativizer in Tetun Terik. In this variety, *mak* can be used as a substitute for the relativizer *ne'ebé*, as shown in (283). Hull and Eccles (2005:48) suggested that this relativizer is especially used in the Tetun Terik regions of Balibó and Suai, close to the Indonesian border. The authors believed that the use of *mak* might have been modelled after the Malay/Indonesian *yang*, as documented in (284), which is equivalent to both *mak*/*maka* 'FOC' and *ne'ebé* 'REL'. Due to the recent influence of Indonesian, some speakers of Tetun Prasa also make use of the contrastive marker *mak*, although this use is considered incorrect by many. However, I have found many instances of *mak* in my written and oral corpus of Tetun Prasa.

(283) Tetun Terik:

[…]   *ita*   *mós*   *bele*   *halo*   *seminar*   *ka*   *kongresu*
      1PI   also   can   make   seminar   or   congress

lian   *Tetun*   *ba*   *profesores*   *Tetun*   *tomak*   *mak*   *iha*
language   Tetun   to   professor.PL   Tetun   all   REL   LOC

Timor-Leste [...]
Timor-Leste

'[...] we can also organize a seminar or a congress about the Tetun language for all the Tetun professors that are in Timor-Leste.'
(0026TS_ATB_41_M_Sec_TT)

(284) Indonesian (Sneddon et al. 2010:295):

orang   **yang**   *membangun*   *rumah*   *saya*
person   REL   build   house   1S

'the person who built my house'

Hull and Eccles (2005:49) described another strategy unique to Tetun Prasa: preposition stranding, i.e. when the relative clause contains a preposition which is placed at the end of the clause:

(285) Tetun Prasa (Hull & Eccles 2005:49):

Ha'u	*buka*	*kadernu*	*ida-ne'ebé*	*nia*	*hakerek*	***ba****.
1S	seek	notebook	one-REL	3S	write	to

'I'm looking for the notebook that she wrote in.'

However, some speakers prefer the strategy with a relativizer *iha-ne'ebé* 'LOC-REL', as shown in (286). When the relativizer is referring to a place, *(ida-)ne'ebé* can be replaced by *iha-ne'ebé* 'where'*,* which is a construction possibly calqued from Portuguese, as in (287).

(286) Tetun Prasa:

Ha'u	*buka*	*kadernu*	***iha-ne'ebé***	*nia*	*hakerek.*
1S	look.for	notebook	LOC-REL	3S	write

'I'm looking for a notebook in which she wrote.'

(287) Portuguese:

Estou	*à*	*procura*	*do*
be.1S.PRS	at.DET.FEM	search	of.DET.MASC

caderno	***em***	***que***	*ela*	*escreveu.*
notebook	in	which	3S.FEM	write.3S.PST

Hull and Eccles (2005:49) further stated that it was also possible to omit the relativizer *ne'ebé* completely and add the determiner *ne'e* 'this' at the end of the clause, as in (288). This strategy was also deemed ungrammatical by my informants.

(288) Tetun Prasa (Hull & Eccles 2005:49):

Sé	*maka*	*koñese*	*malae*	***Ø***	*nia*	*servisu*	*ho*
who	FOC	know	foreigner	Ø	3S	work	with

ne'e?
this

'Who knows the foreigner he works with?'

Lastly, when the relativizer is present, it is possible to omit the preposition:

(289) Tetun Prasa (Hull & Eccles 2005:49):

Kuartu	*ida-ne'ebé*	*ha'u*	*toba*	*Ø*	*fo'er*	*tebetebes.*
room	one-REL	1S	sleep	Ø	dirty	RDP-really

'The room in which I slept was very dirty.'

The Portuguese relative pronoun *que* 'REL' made its way into Tetun Prasa as *ke* 'REL', as shown in (290), which is used mostly in speaking and often occurs with *ne'ebé* as a complex relativizer *ne'ebé ke*, as shown in (291) (Williams-van Klinken et al. 2002a:116).

(290) Tetun Prasa:

Se	*nesesidades*	*país*		*nian*	*ezije*	*krexsimentu*
if	necessity.PL	country		POSS	demand	growth

ekonómiku	*ida*	***ke***	*lais,*	*sustentavel,*	*ita*	*presiza*
economic	this	REL	fast	sustainable	1PI	need

investe	*infraestruturas*		*bázikas.*
invest	infrastructure.PL		basic.FEM.PL

'If the country's necessities demand an economic growth that is fast and sustainable, we need to invest in basic infrastructure.'
(GovernuTL_2010)

(291) Tetun Prasa:

Di'ak,	*hanesan*	*problema*	***ne'ebé ke***		*mosu*	*entre*
OK	like	problem	REL	REL	appear	between

Austrália	*ho*	*Timor-Leste*	*sobre*	*dokumentus*	*ne'ebé ke*	
Australia	and	Timor-Leste	about	document.PL	REL	REL

intel	*Austrália*	*foti*	*ne'e*	[…]
spy	Australia	pick.up	DET	

'Well, regarding the problem that arose between Australia and Timor-Leste about the documents that were collected by an Australian spy […]'
(0008TD_TAB_55_M_PreS_Mam)

6.3.4 Adverbial clauses

Adverbial clauses modify verb phrases or entire clauses. Although they are considered subordinate, they do relate to the main clause as a whole. The verb form or the word order do not change (Thompson et al. 2007:238). Adverbial clauses in Tetun are introduced by subordinating conjunctions. In this section, I will look at conjunctions introducing adverbial clauses of time (6.3.4.1), purpose (6.3.4.2), reason (6.3.4.3), condition (6.3.4.4) and concession (6.3.4.5), analysing the influence of borrowed conjunctions on syntax and semantics of the clauses.

6.3.4.1 Adverbial clauses of time

Adverbial clauses of time (together with those of location and manner) are different from the other four in that they can be substituted by a single, non-anaphoric word, which means that "the semantic relationship between the adverbial clause and the main clause is the same as that between the adverbial word and the main clause" (Thompson et al. 2007: 243-244). In Tetun Prasa, most of the adverbial clauses of time precede the main clause and are introduced by a conjunction. Tetun Prasa possesses a full range of native conjunctions, however that did not stop Portuguese conjunctions from being borrowed as well. For example, *bainhira* 'when' and *kuandu*[167] 'when' (from Port. *quando* 'when'), as shown in (292), coexist almost with the same semantic value, although according to Williams-van Klinken et al. (2002a:110) *kuandu* also carries an element of conditionality, as illustrated in (293). This might have to do with the fact that Port. *quando* can also be translated as 'if' and in that case, it is followed by subjunctive in Portuguese. However, this use of *quando* is so marginal even in Portuguese that it is unlikely that the 'element of conditionality' was borrowed with the conjunction. It is more probable that the conditional meaning was caused by the influence of Indonesian, in which conditional conjunctions can carry the meaning of 'when' (see section 6.3.4.4).

---

[167] *Kuandu*, unlike *bainhira* (or *horibainhira*), cannot be used as an interrogative adverb. When asking questions about present or future, *bainhira* 'when' is used; for the questions about past time, *horibainhira* 'past.when' is employed (Williams van Klinken et al 2002a: 110).

(292) Tetun Prasa:

Tanba	*lian*	*Tetun ne'e,*	*ema*	*lokál*	*sira*
because	language	Tetun this	person	local	PL

kuandu	*ko'alia,*	***bainhira***	*sira*	*la*	*hatene* [...]
when	speak	when	3P	NEG	know

'Because of Tetun, when the locals speak and when they do not understand […]' (0055TD_MAS_45_M_Uni_Mid)

(293) Tetun Prasa:

[...]	***kuandu***	*seidauk*	*iha*	*ema*	*ida*	*ba*
	when	not.yet	EXIST	person	one	to

substitusaun	*nia*	*nafatin*	*komandante*	[...]
substitution	3S	continue	commander	

'[…] when/if there is nobody to substitute him, he will continue as a commander […]' (STL_March_2016)

Just like with other connectors, it is also not unusual, in oral speech, to see *bainhira* followed by *kuandu* as a complex conjunction, as shown in (294). I have also found numerous examples in which subjects precede *kuandu* but not *bainhira*, as illustrated by (295).

(294) Tetun Terik:

Bainhira	***kuandu***	*ha'u-nia*	*aman*	*sira*	*bá*	*halo*
when	when	1S-POSS	father	PL	go	make

kampu,	*lori*	*fatuk*	*obrigatóriu.*
field	carry	stone	obligatory

'When my father and my uncles used to work in the field, it was obligatory to carry stones.' (0039TV_MVJ_51_M_Uni_TT)

(295) Tetun Prasa:

Sira	***kuandu***	*ko'alia*	*bele*	*entende*	*malu*	[...]
3P	when	speak	can	understand	RECIP	

'When they speak, they can understand each other [...]'
(0054TD_NSG_51_M_Uni_TT)

Based on my data, the use of *kuandu* in Tetun Terik of Viqueque and Suai is very marginal. As for Tetun Prasa, I have found *kuandu* only in speech of university-educated speakers or speakers older than fifty, and in the written language.

I also looked at how these connectors, *bainhira* and *kuandu*, combine with TAM markers. As I have mentioned before, tense, aspect and mood are often implied by the context and are not always marked by TAM markers. And although *kuandu* and *bainhira* can introduce past, present and future (general) clauses, these are almost never marked for tense nor aspect. The main clauses that they modify are mostly not marked either. With *kuandu*, I have only found the TAM marker *ona* 'ANT' (5 examples):

(296) Tetun Prasa:

Kuandu	*nia*	*prosesu*	*tama*	*ona*	*iha*
when	3S.POSS	proceedings	enter	INCH	LOC

julgamentu	*to'o*	*hetan*	*desizaun*	[...]
trial	until	get	decision	[...]

'When the legal proceedings reached the trial and before the decision (was made) [...]' (STL_March_2016)

Bainhira combines with TAM markers more readily. Apart from *ona* 'ANT', it is often followed by *atu* 'IRR', as shown in *(297)*. This might have to do with the fact that *bainhira*, unlike *kuandu*, is usually used to describe an event that definitely happened or will happen (Williams-van Klinken et al. 2002a:110). So, in order to express an event that might not come about, an irrealis marker has to be employed:

(297) Tetun Prasa:

Bainhira	*atu*	*taka*	*odamatan*	*boot,*	*derrepente*
when	IRR	close	door	big	suddenly

arguidu sira	*insulta*	*no*	*tolok*	*lisuk*	*lezadu*	*ho*
defendant PL	insult	and	swear	together	plaintiff	with

liafuan [...]
word

'Suddenly, when he was about to open the main door, the defendants insulted and swore at the plaintiff [...]' (STL_March_2016)

There are other Portuguese borrowings used as temporal conjunctions (and, at the same time, as prepositions and adverbs) in Tetun, as illustrated by (298). Often, they do not have a native counterpart, or a more descriptive Tetun expression is used:

(298) a. *enkuantu* 'while' (from Port. *enquanto* 'while');

b. *durante* 'while' (from Port. *durante* 'while');

c. *antes* 'before' (from Port. *antes* 'before');

d. *depois* 'after' (from Port. *depois* 'after');

e. *dezde* 'since' (from Port. *desde* 'since');

f. *momentu* 'while' (from Port. *no momento (de/em que)* 'at the moment (of/in which)').

6.3.4.2 Adverbial clauses of purpose

Unlike temporal clauses, adverbial clauses of purpose, reason, conditionality and concession cannot be substituted by a single word. Purpose clauses "express a motivating event which must be *unrealized* at the time of the main event" (Thompson et al. 2007:250). That is why Tetun Prasa makes use of the irrealis marker *atu* 'IRR' to introduce purposive clauses:

(299) Tetun Prasa:

*Ita   tenke fó   tempu ba   povu **atu**   manán konfiansa.*
1PI   must   give   time to   people IRR   win   confidence

'We must give people time to gain back their confidence.'
(GovernuTL_2010)

Tetun Prasa also borrowed the Portuguese conjunction *para* 'so that', which is used in Portuguese in the form of *para que* + subjunctive or *para* + personal infinitive (Port. *infinitivo pessoal*) to introduce purposive clauses:

(300) Tetun Prasa:

[...] *ami   sai   eskola **para** bele   kaben tiha   de'it*.
       1PE   exit   school so.that can   marry PFV   only

'[...] we left school so that we could get married.'
(0056TD_CS_50_F_Pri_Tet)

As Williams-van Klinken et al. (2002a:112) and Hull and Eccles (2005:215) explained, *para* is more common in speaking and colloquial language than in writing. Indeed, I have found *para* in all the varieties of spoken Tetun, but comparing my three written sources (the daily newspapers *Suara Timor Lorosae* and *Sapo TL* and the official Government page), I noticed that it is only really used in one of them – *Suara Timor Lorosae*. This is interesting, since we usually find a higher ratio of Portuguese borrowings in written, journalistic texts.

It is also not uncommon to see complex conjunctions. Just like the complementizer (*ne'ebé ke*) and relativizer (*katak ke*), native and borrowed conjunctions tend to cluster. In Tetun Prasa, we can find *para atu*, *para hodi* and *atu hodi*.

In section 6.2.4, I have already mentioned that *hodi* has undergone grammaticalization and cannot be considered fully verbal, but rather a prepositional verb when it appears in the second slot of a SVC. In Tetun Fehan, it has even grammaticalized into a coordinating conjunction (see section 6.3.1.1), although according to van Klinken (1999:354) it is very unusual for a verb 'bring, carry' to develop into a clausal coordinator. However, the author did not recognize the function of *hodi* as a subordinating conjunction in Tetun Fehan.

Williams-van Klinken et al. (2002a:113) agreed that *hodi* can express much wider range of semantic relationships in Tetun Terik than in Tetun Prasa, but neither them nor van Klinken (2000) who looked at different functions of *hodi* explained how it could have grammaticalized into a subordinating conjunction in Tetun Prasa. In my oral corpus of Tetun Prasa, *hodi* is used exclusively as a subordinating conjunction:

(301)   Tetun Prasa:

Entaun   *ami*   *tenke*   *halo*   *fali*   *aula*   *seluk*   **hodi**
so   1PE   must   make   again   class   other   in.order.to

kompensa   *aulas*   *ne'ebé ami*   *lakon*   *durante*   *iha*
compensate   class.PL   REL   1PE   lose   during   LOC

krize.
crisis

'So we had to take other classes in order to compensate for the classes we missed during the crisis.' (0057TD_LAOB_30_M_Uni_Tet)

6.3.4.3 Adverbial clauses of reason

Contrary to purpose clauses, reason clauses "express a motivating event which may be *realized* at the time of the main clause event" (Thompson et al. 2007:250-251). Again, these can be introduced by conjunctions of native or Portuguese origin. The most frequent one is *tanba* 'because', both in Tetun Prasa and Tetun Terik:

(302)   Tetun Prasa:

Iha   *loraik*   *ida-ne'e*   *ami*   *kontente*   **tanba**
LOC   afternoon   this.one   1PE   happy   because

hasoru   *kolega ne'ebé di'ak*   *husi*   *nasaun*   *seluk*   [...]
meet   friend REL   good   from   nation   other

'This afternoon we are happy because we got to meet nice friends who come from other nations [...]' (0014TD_FDP_19_F_Sec_Mak)

I have not found an example of *porké* 'because' (from Port. *porque* 'because') in my written nor oral corpus. However, it is mentioned in both Williams-van Klinken et al. (2002a:111) and Hull and Eccles (2005:211). Both works also mention obsolete conjunctions *basá* and *tán*, both 'because', used only in the church register.

Tetun Prasa also borrowed another Portuguese conjunction, namely *komu* 'since' (from Port. *como* 'since'), which has a very limited use in written texts and oral speech:

291

(303)   Tetun Prasa:

*Maibé **komu** la   hatene hakerek,   la   hatene lee   [....]*
but   since NEG know   write   NEG know   read

'But since they didn't know how to write and didn't know how to read
[...]' (0053TD_JLS_44_M_Uni_Wai)

6.3.4.4 Conditional clauses

Conditional clauses in Tetun Prasa (and also in Tetun Terik) can be either formed by simple juxtaposition or by employing the conjunction *se* 'if' (from Port. *se* 'if'), as in (304), or the native adverb *karik* 'maybe'. While *se* always introduces the adverbial clause, *karik* is either placed at the beginning or the end of it, as shown in (305) and (307) respectively. Often, *se* and *karik* can be found in the same clause, either together as a complex conjunction, as in (306)[168], or very often disconnected, as in (307).

(304)   Tetun Prasa:

[...]   *ita   la   bele   halo   komunikasaun   **se**   ita*
       1PI   NEG can   make   communication   if   1PI

*la   hatene língua,   **se**   ita   la   komprende*
NEG know   language   if   1PI   NEG understand

língua.
language

'[...] we cannot communicate if we do not know the language, if we do not understand the language. (0014TD_JBP_26_M_Uni_Mak)

(305)   Tetun Prasa:

*Ó   lalika   hamanas   **karik** di'ak.*
2S   need.not   heat.up   maybe good

'You do not need to heat it up if it is still good.'
(0014TD_JBP_26_M_Uni_Mak)

---

[168] Only documented in two speakers of Tetun Terik in my corpus.

(306) Tetun Terik:

[...]	*se*	*karik*	*ha'u*	*la*	*iha*	*osan*	*atu*	*selu*
	if	maybe	1S	NEG	have	money	IRR	buy

transporte	*tanba*	*ha'u*	*hela*	*fatin*	*dook,*	*entaun*	*ha'u*
transport	because	1S	stay	place	far	so	1S

iha	*uma*	*ne'e,*	*ha'u*	*triste*	*tebetebes.*
LOC	house	this	1S	sad	very

'[...] if I didn't have money to pay for transportation, because I live far away, I would have to stay at home and I would be very sad.'
(0032TS_OA_56_F_Uni_TT)

(307) Tetun Prasa:

Se	*aban-bainrua*	*se*	*ha'u*	*mate*	*tiha*	*karik*	[...]
If	in.the.future	if	1S	die	PFV	maybe	[...]

'If I died one day [...]' (0056TD_CS_50_F_Pri_Tet)

6.3.4.4.1 Conditionals

In general, we can distinguish between two semantically distinct conditionals: reality conditionals and unreality conditionals. Reality conditionals refer to present, habitual and past situations; unreality conditionals refer to unreal situations, either imaginative which make us imagine that something could (have) happen(ed) or predictive which predict that something will occur (Thompson et al. 2007:255). It is quite common for languages to distinguish between these two conditionals by using special markers or morphology. However this distinction does not exist in Tetun.

On the other hand, as I have mentioned before (see section 6.3.4.1), 'if' clauses and 'when' clauses can sometimes have the same form. The conjunction *kuandu* 'when' can be used both in temporal clauses as well as in conditional clauses. This is not unusual, many languages in this region, like Indonesian and some languages of Papua New Guinea, do not make distinction between these two types of clauses (Thompson et al. 2007:257).

In Indonesian, conditional conjunctions like *kalua*, *bila* or *jika* can also be translated as 'when' in case "they indicate that the action of the main clause has occurred or will occur more than once, conditional on what is mentioned in the subordinate clause" (Sneddon et al. 2010:351) (308):

(308)　Indonesian (Sneddon et al. 2010:351):

Kalau	*Lebaran*	*dia*	*selalu*	*pulang*	*ke*	*Jawa.*
if	Lebaran	3S	always	return	to	Java.

'When it's Lebaran, he always returns to Java.'

Hull and Eccles (2005:213) also listed *bainhira* 'when' under conditional conjunctions, although Williams-van Klinken et al. (2002a:111) believed that *bainhira* could not be translated as 'if', claiming that it can only be used for events that happened or will happen. Just like Hull and Eccles (2005:214), I have found examples in which *bainhira* can have a conditional reading:

(309)　Tetun Prasa:

[...]	***bainhira***		*folin*	*mina*	*globál*	*kontinua*		*mantén*
	if		price	oil	global	continue		maintain

iha	*$30*	*ba*	*barríl*	*ida*	*maka*	*tinan*	*ne'e*	*governu*
LOC	$30	to	barrel	one	FOC	year	this	government

TL	*sei*	*hetan*	*de'it*	*osan*	*millaun*	*500* [...]
TL	FUT	get	only	money	million	500

'[...] if the global oil price continues to maintain itself at $30 per barrel, this year the TL government will only receive 500 million [...]'
(STL_March_2016)

Although very marginal, I have found examples of *se bainhira* 'if' in my corpus, too:

(310)　Tetun Prasa:

[...]	*se*	***bainhira***	*ita*	*hakarak*	*atu*
	if	when	2S.HON	want	IRR

servisu iha	*office*	*ruma*	*ita*	*presiza*	*loos*
work LOC	office	some	2S.HON	need	really

língua	*sira*	*ne'e.*
language	PL	this

'[...] if you want to work in an office, you really need (to speak various) languages.' (0014TD_JMB_31_F_Sec_Mak)

Negative conditionals in Tetun Prasa do not have a special form, however, we can find a negative conjunction *selae* 'if not', from *se* 'if' + *lae* 'no', that can be used when a negative conditional clause undergoes ellipsis:

(311)   Tetun Prasa:

Sira	*kuandu*	*ko'alia*	*bele*	*entende*	*malu,*	***selae,***
3P	when	speak	can	understand	RECIP	if.not

tanba	*liafuan*	*sira*	*hosi*	*Tetun*	*ofisiál*	*ne'e*
because	word	PL	from	Tetun	official	this

ninia	*fonte*	*ne'e*	*mai*	*hosi*	*lian*	*rua,*	*ho*
3S.POSS	source	this	come	from	language	two	with

portugés	*ho*	*Tetun Terik.*
Portuguese	with	Tetun Terik

'When they speak, they can understand each other, if not [if they cannot understand each other], it's because the official Tetun words come from two different languages, from Portuguese and from Tetun Terik.' (0054TD_NSG_51_M_Uni_TT)

6.3.4.5 Concession clauses

The last type of adverbial clauses that I am going to discuss are concession clauses. Unlike with other types, it seems like there is no native conjunction to introduce these clauses. According to Williams-van Klinken et al. (2002a:113), concession can be also marked with *bele* 'can' preceding either the predicate or the subject, as shown in (312), or by the clause-final *mós* 'also' in the initial clause. Based on my data, these two strategies seem to be very marginal in all varieties:

(312) Tetun Terik:

Ita	***bele***	*mate*	*ba*	*mais*	*ita-nia*	*naran*
2S.HON	can	die	to	but	2S.HON-POSS	name

la	*lakon.*
NEG	forget

'Even if you die, your name will not be forgotten.'
(0019TS_NA_54_M_Sec_TT)

The borrowed conjunctions that translate as 'although, even though' are:

(313) *biar* (from Indo. *biar* 'although')

mezmu and *mézmuke* (from Port. *mesmo que* 'although')

embora (from Port. *embora* 'although')

maski (from Port. *mas que* 'although')

The meaning of *mas que* 'although' has disappeared from modern Portuguese although its concessive value is included in the old dictionaries of Portuguese by Bluteau (1712-1728) and Silva (1789). In this form (*masque*), the conjunction has entered the Portuguese creoles spoken in South Asia, e.g. Sri Lanka and, in a slightly changed form, in East Asia (Macau Creole Portuguese *masqui*). The same concessive meaning has been also attested in Southeast Asian varieties of Creole Portuguese: Malacca Creole Portuguese *maski*, Macau Creole Portuguese *masqui*) and later was also borrowed by Malay *maski/meski* 'although'. It was also documented in non-Portuguese creoles, e.g. Chinese Pidgin English *maskee*, Tok Pisin *maskí*, Cape Dutch Pidgin *maskie*[169]. (Veiga & Fernández 2012:182-186). It seems probable that this conjunction was borrowed into Tetun Prasa either via one of the Portuguese creoles (Malacca/Macau) or via Malay.

Based on the analysis of the Malay/Indonesian borrowings in my corpus, the most frequently used concessive conjunction *maski* is used predominantly in written texts, as illustrated by (314), while *biar* was found only in one Tetun Prasa speaker, as shown in (315).

---

[169] *Maskin/masque* was also attested in Chabacano but Veiga and Fernández (2012:187) argue against the Portuguese origin and believe that the concessive conjunction was borrowed from Spanish *más que*.

(314) Tetun Prasa:

[...]	**maski**	sira	hasoru	difikuldade	barak,
	although	3P	meet	difficulty	many

maibé	sira	kontinua	hala'o	servisu	ho	di'ak
but	3P	continue	carry.out	work	with	good

iha	distritu.
LOC	district

'[...] even though they encounter many difficulties, they continue to carry out good work in the districts.' (STL_July_2013)

(315) Tetun Prasa:

[...]	maibé	ha'u	hakarak		halo	sira	bá	eskola.
	but	1S	want		make	3P	go	school

Biar		osan	la	barak.
although		money	NEG	many

'[...] but I want them to go to school. Even though we do not have much money.' (0056TD_CS_50_F_Pri_Tet)

As for the Portuguese borrowings, I was unable to find examples of *embora* and *mézmuke* in my corpus, not even in written government texts which usually yield the highest number of Portuguese borrowings. The adverbial *mezmu* was found only in one Tetun Prasa speaker, highly proficient in Portuguese:

(316) Tetun Prasa:

[...]	maibé	ha'u	rasik	la	fiar	ho	ha'u-nia
	but	1S	self	NEG	trust	with	1S-POSS

kualifikasaun	**mezmu**	ha'u-nia	kualifikasaun	di'ak.
qualification	although	1S-POSS	qualification	good

'I myself did not trust my qualification although my qualification was good.' (0057TD_LAOB_30_M_Uni_Tet)

6.3.4.6 Summary

In this section, I have tried to account for the influence of borrowed conjunctions on Tetun syntax. As we have seen, native Tetun conjunctions often coexist with Portuguese and Malay/Indonesian borrowings, but it is difficult to tell what made certain borrowings more prominent than others. It is also interesting to see the complex conjunctions containing one native and one borrowed element.

For a better understanding of the frequency of native and borrowed conjunctions in each type of adverbial clauses, see Table 35:

Adverbial clause	Conjunction	Origin	Written (TP)	Oral (TP)	Oral (TT)
Time	*bainhira*	native	116	14	5
	kuandu	Port.	44	22	22[170]
	bainhira kuandu	native + Port.	0	0	1
Purpose	*atu*	native	45	9	10
	para	Port.	16	21	8
	hodi	native	241	18	24
	para atu	Port. + native	14	6	3
	para hodi	Port + native	0	1	0
	atu hodi	native + native	9	1	6
Reason	*tanba*	native	146	117	71
	porké	Port.	0	0	0
	komu	Port.	0	4	5
Condition	*se*	Port.	10	26	10
	karik	native	10	5	14
	se karik	Port. + native	1	0	10
	se + () + karik	Port. + native	0	2	3
Concession	*bele*	native	3	0	6
	maski	Port./Malay	39	4	1
	biar	Indo.	0	2	0
	mezmu	Port.	0	5	0

Table 35 Distribution of subordinating conjunctions in adverbial clauses

---

[170] Majority of occurrences (18) occurred in one single speaker (0039TV_MVJ_51_M_Uni_TT) with university education and proficiency in Portuguese.

6.4 Conclusions

As I have documented in this chapter, the influence of Portuguese and Malay/Indonesian on Tetun Prasa has been intensive enough to affect the structure of the language. New constructions have been thought to emerge (like the Portuguese gender and number agreement, the copula and the passive voice) and many native conjunctions now coexist with their Portuguese and Malay/Indonesian counterparts. Only time will show if the borrowed structural features will eventually replace the native ones and if features typical for written, journalistic texts will eventually make their way into the spoken language and more conservative dialects of Tetun Terik.

7. PHONOLOGY

Austronesian languages are known to have a more limited phonemic inventory than other world languages, which usually have between 20-37 segments (Maddieson 1984:7). According to Blust (2013:169), 90% of the Austronesian languages only contain around 15-20 consonants and 4-5 vowels, and the total phoneme inventory is between 19-25 segments. Among the Timorese languages, the one with the largest phonemic inventory is Waima'a (31 consonants and 5 vowels) and the one with the most limited phonemic inventory is Dawan (now officially called Uab Meto, spoken in West Timor) with 12 consonants and 7 vowels (Blust 2013:194). Tetun's phonology is often referred to as simple, however, authors working on this language and its varieties do not seem to agree on exactly how many vowels and consonants there are. This is because of the long-lasting influence of Portuguese and Malay/Indonesian, which gave Tetun Prasa a considerable number of foreign phonemes. In this chapter, I will discuss whether these foreign elements can be considered a part of the Tetun Prasa phonemic inventory.

7.1 Methodology

Since there is a lot of variation, I will only list one or two possible pronunciations for each word. As a default, I will use the educated Portuguese-influenced pronunciation for all Portuguese loanwords, although I am aware that a considerable part of Tetun Prasa speakers opt for a more Tetun-like pronunciation. This allophonic variation will be discussed in more detail in sections 7.4 and 7.5. In sections 7.5.2 and 7.5.3, I will present a statistical analysis of the realization of borrowed phonemes[171] and all their possible allophones. To deal with under- and overrepresentation and for easier interpretation, I included both raw numbers (total number of occurrences) and percentages.

---

[171] I refer to all borrowed segments as phonemes, following the standard practice by authors in earlier works on Tetun Prasa phonology. However, I did not test the existence of minimal pairs to decide whether these borrowed segments, especially the palatals and palate-alveolar fricatives, actually represent new phonemes or are just allophones of native phoneme, a discussion already included in Albuquerque (2011).

As for the actual analysis, I uploaded the word lists for each phoneme to Excel and accounted for all the occurrences and variations in pronunciation (Image 8). I did this for individual speakers as well as for individual words.

Image 8 Analysis of borrowed phonemes in Excel

The next step was to look at the possible variables influencing the variation in pronunciation. I combined the phonological data with the sociolinguistic background of every speaker and I was able to create statistics about the variables that could have influenced the realization of these phonemes (see Figure 7). I accounted for age, level of education, period in which the speakers were educated and their proficiency in Portuguese.
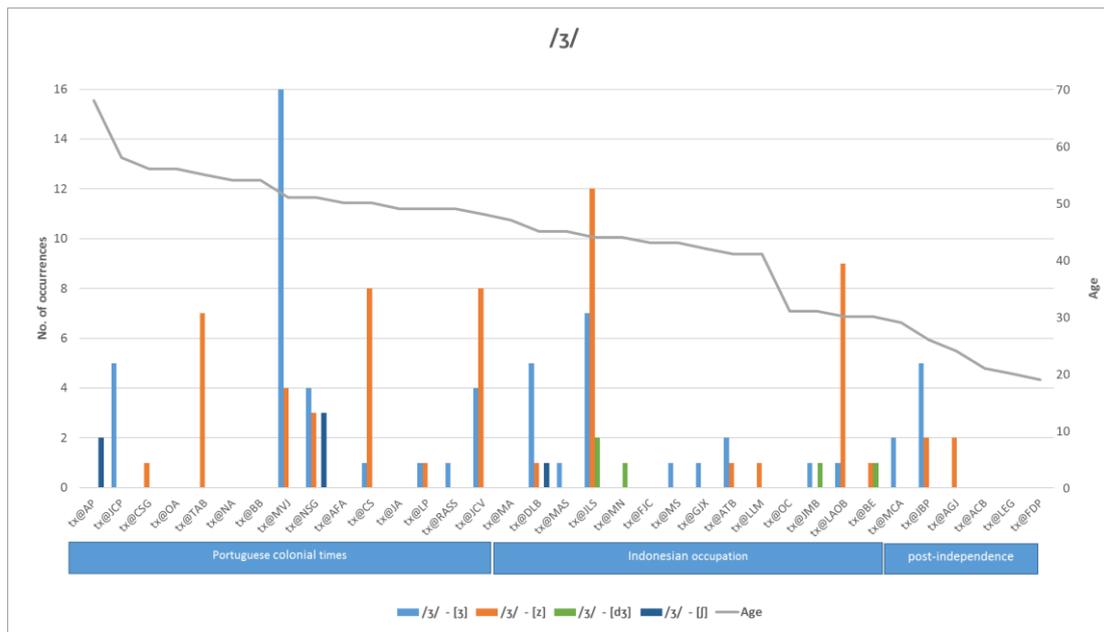
Figure 7 Analysis of borrowed phonemes based on variables

To refer to my speakers, I will use both the full form laid out in section 3.2 but also a shorter form which consists of tx@ + initials of the speaker, as shown in Figure 7. This is necessary due to the limited space in the figures. For the full forms, see Speakers in the List of abbreviations (pg. 19).

7.2 Native phonemes

To better understand the origins and development of the current phonemic inventory of Tetun Prasa, let us look at the various influences Tetun has undergone throughout the centuries. Arthur Capell (1944c) and Geoffrey Hull (2000b) both tried to reconstruct the proto-language from which Tetun originated and determine which phonemes were native to Tetun. Capell (1944c:19) called his proto-language Original Indonesian (OIN) and said that in order to classify Timorese languages as 'Indonesian'[172], "one must demonstrate that their word-store is connected by definite phonetic laws with the Original Indonesian." At the same time, he acknowledged that there were certain regional linguistic types and he focused on sound laws and consonant changes in languages of Kupang, West Timor and Portuguese East Timor. According to Capell (1944c:25), the 'Indonesian' languages of Timor-Leste were a relatively homogenous group, although every language presented its own peculiarities. For example, while, in general, certain final consonants were lost in many Timorese languages, in Tetun they were often kept or replaced. Looking only at the Tetun data, we can determine that according to Capell's reconstruction of OIN[173] and its reflexes in Tetun[174] (see Table 36), there were 11 native consonants: /t/, /d/, /k/, /s/, /m/, /n/, /f/, /r/, /h/, /l/, /w/. The list is

---

[172] Capell (1944b:311) distinguished between 'Indonesian' and 'non-Indonesian' languages spoken in Timor. By 'Indonesian' he referred to languages that are nowadays termed 'Austronesian', 'non-Indonesian' languages, on the other hand, are sometimes also referred to as 'Papuan' or 'non-Austronesian' languages.

[173] I decided not to use the IPA symbols when describing the OIN phonemes found in Capell (1944c:22) but rather follow the author's own representations since, in a few cases, I was uncertain about the quality of some OIN sounds. Just to make it clear, I do not believe there were any ejectives in OIN, since this type of consonants is rarely found in Austronesian languages (in fact, we only know of one, Yapese), so I doubt /d'/, /t'/ and /g'/ represent ejectives.

[174] It is unclear which variety Capell (1944b:313) looked at. He mentioned that the northern Tetun (although incorrectly labelled as Tetun Terik) was somehow corrupt and that the southern Tetun was taken as standard. We can thus assume the author chose the southern variety in his description, which is reinforced by the fact that he lists /w/ as one of the phonemes in Tetun.

almost identical to the one Hull (2000b) proposed, using the reconstruction of Proto-Timoric, however Capell (1944c) did not account for the development of Tetun /b/ and /'/ which Hull (2000b) considered native consonants too. This might be because, as Capell (1944c:22) himself admitted, he did not have adequate material to come to a conclusion on certain sounds. For example, he noticed OIN /g/ was adopted as a glottal stop in the Roti language, but did not find any further evidence of this change in other languages.

Native consonants	Examples
OIN *p > T /h/	*ə(m)pat > haat* 'four' *pitu > hitu* 'seven'
OIN *b > T initial /f/ > T medial /h/	*bulan > fulan* 'month' *libu > rihun* 'thousand'
OIN *t > T /t/ except as final	*mata > matan* 'eye'
OIN *d > T /r/	*dəŋəɣ > rona* 'listen'
OIN *ɗ > T /d/	*d'alan > dalan* 'way' *hud'an > udan* 'rain'
OIN *k > T /k/ or lost	*kəbav > karau* 'buffalo' *kita > ita* '1PI'
OIN *g > T not documented	
OIN *m > T /m/	*lima > liman* 'hand'
OIN *n > T /n/	*ina > inan* 'mother'
OIN *ŋ > T /n/	*haŋin > anin* 'wind'
OIN *ť > T /s/	*t'iva > sia* 'nine' but exception: *əťa > ida* 'one'
OIN *y in final /-ay/ > T /e/ medial > T lost	*vayɣ > wee/bee* 'water' (in this case, the final syllable was lost)
OIN *ɣ > T usually lost	*ɣumah > uma* 'house'
OIN *h > T lost	*haŋin > anin* 'wind'
OIN *g' > T /r/ or /l/	*pig'a > hira* 'how many' *ag'i > alin* 'younger sibling'
OIN *l > T /l/	*lima > liman* 'hand'
OIN *v initial > T /w/ medial > T lost	*vayɣ > wee/bee* 'water' *duva > rua* 'two'

Table 36 Reconstruction of the OIN phonemic inventory and its reflexes in Tetun (Capell 1944c:25-27)

Hull (2000b), in his study on the historical phonology of Tetun, took a similar approach. He connected the modern Tetun phonemes with their reconstructed Proto-Timoric (PT) and Proto-Moributonic (PMB) etyma (see Table 37). Originally, Proto-Timoric had a consonant system of 28 phonemes. Due to phonetic changes, Tetun lost

dental, palatal, uvular and pre-nasalized phonemes and was left with only 13 native consonants: /h/, /f/, /m/, /n/, /w/, /t/, /r/, /s/, /d/, /l/, /k/, /b/ /'/. Tetun Prasa, due to the loss of the glottal stop and assimilation of /w/ to /b/, has even fewer, only 11 native consonants (Hull 2000b:167-189). Hull (2000b), unlike Capell (1944c), considered /b/ a native consonant which developed from PT *mp, *mb and medial *m.

As for the vowel inventory, Tetun inherited the PMB and PT system of 5 vowels which could be either tonic or atonic: /a/, /e/, /i/, /o/, /u/ (Hull 2000b:159).

PT reconstructions > T reflexes	Examples
*p > /h/	*pija > hira 'how many'
*b initial > /f/ medial > /h/	*bulan > fulan 'month' *babuy > fahi 'pig'

*m initial and medial > /m/
  medial > sometimes /b/
  final > /n/
	*manu > manu 'fowl'
	*ʀamut > amut > abut 'root'
	*dalem > laran 'inside'

*mp > /b/	*mpesi > besik 'near'
*mb > /b/	*mbuca > busa 'cat'
*w > TT /w/ > TP /b/	*wahiʀ > *wai > TT wee 'water' > TP bee 'water'
*t initial and medial > /t/ final > unreleased [t̚]	*tasi > tasi 'sea' *kulit > kulit [kulit̚] 'skin'
*d > /r/	*duwa > rua 'two'
*s > /s/	*siaw > sia 'nine'
*n > /n/	*manu > manu 'fowl'
*nt > /t/ *nt > *nd > /d/	*lante > leten 'top' *punti > hudi 'banana'
*nd > /d/	*inda 'piece' > ida 'one'
*ns > /s/	*ʀansa > *asa > aas 'high'
*r > /r/	*kakuraw > kakorok 'throat'
*l all positions > /l/, if final > sometimes /r/	*saliw 'exchange' > selu 'pay' dikul > dikur 'horn'
*c > /s/ or /t/	*mbuca > busa 'cat' *meca 'watercourse' > mota 'river'
*z realized as /ʝ/ > /d/ initial > lost	*zalan > dalan 'road' *zaquat > aat 'bad'

*nz > /nd/ only one example	*panziʀi > handii 'visit'
*j /j/ > /ʎ/ > /r/	*qalejaw > *qaleʎaw > *'aleraw > loron 'sun' *pija > hira 'how many'
*k initial > /k/ or lost medial > [ʔ]	*kempuŋ > kabun 'stomach' *kami > 'ami > ami '1PE' *aku > ha'u '1S'
*g > /k/ > sometimes [ʔ]	*gege 'move' > koko 'try' *wagay > *waki > wa'in 'big'
* ŋ > /n/	*deŋeʀ > rona 'hear'
*ŋk > /k/	*baŋku > baku 'beat'
*ng > /k/	*laŋgadaw 'gourd' > lakeru 'pumpkin'
*q /q/ > /h/ > lost before /a/ and /u/	*qentaw > hotu 'finish' *quaʀ > *hau > au 'bamboo'
PMB *ʀ [ʁ] > PT [ɦ] > lost	*daʀa > raan 'blood'
*' [ʔ] medial > /h/	*ka'en > han 'eat'
*h > lost	*hisi 'flesh' > isin 'body'

Table 37: Proto-Timoric reconstructions and reflexes in Tetun (Hull 2000b)

When it comes to the realization of native consonants, there is very little allophonic variation. Except for [w] ~ [b] variation in Tetun Terik and Tetun Prasa, and possible loss of the glottal stop in Tetun Prasa, which are discussed in sections 7.5.2.4 and 7.5.2.3 respectively, there is no phonetic adaptation of native consonantal phonemes.

7.3 Malay-derived phonemes

At the beginning of the 15th century, the first Malay traders reached the northern coast of Timor, where, most probably, Tetun was already spoken. Since Malay was used as a trade language, it possibly loaned numerous words to Tetun, covering semantic domains like business and maritime trade. This is also evidenced by the phonological analysis of Malay loanwords in Tetun Prasa. For example, Malay [w] was substituted by [b] in the early Malay borrowings because [w] was probably not part of the phonemic inventory of Tetun spoken on the northern coast (Hull 2005a:102). Hull (2005a) further

supported this claim with the fact that many of these words (especially maritime and fishing terms) were unknown in the inland dialects. At this point, we can't yet talk about 'Tetun Prasa', which has started to really develop only after the capital was moved to Dili in 1769.

However, looking at the phonology of Malay borrowings, Hull (2005a:102-103) distinguished between Malay words borrowed at an early stage, between the 15th and 17th (so called 'old Malayisms') century and Malay loanwords borrowed later (so called 'new Malayisms'). The difference between these two groups is that foreign phonemes borrowed at the early stage were assimilated and replaced by native Tetun Prasa ones.

Phonemes found in Malay that were replaced by native Tetun phonemes in the early borrowings included: <c> /ʧ/, <j> /ʤ/, <ng> /ŋ/, <ngg> /ŋg/, <ny> /ɲ/, <w> /w/ and <r> /ʁ/. They were replaced in the following way:

Malay > Tetun	examples (Hull 2005a)
/ʧ/ > /s/	*cucuk* 'to prick' > *susuk* 'mosquito'
/ʤ/ > /d/	*jerok* 'citrus fruit' > *derok* 'lemon/lime'
/ŋ/ > /n/	*lenga* > *lena* 'sesame'
/ŋg/ > /k/	*pinggan* > *bikan* 'plate'
/ɲ/ > /n/	*kunyit* > *kinur* 'turmeric'
/w/ > /b/	*kawin* 'to marry' > *kaben* 'legally married spouse'
/ʁ/ > /r/	*darat* [da.ʁat] > *raat* [ra.at] 'shore'
/p/ > /b/ or /h/	[b] in initial and medial position: *dapur* > *dabur* 'kitchen' [h] in the very early borrowings in medial position: *tepi* > *tehen* 'edge'

Table 38 Malay consonants adopted at an early stage by Tetun

Apart from new sounds, Malay also introduced wider possibilities for the distribution of certain consonants. Originally, only these seven consonants were allowed in final position: /t/, /s/, /n/, /r/, /l/, /w/, /k/ and in the earliest borrowings, all foreign phonemes were replaced with these native counterparts. As for initial and medial (intervocalic) position, the same phenomenon happened (Hull 2005a:103).

Thomaz (1974:194-195) also noticed the tendency of Tetun (and other Timorese languages) to reduce the number of plosives (and affricates) and create an asymmetry within the consonantal system:

	t	k	'
b	d		

p	t	c	k
b	d	j	g

Table 39 Tetun (on the left) versus Malay (on the right) plosives and affricates (Thomaz 1974:196-197)

The only exception, in which a plosive/affricate was preserved, was the replacement of the Malay affricate /dʒ/ by the Tetun plosive /d/, e.g. M. *hujan* > T. *udan* 'rain'. Thomaz (1974:196) concluded that the consonantal system of Tetun resembled more the Polynesian languages, which showed the same tendency (Tahitian only has three plosives: /p/, /t/ and /'/) but unlike them, it also contained voiced plosives (/b/ and /d/). However, compared to Malay and Indonesian, Tetun presented a more restricted consonantal system.

When the influence of Malay began, Tetun only had five vowels: /a/, /e/, /i/, /o/, /u/. On the other hand, Standard Malay language had six vowels: /a/, /e/, /i/, /o/, /u/ and /ə/. Malay /ə/ (or a pretonic /e/), was thus, in many cases, replaced by /a/, as shown in (317). However, Hull (2005a:104) mentioned that this change was already happening in Malay itself and was a frequent phenomenon in other varieties of Malay, namely Kupang Malay and Ambonese Malay, as is illustrated in (318).

(317)   M. *keladi* 'yam' > T. *kaladi* 'yam eaters'
        M. *serani* > T. *sarani* 'Christian'

 (318)   (van Minde 1990:73)

        M. *celana* > AmbM. *calana* 'trousers'
        Port. *pepino* > AmbM. [paˈpi.ɲu] or [paˈpi.ɲo] 'cucumber'

This leaves us wondering whether the Malay loanwords were borrowed already with the low vowel /a/ via one of these varieties in which the shift happened first or the vowel was lowered and fronted only after it entered Tetun. Considering that Hull (2005a:84) himself suggested that Tetun spoken on the northern coast was, later in the 15th century, in contact with Ambonese Malay, which also contains only a five-vowel system (Paauw 2013), it is possible that Malay words were already borrowed with /a/.

On the other hand, tonic [ə] was replaced by [o] or [e] in Tetun (Hull 2005a:105):

(319)   M. *tempur* > T. *tobur* 'to hit against'

M. *besi* [bəsi] > T. *besi* [besi] 'iron'

With time (roughly the beginning of the 18th century) and with more and more locals speaking the local variety of Malay, Malay loanwords started to be adopted in their original form and new phonemes entered Tetun: [g], [p], [ʤ], [ŋ], [ŋg], as shown in Table 40 (Hull 2005a:103). This was even more intensified during the Indonesian occupation (1975-1999) when the Indonesian language was forced onto the whole population and Indonesian borrowings were pronounced in their modern standard way, including <c> as [ʧ]. So far, I was unable to find any Indonesian borrowings containing <ny> [ɲ], which, of course, does not mean that they do not exist.

Malay/Indonesian > Tetun	Examples (Malay examples from Hull 2005a)
[g] > [g]	M. *sagu hati* 'cordial sago' > T. *saugati* 'free of charge' Ind. *gudang* > T. *gudang* 'warehouse'
[p] > [p]	M./Ind. *bapak* 'father' > T. *bapa* 'dad' or 'Indonesian person'
<ng> [ŋ] > [ŋg]	M. *mengemam* 'to mumble' > T. *mangame* 'to rave' Ind. *warung* > T. *warung* [waruŋg]'Indonesian-style restaurant'
<ngg> [ŋg] > [ŋg]	M. *runggu-rangga* 'bristly' > T. *rungu-ranga* 'disturbance'
<ngk> [ŋk] > [nk]	Ind. *bengkel* > T. *bengkel* [benkel] 'workshop'
[ʤ] > [ʤ]	M. *jambu* 'rose apple' > T. *jambua* 'pomelo'
M. [ɲ] > [n] Ind. [ɲ] > not attested	Ind. not attested
M. [ʧ] > T. [s] Ind. [ʧ] > T. [ʧ]	(M. see Table 38) Ind. *capcai* > T. *capcai* 'stir-fried vegetables'

Table 40 Malay and Indonesian consonants adopted at a later stage by Tetun

One variety of Tetun that has preserved more Malay borrowings (especially from the non-standard varieties spoken in Timor) than any other is Tetun Fehan, also due to the fact that it is spoken in West Timor and has not been in much contact with Portuguese. According to van Klinken (1999:18), there was a tendency to omit the final glottal consonants /h/ and /'/ in Malay loanwords (e.g. M. *sekolah* 'school' or M. *pa'* 'mister') and to replace the final [ŋ] with [n] (e.g. M. *senáŋ* 'happy'). However, the degree of the

phonological assimilation of these loans varies from speaker to speaker (van Klinken 1999:27).

7.4 Portuguese-derived phonemes

Although Malay influence was significant especially in the early period, the language that has introduced the majority of foreign phonemes in Tetun Prasa is undoubtedly Portuguese. Although the number of Portuguese nationals or Portuguese-speaking natives was never considerably high, Portuguese words started entering Tetun Prasa and other native languages with the establishment of Portuguese administration and trade, later also with the introduction of Christianity. Portuguese missionaries resorted to introducing many Portuguese loanwords in order to introduce various concepts of their faith and with these Portuguese words also came Portuguese phonemes. The locals have often struggled to pronounce some of them and subsequently, there has always been a lot of variation. With the large number of Portuguese borrowings, phonetic realizations that were formerly introduced by Malay (/p/, /g/, and /ɲ/) finally found a steady place in Tetun Prasa phonology (Hull 2005a:103). Apart from newly introduced consonants /ʒ/, /ʎ/, /ʁ/, /v/, /ʃ/, and /z/ (see Table 41), Tetun Prasa also adopted nasal vowels [ã], [ẽ], [ĩ], [õ], [ũ] and diphthongs [aj], [ej], [oj], [uj], [aw], [ew], [ow].

Portuguese > Tetun Prasa	Examples
/ʒ/ > [ʒ] or [z]	*ajuda* 'help' [aˈʒu.da] or [aˈzu.da]
\<lh\> /ʎ/ > \<ll\> [ʎ] or [l]	*konsellu* 'advice' [konˈse.ʎu] or [konˈse.lu]
\<nh\> /ɲ/ > \<ñ\> [ɲ] or [n]	*señora* 'lady' [seˈɲo.ra] or [seˈno.ra]
\<rr\> or \<r\> in word-initial position and following /n/, /l/, /z/, or /ʒ/ > \<rr\> [ʁ] or [ɾ]	*rádiu* 'radio' [ˈʁa.di.u] or [ˈɾa.di.u]
/v/ > [v] or [b]	*avó* 'grandfather' [aˈvo] or [aˈbo]
/ʃ/ > [ʃ] or [s]	*depois* 'after' [deˈpoiʃ] or [deˈpois]
/z/ > [z] or [s]	*azúl* 'blue' [aˈzul] or [aˈsul]

Table 41 Portuguese consonants adopted into Tetun Prasa

7.5 Phonology of Tetun Prasa and other varieties

Pronunciation in Tetun Prasa varies considerably due to several factors. One of them is the amount of foreign words that Tetun Prasa speakers adopted and use in everyday speech. Originally, Tetun pronunciation was similar to that of other Austronesian languages. As Dores (1907:9) put it: "A pronuncia do teto é suave como a do malaio, e comquanto não me pareça tão harmoniosa, ainda assim encontrei-lhe varios pontos de contacto."[175] Nowadays, however, pronunciation varies from speaker to speaker. Corte-Real (1990:54) studied the phonological features of Tetun[176] and divided speakers of Tetun into two categories: non-Portuguese-influenced speakers (NPI), mostly illiterate villagers, and Portuguese-influenced speakers (PI). In his thesis, the author noticed that the consonant inventory for each of the groups was remarkably different, with the NPI speakers lacking the Portuguese-borrowed phonemes.

Similar observations were made by Hull and Eccles (2005:225-226) who identified five main 'varieties' of Tetun spoken in Timor-Leste:

- two 'high varieties': an 'acrolectal' variety of Tetun Prasa spoken in Dili (spoken by the Dili inhabitants who are also fluent in Portuguese) and an 'acrolectal' variety of Tetun Prasa spoken outside of Dili (spoken by other Timorese around the country who are also fluent in Portuguese);
- one 'middle variety': a 'mesolectal' variety of Tetun Prasa spoken by educated people not fluent in Portuguese and by many people fluent in Portuguese living outside of Dili (this variety lacks certain Portuguese diphthongs and foreign consonants);
- one 'low variety': a 'basilectal' variety of Tetun Prasa used by uneducated people who do not speak Portuguese and, in the case of speakers of other native languages, who are often not completely fluent in Tetun either (in

---

[175] "The pronunciation of Tetun is suave as that of Malay, and although I do not find it so harmonious, I still have managed to find various contact points."

[176] Corte-Real did not specify which variety he was concerned with; he referred to the language he studied as Tetun: "This study, however, limited itself to Tetun widely spoken in the eastern part of the island, East Timor, and dealt only with the modern Tetun which suffered some influence of Portuguese language, yet understood and recognized as Timorese language or the so-called Tetun." (Corte-Real 1990:13). From this definition, it seems like Corte-Real is referring to Tetun, the lingua franca, which is, in our definition, referred to as Tetun Prasa.

its most conservative form, this variety lacks all the Portuguese diphthongs and all the foreign consonants).

The last 'variety' is the dialects of Tetun Terik, spoken in the areas of Balibo, Suai, Samoro, Alas, Lacluta, Luca and Viqueque. In addition to Portuguese, the five varieties were also influenced by the Malay and Indonesian languages, especially with respect to the lexicon.

In my analysis, I was interested in determining whether variables like proficiency in Portuguese and urban/rural environment had any influence on the pronunciation of foreign phonetic realizations. I added some more variables: age, the period in which the speakers were educated (Portuguese colonial times, Indonesian occupation, post-independence) and level of education. I looked at stress, syllable structure but also analysed Tetun Terik phonetic realizations, foreign consonants and some vowel variation.

7.5.1 Stress and syllable structure

In native words, stress is always on the penultimate syllable (most of the lexical words have two to four syllables):

(320) *hemu* 'to drink'   [ˈhe.mu]

       hamutuk 'together'   [ha.ˈmu.tuk]

However, in borrowings, stress can fall on the final, penultimate or antepenultimate syllable. Portuguese loanwords keep their original stress, which is, in the official orthography, marked with the acute accent if it does not fall on the penultimate syllable:

(321) *devér* 'duty'   [de.ˈve:r]   (from Port. *dever* 'duty')

       máskara 'mask'   [ˈma:ʃ.ka.ra]   (from Port. *máscara* 'mask')

       pasaporte 'passport' [pa.sa.ˈpor.te] (from Port. *passaporte* 'passport')

I could also hear speakers misplacing the stress on the penultimate syllable possibly due to the fact that they are unaware that certain words are borrowings with a

different stress pattern (see section 2.10). This case could be possibly considered a part of the process of phonological adaptation:

(322) *balu   ema   índia*
       [ˈbalu ˈema   inˈdia]
       some   person India
       'some are Indians' (0050TV_JCV_48_M_Sec_TT)

Syllable structure in Tetun Prasa is very different from other varieties. While in Tetun Fehan the syllable structure is (C)V(C) (van Klinken 1999:14) with almost non-existent consonant clusters (except for word-medial consonant clusters in compounds and reduplicated words and clusters with kC-), in Tetun Prasa previously inadmissible consonant clusters were introduced with loanwords of Portuguese and, to a lesser extent, of Malay origin. The syllable template of Tetun Prasa words is (C)(C)V(V)(C), but in the case of borrowings, the longest consonant clusters can contain, although sparsely, up to three consonants in word-initial and word-medial position, as shown in (328). No consonant clusters are allowed in the word-final position.

In a broader sense, a consonant cluster can be defined as an "uninterrupted sequence of two or more consonants within some well-defined unit of language, such as a syllable, word or a phrase" (Vennemann 2012:12). Although in most cases in Tetun Prasa the clusters go over the syllable boundary (so-called intersyllabic clusters), I will only consider those that are found within a syllable, as in (323a). Some word-initial consonant clusters containing three segments are only phonetical, not phonological, and are the result of deletion of [e] or [i], as happened in (323b). The same phenomenon exists in colloquial European Portuguese and the initial vowel was also often deleted in Portuguese borrowings in restructured varieties of Portuguese and Malay.

(323) *estranjeiru* 'foreigner'   a.   [eʃ.t̪ran.ˈʒej.ru] or [es.t̪ran.ˈzei.ru]

                                    b.   [ʃt̪ran.ˈʒej.ru]

Consonant clusters were originally more prominent in the dialects of Tetun Terik, mainly due to the subject marking on verbs. As I have explained in section 5.2.1, when a verb is marked for a subject, the initial consonant is either replaced or preceded by a prefix (*k-* for 1S, *m-* for 2S, *n-* for 3S and *r-* for 3P):

(324) Tetun Terik

ha'u krona 'I 1S.hear' from *rona* 'hear'

The most common consonant cluster in native words contains initial /*kC*-/. The second element can be any consonant but a glottal or /*k*/ itself. Historically, /*k*-/ might have been a derivational prefix attached to a C-initial root. This word-initial cluster is, however, very unstable. Many Tetun Terik and Tetun Fehan words now occur without it, especially in casual speech (e.g. *ktodan* vs. *todan* 'difficult', *kdook* vs. *dook* 'far'). Consonant clusters can be also sometimes broken up through the insertion of an epenthetic vowel (Williams-van Klinken et al. 2002a:13):

(325) *metru* 'meter'   [ˈmet.ru] ~ [ˈme.tu.ru]
       (0050TV_JCV_48_M_Sec_TT)

To facilitate the pronunciation, speakers of Tetun Prasa would not only insert a vowel to break a consonant cluster but also often delete the initial /*k*-/, as is the case of (326), though some common Tetun Terik consonants clusters, e.g. /*kl*/, /*kn*/, /*kr*/, are still preserved in Tetun Prasa, as shown in (327) (Williams-van Klinken 2010b).

(326) *kmanek*   'wonderful'   [ˈma.nekˀ]

(327) a. *klaran*   'centre' (in opposition to *laran* 'inside')

       b. *knaar*   'duty'

       c. *kraik*   'downhill'

On the other hand, consonant clusters are often heard in fast speech, when certain unstressed vowels are deleted, as shown above in (323b). In careful speech, one can find initial and medial consonant clusters made of up to three consonants only in borrowings, as in (328). Syllable-initial consonant clusters containing two consonants are very frequent, especially in Portuguese borrowings, in which C1 can contain /p, b, t, d, k, g, f/, followed by C2 /ɾ/ or /l/, as in (329).

(328) a.   CCC-

       stres [ˈstres] 'stressed' (from Eng. *stress* probably via Port.)

       b.   -CCC-

       abstratu [abˈʃtra.tu] 'abstract' (from Port. *abstrato*)

administradór [ad.mi.ni.ʃtra.ˈdor] 'administrator, officer' (from Port. *administrador*)

(329) a.   CC-

brutalidade [bru.ta.li.ˈda.de] 'brutality' (from Port. *brutalidade*)

droga [ˈdro.ga] 'drug' (from Port. *droga*)

b.   -CC-

estrada [eʃ.ˈtra.da] 'road' (from Port. *estrada*)

estraga [eʃ.ˈtra.ga] 'destroy' (from Port. *estragar*)

7.5.2 Consonants

Hull (1993:xiv-xviii) asserted that the maximal consonantal system of Tetun Prasa has 22[177] phonemes: 11 univalent with only one possible pronunciation and 11 plurivalent with alternative pronunciations, as shown in Table 42.

univalent	b, f, h, l, m, n, s, t, d, k, r
plurivalent	p, v, g, j, x, z, rr, ll, ñ, w, ', (d)

Table 42: Consonant inventory in Tetun Prasa (Hull 1993:xiv-xviii)

However, the consonantal system of Tetun Prasa, and especially the Malay- and Portuguese-derived consonants, still suffer a lot of variation. According to Williams-van Klinken et al. (2002a:8), the changes in pronunciation are caused by the following social variables: the level of education; social class; and urban/rural place of living. Educated people living in the urban areas, who speak both Tetun Prasa and Portuguese, show the widest inventory of phonemes. On the other hand, many speakers, especially those not proficient in Portuguese, replace foreign phonemes with native Tetun ones. In the analysis I conducted, this generalization was shown not to be completely true, considering variables like level of education, proficiency in Portuguese and age. For more detailed analysis of individual phones, see subsections of this section.

---

[177] The official publication of the National Institute of Linguistics *Ortografia Padronizada do Tétum* (2004) only lists 21 consonants, leaving out – for no logical reason – the phoneme /f/.

Variation concerning foreign consonants can be summarized as follows:

Phoneme	*Example*
/v/ > [v] or [b]	*avó* 'grandfather' > [aˈvo:] or [aˈbo:]
/ʒ/ > [ʒ] or [z] or [dʒ] or [d] or [ʃ]	*ajuda* 'help' > [aˈʒu.da] or [aˈzu.da] or [aˈdʒu.da] or [aˈʃu.da] *jerasaun* 'generation' > [ʒe.ra.ˈsãũn] or [de.ra.ˈsãũn]
/ʃ/ > [ʃ] or [s]	*eskola* 'school' > [eʃ.ˈko.la] or [es.ˈko.la]
/ʎ/ > [ʎ] or [il] or [l] or [j]	*traballu* 'work' > [tra.ˈba.ʎu] or [tra.ˈbai.lu] *fillu* 'son' > [fi.ˈʎu] or [fi.ˈlu] or [fi.ˈju]
/ɲ/ > [ɲ] or [n]	*señór* 'mister' > [se.ˈɲor] or [se.ˈnor]
/p/ > [p] or [b]	*pregu* 'nail' > [pregu] or [breku] (Hull & Eccles 2005:239)[178]
/g/ > [g] or [k]	*gaveta* 'drawer' > [gaveta] or [kabeta] (Hull & Eccles 2005:239)
/z/ > [z] or [s] or [ʒ]	*uza* 'use' > [ˈu.za] or [ˈu.sa] or [ˈu.ʒa]
/r/ > [ʁ] or [ɾ] or [ʀ]	*rona* 'hear' > [ˈɾo.na] or [ˈʀo.na]
/d/ > [d] or [ɾ]	*daudaun* 'PROG' > [da(u).ˈda.un] or [ra.ˈra.un]
/w/ > [w] or [b]	*wainhira* 'when' > [wa.in.ˈhi.ra] or [ba.in.ˈhi.ra]
/'/ > [ʔ] or unrealized	*ha'u* '1S' > [ˈhaw] or [ˈhaʔu]

Table 43: Different realizations of Tetun Prasa plurivalent consonants (unless indicated otherwise, all examples are taken from my oral corpus)

Corte-Real (1990:55-59) called these alternations distortions and listed (1990:6) one more important variable: multilingualism. L2 speakers of Tetun Prasa might find it easier to pronounce those 'foreign' phonemes that are also found in their mother tongues. For example, Galolen has 13 native phonemes, including /w/, /g/ and /'/ (Hull 2003c:4), Makasae has 14 native consonants including /w/, /f/, /g/ and /'/ (Hull 2005b:5), Baikeno has 12 native consonants including /p/, /f/, /ʐ/ and /'/ (Hull 2003b:5) and Fataluku has 16 native consonants, including /p/, /'/, /f/, /v/, /j/, /cç/ and /ɟj/, although it does not contain any voiced stops (Hull 2005c:6). I have found a couple of examples of this kind of variation, which are discussed in section 7.5.2.

---

[178] I did not account for [p] ~ [b] nor [g] ~ [k] variation in my oral corpus.

Albuquerque (2011:85-86) believed that some of the borrowed phonemes have never been fully incorporated into the Tetun Prasa phonology. Unlike /p/ and /g/, which are now part of the phonemic inventory of most of the speakers, /v/ and /z/ are still found only in the recent Portuguese borrowings. According to the author, sounds like [ʃ], [ʒ], [ɲ], [ʎ], which were borrowed from Portuguese, do not make part of the Tetun Prasa phonemic inventory since they are found only in Portuguese borrowings. They present various allophones and only a small fraction of highly educated people pronounce them according to the European norm. This is why the palatal consonants are considered only allophones of their respective alveolar counterparts and the author assigned two sociolinguistic factors to them: [+older age] [+educated] (Albuquerque 2011:87-88). I believe that these phonetic realizations *should* be considered part of the phonemic inventory of Tetun Prasa speakers which is documented by my analysis in the subsections of this section.

Variation can also be found in other varieties of Tetun. In Tetun Fehan, different realizations are caused by dialect mixing and cliticization. Most of the variation occurs in the final consonants: [n] ~ [k] (*fatin ~ fatik* 'place'), [n] ~ no coda (*ulun ~ ulu* 'head'), [k] ~ no coda (*hotuk ~ hotu* 'all'), which might be morphologically motivated, and [r] ~ no coda (*emar ~ ema* 'person'). As for other consonants, there are some examples of variation [d] ~ [r], [b] ~ [f], [s] ~ [f], [b] ~ [w], dropping of the glottal stop or other word-medial consonants (van Klinken 1999:41-42).

In Tetun Terik, the deletion of word-final coda is also very common, especially in the case of stops /k/ and /t/[179], as shown in (330). In Tetun Prasa, this strategy is found mostly in a fossilized form (words occur without the final coda) but it is not really productive.

(330)   Tetun Terik:

ladi'ak   'not good'   [la.ˈdi.ak] ~ [la.ˈdi.ak̚] ~ [la.ˈdi.a][180]
(0050TV_JCV_48_M_Sec_TT)

de'it   'only'   [ˈde.it] ~ [ˈde.it̚] ~ [ˈde.i] ~ [ˈdej]
(0023TS_LLM_41_M_Sec_TT)

---

[179] /p/ does not occur in word-final position.
[180] Out of 7 occurrences of *ladi'ak* in my oral corpus, none was pronounced with a glottal stop.

Despite the heavy influence of Portuguese, distribution of certain phonemes is also quite limited in Tetun Prasa. All Tetun Prasa phonemes can occur in the medial position, but only some can occur in the final position. Out of 22 phonemes, only seven can occur in all three positions: initial, medial and final - /n/, /r/, /l/, /t/, /s/, /ʃ/, and /k/. The Portuguese /ʃ/ is also the only foreign segment that can occur in final position.

In the following subsections, I will closely examine the foreign consonants and their phonetic realizations and adaptations in Tetun Prasa and Tetun Terik.

7.5.2.1 [l] > [r]

According to Hull (2000b:180), the final [l] has been replaced by [r] in many but southern dialects. However, this variation does not seem to be contact-induced, a fact already observed by Mendes (1935:189): "Em tétum as terminações *al* e *el* variam de região para região, mudando-se em *ar* ou dizendo-se indiferentemente *al* ou *ar*, *el* ou *er.*"[181] Thomaz (1974:240) observed the same phenomenon and commented that the final [l] and [r] merge into one liquid archiphoneme |L| that can be realized in two ways, as [r] or as [l], as shown in (331). However, I have not found examples of this kind of variation in my corpus.

(331)   (Thomaz 1974:240)
   dikul ~ dikur 'horn'
   akar ~ akal 'sago'

7.5.2.2 [d] > [r] > [d]

On the other hand, I accounted for the [d] > [r] > [d] variation. Although it is not as frequent as other phonetic adaptations, it is often heard in words like *daudaun* 'PROG', as in (332), or *horibainhira* 'when in the past', as in (333).

---

[181] "In Tetun, the terminations *al* and *el* vary from region to region, they change into *ar*, or are pronounced indifferently *al* or *ar*, *el* or *er.*"

(332) *daudaun*   'PROG'   [da(u).ˈda.un] ~ [ra.ˈra.un]
       (0033TS_AGJ_24_M_Sec_Bun)

(333) *horibainhira* 'when in the past'   [ˌho.ri.bajn.ˈhi.ra] ~
       [ˌho.di.bajn.ˈhi.ra] (0053TD_JLS_44_M_Uni_Wai)

7.5.2.3 glottal stop [ʔ]

Originally, the glottal stop represented the lost proto-Malayo-Polynesian /k/, which is also the reason why /'/ is still present in the standard Tetun Prasa orthography, largely based on the etymological principle (Hull 1993:xviii). For most of the outsiders, glottal stop was a foreign sound[182]. Silva (1889:41) commented on it, saying: "Sem se ouvir não se pode comprehender."[183] It is also one of the two phonemes that Tetun Prasa presumably eliminated from its inventory during its development. The main reason that is usually cited by authors (Williams-van Klinken 2010b, Albuquerque 2011) is the absence of this phoneme in Mambae, a language that was originally spoken in Dili, where Tetun Prasa developed. The glottal stop, however, still exists and is pronounced in Tetun Terik and Tetun Fehan.

To understand the extent to which the glottal stop is still used in Tetun Terik and to verify whether Tetun Prasa speakers indeed do not possess this segment, I analysed my oral corpus which contains data from two Tetun Terik regions (Viqueque and Suai) and from Dili where Tetun Prasa is spoken. In this analysis, as in the analyses of other segments, I looked at all the possible allophones of a particular phoneme. For the analysis of the glottal stop, I included all the words written with a glottal stop in the official orthography but I observed that a glottal stop in the orthography is not indicative of the presence of this phoneme in Tetun Terik. That is also the reason why the numbers for an unrealized glottal stop were so high.

As I had expected, I found a limited number of occurrences of the glottal stop in my oral Tetun Prasa corpus (see Figure 8) and the percentage for each speaker was very low (see Figure 9).

---

[182] According to van Minde (1990:65), Ambonese Malay has no phonemic glottal stop, but "a phonetic glottal stop is realized word-initially, morpheme-initially after a vowel, and morpheme-medially between like vowels in morphemes which are obviously loans from indigenous languages."

[183] "Unless you hear it, you can't understand it." (my translation)
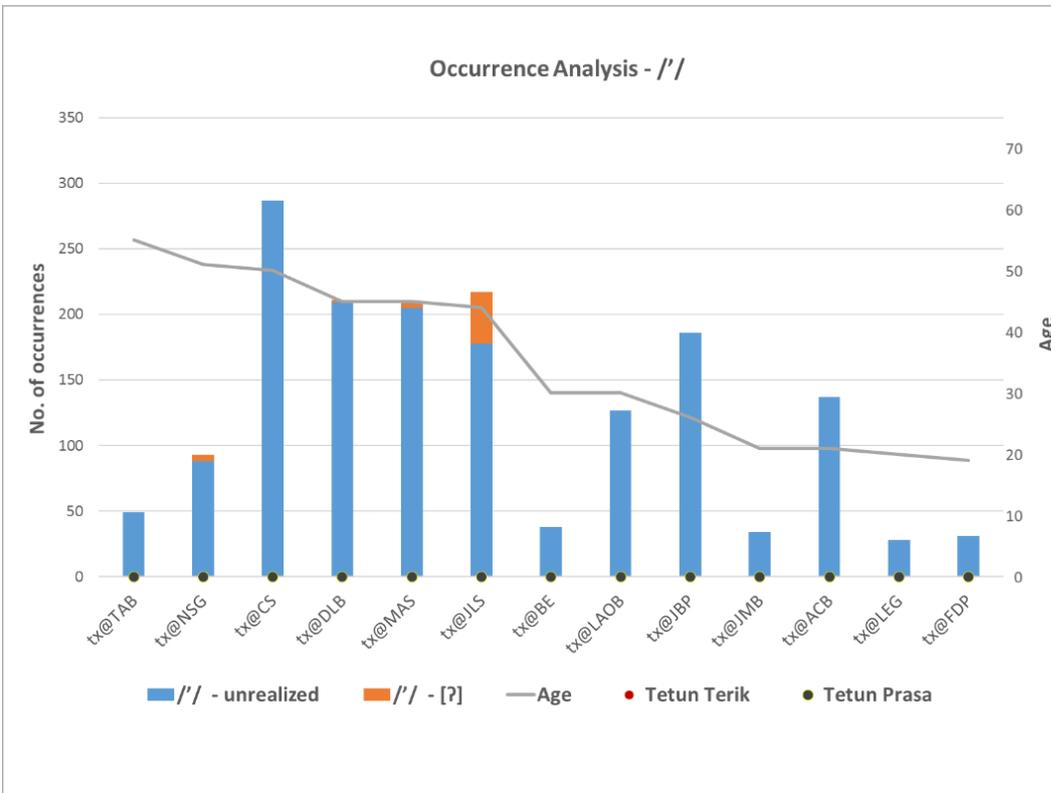
Figure 8 Use of the glottal stop among Tetun Prasa speakers according to number of occurrences
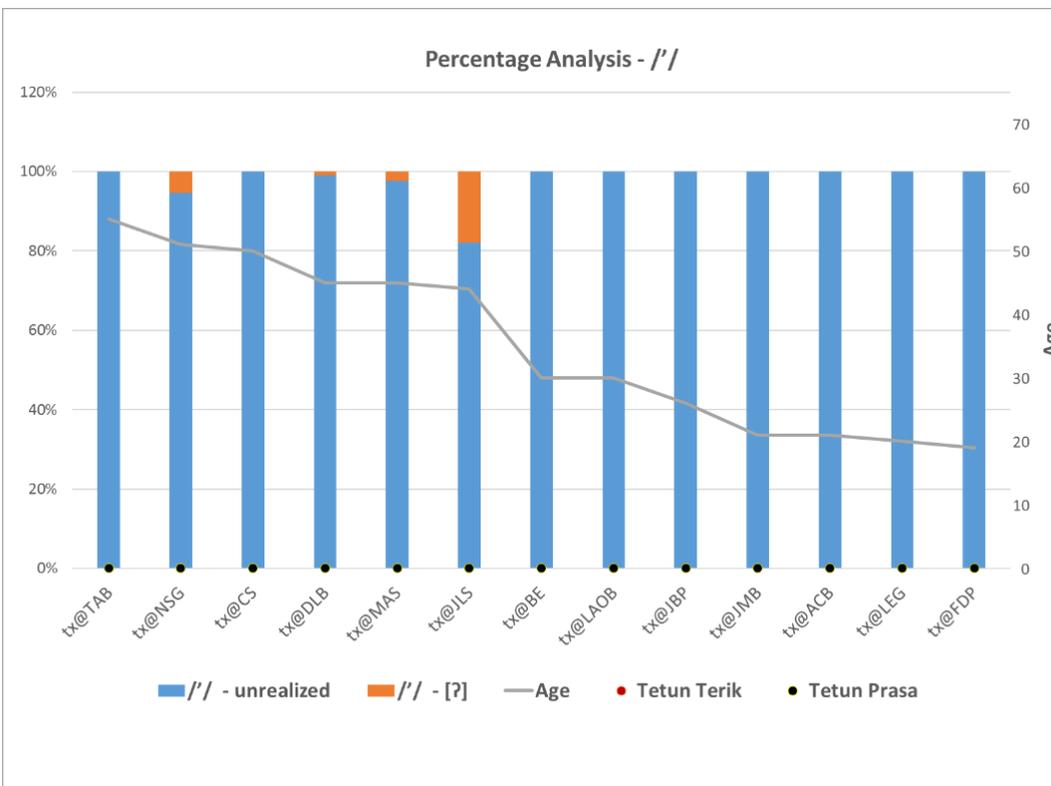
Figure 9 Use of the glottal stop among Tetun Prasa speakers (in %)

319

All the speakers that *did* pronounce the glottal stop in these few cases were L2 speakers of Tetun Prasa, with different mother tongues:

- Waima'a (spoken by tx@JLS)
- Tetun Terik (spoken by tx@NSG)
- Midiki (spoken by tx@MAS)
- Kemak (spoken by tx@DLB)

From the data available to me, at least Waima'a and Tetun Terik contain /'/, which could explain the presence of glottal stop in the speech of these speakers. However, only one Tetun Prasa speaker (tx@JLS) appears to pronounce glottal stops regularly. His mother tongue is Waima'a, an Austronesian language known for glottalized segments, containing the glottal stop as well as ejective stops (Hajek & Stevens 2005:2889). Since he is aware of the existence of the glottal stop in certain varieties of Tetun and his phonemic inventory does contain this phoneme, he might have decided to pronounce this segment also when speaking Tetun Prasa. Another speaker (tx@NSG), who lives in Dili but is originally from Fohoren (Cova Lima district) and speaks Tetun Terik as a mother tongue, was asked to explain the differences between Tetun Prasa and Tetun Terik. The interview was conducted in Tetun Prasa and the speaker did not use glottal stop unless he was giving examples of Tetun Terik structures.

Based on my analysis of Tetun Terik, spoken both in Suai and Viqueque, it can be concluded that the glottal stop is still used, regardless of the age of the speakers (see Figure 10 and Figure 11). Speakers who do not use glottal stops can be divided into two categories:

- Speakers who grew up and live in a Tetun Terik area but have a different mother tongue (e.g. tx@OC and tx@AGJ are speakers of Bunak)[184];
- Speakers who grew up in a Tetun Terik area but due to their work and environment shifted to Tetun Prasa (e.g. tx@MVJ).

---

[184] The reason I decided to include these speakers in my Tetun Terik analysis is to show that L2 speakers do not necessarily acquire glottal stop despite living in the environment where it is used.
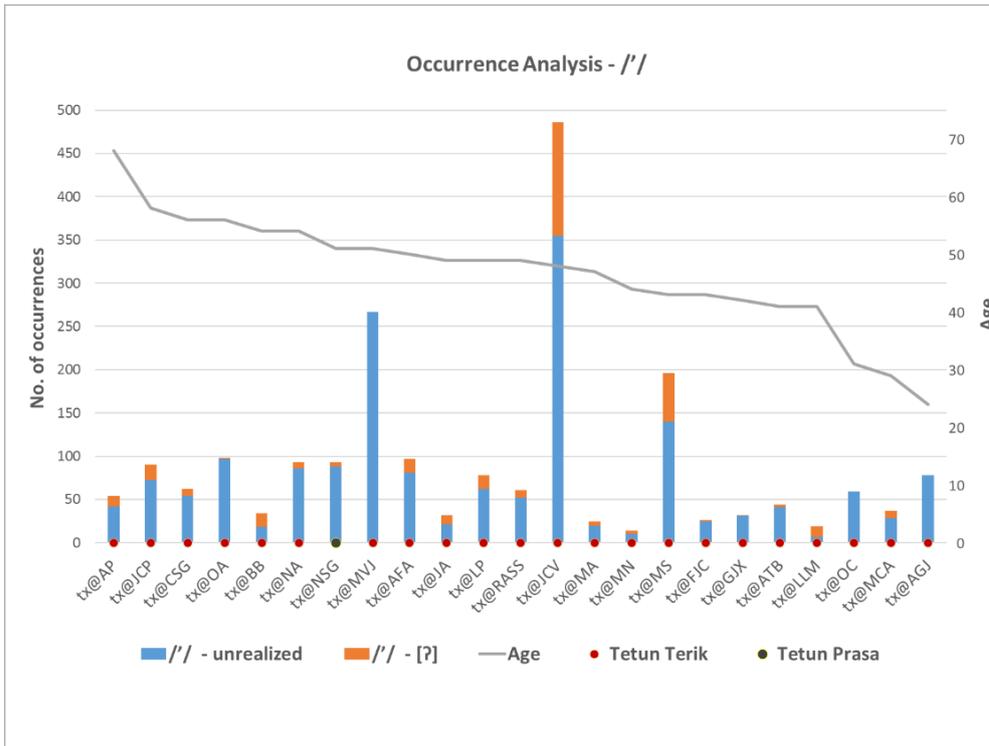
Figure 10 Use of the glottal stop among Tetun Terik speakers according to number of occurrences[185]
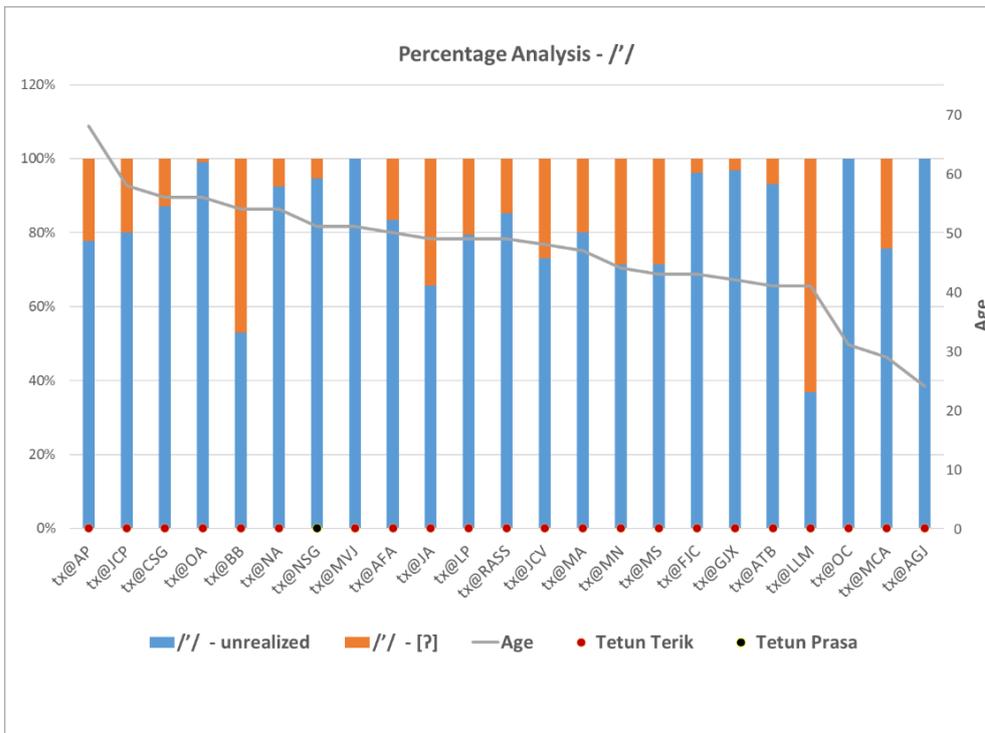
Figure 11 Use of the glottal stop among Tetun Terik speakers (in %)

---

[185] One of the speakers (tx@NSG) is included in both Tetun Terik and Tetun Prasa analysis, since he is a native Tetun Terik speaker but has lived in Dili for a considerable part of his life and was part of my Tetun Prasa group.

Two speakers (i.e. tx@OA and tx@GJX) made a conscious decision to speak "a more developed variety of Tetun", as the speakers put it, so the occurrences of glottal stop are lower in their case, too. This could mean that speakers in contact with Tetun Prasa tend to make conscious or unconscious decisions to drop this phoneme, although I would not go as far as saying that the use of glottal stop is stigmatized. Still, looking at the number of occurrences and percentages, it seems like the use of glottal stop in Tetun Terik is becoming marginal, too. To see if this was really true, I was interested to find out whether <'> in the official orthography was really indicative of this phoneme in the actual speech and whether certain words were more likely to be pronounced with a glottal stop than others. I looked at 20 most frequent words with <'> in my corpus (both Tetun Terik and Tetun Prasa), as shown in Table 44. Out of these, half scored zero or one occurrence; on the other hand, some words were more often pronounced with a glottal stop than without it. Function words (highlighted) were more likely to be pronounced without a glottal stop, e.g. words like *ne'e* 'this', *de'it* 'only' or *nu'udar* 'as' were never pronounced with a glottal stop. I also noticed that speakers use the glottal stop in a more careful pronunciation but in fast speech, the glottal stop is often inaudible.

Translation	Word List	Occurrences	[ʔ]	%	unrealized	%
this	*ne'e*	1066	0	0%	1066	100%
1S	*ha'u*	693	90	13%	600	87%
REL; which, where	*ne'ebé*	383	0	0%	383	100%
1S-POSS	*ha'u-nia*	200	23	12%	177	89%
only	*de'it*	159	0	0%	159	100%
arrive; until	*to'o*	125	55	44%	69	55%
good	*di'ak*	117	1	1%	116	99%
there	*ne'ebá*	104	0	0%	104	100%
speak	*ko'alia*	104	1	1%	103	99%
this one	*ida-ne'e*	77	0	0%	77	100%
priest	*na'i-lulik*	44	30	68%	14	32%
CLF:human-two	*na'in-rua*	31	9	29%	22	71%
as, like	*nu'udar*	31	0	0%	31	100%
walk	*la'o*	31	16	52%	15	48%
climb, rise	*sa'e*	29	15	52%	14	48%
CLF:human	*na'in*	28	9	32%	19	68%
thus; like this	*nune'e*	28	0	0%	28	100%
small	*ki'ik*	27	8	30%	19	70%
NEG	*la'ós*	23	16	70%	7	30%
say	*na'ak*	20	0	0%	20	100%

Table 44 Twenty most frequent words written with a glottal stop (both Tetun Prasa and Tetun Terik speakers)

However, this table does not tell us much about the preference for glottal stops in Tetun Terik. Instead of looking at the orthography, I selected 15 most frequent words that were produced with a glottal stop in Tetun Terik (see Table 45) to see if any generalization could be made.

Translation	Word List	Occurrences	[ʔ]	%	unrealized	%
1S	*ha'u*	337	83	25%	254	75%
arrive; until	*to'o*	80	45	56%	35	44%
1S-POSS	*ha'u-nia*	75	20	27%	55	73%
priest	*na'i-lulik*	44	30	68%	14	32%
CLF:human	*na'in*	25	8	32%	17	68%
small	*ki'ik*	24	7	29%	17	71%
climb, rise	*sa'e*	19	15	79%	4	21%
FOC	*ha'e*	18	9	50%	9	50%
CLF:human-two	*na'in-rua*	18	8	44%	10	56%
rice	*to'os*	18	17	94%	1	6%
cook	*te'in*	10	9	90%	1	10%
meat	*na'an*	6	5	83%	1	17%
walk	*la'o*	6	2	33%	4	67%
sunrise	*Loro-Sa'e*	4	4	100%	0	0%
	Grand Total	**684**	262	38%	422	62%

Table 45 Fifteen most frequent words with a possible glottal stop in Tetun Terik speakers

As you can see, some words are more likely to be pronounced with a glottal stop than without it. Again, function words are more frequently pronounced without a glottal stop, possibly because they are more frequent in everyday speech than the content words mentioned in Table 45.

7.5.2.4 [w] > [b]

As I have noted before, Tetun [w] is often replaced by [b] in Tetun Prasa, especially in two words: *wainhira* 'when' and *wee* 'water'. This has been assigned to the fact that Tetun Prasa developed in Dili, surrounded by Mambae language which does not contain [w]. However, [w] can still be found in the church register and widely throughout the Tetun Terik area where Mambae had no influence on the pronunciation (Hull & Eccles 2005:235). Also in Tetun Fehan, /w/ is still widely used and contrasts with /u/ as well as

with /b/, e.g. *wé* 'water' – *bé* 'also'. However, there are few words even in Tetun Fehan, where either [b] or [w] are heard, e.g. *bei ~ wei* 'however' (van Klinken 1999:28-29).

Based on the analysis of my corpus, I was able to attest that both speakers of Tetun Prasa and Tetun Terik often freely alternate between [w] and [b]. In the case of Tetun Terik speakers, this can be due to the influence of Tetun Prasa, which they all speak too; in the case of Tetun Prasa speakers, this may be due to the influence of other native Timorese languages that contain the phoneme /w/, as discussed below.

In my analysis, I have originally considered all words containing this phoneme but I was only able to find the [w] ~ [b] variation in three instances: *wainhira/bainhira* 'when', as shown in (334), *wee/bee* 'water' and *waiwain/baibain* 'usually', so I decided to focus only on these three words. Figure 12 shows that some speakers of Tetun Prasa do pronounce [w] in these three words.
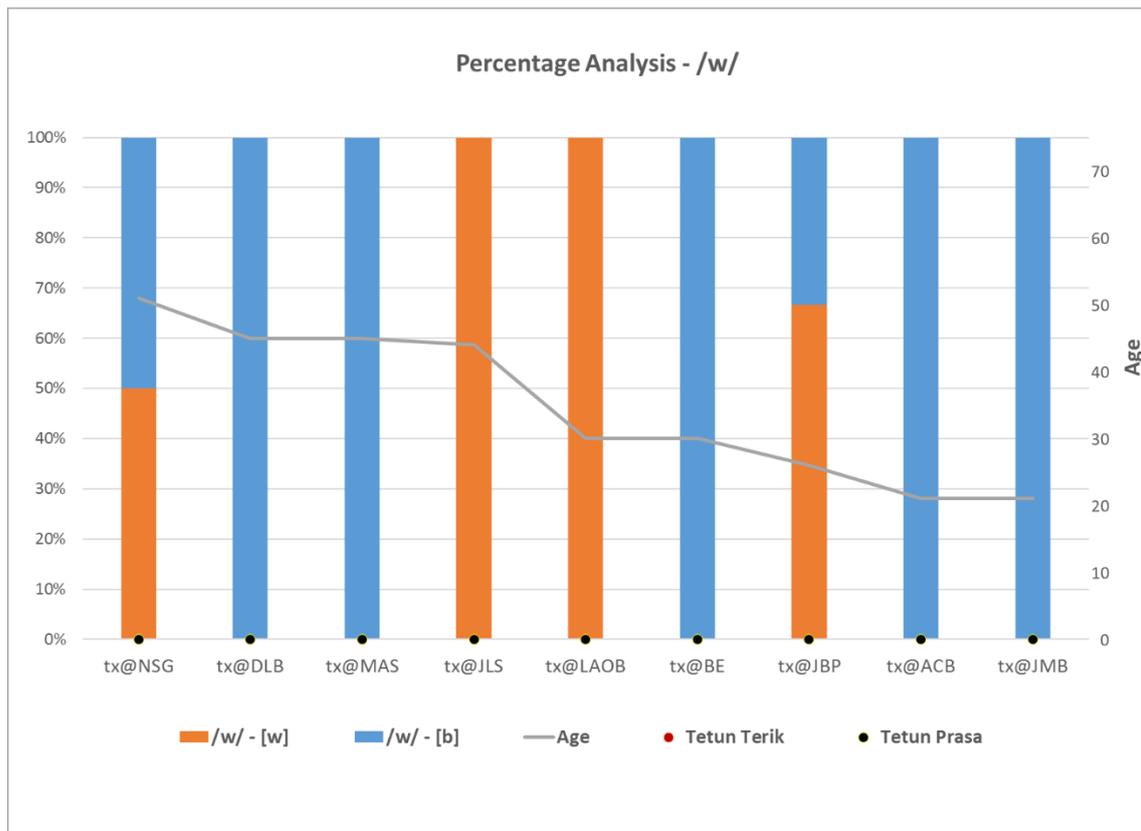
Figure 12 Preferences for [w] ~ [b] by Tetun Prasa speakers

As a possible source of this variation I have, again, identified the various mother tongues of my speakers. The ones that pronounce [w] also speak other Timorese languages (either as their mother tongue or a mother tongue of one of their parents) and do not find it foreign to their inventory. The fact that both of these realizations have become acceptable in the Tetun Prasa variety and are interchangeable is further supported by the language of the media, where spelling alternates between *bainhira* and *wainhira*:

(334)   Tetun Prasa:

	a.	***Wainhira***	*ha'u*	*rona*	[…]
		when	1S	hear	

'When I hear […]' (0005TD_JBP_26_M_Uni_Mak)

	b.	[…]	***bainhira***	*ita*	*survei*	*tiha*	[…]
			when	1PI	survey	PRF	

'[…] when we finish (this kind of) survey […]'
(0005TD_BE_30_M_Uni_Tet_Mak)

I got similar results analysing Tetun Terik. There are speakers who prefer to pronounce [w], as in (335a), and speakers who prefer [b], as in (335b), also illustrated by Figure 13.
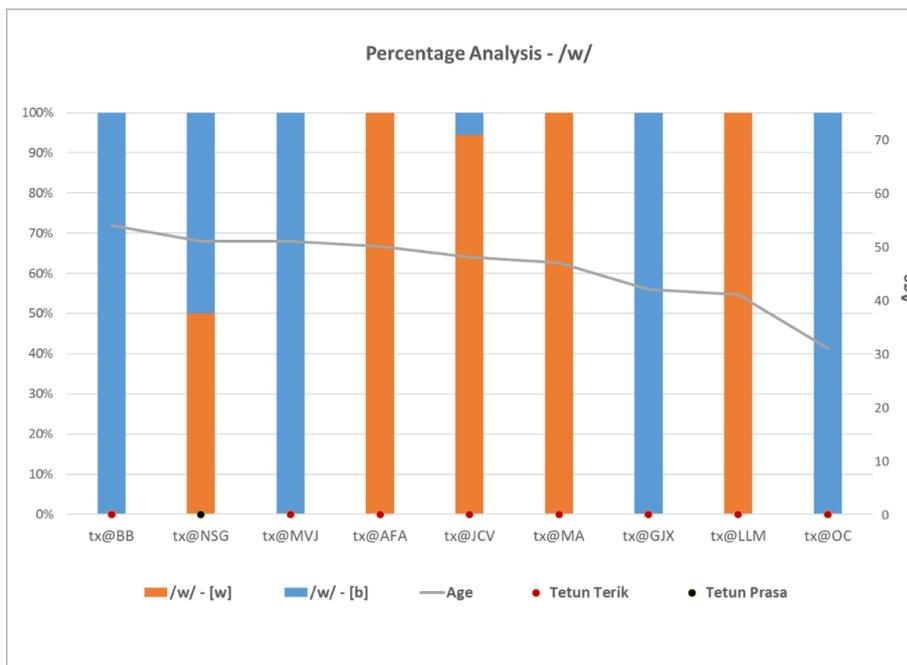
Figure 13 Preferences for [w] ~ [b] by Tetun Terik speakers

The two speakers (i.e. tx@OA and tx@GJX), who show preference for [b], made a deliberate choice to speak "a more developed variety of Tetun", as mentioned before. Two other speakers (i. e tx@MVJ and tx@NSG) live or used to live in an urban environment, where Tetun Prasa is the norm.

(335) Tetun Terik:

a. ***Wainhira*** *nia* *bá* *tuur* *heli* […]
   when 3S go sit hide

   'When he went to hide […]' (0046TV_JCV_48_M_Sec_TT)

b. ***Bainhira*** *kuandu* *ha'u-nia* *aman* *sira* *bá*
   when when 1S-POSS father PL go

   halo *kampu*
   do field

   'When my fathers (father and his brothers) went to work in the field […]' (0039TV_MVJ_51_M_Uni_TT)

7.5.2.5 [ŋ]

According to Blust (2013:196), most Austronesian languages of the Lesser Sunda have three nasals: labial, alveolar and velar. However, some Timorese languages, like Tetun, have only /m/ and /n/. Unlike in Malay/Indonesian language, where /ŋ/ is very common, in Tetun this velar nasal is used only as a free variant of /n/ in a final position or before a velar stop (Corte-Real 1990:60). I have noticed this variation especially in my data collected in Viqueque, where Tetun Terik is spoken, as shown in (336) and (337), but also in Tetun Prasa, mostly in the speech of young people, as shown in (338) and (339).

(336) Tetun Terik from Viqueque: *metan* '2S.get; 2S.find' [ˈme.tãŋ]

[...] *metan* *buat* *ladi'ak* *ruma* [...]
2S.find thing not.good some

'[…] find some bad things […]' (0047TV_AFA_50_M_Sec_TT)

(337) Tetun Terik from Viqueque: *naran* 'name' [ˈna.rãŋ]

Ha'u   naran jentiu   naran Rubi   Fuluk […]
1S   name   pagan   name   Rubi   Fuluk

'My native name is Rubi Fuluk [...]' (0049TV_MS_43_M_Mid_TT)

(338) Tetun Prasa: *da'an* 'boil' [ˈdã.ãŋ]

[...] *ami   da'an atu   prepara   ba   meiudia   nian.*
       1PE   boil   IRR   prepare   to   noon   POSS

'[…] we cook (corn) to prepare it for lunch.'
(0014TD_LEG_20_F_Sec_Mak)

(339) Tetun Prasa: *uitoan* 'a little' [uj.ˈtu.ãŋ]

Da'an hotu   tiha   ami   tuur   deskansa   uitoan.
boil   finished   PFV   1PE   sit   relax   a.little

'Having finished cooking, we sit down and relax a bit.'
(0014TD_JMB_31_F_Sec_Mak)

7.5.2.6 [ʃ] > [s] > [ʃ]

An interesting phenomenon happens when it comes to the pronunciation of [ʃ] and [s]. In Portuguese borrowings, [ʃ] is represented by <x> when followed by a vowel or by <s> when followed by a voiceless consonant. In general, less Portuguese-influenced speakers would pronounce /ʃ/ as [s], while speakers in contact with Portuguese might prefer to follow the pronunciation of European Portuguese, as shown in (340). They might also pronounce the final <s> as [ʃ], although even in European Portuguese, the pronunciation of word-final <s> can differ depending on the phonetic context.

(340) *estranjeiru* 'foreigner'

a. [ʃtɾã.ʒɐj.ɾu] – pronunciation in modern European Portuguese

b. [eʃ.tɾan.ʒei.ɾu] – pronunciation in Tetun Prasa

c. [es.tɾan.zei.ɾu] – pronunciation in Tetun Prasa

However, this is not universal and sometimes it is difficult to say what the motivation behind speakers' preferences is. For example, one of my speakers (tx@TAB),

who is proficient in Portuguese and received primary education in Portuguese, pronounces [s] in all positions when speaking Tetun Prasa. I decided to conduct an interview with him in Portuguese to see whether this kind of variation also occurs in his Portuguese. Interestingly, when speaking Portuguese, this speaker varies between [ʃ] and [s] freely, pronouncing mostly [s] in all positions.

I was interested to see whether word class has any influence on the pronunciation, so I looked at the pronunciation of *eskola* (from Port. *escola* 'school'), which can be used as a noun ('school') or as a verb ('to attend school'). In most of the cases (139 out of 177 occurrences), the speakers opted for a post-alveolar fricative [ʃ], and they were equally likely to do so in case of verbs as well as nouns, as shown in Table 46.

VERB		NOUN	
[ʃ]	[s]	[ʃ]	[s]
54	15	85	23
78%	22%	78%	22%

Table 46 Pronunciation of eskola 'school'

However, not all speakers made this distinction, as is clear from Table 47. In the case of the noun, the speakers were much more likely to pronounce [ʃ] – in fact, only three speakers of Tetun Prasa and one speaker from Viqueque showed variation in pronunciation in this case. This was, however, not the case of the verb *escola*. For example, all speakers of Tetun Terik from Viqueque showed variation in this regard and most of the speakers of Tetun Prasa and Tetun Suai were more likely to opt for [s].

		VERB				NOUN			
		[ʃ]	[s]	[ʃ]	[s]	[ʃ]	[s]	[ʃ]	[s]
TP	0014TD_ACB_21_M_Uni_Mak.txt		3		100%	3		100%	
TP	0056TD_CS_50_F_Pri_Tet.txt	16	13	55%	45%	29		100%	
TP	0051TD_DLB_45_M_Uni_Kem.txt		2		100%	1	1	50%	50%
TP	0014TD_JBP_26_M_Uni_Mak.txt		1		100%	1		100%	
TP	0053TD_JLS_44_M_Uni_Wai.txt	11	6	65%	35%	1	16	6%	94%
TP	0057TD_LAOB_30_M_Uni_Tet.txt	8	35	19%	81%	26	17	60%	40%
TP	0055TD_MAS_45_M_Uni_Mid.txt		1		100%	1		100%	
TS	0026TS_ATB_41_M_Sec_TT.txt		1		100%	1		100%	
TS	0019TS_BB_54_F_Sec_TT.txt	1	4	20%	80%	5		100%	
TS	0018TS_JA_49_M_Sec_TT.txt		1		100%	1		100%	
TS	0023TS_LLM_41_M_Sec_TT.txt		2		100%	2		100%	
TS	0017TS_MCA_29_F_Sec_TT.txt	1	4	20%	80%	5		100%	
TS	0017TS_MN_44_F_Sec_TT.txt	1	5	17%	83%	6		100%	
TS	0019TS_NA_54_M_Sec_TT.txt	6	10	38%	63%	16		100%	
TS	0032TS_OA_56_F_Uni_TT.txt		2		100%	2		100%	
TS	0018TS_RASS_49_M_Uni_TT.txt		2		100%	2		100%	
TV	0043TV_AP_68_M_NA_TT.txt	3	2	60%	40%	5		100%	
TV	0044TV_JCP_58_M_Sec_TT.txt	2	2	50%	50%	4		100%	
TV	0050TV_JCV_48_M_Sec_TT.txt	1	1	50%	50%	2		100%	
TV	0045TV_LP_49_M_Sec_TT.txt	7	5	58%	42%	8	4	67%	33%
TV	0042TV_MS_43_M_Mid_TT.txt	5	3	63%	38%	8		100%	
TV	0039TV_MVJ_51_M_Uni_TT.txt	7	3	70%	30%	10		100%	

Table 47 Pronunciation of eskola 'school' per speaker

To see whether phonetic context played any role in different realizations, I analysed one to the most frequent words in my corpus – *depois* 'later; after'. In European Portuguese, the final <s> in *depois* is pronounced as [ʃ] when the following word starts with a consonant or before a stop. In case of vowel-initial words, the final <s> is realized as [z]. This rule does not officially exist in Tetun Prasa but I was still interested to see whether the pronunciation of final <s> may vary depending on the context.

As we see in Table 48, in most of the cases the final <s> is pronounced as [s]. However, we see higher percentages of this allophone in case of following V-initial (79%) or /s/-initial (92%) words. I have documented only one case, in which the final <s> was pronounced as [z] and in two cases, the final <s> was not realized at all.

C-initial				V-initial				/s/-initial			
[ʃ]	**[s]**	[z]	no realization	[ʃ]	**[s]**	[z]	no realization	[ʃ]	**[s]**	[z]	no realization
24	**51**	0	1	4	**22**	1	1	0	**11**	0	1
32%	**67%**	0%	1%	14%	**79%**	4%	4%	0%	**92%**	0%	8%

Table 48 Pronunciation of final <s> in depois *depending on context*

Subsequently, I was interested to see whether age, education or proficiency in Portuguese influence the pronunciation of /ʃ/ in Tetun Prasa. From Figure 14 and Figure 15 we can see that age of the speakers seems to play some role although we can find speakers who received education in Portuguese (either during the colonial times or post-independence) whose phonemic inventory does not contain this sound. On the other hand, some speakers educated during the Indonesian times tend to opt for European Portuguese pronunciation in certain cases.
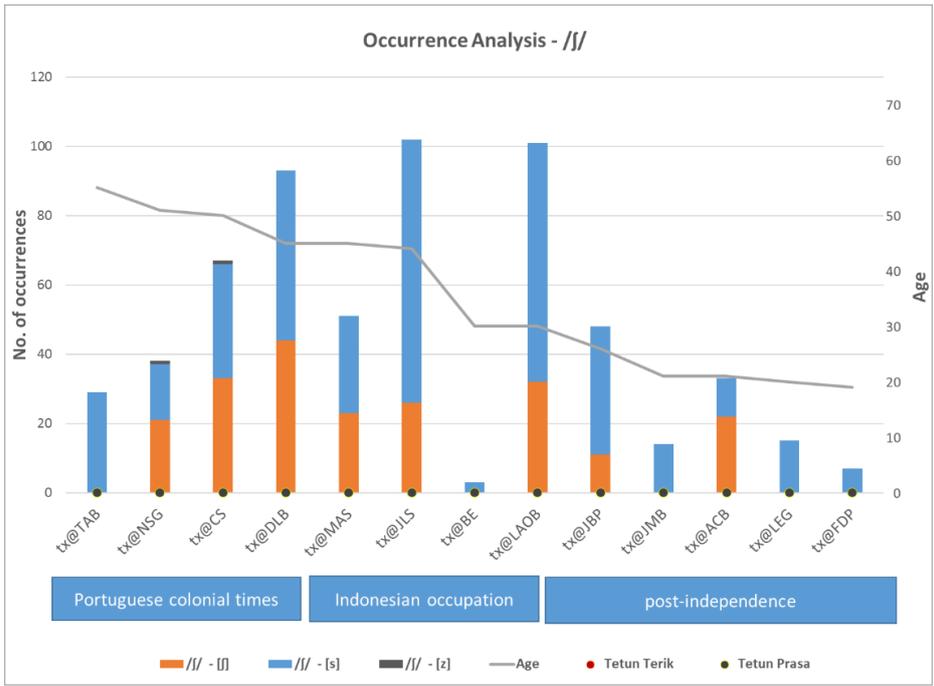
Figure 14 The pronunciation of /ʃ/ based on the period in which the speakers of Tetun Prasa were educated according to number of occurrences
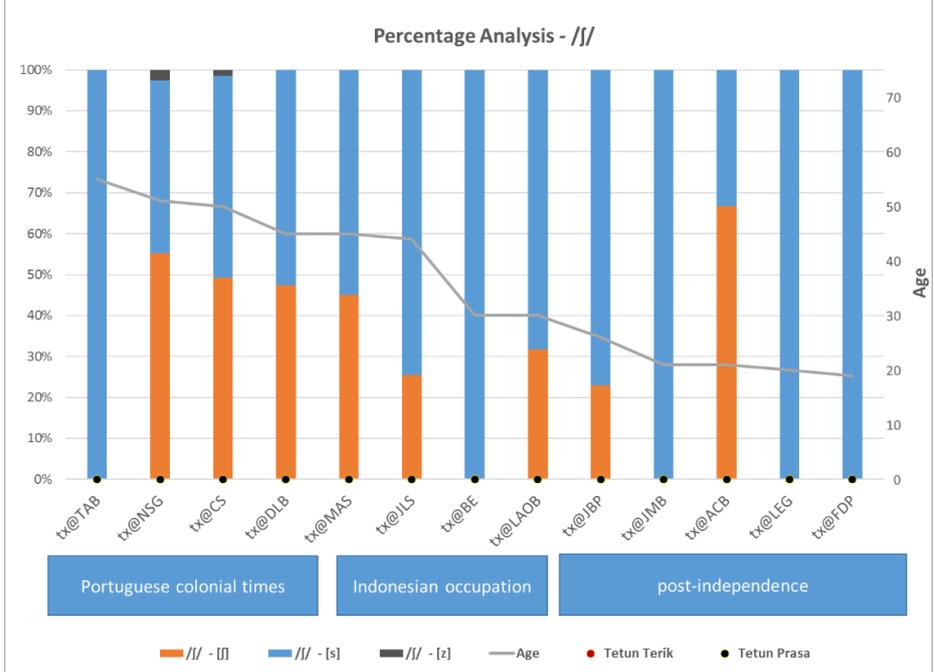
Figure 15 The pronunciation of /ʃ/ based on the period in which the speakers of Tetun Prasa were educated (in %)

The fact that age or the period in which speakers were educated are not the decisive factors is quite expected. This conclusion does not only apply to Tetun speakers, but can be applied globally to L2 speakers (which most of the Tetun Prasa speakers are) of any language:

"Indeed, the amount of variation in pronunciation attainment among adult learners, from largely unintelligible to native-like, suggests that age cannot be the only influence on attainment. Instead, other causes such as exposure to the target language and social influences may be central to ultimate attainment. (...) Sociolinguistic research in SLA has documented that social, cultural and psychological factors affect language acquisition." (Levis & LeVelle 2012:1).

Indeed, social factors, like exposure to Portuguese language and level of education can play a role in the pronunciation of Portuguese sounds. So I took a look at another variable that might shed some light on the pronunciation of /ʃ/: the level of proficiency in Portuguese. As I noted in the methodology section, all speakers self-assessed their knowledge of Portuguese. I divided them into three groups: a. those who speak Portuguese, b. those who can understand it and c. those who have no knowledge of it. As Figure 16 and Figure 17 show, even some speakers proficient in Portuguese or with some understanding of the language do not pronounce [ʃ]. It is unclear whether they make a conscious decision not to pronounce the borrowed phone, e.g. they find the pronunciation difficult, they dislike Portuguese, etc.
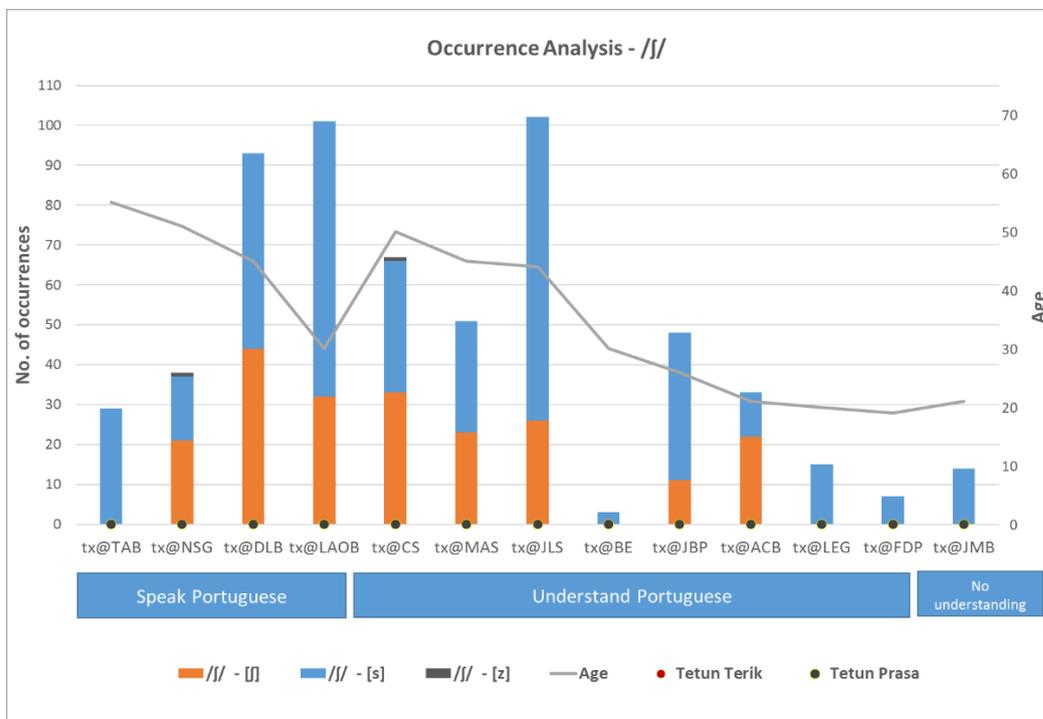
Figure 16 Pronunciation of /ʃ/ in Portuguese borrowings by Tetun Prasa speakers based on their proficiency in Portuguese according to number of occurrences
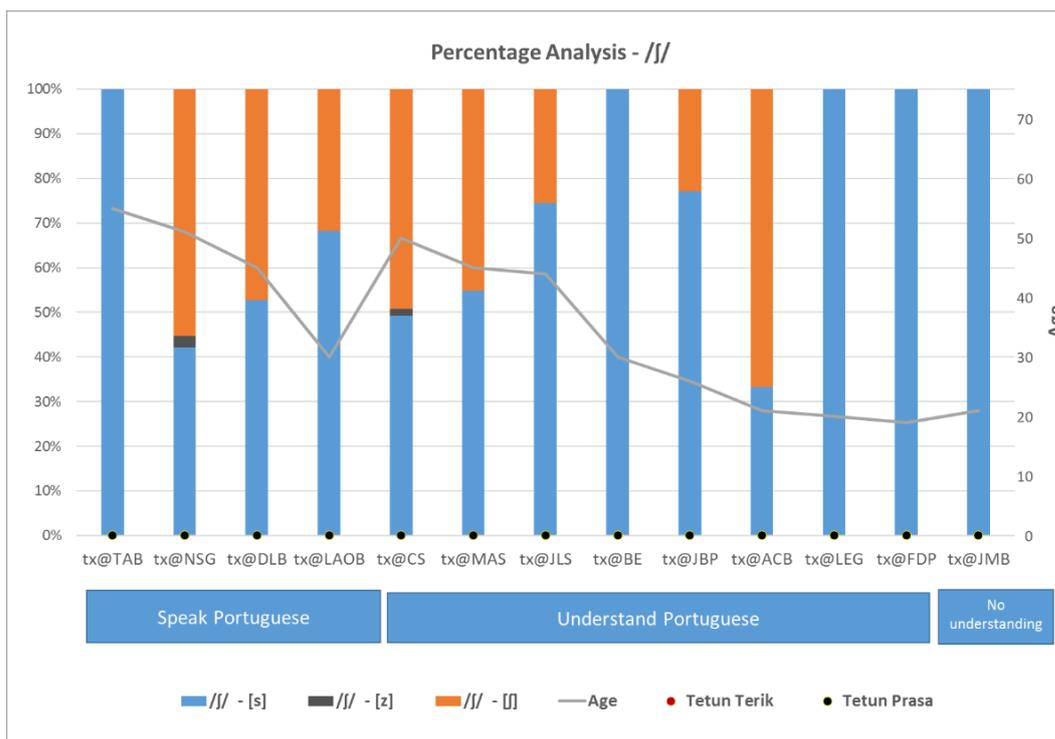
Figure 17 Pronunciation of /ʃ/ in Portuguese borrowings by Tetun Prasa speakers based on their proficiency in Portuguese (in %)

One variable that is more indicative of this distribution is the level of education. In general, speakers who receive better education are consequently more likely to have jobs in government, public sector or international companies, where technical language containing many Portuguese borrowings, is widely used. Again, I divided my speakers into three groups: a. speakers who received higher education (e.g. university course or a post-graduate specialization, in my case, all are young or middle-aged men), b. speakers who have secondary education (in my case, all are young women) and c. speakers who received only primary education (in my case, older people, as no further education was available to them). As Figure 18 and Figure 19 show, all speakers who received higher education (except for @BE, who I decided to disregard because of the limited number of occurrences) follow the Portuguese pronunciation to some extent. On the other hand, all speakers who received only secondary education fail to pronounce [ʃ], despite the fact that all of them were educated in the post-independence period and most of them have some understanding of Portuguese. However, we also have to consider the fact that their numbers are relatively low and this could have accounted for these results.
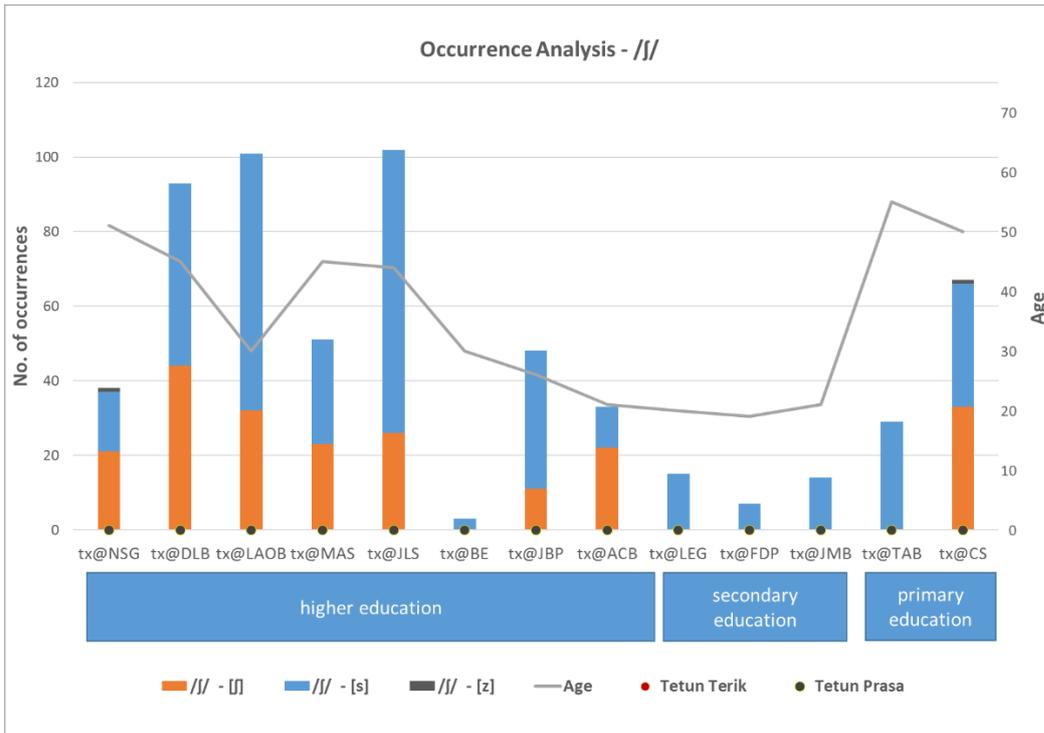
Figure 18 Pronunciation of /ʃ/ by Tetun Prasa speakers based on the level of education (by no. of occurrences)
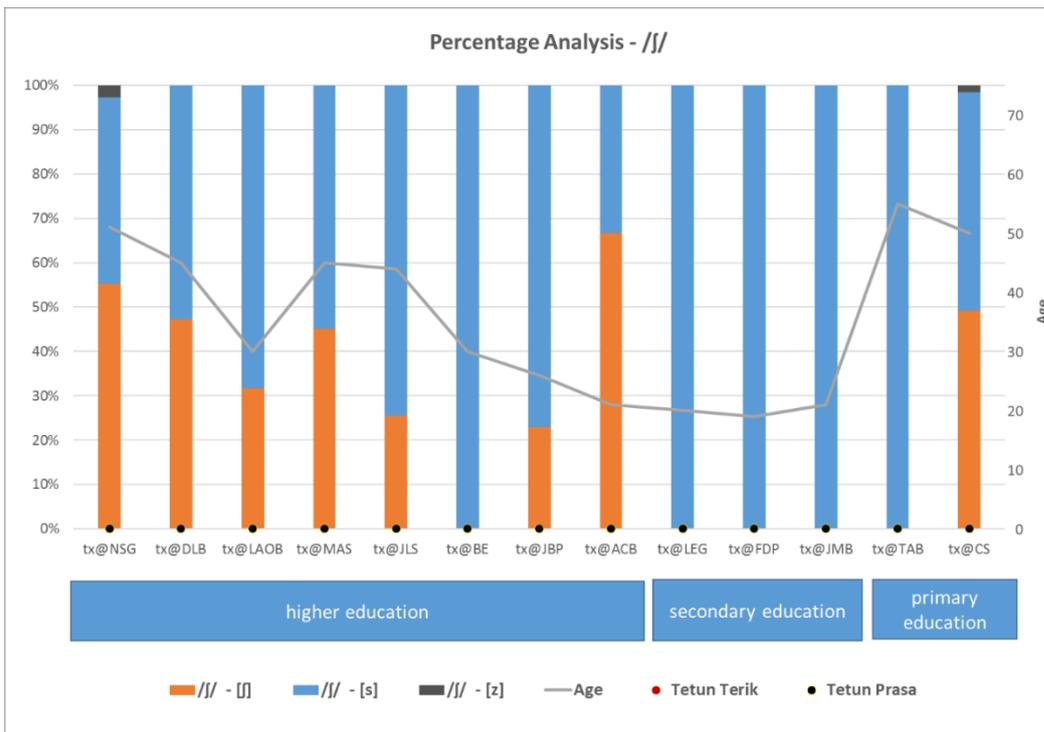
Figure 19 Pronunciation of /ʃ/ by Tetun Prasa speakers based on the level of education (in %)

The two speakers who only received limited primary education and were born and educated in the Portuguese era present a very interesting case. The first speaker (tx@TAB) is fluent in Portuguese and works in local government, where he is in contact with Portuguese on a daily basis but does not pronounce the borrowed phoneme. On the other hand, the second speaker (tx@CS) has a limited understanding of Portuguese and she works as a cleaner for (Portuguese-speaking) expat families. Both their families come from a Mambae-speaking region, but only tx@CS is a native Tetun Prasa speaker. This is why I believe that great attention needs to be paid to individual differences and environment these speakers grow up, are educated, work and live in. As we can see, the background of Tetun Prasa speakers can be incredibly diverse.

I did the same analysis with my Tetun Terik data and, at the first glance, the outcome looks quite surprising. As Figure 20 and Figure 21 reveal, it might seem that the pronunciation of /ʃ/ is much more frequent in this variety than in Tetun Prasa.
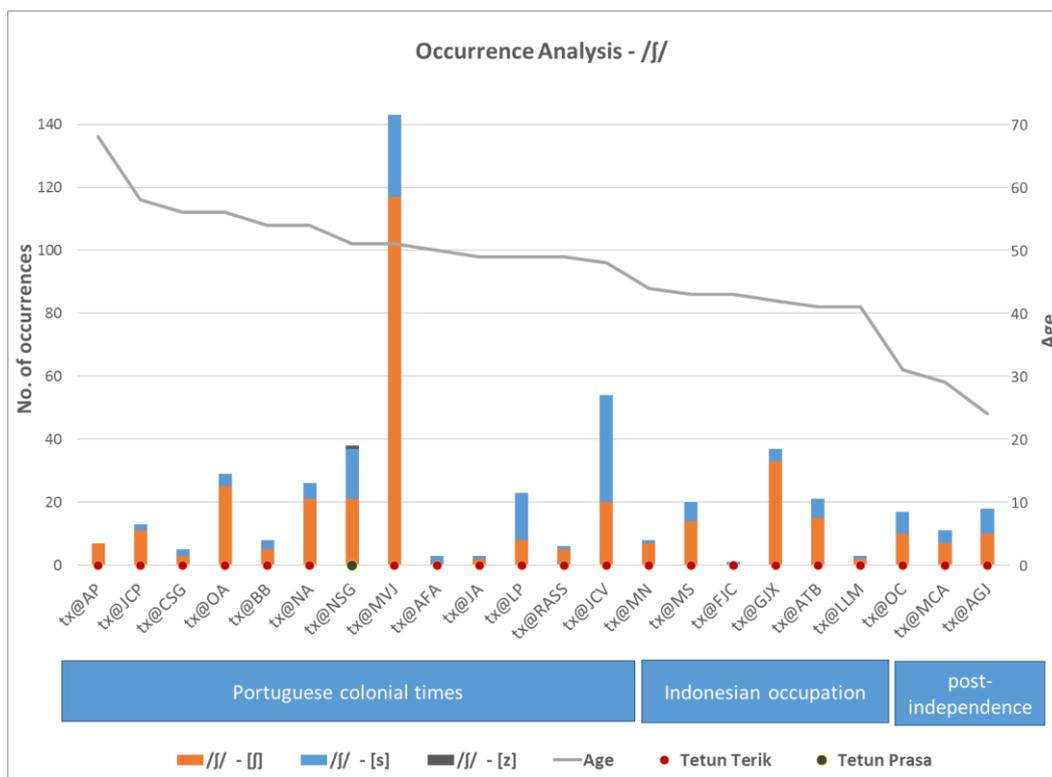
Figure 20 The pronunciation of /ʃ/ based on the period in which Tetun Terik speakers were educated (by no. of occurrences)
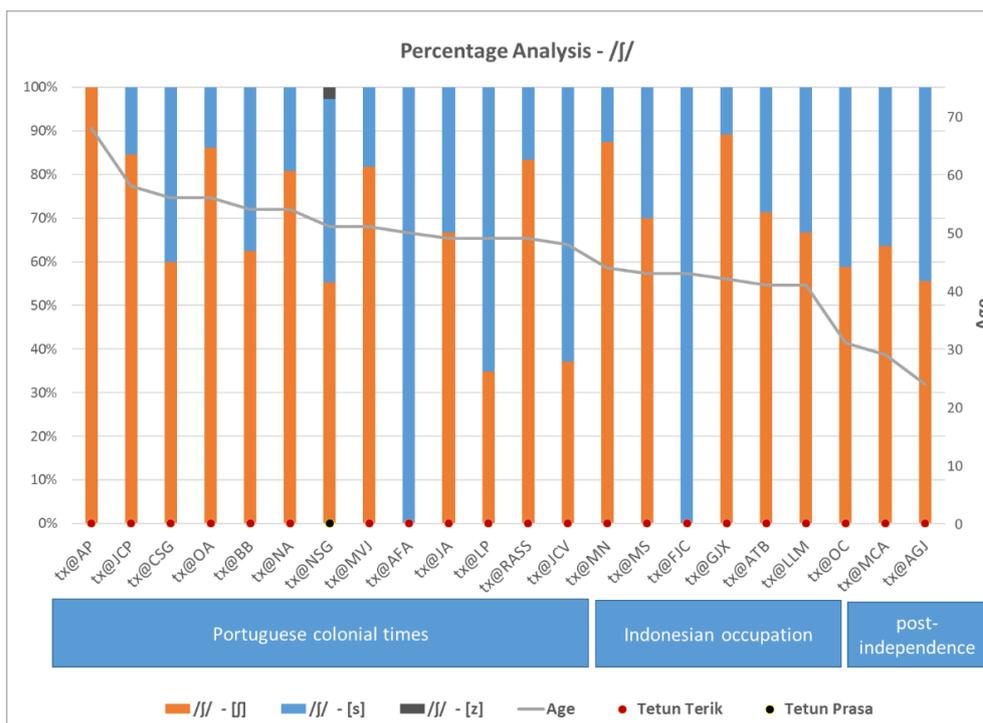
Figure 21 The pronunciation of /ʃ/ based on the period in which Tetun Terik speakers were educated (in %)

To understand why the results were contrary to my hypothesis, I looked at three factors: number of occurrences, the general topic of the interviews and the composition of the vocabulary used by Tetun Terik speakers. As you can see, in most of the cases, the number of occurrences was less than 20. At the same time, interviews conducted in Suai and Viqueque had a limited range of topics: family, childhood, current job situation, traditions and traditional stories. This is also reflected in the composition of my wordlist. I looked at the 20 most frequent words, which made up almost 70% of all the occurrences (see Table 49). Most of them are connected to school environment, where Portuguese is used as one of the languages of instruction (*eskola* 'school', *istória* 'history; story', *alunus* 'students', *mestra* 'female teacher', *mestre* 'male teacher', *esplika* 'explain'). Some words have been part of Tetun syntax (as conjunctions or in adverbial clauses) for a long time (*depois* 'after', *mas/mais* 'but').

Translation	Word List	Occurrences	[ʃ]	%	[s]	%
school	*eskola*	81	73	90%	7	9%
after, then	*depois*	41	20	49%	17	41%
history, story	*istória*	32	26	81%	6	19%
but	*mas*	28	22	79%	4	14%
Students	*alunus*	17	12	71%	4	24%
but	*mais*	16	5	31%	9	56%

hour	*oras*	11	1	9%	10	91%
teacher.FEM	*mestra*	8	8	100%	0	0%
like	*gosta*	7	7	100%	0	0%
Portuguese	*Portugés*	7	1	14%	6	86%
rest	*deskansa*	7	3	43%	4	57%
Gaspar	*Gaspár*	7	0	0%	7	100%
indeed	*pois*	6	6	100%	0	0%
teacher.MASC	*mestre*	5	4	80%	1	20%
means	*meius*	5	4	80%	1	20%
district	*distritu*	5	4	80%	0	0%
so-so	*maizumenus*	5	1	20%	4	80%
explain	*esplika*	5	4	80%	1	20%
nine hundred	*novesentus*	4	2	50%	2	50%
unknown	*deskoñesidas*	4	2	50%	2	50%

Table 49 Twenty most frequent words containing /ʃ/ in my Tetun Terik corpus

I then decided to compare the pronunciation of these words found in Tetun Prasa and the results were surprising. Fourteen of these words were found in both varieties and according to my analysis, [ʃ] is more common in the more conservative Tetun Terik, as documented by Table 50 and Table 51. Unfortunately, I have not been able to find an explanation for these unexpected results but it is a proof that the gap between these two varieties is narrowing.

Translation	Word List	Occurrences	[ʃ]	%	[s]	%	[z]	%
school	*eskola*	96	62	65%	34	35%		
after, then	*depois*	74	10	14%	61	82%		
Potuguese	*portugés*	37	7	19%	30	81%		
but	*mas*	28	8	29%	20	71%		
district	*distritu*	11	5	45%	6	55%		
but	*mais*	11	4	36%	7	64%		
hour	*oras*	7	0	0%	7	100%		
explain	*esplika*	7	2	29%	5	71%		
indeed	*pois*	4	0	0%	3	75%	1	25%
history, story	*istória*	4	1	25%	3	75%		
nine hundred	*novesentus*	3	2	67%	1	33%		
like	*gosta*	3	1	33%	2	67%		
teacher	*mestre*	3	2	67%	1	33%		
rest	*deskansa*	3	0	0%	3	100%		
	Total	**291**	**104**	**36%**	**183**	**63%**	**1**	**0,2%**

Table 50 Pronunciation of /ʃ/ in Tetun Prasa

Translation	Word List	Occurrences	[ʃ]	%	[s]	%
school	*eskola*	81	73	90%	7	9%
after, then	*depois*	41	20	49%	17	41%
history, story	*istória*	32	26	81%	6	19%
but	*mas*	28	22	79%	4	14%
but	*mais*	16	5	31%	9	56%
hour	*oras*	11	1	9%	10	91%
like	*gosta*	7	7	100%	0	0%
Portuguese	*Portugés*	7	1	14%	6	86%
rest	*deskansa*	7	3	43%	4	57%
indeed	*pois*	6	6	100%	0	0%
teacher.MASC	*mestre*	5	4	80%	1	20%
district	*distritu*	5	4	80%	0	0%
explain	*esplika*	5	4	80%	1	20%
nine hundred	*novesentus*	4	2	50%	2	50%
	Total	**291**	**178**	**61%**	**67**	**23%**

Table 51 Pronunciation of /ʃ/ in Tetun Terik

Additionally, I have found two other situations, in which [ʃ] is pronounced:

1) When <s> (in case of Tetun <z>) is followed by a voiced consonant, according to the European Portuguese phonological rules, the resultant sound is [ʒ], but in Tetun Prasa we can find [ʃ] as well:

(341) *dezde* 'since'   [dezde] but also [deʃde] from Port. *desde* 'since'
        (0054TD_NSG_51_M_Uni_TT)

2) Tetun speakers sometimes pronounce the final /s/ as [ʃ] even in native words. It is possible that speakers who are proficient in Portuguese (for example one of my Tetun Terik speaker tx@NA) might be applying hypercorrection to Tetun words, as in (342) but I have found instances of this hypercorrection also in speakers who claim not to speak Portuguese and who never or rarely pronounce [ʃ] in Portuguese borrowings. One of my Tetun Prasa speaker (tx@JMB) opted for [s] in all 14 borrowings, in which [ʃ] should be pronounced according to the European Portuguese norm, but opted for [ʃ] when pronouncing native *loos* 'right'.

(342) *loos* 'right'   [lo:ʃ] instead of [lo:s]
        (0019TS_NA_54_M_Sec_TT; 0014TD_JMB_31_F_Sec_Mak)

7.5.2.7 [ʒ] > [z] > [ʒ]

A similar phenomenon occurs also with <z>, although this one does not have any basis in the European Portuguese pronunciation. In Portuguese borrowings, the voiced palate-alveolar fricative [ʒ], spelled as <j> is often pronounced as the voiced alveolar fricative [z] especially by speakers not influenced by the Portuguese language. Sometimes, hypercorrection can be found in the speech of younger Timorese who have been influenced by Indonesian orthography, in which <j> is pronounced as [dʒ] and not as [ʒ]:

(343) *ajuda* 'help' [aˈʒu.da] but also [aˈdʒu.da]
(0017TS_MN_44_F_Sec_TT)

Some, especially non-Portuguese speakers, might have been also influenced by English, in which <ge> is often pronounced as [dʒ], e.g. *general* [ˈdʒe.nə.rəl]:

(344) *jerál*   'general'   [ʒeˈral] but also [dʒeˈral]
        (0053TD_JLS_44_M_Uni_Wai)

ajénsia 'agency'   [aˈʒen.si.a] but also [aˈdʒen.si.a]
        (0005TD_BE_30_M_Uni_TP_Mak)

In my corpus, I have also come across [ʒ] in Portuguese borrowings, where [z] should be pronounced according to the European Portuguese norm:

(345) *uza* 'use'   [uza] but also [uʒa] (0054TD_NSG_51_M_Uni_TT)

Although the orthography is not always the perfect representation of pronunciation, we see words in journalistic texts containing [z] written with <j>, which may be the result of this shift ([z] > [ʒ]):

(346) *razaun* 'reason'   <rajaun>   (STL_March_2016)

presiza 'need'   <presija>   (STL_March_2016)

zona 'zone'   <jona>   (STL_March_2016)

7.5.2.8 [ɲ] > [n] and [ʎ] > [l], [il] or [j]

The palatal sounds [ɲ] and [ʎ] are also rather marginal in Tetun. I have only found 54 occurrences of [ɲ] (in 14 different Portuguese borrowings) and 7 occurrences of [ʎ] (in 4 Portuguese borrowings). In case the of /ʎ/, I came across three allophones: [ʎ], [il] and [j]. The allophone [il] was found in the word traballu 'work' in two different Tetun Prasa speakers. The allophone [j] was found only once in the word filla 'daughter'. However, due to the limited number of occurrences, it is hard to draw any final conclusions.

7.5.2.9 [v] > [b]

This consonant variation is very common also outside of Timor. In the northern dialects of Portuguese spoken in Portugal as well as in the Galician language, /v/ is often pronounced as [b], e.g. *vaca* [ˈba.ka] 'cow', *vinho* [ˈbi.ɲu] wine. Although this phonetic adaptation is not very frequent in Tetun, I found a few instances in both Tetun Terik, as shown in (347), and Tetun Prasa, as shown in (348):

(347)   Tetun Terik:

   avó[186] 'grandfather'   [a.ˈvo:] ~ [a.ˈbo:] (0046TV_JCV_48_M_Sec_TT)

(348)   Tetun Prasa:

   vida   'life'   [ˈvi.da] ~ [ˈbi.da] (0056TD_CS_50_F_Pri_Tet)

7.5.2.10   /r/ > [ʁ], [ʀ] or [ɾ]

In Portuguese, when <r> occurs word-initially or following /n/, /l/, /z/, or /ʒ/, and when <rr> occurs word-medially, the standard pronunciation is the uvular fricative [ʁ] with the uvular trill [ʀ] and alveolar trill [r] as alternatives. In all other position, <r> is pronounced as the alveolar flap [ɾ]. Based on my analysis of Tetun Prasa and Tetun Terik, it seems like [ɾ] is pronounced in 99% of cases in which [ʁ] or [ʀ] would be realized in

---

[186] In Tetun, *avó* refers to Port. *avô* 'grandfather'. The word for Port. *avó* 'grandmother' is *avó-feto* (lit. 'grandfather-woman').

European Portuguese. I have noticed these sounds only in two of my Tetun Prasa speakers, both university-educated and with knowledge of Portuguese:

(349)   Tetun Prasa:

rona   'hear'   [ˈɾo.na] ~ [ˈʀo.na]
(0055TD_MAS_45_M_Uni_Mid)

remata   'finish'   [ɾe.ˈma.ta] ~ [ʁe.ˈma.ta]
(0057TD_LAOB_30_M_Uni_Tet)

7.5.3   Vowels and diphthongs

In this section, I will discuss Tetun Prasa vowels, including nasal vowels, vowel length and diphthongs as well as different approaches to identification of diphthongs in native words and their simplification in borrowed words.

7.5.3.1 Vowels

Compared to Portuguese, which has a rich vowel phonology with nine vowels used in stressed and unstressed syllables (Cunha & Cintra 1999), Tetun Prasa has a simple five-vowel system: /a/, /e/, /i/, /o/, /u/.

In general, /a/, /i/ and /u/ are relatively fixed sounds and do not undergo much variation, however, when unstressed, they can be weakened. For example, /a/ can be raised to [ɐ] or even to [ə] in an unstressed syllable, as shown in (350), or altogether deleted, which creates the possibility for consonant clusters (Hull & Eccles 2005:227).

The two vowels that are affected in pre-tonic or final/post-tonic position are /o/ and /e/, respectively. The close-mid front vowel /e/ can be lowered to open-mid [ɛ], as shown in (351), or, in the case of Portuguese borrowings, it can also be realized as [ə] in an unstressed position, as in (352) (Albuquerque 2011:83). On the other hand, following the alternative pronunciation found in Brazilian and African varieties of Portuguese, in the Portuguese-based creoles as well as in some dialects of European Portuguese, /e/ can be raised to [i] in a post-tonic position, as shown in (353).

(350)   (Williams van Klinken et al. 2002a:11)

  oras   'time'   [ˈo.ɾas] ~ [ˈo.ɾɐs] ~ [ˈo.ɾəs]

(351)   (both examples Albuquerque 2011:82)

  mane   'man'   [ˈmã.nɛ]

  señór   'mister'   [sɛ.ˈɲɔr]

(352)   (Albuquerque 2011:83)

  pergunta   'question'   [pəɾ.ˈgũn.tɐ]

(353)   *xefe*   'chief' [ˈʃe.fi]

According to Albuquerque (2011:84), the vowel /o/ is mostly realized as an open-mid back [ɔ], as in (354), but can be raised to [o] in unstressed syllables or when followed by close vowels /u/ and /i/, as in (355). However, Williams-van Klinken (2002a:11) claimed that in this case speakers show a considerable variation and while for some the basic allophone is [ɔ], for others it can be a close-mid allophone [o]. In Portuguese borrowings, /o/ is often pronounced as [u] in unstressed syllables, as shown in (356).

(354)   (Williams-van Klinken et al. 2002a:11)

  tohar   'broken'   [ˈtɔ.har]

(355)   (Albuquerque 2011:85)

  kotuk   'previous'   [ˈko.tuk]

(356)   *governu*   'government' [go.ˈver.nu] ~ [gu.ˈver.nu]

Since I am especially interested in language contact, I have looked at the variation in the pronunciation of unstressed /e/ and /o/ in Portuguese borrowings. In general, post-tonic unstressed /e/ is pronounced either as [e] in the standard pronunciation or as [i] by rural speakers:

(357)   *tenke* 'must'   [ˈten.ke] or [ˈten.ki]

In the case of unstressed /o/, Tetun Prasa standard pronunciation follows the standard European Portuguese pronunciation, in which /o/ is pronounced as [u]. In the popular pronunciation, unstressed /o/ is realized as [o]:

(358) *Portugál* 'Portugal'   [pur.tu.ˈgal] or [por.tu.ˈgal]

For my analysis, I decided to take a closer look at the pronunciation of /e/ in Portuguese borrowings in the post-tonic position, but I was also interested in Tetun words, like *sa'e* 'rise', *malae* 'foreigner' and *soe* 'throw' as well as names and toponyms. While there was considerable variation in case of the toponyms (Vemasse [Ve.ˈma.se] ~ [Ve.ˈma.si]), I have not documented any variation in native words.

I tried to see whether age had any influence on the pronunciation of post-tonic /e/ but looking at Figure 22 and Figure 23 we can see that [i], although not prevalent, is realized by Tetun Prasa speakers across all age groups.
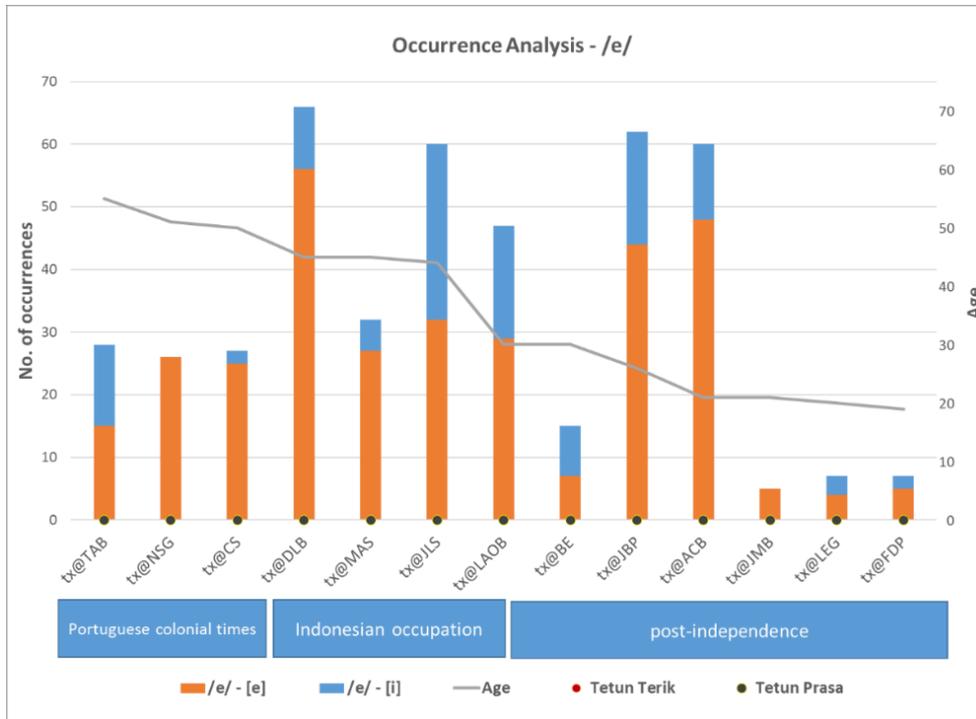
Figure 22 The pronunciation of /e/ in Portuguese borrowings and some Tetun words by Tetun Prasa speakers based on the period in which they were educated (in no. of occurrences)
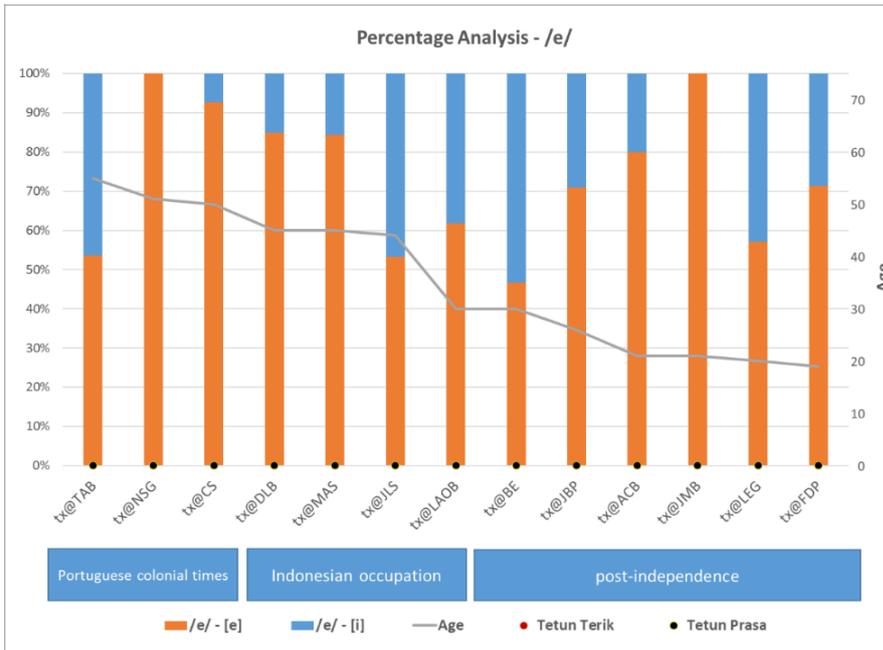
Figure 23 The pronunciation of /e/ in Portuguese borrowings and some Tetun words by Tetun Prasa speakers based on the period in which they were educated (in %)

In case of Tetun Terik speakers, the percentages might again seem a little misleading, showing the prevalence of [e] (see Figure 24), but the number of occurrences in some speakers was sometimes as low as 1 (see Figure 25), so we need to be careful when making any conclusions.
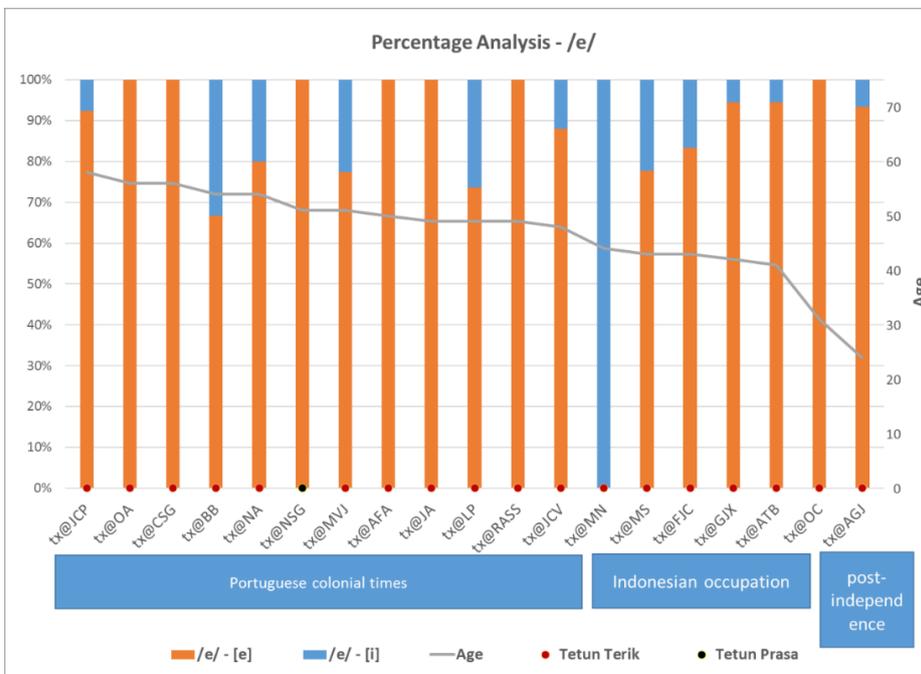
Figure 24 The pronunciation of /e/ in Portuguese borrowings and some Tetun words by Tetun Terik speakers based on the period in which they were educated (in %)
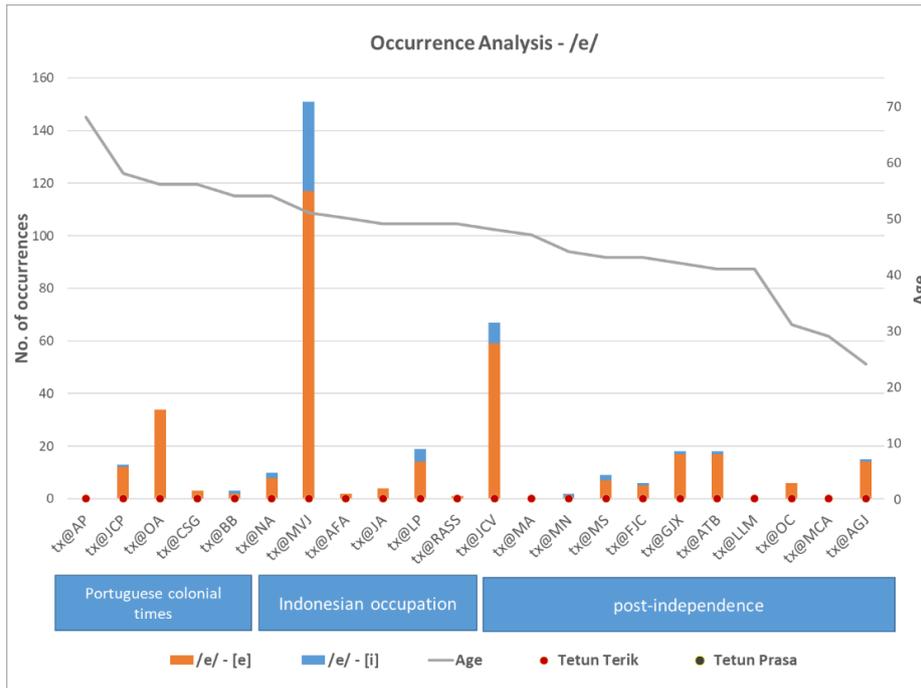
Figure 25 The pronunciation of /e/ in Portuguese borrowings and some Tetun words by Tetun Terik speakers based on the period in which they were educated (in no. of occurrences)

7.5.3.2 Diphthongs

Originally, diphthongs did not exist in Tetun. Presently, when there is a sequence of two different vowels within a native word, these are considered to belong to two different syllables (Hull 1993:xviii). In careful speech, these can still be heard, but in fast speech, they are often pronounced as a diphthong (Williams-van Klinken et al. 2002a:12):

(359) *ain* 'leg' [ˈa.in] ~ [ˈajn]

  saida 'what' [sa.ˈi.da] ~ [ˈsaj.da]

Tetun Prasa adopted diphthongs with Portuguese borrowings, though in popular pronunciation there is a tendency to simplify them. Thus, pronunciation of diphthongs can vary. According to Hull and Eccles (2005:231), when a Portuguese diphthong occurs in the penultimate syllable, it can be pronounced either as a diphthong or it can be simplified into a single vowel or transformed into a vowel sequence:

(360) a. T. *kadeira* 'chair' (from Port. *cadeira*)

  diphthong [ka.ˈdej.ra]   single vowel [ka.ˈde.ra]

344

b. T. *kaixa*   'box'   (from Port. *caixa*)

diphthong [ˈkaj.ʃa]   vowel sequence [ka.ˈi.sa]

The final /iu/ and /ia/ in Portuguese borrowings can be shortened to [i] in popular speech (Hull 1993:xiv), as in (361). This might be possibly caused by the Indonesian influence. In Indonesian, foreign words are borrowed mostly from English and Dutch but their final derivational suffixes are also often simplified or omitted, e.g. Ind. *ekonomi* 'economy', *administrasi* 'administration' (from Dutch *administratie*), *konsentrasi* 'concentration' (from Dutch *concentratie*). Similar affixes (*-ção*) are found in Portuguese, including the word-final vowel sequence <iu> and <ia>. The examples below were taken from my corpus which contains numerous cases in which Portuguese borrowings were either replaced by Indonesian borrowings, possibly due to their easier pronunciation (as in [is.ˈto.ri]) or the Portuguese endings were simplified to resemble Indonesian pronunciation (as in [ar.ˈma.ri]).

(361) *armáriu* 'wardrobe'   [ar.ˈma.ri.u] ~ [ar.ˈma.ri] (cf. M. *almari* 'wardrobe')

istória 'story'   [is.ˈto.ri.a] ~ [is.ˈto.ri]   (cf. Ind. *istori* 'argue')

7.6 Summary and conclusions

To summarize, I decided to compare the ratio of Portuguese phonetic realizations versus native phonetic realizations in various groups of speakers, i.e. preference for standard Portuguese-like pronunciation in Portuguese borrowings. I considered the following Portuguese phonemes: /ʒ/, /ʃ/, /ʎ/, /ɲ/ and a post-tonic /e/ and compared them against the native allophones [z], [s], [il], [j], [n] and post-tonic [i]. The figures below (Figure 26, Figure 27 and Figure 28) show the preferences of Tetun Prasa and Tetun Terik speakers, respectively, for Portuguese and native sounds according to the period in which they were educated. Unsurprisingly, the preference for Portuguese realizations is the highest in speakers educated during the Portuguese colonial times.
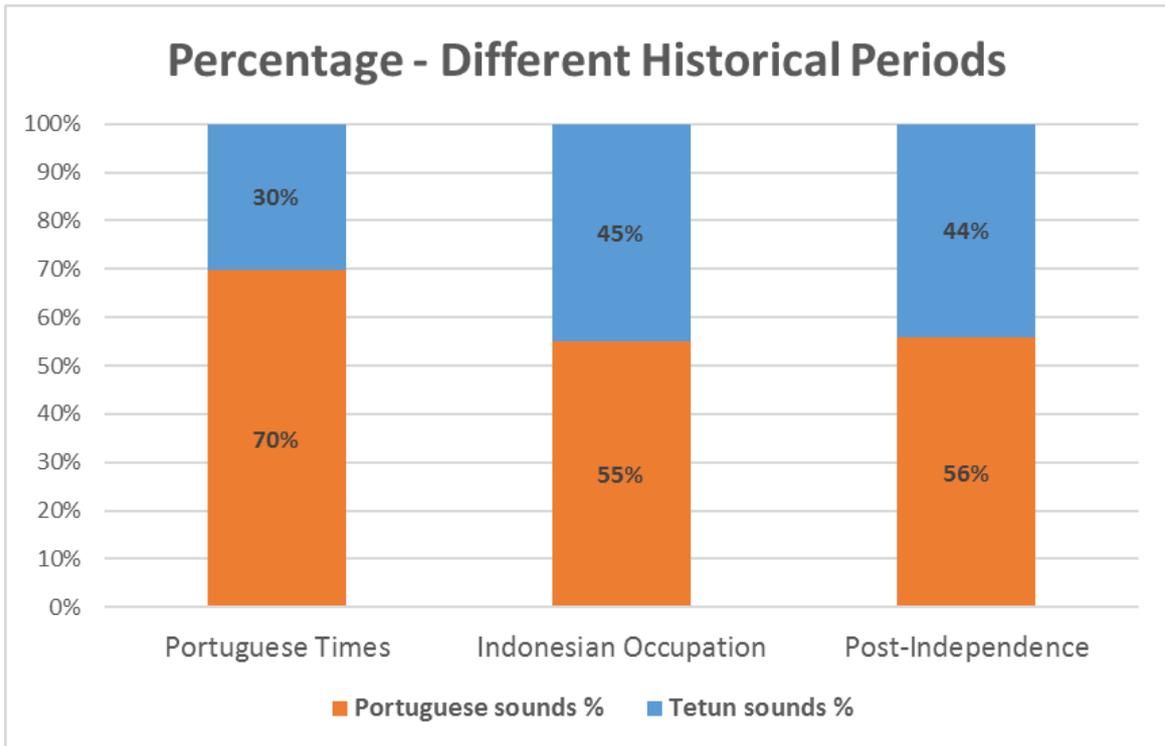
Figure 26 Overall representation of the use of Portuguese and native sounds by Tetun Terik and Tetun Prasa speakers according to the period in which they were educated (in %)
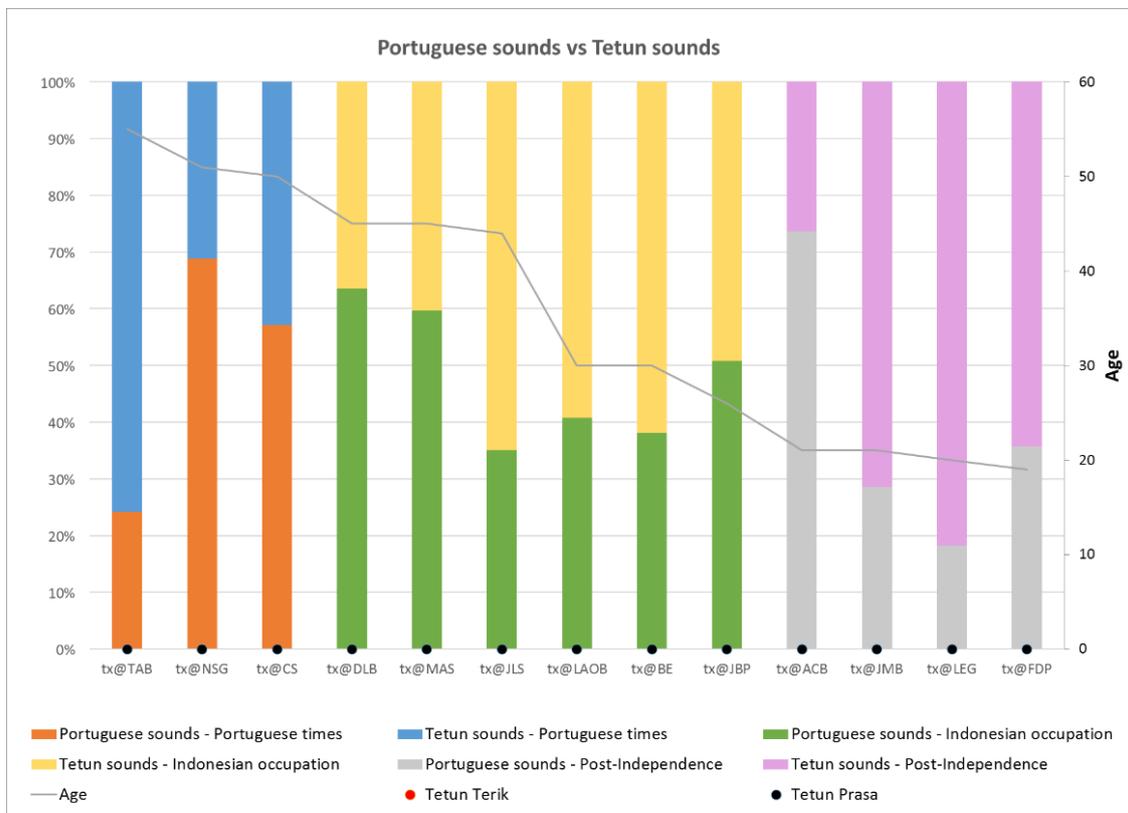
Figure 27 Comparison of the pronunciation of Portuguese sounds versus native sounds by Tetun Prasa speakers according to the period in which they were educated (in %)
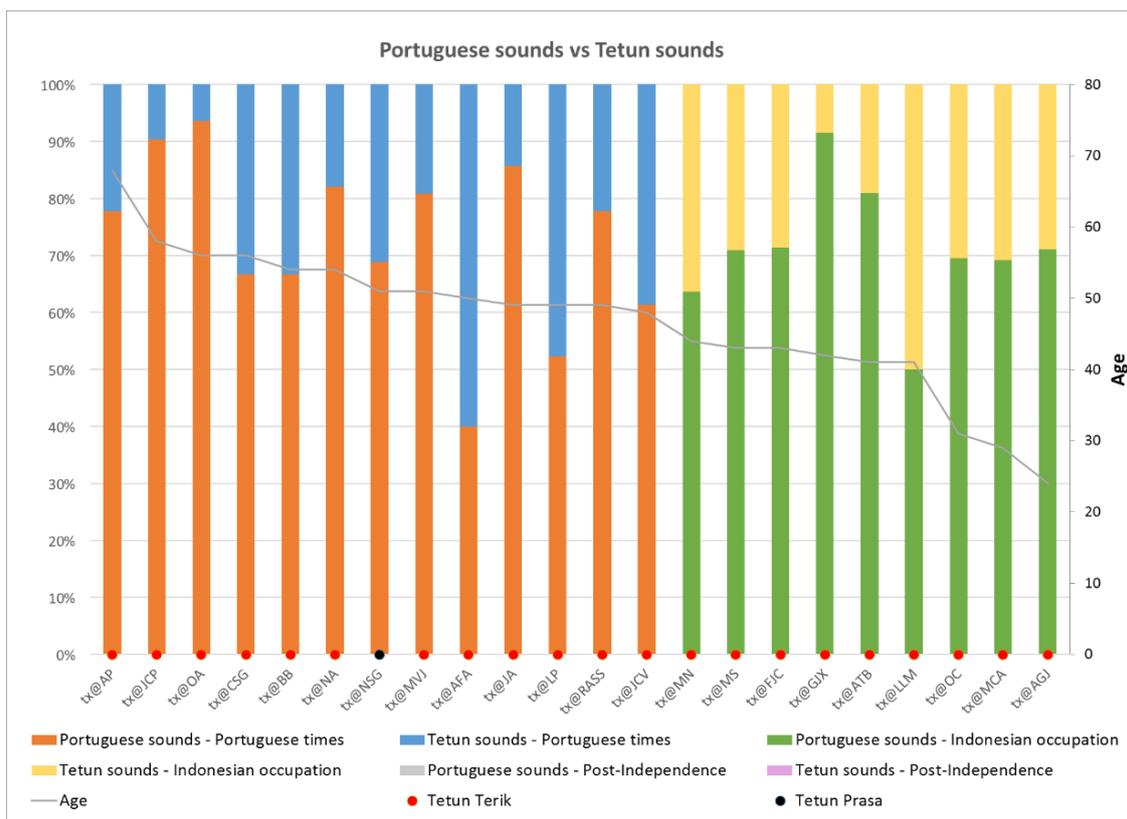
Figure 28 Comparison of the pronunciation of Portuguese sounds versus native sounds by Tetun Terik speakers according to the period in which they were educated (in %)

The second variable was the level of education, as shown in Figure 29. It might seem that speakers with no education present the highest ratio of Portuguese realizations but it needs to be noted that this figure is represented by only one Tetun Terik speaker. However, percentages for the other three levels of education (primary, secondary and higher) are what we would have expected. Speakers with lower education still show almost 50% ratio of Portuguese sounds which could be explained by the fact that these are the older speakers educated during the Portuguese times. The highest ratio of Portuguese realizations is found in speakers with higher education, which is quite expected. If we compare the two varieties, it might seem that Portuguese realizations are preferred by Tetun Terik speakers, as shown in Figure 30 and Figure 31. This, again, might have to do with the fact that there were fewer occurrences of Portuguese realizations in Tetun Terik and these were usually found in words that are very common across the whole of Timor, but it is also a clear sign that the two varieties are approximating.
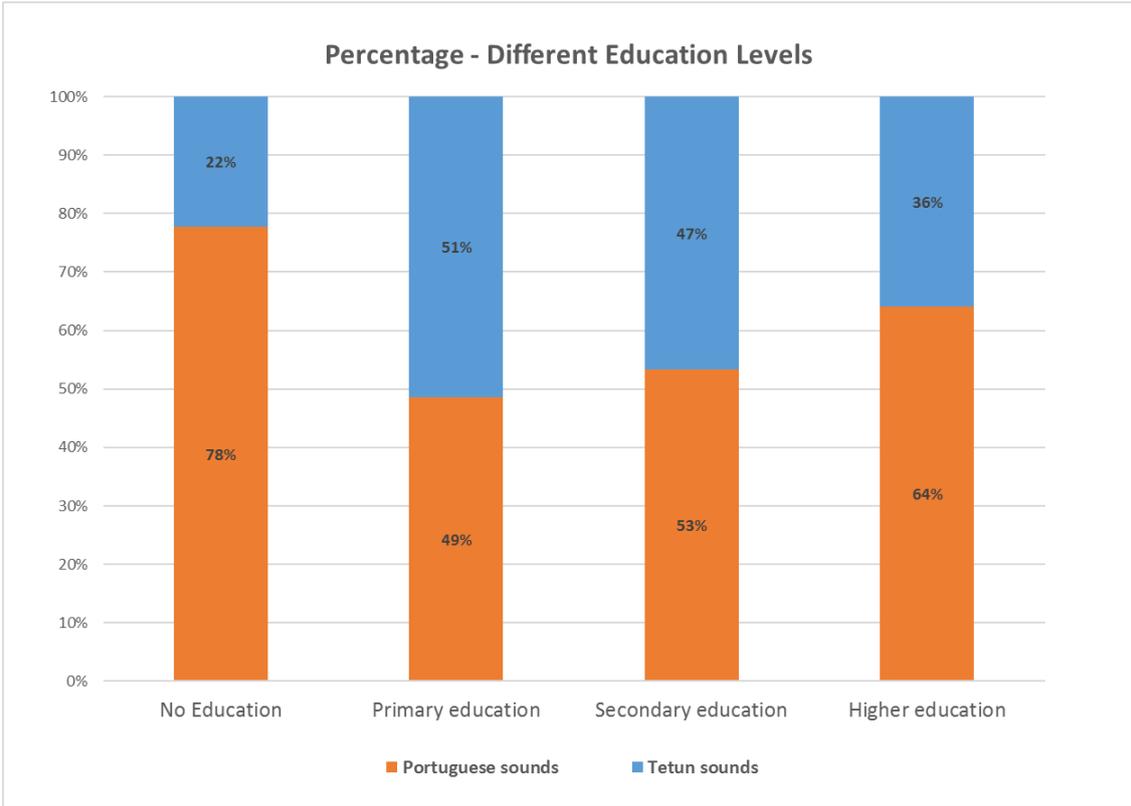
Figure 29 Overall representation of the use of Portuguese and native sounds by Tetun Terik and Tetun Prasa speakers according to the level of education (in %)
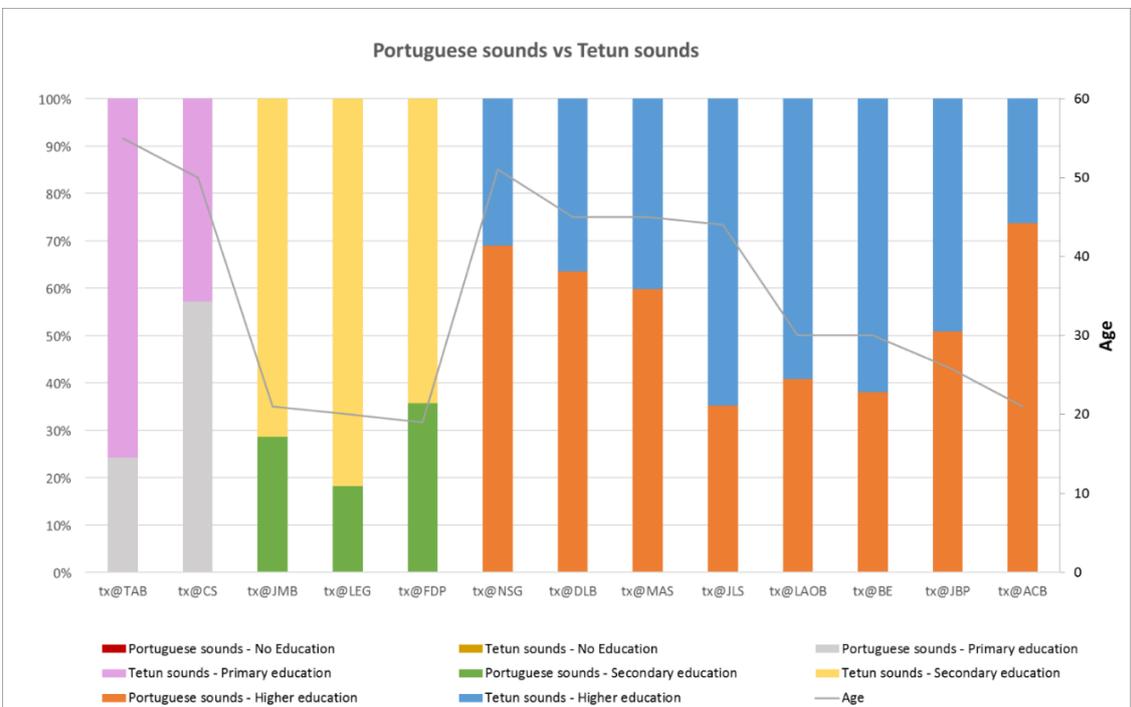
Figure 30 Comparison of the pronunciation of Portuguese sounds versus native sounds by Tetun Prasa speakers according to their level of education (in %)
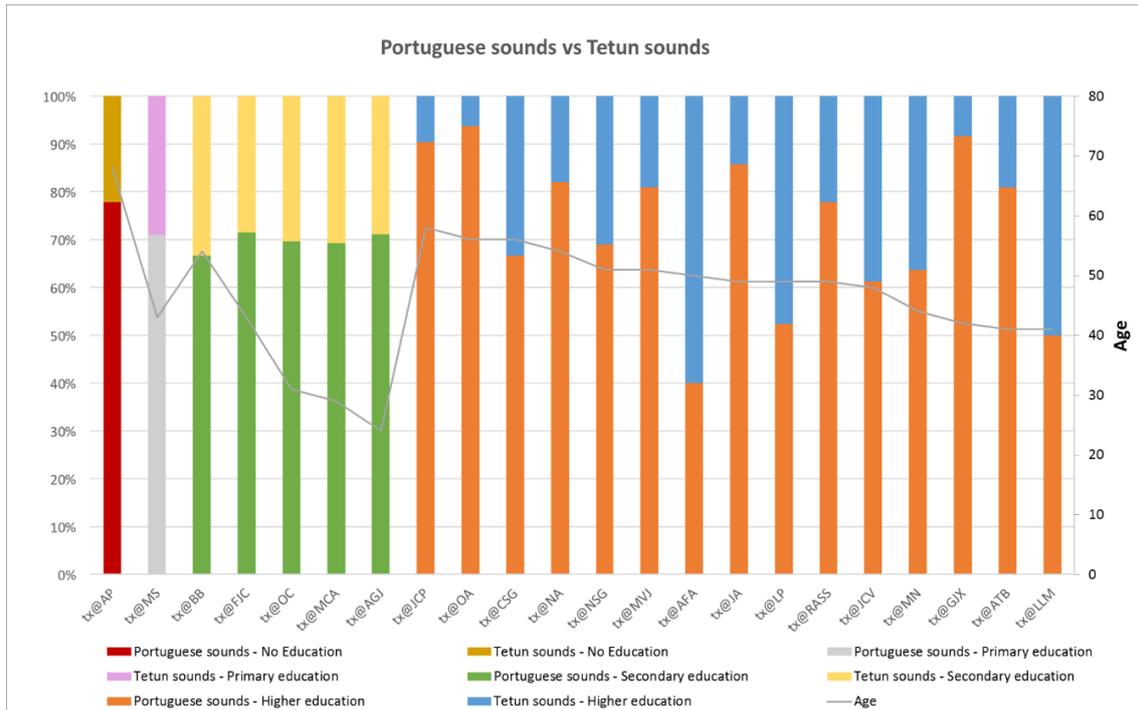
Figure 31 Comparison of the pronunciation of Portuguese sounds versus native sounds by Tetun Terik speakers according to their level of education (in %)

The third variable was the level of proficiency in Portuguese. As noted in the section of Methodology, speakers self-assessed their knowledge being able to list this language as a 'Other language I speak' (*Lia seluk ne'ebé ko'alia*) or 'Other languages I understand' (*Lia seluk ne'ebé ko'alia*) (see Annex 1). Those who did not list Portuguese as one of the languages they speak/understand were considered not to have any knowledge of it. This self-assessment might have led to some misrepresentations, as is clear from Figure 32, which shows that speakers with no understanding of Portuguese are more likely to opt for Portuguese sounds than speakers who understand Portuguese (61% vs. 50%). As expected, speakers fluent in Portuguese show the highest ratio of Portuguese phonetic realizations (68%). Most of the Tetun Prasa speakers claimed a certain level of proficiency, as is documented in Figure 33, unlike Tetun Terik speakers, who seemed to either speak Portuguese or have no knowledge of it, which makes us wonder whether they underestimated/overestimated their actual knowledge (see Figure 34).
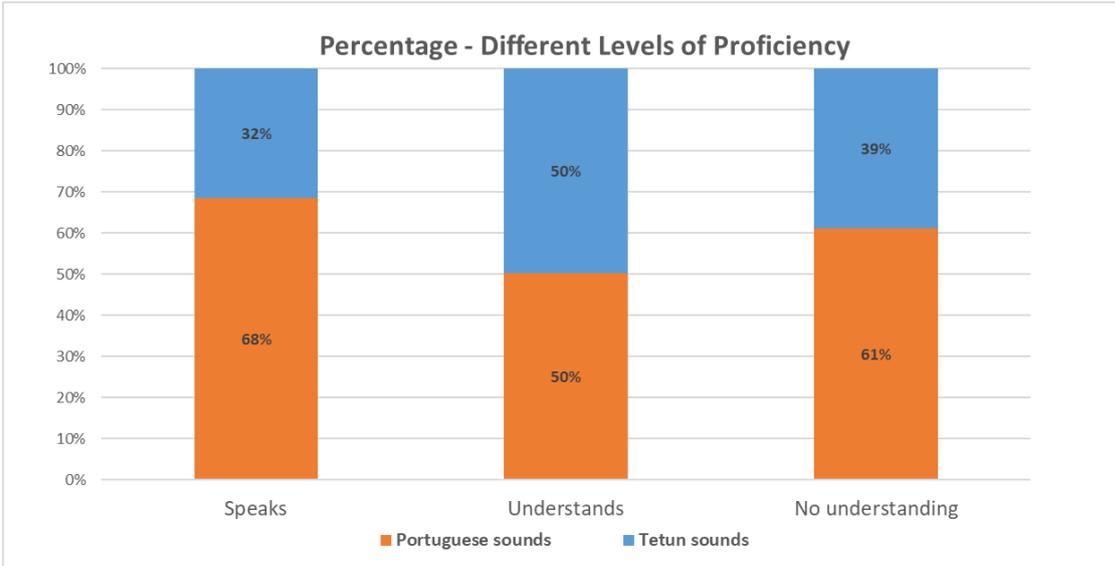
Figure 32 Overall representation of the use of Portuguese and native sounds by Tetun Terik and Tetun Prasa speakers according to their proficiency in Portuguese (in %)
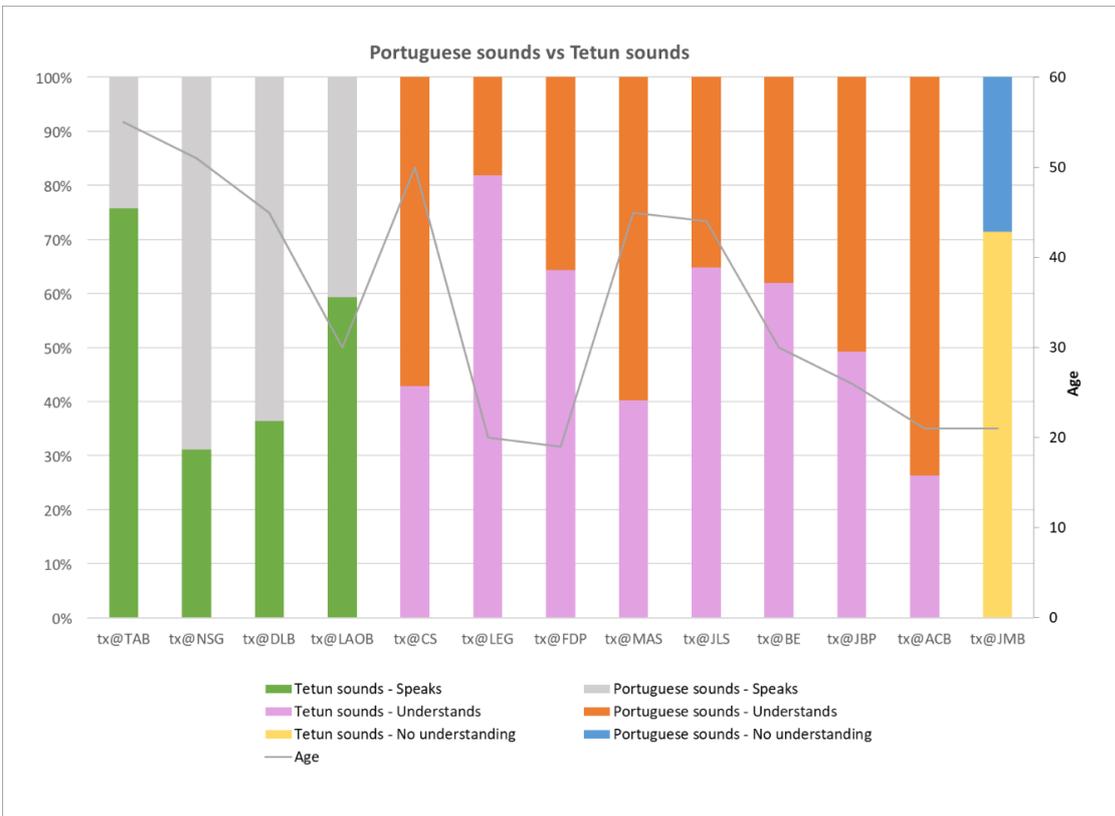
Figure 33 Comparison of the pronunciation of Portuguese sounds versus native sounds by Tetun Prasa speakers according to their proficiency in Portuguese (in %)
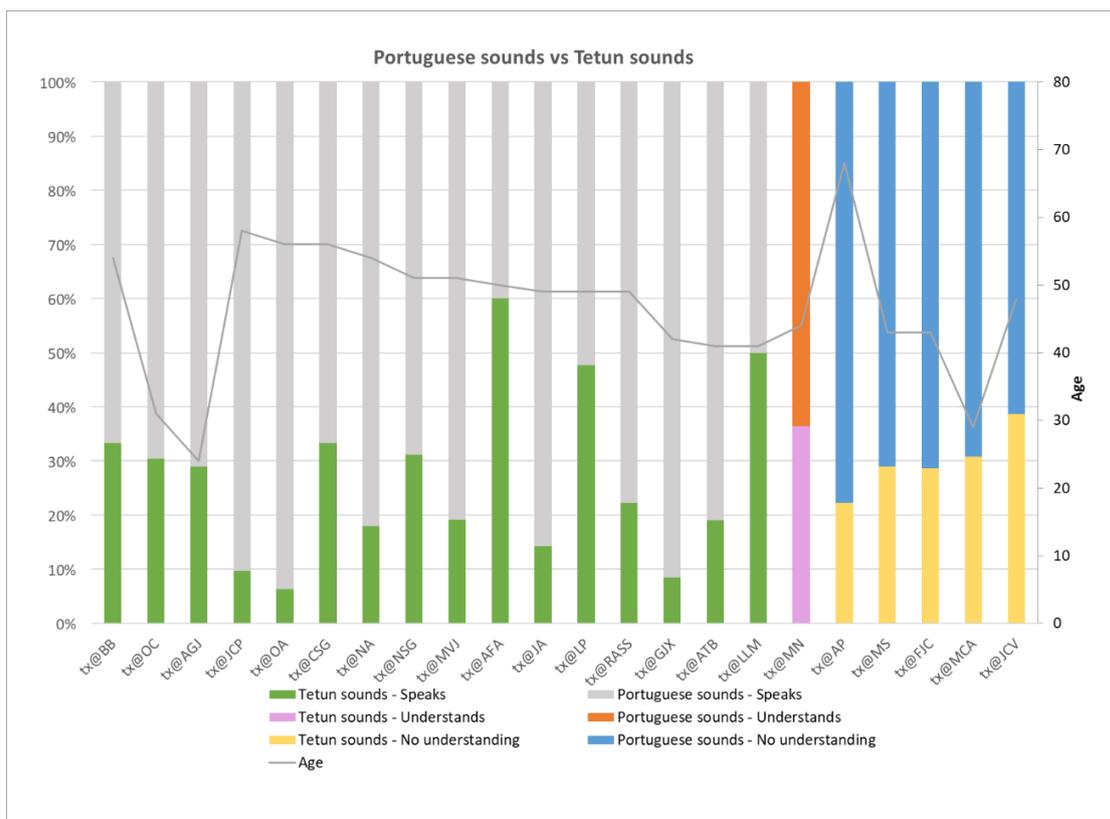
Figure 34 Comparison of the pronunciation of Portuguese sounds versus native sounds by Tetun Terik speakers according to their proficiency in Portuguese (in %)

We can see that speakers of Tetun Prasa, regardless of age, proficiency in Portuguese and level of education use both Portuguese and native sounds and since they all live in urban environment and are in daily contact with technical language, either at work or through media, they are familiar with Portuguese borrowings and their Portuguese pronunciation. Surprisingly, the fact that some speakers possess active and some passive knowledge of Portuguese does not seem to be the decisive factor. On the other hand, higher education and age above 46 favour the Portuguese pronunciation.

Based on my data, it seems like Tetun Terik speakers from Suai and Viqueque are also very familiar with Portuguese pronunciation, regardless of age, proficiency in Portuguese or education. Although they live in a rural environment, contact with and influence of Tetun Prasa is undeniable also in more remote areas. The spread of radio, which broadcasts in Tetun Prasa and Portuguese, has left its mark also on the more conservative varieties of Tetun. To summarize, knowledge of Portuguese or higher education are no prerequisites of the ability to acquire foreign sounds which were originally not part of the phonemic inventory of Tetun Prasa and Tetun Terik speakers.

8. LEXICON

"Unidade, Asaun, Progresu[187]"
      – the official motto of the Democratic Republic of Timor-Leste

In this chapter, I will look at loanwords borrowed from Portuguese as well as Malay/Indonesian and creoles based on them. I will also discuss different registers and provide some statistical analysis of my written and oral corpus.

8.1 Lexical borrowings

A lexical borrowing, or a loanword, is a word that, at some point in history, was borrowed from one language into another. In this chapter, I will use the terms 'borrowing' and 'loanword' interchangeably. I will refer to the language that was the source of borrowings as a 'source language' and the language that adopted borrowings as a 'recipient language'. In my analysis, I will also include structural borrowings, although their use has been discussed in previous chapters (see section 5.3 for derivational morphemes and chapter 6 for function words).

One of the very interesting debates in comparative linguistics focused on why some words were more likely to be borrowed than others. It is true that all languages have their internal mechanisms to create new words using their own resources instead of borrowing them from a foreign language. However, there are two main factors that influence the likelihood of lexical borrowings (Haspelmath 2009:35):

- social factors (e.g. prestige of the source language) and attitude (e.g. puristic attitudes);
- grammatical factors (e.g. some words classes, like verbs, are more difficult to borrow than nouns since they need more grammatical adaptation).

It is also important to distinguish between cultural borrowings and core borrowings[188]. Cultural borrowings describe new concepts introduced by source language

---

[187] "Unity, Action, Progress" (my translation)
[188] Core borrowings do not only occur in core vocabulary.

and culture, while core borrowings duplicate or replace native words that already exist in the recipient language. Both the factors influencing the borrowing process and the distinction between core and cultural borrowings will be discussed later in this chapter.

In the case of Tetun Prasa, I will only discuss borrowings from Portuguese as well as Malay/Indonesian and creoles based on them[189]. I will also look at English borrowings, although this category is marginal, since most of the English loanwords were probably introduced via Portuguese or Indonesian. Since many of these borrowings occur only once in my corpus, it is necessary to discuss the possibility of them being single-word code-switches, also known as nonce borrowings, rather than loanwords. According to Haspelmath (2009:40), the easiest and most reliable criterion to distinguish loanwords from code-switches is to see whether a particular word can be used in a situation where no code-switching takes place, that is, in the speech of monolinguals. However, considering that multilingualism is the norm in Timor-Leste, this does not help in solving my dilemma. We also cannot fully rely on frequency criteria. The fact that a certain word occurs only once in my corpus does not mean it is infrequent, especially if it can be found in monolingual and bilingual dictionaries of Tetun. The phonological adaptation criterion is helpful in the case of older borrowings, but recent borrowings usually preserve their original phonological form. As I discussed in Chapter 7, the phonological inventory of some speakers can contain as many as 22 consonants. And since Tetun Prasa is a rather isolating language, there is hardly any morphological adaptation of the loans, which makes it even more difficult to distinguish them from single-world code-switches. Also, if a large number of words is borrowed from one single source language (in the case of Tetun Prasa, Portuguese), there is less need for adaptation and a higher chance that the source language patterns will also be imported (Haspelmath 2009:43). I thus decided to treat all single-word occurrences as loanwords, adapting the Portuguese ones to the official orthographic conventions and preserving the original form in the case of Indonesian and English ones, as this seems to be the convention.

---

[189] Apart from loanwords, Tetun Prasa also makes use of calques, i.e. lexical items created by word-for-word translation of a source unit. One of the examples of calques are loanblends which have been described in section 5.4.

Because Tetun Prasa has a long history of contact, and also because of its rather isolating character, the adaptation, or the integration of new words has been rather straightforward. For example, Portuguese verbs are always borrowed in their 3S present indicative form (few exceptions include *fiar* 'believe, trust' and *pasiar* 'go for a stroll' in the infinitive) and do not take any inflections. Although identifying the origin of a word is easy in the case of Portuguese borrowings, identification of the language that introduced it can be tricky. As there were a number of languages spoken in the region throughout the history, sometimes it is impossible to say if it was Portuguese (or Portuguese-based creoles) or Malay (or Malay-based creoles) that was the source language. In cases when it is impossible to tell, I will list both the origin of the word and the language that most likely introduced it in Tetun Prasa.

So how does a word get borrowed into one's language? As detailed in section 3.1.1, there needs to be a contact between two groups speaking different languages, and one of the decisive factors that influences the extent of borrowings in the intensity of this contact. If the contact is casual, only non-basic vocabulary gets borrowed, but if the intensity increases, other kinds of borrowings are likely to enter the recipient language. But beside this social criterion, there is a linguistic one too: "(...) less tightly structured features are easier to borrow than features that fit into tightly integrated closed structures. So non-basic vocabulary items are the easiest to borrow, since in most languages a new noun (for instance), or even a verb, can be inserted readily into existing constructions." (Thomason 2001:69).

The first elements to be borrowed are cultural borrowings. In Tetun Prasa, these include everyday items like *janela* (from Port. *janela* 'window'), *meza* (from Port. *mesa* 'table'), *governu* (from Port. *governo* 'government'), *obrigadu* (from Port. *obrigado* 'thank you'). However, the incorporation of cultural borrowings goes both ways. For example, Portuguese, during its long-lasting presence in Southeast Asia, also borrowed many words from Austronesian languages, although only a limited number from Tetun[190].

---

[190] For example, *barlaque* (from M. *berlaki*) that was used in the Portuguese Timor to signify "compra de mulher segundo o rito gentílico" (Dalgado 1919:100) – buying a woman according to the pagans' customs; *suco* (from T. *suku*) to describe the lower administrative divisions of kingdoms, nowadays *postus administrativus* (Dalgado 1919:321); or *liurai* (from T.) to refer to a Timorese king or a ruler (Dalgado 1919:503).

The other category of loanwords are core borrowings. One of the reasons to adopt core borrowings is the cultural pressure of the dominant language, and the longer and the stronger this cultural pressure lasts, the more likely it is for the recipient language to also borrow structural features. In the case of lexical borrowings, it is not necessary for the speakers of the recipient language to be bilingual. In the case of other features, bilingualism is usually widespread (Thomason & Kaufmann 1991:37). Also, the introduction of structural borrowings does not happen within a few generations, but can take centuries. It depends mainly on three important social variables: relative population size, length of contact and degree of bilingualism.

8.2 Borrowability

It is very difficult to say what makes one language more likely to borrow a large number of lexical words. Based on the results of the Loanword Typology project (2004-2008), the authors concluded that all 41 languages considered exhibited very different typological and sociolinguistic types and there could not be any generalizations formulated about the ten languages with the highest and lowest borrowing rates.

In my analysis of Tetun Prasa, I tried to detect what kind of words were most likely to be borrowed: I looked at different word classes, especially nouns and verbs, content words and function words, and different semantic fields. But first I ran two standard tests to look at the borrowability of basic vocabulary. The most famous list of basic vocabulary is the Swadesh list, produced by the American linguist Morris Swadesh, which was published in a longer version containing 200 words and a shorter version containing 100 words (see Table 52). However, the items on the list are based purely on the author's intuition. I filled in the 100-item list and failed to find a single lexical borrowing, although one word was rather controversial: green. In Tetun, there is a word *modok* which could be translated as yellowy green, greenish or yellowish. To solve this dilemma, Tetun Prasa adopted a Portuguese loanword *verde* (from Port. *verde*) which describes the prototypical green colour. However, when talking about green leaves, the Timorese say *ai-tahan matak* 'green leaves' and, at the same time, *matak* 'raw, unripe' can also be translated as green, as is common in other languages like Portuguese.

Recently, a new list of basic words has been developed – the Leipzig-Jakarta list (see Table 52). It resulted from the Loanword Typology project and Tadmor (2009:72-73) described it as more reliable for scientific purposes as it is based on computational linguistic research and academic collaboration with a strong empirical foundation. The two lists share 62 items, which only shows that Swadesh's intuition was very good. Filling in the Leipzig-Jakarta list, I reached the same results as in the case of Swadesh list – all items were of native Tetun origin.

Swadesh's 100-word list (Swadesh & Sherzer 2017:283)			Leipzig-Jakarta List Tadmor (2009:69-71)		
1S pronoun	*ha'u*		fire	*ahi*	
2S pronoun	*ó* (2S)	*ita* (2S.HON)	nose	*inus*	
we	*ita* (1PI)	*ami* (1PE)	to go	*bá*	
this	*ne'e*		water	*bee*	
that	*ne'ebá*		mouth	*ibun*	
who?	*sé?*		tongue	*nanál*	
what?	*saida?*		blood	*raan*	
not	*la*		bone	*ruin*	
all	*hotu*		2S pronoun	*ó* (2S)	*ita* (2S.HON)
many	*barak*		root	*abut*	
one	*ida*		to come	*mai*	
two	*rua*		breast	*susun*	
big	*boot*		rain	*udan*	
long	*naruk*		1S pronoun	*ha'u*	
small	*ki'ik*		name	*naran*	
woman	*feto*		louse	*kutun*	
man	*mane*		wing	*liras*	
person	*ema*		flesh/meat	*na'an*	
fish	*ikan*		arm/hand	*liman*	
bird	*manu*		fly	*semo*	
dog	*asu*		night	*kalan*	

louse	*kutun*		ear	*tilus*	
tree	*ai*		neck	*kakorok*	
seed	*fini*	*musan*	far	*dook*	
leaf	*tahan*		to do/make	*halo*	
root	*abut*		house	*uma*	
bark	*ai-kulit*		stone/rock	*fatuk*	
skin	*kulit*		bitter	*moruk*	
flesh	*na'an*		to say	*dehan*	
blood	*raan*		tooth	*nehan*	
bone	*ruin*		hair	*fuuk*	
grease	*bokur*	*mina*	big	*boot*	
egg	*tolun*		one	*ida*	
horn	*dikur*		who?	*sé?*	
tail	*ikun*		3SG pronoun	*nia*	
feather	*manu-fulun*		to hit/beat	*baku*	
hair	*fuuk*		leg/foot	*ain*	
head	*ulun*		horn	*dikur*	
ear	*tilus*		this	*ne'e*	
eye	*matan*		fish	*ikan*	
nose	*inus*		yesterday	*horisehik*	
mouth	*ibun*		to drink	*hemu*	
tooth	*nehan*		black	*metan*	
tongue	*nanál*		navel	*husar*	
fingernail	*liman-kukun*		to stand	*hamriik*	
foot	*ain*		to bite	*tata*	
knee	*ain-tuur*		back	*kotuk-laran*	
hand	*liman*		wind	*anin*	
belly	*kabun*		smoke	*suar*	
neck	*kakorok*		what?	*saida?*	

breasts	*susun*		child (kin term)	*oan*	
heart	*fuan*		egg	*tolun*	
liver	*aten*		to give	*fó*	
drink	*hemu*		new	*foun*	
eat	*han*		to burn (intr.)	*sunu*	
bite	*tata*		not	*la, lae*	
see	*haree*		good	*di'ak*	
hear	*rona*		to know	*hatene*	
know	*hatene*		knee	*ain-tuur*	
sleep	*dukur*		sand	*rai-henek*	
die	*mate*		to laugh	*hamnasa*	
kill	*oho*		to hear	*rona*	
swim	*nani*		soil	*rai*	
fly	*semo*		leaf	*tahan*	
walk	*la'o*		red	*mean*	
come	*mai*		liver	*aten*	
lie	*toba*		to hide	*helik*	*subar*
sit	*tuur*		skin/hide	*kulit*	
stand	*hamriik*		to suck	*susu*	
give	*fó*		to carry	*lori, hodi*	
say	*dehan*	*hateten*	ant	*nehek*	
sun	*loro*		heavy	*todan*	
moon	*fulan*		to take	*lori*	
star	*fitun*		old	*tuan*	
water	*bee*		to eat	*han*	
rain	*udan*		thigh	*kelen*	
stone	*fatuk*		thick	*mahar*	
sand	*rai-henek*		long	*naruk*	
earth	*rai*		to blow	*huu*	

358

cloud	*kalohan*		wood	*ai*	
smoke	*suar*		to run	*halai*	
fire	*ahi*		to fall	*monu*	
ash	*latun*	*ahi-kesan*	eye	*matan*	
burn	*sunu*		ash	*latun*	*ahi-kesan*
path	*dalan*		tail	*ikun*	
mountain	*foho*		dog	*asu*	
red	*mean*		to cry/weep	*tanis*	
green	*modok*	*verde* (PT)	to tie	*kesi*	
yellow	*kinur*		to see	*haree*	
white	*mutin*		sweet	*midar*	
black	*metan*		rope	*tali*	
night	*kalan*		shade/shadow	*mahon*	
hot	*manas*		bird	*manu*	
cold	*malirin*		salt	*masin*	
full	*nakonu*		small	*ki'ik*	
new	*foun*		wide	*luan*	
good	*di'ak*		star	*fitun*	
round	*kabuar*		in	*iha*	
dry	*maran*		hard	*susar*	*todan*
name	*naran*		to crush/grind	*de'ut*	

Table 52 Swadesh's list (on the left) and Leipzig-Jakarta list (on the right)

As we can see, Tetun Prasa's basic vocabulary is resistant to borrowings and the Austronesian nucleus is almost intact. Together with the sociolinguistic history of Tetun described in Chapter 2, this should be enough evidence for arguing against Tetun Prasa being a Portuguese-based creole. However, it is undeniable that Tetun Prasa borrowed a fair share of words. Let us now look at the different periods of contact with Malay, Portuguese and Indonesian. As it is almost impossible to say which variety of Malay and Portuguese was the source of borrowings in many cases, I will refer to them as 'M.' for

all local and restructured varieties of Malay and 'Port.' for all local and restructured varieties of Portuguese.

8.3 Malay influence

Malay influence has been undeniable in the whole of Southeast Asia. It was the language of the maritime trade in former Insular Southeast Asia in the 15th century and it is not surprising that Tetun borrowed many words from it. However, sometimes it can be difficult to identify these borrowings since Malay and Tetun both share a considerable Austronesian lexicon. Because we lack any historical documents on the Malay trade in Timor, we need to look at the nature of these borrowings. What we know about those early dates is that the traders most probably started visiting Timor in the 15th century and came from Ambon and central Maluku (Hull 2005a:84), and their presence was almost exclusively commercial along the coast. There were several ports along the northern coast, which were visited and documented by Chinese traders (see Map 4) also in the 15th century. Because of the limited interaction between the traders and the local population, the oldest loanwords from Malay are mostly cultural borrowings and belong to specific semantic domains. Looking at their phonology, Hull (2005a:103) divided Malay loanwords into two categories according to two historical periods: old Malayisms (15th-17th cent) and new Malayisms (not specified by Hull). In total, he managed to identify 450 of them, mostly cultural borrowings. Many of these words were restricted to the coastal dialects of Tetun as Lifau, and later Dili, became the main port of entry for Malay loanwords. However, taking a closer look at the Malay borrowings Hull (2005a) identified, I have considerable doubts about their foreign origin. Some, like *mina* 'oil' (M. *minyak*) or *tua* 'palm wine' (M. *tuak*) seem to be of Austronesian origin rather than Malay borrowings. In other cases, the suggested sound change seems just unimaginable, e.g. M. *daging* 'meat' > T. *na'an* 'meat', or *lebih* 'more' > T. *lai* 'first' (Aone van Engelenhoven, p.c.). One way of resolving this would be to look at other Timorese languages and Austronesian languages spoken on the neighbouring islands and see if they contain similar cognates. But because this data is not available to me, I decided to adopt Thomaz's more conservative approach to identifying Malay borrowings. Thomaz

(1974:230) examined 80 of the most obvious Malay loans which could be categorized as cultural loans. He based his analysis on linguistic and extra-linguistic criteria: we can only be sure about those Malay borrowings that are of Sanskrit, Arabic or other origin, or which contain foreign phonemes (like /p/ and /g/), or which describe an object only recently introduced in Timor, or, with less certainty, which are only used on the coast.

Albuquerque (2011:70-71) also believed Malay influence consisted of two layers: old borrowings from Bazaar Malay that were adapted to the phonological rules of Tetun Prasa and that were readily used across all sociolinguistic groups, and recent borrowings that were introduced during the Indonesian period via the Indonesian language. These new borrowings usually belong to specific semantic fields (military, politics, food) and are only used by certain sociolinguistic groups (speakers with a lower education and speakers between 20 and 40 years of age).

When it comes to old borrowings, the most numerous semantic fields are (Hull 2005a, Thomaz 1974):

- nature: *karán* 'reef' (from M. *karang* 'reef');
- substances: *besi* 'iron' (from M. *besi* 'iron'), *kaleen* 'tin' (from M. *kaleng* 'tin');
- animals: *kuda* 'horse' (from M. *kuda* 'horse' of Tamil origin), *kurita* 'octopus' (from M. *gurita* 'octopus' of Sanskrit origin);
- maritime: *lepa* 'small boat without outriggers' (from M. *lepap* 'boat with flat bottom');
- people: *durubasa* 'interpreter' (from M. *juru bahasa* 'interpreter' of Sanskrit origin), *nona* 'unmarried girl; mistress of a European' (from M. *nona* 'young lady').

Many of these loanwords were lost due to substantial Portuguese influence or they became obsolete or archaic. However, Portuguese was never able to eradicate the Malay lexical element completely.

The reason why it is often so difficult to identify the source language is quite clear. Malay and Portuguese have been in contact for centuries and local and creole varieties of both languages emerged across the region. The trade language Bazaar Malay (or Pasar

361

Melayu) and its varieties (like Ambon Malay and Kupang Malay), together with Malacca Creole Portuguese, Macau Creole Portuguese and Batavia Creole Portuguese, all shared many words. Malay began borrowing from Portuguese at the beginning of the 16th century, when the Portuguese captured Malacca (Hull 2005a:117). Although the Portuguese presence lasted only for just over a century, hundreds of Portuguese loanwords have survived in modern Malay[191]; others were replaced by Dutch counterparts in modern Indonesian. However, many of them were preserved in the restructured varieties that Tetun was in contact with. Ambonese Malay retained many Portuguese borrowings that are not used in other varieties: *forsa* 'force, strength' (from Port. *força* 'force, strength' vs. Indo. *kekuatan*), *kadera* 'chair' (from Port. *cadeira* 'chair' vs. Indo. *kursi*), or *kalsan* 'trousers' (from *calças* 'trousers' vs. Indo. *celana*) (Teixeira 1963:471). One of the reasons why the influx of Portuguese words into Malay was so extensive was the fact that although the Portuguese were expelled from Sulawesi and Moluccas, their language was retained by the so-called Mardijerks as a colloquial language in Batavia (present-day Jakarta). It was due to the influence of Batavia Malay that many of the Portuguese loanwords became incorporated in the modern newspaper language. Originally, the influence of Portuguese was local (e.g. in Moluccas and Batavia) with the written Malay language almost unaffected, but with time, Batavia's influence on colloquial Malay grew, together with the number of loanwords, also due to the influence of the powerful Malay-Chinese press (Teixeira 1963:468-472).

According to Hull (2005a:117-118), it was via Malay that the earliest lusisms entered Tetun, especially words referring to Catholicism and material culture of the Portuguese: *karreta* 'car' (via M. *kereta* from Port. *carreta* 'cart'), *kreda* 'church' (via M. *gereja* from Port. *igreja* 'church'), *maski* 'although' (via M. *meski* from Port. *por mais que* 'although') The same forms are found also in other SE Asian varieties of creole Portuguese, e.g. Malacca Creole Portuguese has *kareta* 'car', *greza* 'church' and *maski* 'although' (Baxter & de Silva 2004), Batavia/Tugu Creole Portuguese has *gredja* 'church' and *kareta/karetta* 'car' (Schuchardt 1891) but very few of these words were preserved in this form after the Portuguese settled down in Timor. Due to the long-lasting

---

[191] In 1913, Dalgado identified 431 Portuguese borrowings in Malay (see Cardoso 2016a:74). Also, basing his numbers on the work of Santa Maria (1967), Baxter (1996:305) confirmed that there are 312 words of definite Portuguese origin and 95 words of possibly Portuguese origin in Malay. Most of them are nouns.

Portuguese influence, they assimilated back to their original etyma and in modern Tetun Prasa, *karreta* 'car' and *igreja* 'church' are now the norm. On the other hand, many Malay borrowings were preserved in their archaic form and no longer exist in modern Malay, e.g. *sasate* 'satay' (now *satai* or *sate* in M.) or *lalakun* 'manner of speaking' (now *lagu* 'song, tune' in M.) (Thomaz 1981:8). Thomaz (1974:242) concluded that the number of Malay borrowings in Tetun was a sign of the fact that, although not very intense, Malay presence was rather long-lasting, especially in the coastal areas. The variety of words also shows that the contact was not only carried out via commerce but there must have been a considerable number of bilinguals who helped to spread the language.

It is much easier to identify the borrowings that were introduced into Tetun Prasa from Indonesian, since they have not changed in form or pronunciation. During the Indonesian occupation (1975-1999), it was forbidden to speak Portuguese, and the Indonesian government did a tremendous job, through aggressive Indonesianisation, of turning most of the population into Indonesian speakers. Although Tetun was still used as a lingua franca, education was carried out in Indonesian. Portuguese was also replaced in media and administration (Hajek 2000b:405). Indonesian borrowings covered specific semantic domains:

- military: *intel* 'spy' (from Indo. *intel*), *tentara* 'Indonesian soldier, army' (from Indo. *tentara* 'army');
- politics: *integrasi* 'integration' (from Indo. *integrasi*);
- food: *warung* 'Indonesian restaurant' (from Indo. *warung), bakso* 'soup with meatballs' (from Indo. *bakso), eskrim* 'ice cream' (from Indo. *eskrim*[192]);
- numbers: when talking about price in the market or when specifying dates, people still often use Indonesian numbers, e.g. *satu* 'one', *dua* 'two', *tiga* 'three'.

However, after independence, the government (and the National Institute of Linguistics) decided that, although Indonesian was one of the working languages,

---

[192] *Eskrim* is clearly of English origin (*ice cream*) but it is more likely that it was borrowed into Tetun via Indonesian, since other Indonesian borrowings referring to frozen snacks were borrowed together with it, e.g. *es* 'ice lollipop', *es campur* 'ice mixed with fruits and syrup'.

Indonesian borrowings should be slowly replaced by either Tetun Terik forms or Portuguese loanwords. Nowadays, Indonesian is not heard in everyday life anymore, but Indonesian borrowings are still present, especially in the speech of Timorese who were born and educated during the Indonesian times. Television has also had influence on the spread of the Indonesian language. Currently, there are only few TV stations in Tetun Prasa and Portuguese (*Televizaun Timor Lorosae*, *Televisao Edukasaun Timor*, *STL TV*, *TV Maubere*), while the rest of the channels that are accessible via satellite transmit in Indonesian.

8.4 Portuguese influence

Although the Portuguese influence can be visible at first sight, documenting it can be rather difficult due to the lack of older documentation, as well as the coexistence of many L1 and L2 varieties of Portuguese in this region. Albuquerque (2011:73) divided Portuguese influence into five historical periods:

- 1st period (16th – 18th century): influence via Malacca Creole Portuguese, which was used as a lingua franca although there are no written sources from this period;
- 2nd period (18th – 19th century): influence of Macau Creole Portuguese, which is well documented by historical and primary linguistic sources;
- 3rd period (1898 - 1974): establishment of the Colégio de Soibada marks the implementation of Portuguese-based education in Portuguese Timor;
- 4th period (1975 - 1999): during the Indonesian occupation, Portuguese was used only in the rural and mountainous regions by the members of the resistance movement;
- 5th period (2002-present): Portuguese became the official language which led to 'renovation and modernization of the Tetun lexicon'.

It is difficult to see how Malacca Creole Portuguese could have influenced Tetun before the Portuguese administration settled down in Lifau in 1702 (see Chapter 2 for historical overview). The Portuguese spoken in this area before the 18th century most

probably came from the Portuguese settlements in Flores (Larantuka, Ende) and Solor (Alan Baxter, p.c.) (see Map 5). When the Portuguese finally settled in Dili in 1769, the Portuguese language was already losing its importance in the Far East. Tetun, on the other hand, was already a language of contact among local peoples. The limited number of Portuguese and other foreigners prevented Portuguese from spreading and replacing Tetun as a lingua franca. However, Portuguese influence on Tetun was significant, especially with the development of Dili as a commercial and administrative centre. The lack of modern vocabulary was the cause of massive borrowing from Portuguese and Portuguese-based creoles, especially Macau Creole Portuguese, due to a thriving sandalwood trade, which was coordinated from Macau. According to Thomaz (1974:251), the number of Portuguese settled in Timor was always very low. In the second half of the 19th century, there were only 100 Portuguese in Timor-Leste. Until the end of the 19th century, these were mostly Dominican missionaries, Macau traders and administrators, and deportees.

Missionaries working in the districts soon adopted local languages to spread the Catholic religion, so, in general, Portuguese was used as a vehicular language only in the city[193]. It was thus a language of administration and of an imported culture, and there was only a small group of people who were bilingual in both Tetun and Portuguese. In 1876, Affonso de Castro already commented on the numerous loanwords in Tetun: "O Teto usa muitas palavras portuguezas, que os naturaes de Timor adoptaram, à falta de termos para exprimir objectos que lhes eram desconhecidos antes da conquista."[194] (Castro 1867:328).

The first Tetun dictionary, compiled by Silva (1889), reflected this situation, as it contained more than 1000 Portuguese loanwords, which Silva justified by saying that "[n]ão se encontrarão todas as palavras que em portuguez trazem os nossos diccionarios, por não as haver correspondentes em *tétum*, pois não as ha senão para exprimir o que é conhecido em Timor"[195] (Silva 1889:iv). Silva especially struggled with finding

---

[193] According to Silva (1885:7), some missionaries also gave "muitas palestras no algaraviado portuguez d'aquelle paiz." ["many speeches in the gibberish Portuguese of that country"] (translation in Baxter & Cardoso 2017:267)

[194] "Tetun uses a lot of Portuguese words, that the local Timorese adopted, because they lacked words to describe objects that were alien to them before the conquest." (my translation)

[195] "not all the words that are found in our Portuguese dictionaries are found here, because they do not have a counterpart in Tetun, which can only express notions known to Timor." (my translation)

counterparts to abstract nouns. As he said, there were none in Tetun. Before publishing his dictionary, he compiled a catechism in Tetun (*Catecismo da Doutrina Christã em Tétum,* 1885), the first known publication in this language. He had to use many Portuguese borrowings to be able to introduce concepts of the Catholic faith to the Timorese. Excerpting the first ten pages of his translation of the Catechism, we find nouns like *fé* 'faith', *cruz* 'cross', *diabo* 'devil', *espírito* 'spirit', *castigo* 'punishment', *inferno* 'hell', *anjo* 'angel', *jejum* 'fasting' and verbs like *fiar* 'to trust', *precisa* 'to need', *reza* 'to pray', *adora* 'to worship', *obedece* 'to obey' (all in original spelling), as shown in Table 53. By no means does this mean that these words did/do not exist in the native Tetun vocabulary. It is more than possible that native words just weren't available to Aparício da Silva at that time or he wanted to avoid words with a pagan connotation, so he opted for Portuguese borrowings. However, if we consult the modern Tetun dictionaries, we see that all these words, except for *obedece*, are now part of the Tetun Prasa lexicon, often without a native counterpart.

Portuguese	Tetun	English
fiar (v.)	fiar	to trust
fé (n.)	fé	faith (in God)
precisar (v.)	presiza	to need st.
rezar (v.)	reza	pray in God; but pray for sb. = harohan
		pray in animist religion = hamulak
cruz (n.)	krús	cross
diabo (n.)	diabu	devil
espírito (n.)	espíritu	spirit
castigo (n.)	kastigu	punishment
inferno (n.)	infernu	hell
adorar (v.)	adora	to worship
anjo (n.)	anju	angel
obedecer (v.)	halo tuir	obey
jejum (n.)	jejún	fasting
contra (prep.)	kontra	against; disobey

Table 53 Some Portuguese borrowings used by Aparício da Silva in his Tetun translation of the Catechism

But Silva was not the only one to struggle with the lack of words to describe abstract notions. Another missionary, Father Abílio José Fernandes, also tried to justify the introduction of Portuguese loanwords:

"Estas línguas, ou dialectos, se o leitor prefere, são tôdas ricas em termos e expressões referentes a objectos ou trabalhos materiais, mas pobríssimas em tudo o que diz respeito ao espírito. Não possuem substantivos abstractos. Daí a necessidade de introduzir termos portugueses para explicar a doutrina cristã. E, caso raro, talvez único no globo, não há em nenhuma destas línguas palavra ou expressão de gratidão. Todos introduziram a nossa expressão: *obrigado*."[196] (Fernandes 1931:20).

However, not all the words explaining the Christian faith were borrowed from Portuguese. Tetun also created new words from its own linguistic resources. Portuguese missionaries adopted native *Maromak* for God which originally meant 'bright, shining', or *na'i-lulik* 'priest' from *na'i* 'lord' and *lulik* 'sacred' (Thomaz 1981:9).

Raphael das Dores, the author of the second Tetun dictionary, adopted a more purist attitude. He argued that Portuguese missionaries were using too many Portuguese words while giving their sermons in Tetun:

"Essa introducção tem-se feito sempre devido aos empregados que vão em serviço ao interior da ilha, e que á falta de outro expediente as vão usando no seu trato com os naturaes, e principalmente aos missionarios que as introduzem não só no seu trato, mas ainda nas suas praticas e sermões feitos na lingua do pais, mesmo antes de a saberem bem, com o que provam a sua dedicação pela missão religiosa, sem lhes importar com a pureza de linguagem."[197] (Dores 1907:6).

So, in his dictionary, Dores decided to include only those Portuguese loanwords that the author, with the help of his friends, documented between 1871 and 1873, during his first and longest posting in Timor, and which were already fully adopted by the local speakers. The rest of the Portuguese borrowings, exclusively collected by Silva,

---

[196] "These languages, or dialects, if you prefer, are all rich in terms and expressions referring to material objects and tasks, but very poor in everything connected with spirituality. They do not have abstract nouns. That is why there is a necessity to introduce Portuguese words to explain the Christian doctrine. They are also a rare case, maybe unique in the whole world, because none of these languages has a word for expressing gratitude. They all adopted our expression: *obrigado* ['thank you']." (my translation)

[197] "This introduction was done by employees who went to work inland, and who kept using [the Portuguese loanwords] because they lacked a different medium to communicate with the locals. Especially the missionaries used them not only in their dealings, but also in their practice and sermons conducted in the language of the country, even though they didn't speak it well, which only shows how dedicated they were to their religious mission, not caring about the purity of the language." (my translation)

constituted a separate list since they were thought to be recent loans although some of them were already established borrowings in Tetun, as the author noted himself (Cardoso 2017:19).

Dalgado (1913) was one of the first linguists who tried to map the Portuguese influence on Asian languages. He studied 52 languages spoken across the whole Asian continent and compiled a list of lexical borrowings for every one of them. According to his counts, Tetun was the language with the second highest number of Portuguese loanwords: 774, preceded by Konkani (spoken in Goa) with 1768 loanwords and followed by Malay with 431 loanwords. It is believed these numbers were lower than the actual state, since the author did not have full access to the etymological information of all these languages at the time (Cardoso 2016a:74). Or, as is the case of Tetun, he decided not to include all the Portuguese words found in Tetun dictionaries published in those times:

> "Sendo Timor uma possessão portuguesa e teto um idioma muito pobre, é óbvio que o seu vocabulário está inçado de termos portugueses, mais ou menos vulgarizados, conforme o maior ou menor contacto dos seus povos com a civilização europeia. Não reproduzo no vocabulário todas as palavras portuguesas consignadas nos dicionários do teto e do galóli, muitas das quais são evidentemente dadas para suprir a falta de correspondentes vernáculos; mas incluo-as nos seus respectivas índices."[198] (Dalgado 1913:64-65).

Indeed, no language is resistant to language contact, especially when it is spoken in a multilingual environment. But the foreigners did not only introduce new words, but new customs too. Greetings are a good example of this. Originally, the locals used to greet each other with gestures rather than words. But with the arrival of the Portuguese, they adopted expressions like *bondia* (from Port. *bom dia* 'good morning') and *bonoite* (from Port. *boa noite* 'good evening/night'), although Tetun did possess words like 'good' (*di'ak*), 'morning' (*loron*) and 'night' (*kalan*) (Dores 1907:7).

---

[198] "Timor being a Portuguese possession and Teto, a very poor language, it is obvious that its vocabulary is laden with Portuguese terms, which are current in a lesser or greater degree, according to the extent of the contact of the peoples with European civilization. I do not mention in this work all the Portuguese words that find a place in dictionaries of Teto or Galoli, many of which are entered evidently to make up for the want of the corresponding vernacular terms; but I am finding a place for such in their respective indices." (Dalgado 1936:xciii)

Suddenly, there was a great need for Tetun to adapt to the new cultural concepts (with the introduction of the Christian calendar, people started celebrating *Natál* 'Christmas', from Port. *Natal*, and *Páskua* 'Easter', from Port. *Páscoa*, etc.) as well as to the ever-changing modern civilization. Thomaz (1981:14) stressed the interdependence of linguistic and historical evolution, and I will present a few examples to document how the arrival of foreigners not only influenced the language but also changed Timorese society.

Foreigners visiting or living in Timor are often referred to as *malae* 'foreigner'. Just like the Indonesian *bule* 'white person' or Thai *farang* 'a person of white race', *malae* is a non-derogatory term referring to a foreigner. In this case, the foreigner does not necessarily need to be Caucasian, as this term is also applied to nationals of other SE Asian countries. Indeed, the origin of the word is clearly Malay[199] (from M. *Melayu* 'Malay'), as the Malay traders were the first ones to land on the Timorese coast with the intention of carrying out commerce. However, according to Sá (1961:200), the use of *malae* was restricted to the Portuguese, possibly because they were the most numerous group:

> "Os indigenas designam assim todos os individuos que não são timorenses, com este pequeno pormenor: aos «Portugueses» chamam simplesmente *malai*, mas para os individuos de outras nacionalidades acrescentam a este termo o designativo patronimico, ou qualquer expressao designativa duma nota caracteristics: *malai-ingles*, «o ingles», *malai-balanta*, «o holandes», *malai-sina*, «o chines»; *malai-metan*, «estrangeiro preto», «o africano»; *malai-matan-bubu,* «estrangeiro de olhos sapudos», «o japones»".[200] (Sá 1961:200).

Presently, to specify the race, it is possible to use *malae-mutin* 'white foreigner' or *malae-metan* 'black foreigner'. Indonesians can also be referred to as *malaes*, but both

---

[199] According to Hull (2005a:94), the word *Melayu* also has the meaning 'foreigner' in Ambonese Malay and based on the /e/ > /a/ change, it is likely that the word was introduced into Tetun via this variety, as discussed in section 7.3.

[200] "The natives called [malai] all the people who were not Timorese, with a small detail: the Portuguese were called simply *malai*, but individuals of other nationalities carried also a term indicative of their origin, or any expression referring to some of their distinctive features: *malai-ingles*, «Englishman», *malai-balanta*, «Dutch», *malai-sina*, «Chinese»; *malai-metan*, «black foreigner», «African»; *malai-matan-bubu,* «foreigner with slanted eyes», «Japanese»". (my translation)

men and women are also often called simply *bapa* (from Indo. *bapak* 'father, older man').
Malae can also be used as a modifier in nominal compounds, creating thus new words
referring to non-native concepts: *bibi-malae* 'sheep' (from T. *bibi* 'goat'), or *ai-ata-malae*[201] 'soursop' (from T. *ai-ata* 'custard apple'). According to Thomaz (1981:11), the
word *makau* 'Macau; foreign' used to have the same connotation as *malae*. It was
preserved only in one expression: *fehuk-makau*[202] 'potato' (from T. *fehuk* 'sweet potato').

The introduction of the Christian calendar meant that a year was divided into
months, as shown in (362), and a week into days, as shown in (375). All these words were
borrowed from Portuguese, together with the main Christian holidays, which the
Timorese started celebrating. They were also given Christian names, so called *naran-sarani* (from T. *naran* 'name' and M. *serani*[203] 'Christian'), that they used along with
their native names, *naran-jentiu* (from Port. *gentio* 'pagan').

(362)			(363)		
Janeiru	'January'			*segunda*	
Fevereiru	'February'			*segunda-feira*[204]	'Monday'
Marsu	'March'			*loron-segunda*	
Abríl	'April'				
Maiu	'May'			*tersa*	'Tuesday'
Juñu	'June'			*kuarta*	'Wednesday'
Jullu	'July'			*kinta*	'Thursday'
Setembru	'September'			*sesta*	'Friday'
Outubru	'October'			*sábadu*	'Saturday'
Novembru	'November'			*domingu*	'Sunday'
Dezembru	'December'				

Portuguese numbers are also frequently used, especially for dates and large
numbers. Again, large numbers do exist in Tetun, but for practical reasons, they are
seldom used.

---

[201] Custard apple is native to Mexico and the Tetun term *ata* 'sugar apple' comes from an indigenous language of Mexico. Soursop is also native to Central and South America but was probably brought over at a later stage.

[202] Also called *fehuk-malae* or *fahuk-ropa* (*ropa* from Port. *Europa*)

[203] It is interesting that the word for 'Christian' is derived from Arabic *naṣārā* 'Christian', literally 'Nazarene' (Hull 2005a:94).

[204] *-feira* from Pr. *feira* 'fair' is used with Monday to Friday, just like in Portuguese. On the other hand, T. *loron* 'day' can be used with any day of the week.

When the Portuguese arrived in Timor in the 16th century, Timor was divided into kingdoms (*rai*) ruled over by *liurais* (traditional Timorese rulers). The kingdoms were further divided into *sukus* 'village' lead by *datos* 'chiefs'. [205] Each *suku* consisted of several *knuas* 'hamlet' and it was run by a *katuas* 'older man' (Sá 1961:154). At the end of the 19th century, the Portuguese Crown introduced its own administrative structure. The eastern part of the island was then divided into 13 districts (T. *konsellu* from Port. *concelho*), and subsequently subdistricts (T. *postu* from Port. *posto* 'post'), villages (T. *suku*) and hamlets (T. *aldeia* from Port. *aldeia* 'village'). The terminology has changed a couple of times throughout history. During the Indonesian times, Timor-Leste (*Timor Timur* or *Tim-Tim* in Indonesian) became the 27th Indonesian province and, while the administrative division was preserved, the nomenclature changed: *konsellus* were renamed *kabupatens* (from Indo. *kabupaten* 'regency') and *postus* were called *kecamatan* (from Indo. *kecamatan* 'subdistricts'). After independence, *kabupatens* came to be called *distritus* (T. *distritu* from Port. *distrito* 'district'), and were divided into *subdistritus*. Recently, the nomenclature changed again, districts are now called *munisípius* (from Port. *município* 'municipality'), subdistricts are again referred to as *postus administrativus* (from Port. *postos administrativos* 'administrative posts'), subdivided into *sukus* and subsequently into *aldeias*.

Apart from the administrative division, the Portuguese also introduced a new organization of society. Originally, Timorese society was based on ranks and not social classes, the important aspects being nobility, power and authority, but not wealth (Thomaz 1981:12). Therefore, Tetun lacked words for rich (*riku* from Port. *rico* 'rich') and poor (originally from T. *kiak* 'orphan', now meaning 'poor'). Due to these social changes (development of urban society, increased social mobility, rise of new social classes), the traditional courtesy language ceased to be widely used. Nowadays, its use in Tetun Prasa is rare and is only restricted to God or saints in the prayers (see section 1.3) (Thomaz 1981:12).

---

[205] It is interesting that the Indian concept of state, power and kinship (like the Sanskrit title *raja*) is found all over the archipelago but not east of Lombok. Tetun Prasa uses either native *rai* 'country, region' or *reinu* (from Port. *reino* 'kingdom') (Thomaz 1981:7).

Timorese society was also illiterate (Thomaz 1981:13) and writing and education was introduced only with the arrival of the Portuguese. Words like *lee* 'read' (from Port. *ler* 'read'), *hakerek* 'write' (originally T. 'to draw, carve') and *surat* 'paper' (from M. *surat* 'letter, document') where introduced or acquired new meaning together with the introduction of writing (Thomaz 1981:13). The first schools, *kolégius* 'boarding schools' in Soibada and Manatuto, were founded by the Catholic mission, followed by a few *eskolas primárias* 'primary schools'. The language of instruction was Portuguese (see section 2.7) and, as we will see later in my analysis, words related to school and education are almost entirely borrowed from Portuguese and are also abundant in the speech of Tetun Terik speakers who are not proficient in Portuguese.

8.5 Current state (post-1999)

The newly-earned independence posed many challenges to the local language policies. The choice of Tetun and Portuguese as the official languages, with English and Indonesian as working languages, gave way to another wave of new loanwords entering Tetun, and Tetun Prasa started to diverge even more from Tetun Terik. And although more than 80% of the population speaks Tetun Prasa as an L2, Tetun Terik in its purest form would be hardly comprehensible to most of them. According to Hull (1999:x), this is because Tetun Prasa:

> "(...) has been impoverished by the loss of many words referring to agricultural life and indigenous culture, the rich range of synonyms (especially verbs) and the traditional poetic register of Tetun-Terik. In addition to this, many of the basic hereditary lexemes still used in Tetun-Terik have been replaced by lusisms."

In his dictionary, Hull (1999:xi) divided Tetun vocabulary into five lexical classes (words of indigenous origin common to all varieties; words typical of Tetun Prasa; words common to all varieties of Tetun Terik; words restricted to certain dialects of Tetun Terik; and words of recent Portuguese origin not belonging to previous classes) while adding that it is often difficult to draw a dividing line between Tetun Prasa and Tetun Terik.

Williams-van Klinken et al. (2002a:1) also claim that the two varieties (Tetun Prasa and Tetun Terik) are considered mutually unintelligible because of the number and source of the borrowings. According to the authors, Tetun Terik shows much greater influence of Malay and Indonesian with relatively limited contact with Portuguese. As I have mentioned, Tetun Prasa has been in contact with Malay for centuries and imported many words from Indonesian, however, there has been a conscious effort to 'purify' Tetun Prasa of recent Indonesian borrowings. On the other hand, some dialects continue to be in close contact with Indonesian because they are spoken in border regions and there is no pressure on them to adapt to the puristic attitude.

In general, loanwords in Tetun Prasa can be divided into four categories, depending on how deeply rooted they are in the language (Williams-van Klinken et al. 2002a:4):

- loans which everyone uses and accepts as Tetun (T. *kama* 'bed' from Port. *cama,* T. *kafé* 'coffee' from Port. *café*);
- loans used mostly by Portuguese-influenced speakers but understood by the majority of the population (T. *sín* 'yes' from Port. *sim,* T. *komu* 'because' from Port. *como*);
- loans used by educated Portuguese-Tetun bilinguals, typically in more formal and intellectual contexts, which are not necessarily widely understood (T. *sobre* 'about' from Port. *sobre* vs. T. *kona-ba*);
- loans used exclusively by educated Portuguese-Tetun bilinguals which are not understood or accepted by others (T. *alende* 'besides, apart from').

I believe that it is just a matter of time until the borrowings that are still not fully accepted by the wider population become widespread and understood by most of the Tetun Prasa speakers. To document the extent of the Portuguese influence, I created a list of Portuguese borrowings found in both Tetun Prasa and Tetun Terik oral and written corpora (see Annex 4). The list includes Portuguese borrowings in their basic form (masculine and singular) except for animate nouns which, if appeared, are listed in both their masculine and feminine form. Certain categories of words which were included in the analysis, were excluded from the list: loanblends (e.g. *loron-kinta*), Portuguese loans with native derivational affixes (e.g. *avón, haforsa*), Portuguese loans with Portuguese

373

derivational affixes (e.g. *vise-prezidente*), in which case the prefix and the noun were represented as two separate borrowings, Portuguese borrowings with alternative spelling (e.g. *asume* v. *asumi*, *servisu* v. *serbisu*), words mentioned in (375) and other words belonging to the these categories, as well as words with a seemingly Portuguese form but without a Portuguese equivalent (e.g. *insentimente*, *monolíngua*). On the other hand, homonymous loanwords belonging to two different word classes were included twice (e.g. *eskola* as a noun and *eskola* as a verb). It would be interesting to compare my list of Portuguese borrowings with the dictionaries of Aparício da Silva (1889) and Raphael das Dores (1907) to identify older and more recent borrowings. However, because the above-mentioned dictionaries only exist in a book form and/or digitalized, this task would be rather time-consuming.

8.6 Analysis

Based on previous studies on borrowability and lexical borrowings in Tetun Prasa, I expected to confirm these hypotheses:

- the number of borrowings in Tetun Prasa is not sufficient for the language to be classified as a creole;
- the number of borrowings is higher in written texts than in oral speech;
- the proportion of borrowed content words is higher than that of function words;
- the number of borrowings is higher in Tetun Prasa than in Tetun Terik;
- speakers proficient in Portuguese show a higher number of (Portuguese) borrowings;
- speakers with higher education show a higher number of (Portuguese) borrowings;
- speakers whose language of instruction at school was Portuguese show a higher number of (Portuguese) borrowings;
- speakers from the Suai region (bordering Indonesia) show a higher number of Indonesian borrowings;

- certain semantic domains like religion, politics, economy, technology are more likely to contain a high number of borrowings;
- some borrowings have undergone semantic change.

The aim of the following statistical analysis is to offer more detailed quantitative data to support the general claims about the high number of borrowings, especially in Tetun Prasa. In order to do so, I decided to conduct two separate analyses of the lexical borrowings: one based on word tokens and another one based on word types. While the analysis of word tokens takes in consideration the frequency of native and borrowed words, the analysis of word types gives us more insight into the extent of both lexicons.

All the words in my corpora were classified for number of occurrences, their presence in oral/written corpus, language of origin, word class(es), semantic class and semantic type (core or cultural borrowing). Some words were classified as originating in two different languages:

- single-word blends, e.g. *armari* 'cupboard' (from Port. *armário* and M. *almari*);
- derived words, e.g. *komitementu* 'commitment' (from Eng. *commitment* and Port. *-mento*);
- compounds, e.g. *loron-kinta* 'Thursday' (from T. *loron* 'day' and Port. *kinta* 'Thursday').

Malayo-Portuguese loans (e.g. *karreta* 'car', *maski* 'although') that are found in most of the restructured SE Asian Malay and Portuguese varieties were classified as Portuguese borrowings, since they are of Portuguese origin and we can only hypothesize which language introduced them into Tetun Prasa.

To distinguish Malay borrowings from Indonesian borrowings, I used two criteria: words that have undergone sound changes but can still be traced back to Malay or its restructured varieties (e.g. *kaben* 'marry' from M. *kawin* 'marry'), especially if they are themselves of foreign origin (e.g. *sasin* 'witness' via M. *saksi* from Sanskrit) and words that have not changed in form but refer to realities that were known to Timor before the Indonesian occupation (e.g. *kuda* 'horse' via M. *kuda* from Tamil, *budu* 'pickle' from M. *budu* 'salt fish') were classified as Malay borrowings. On the other hand, Indonesian

loanwords were borrowed in their original form and usually refer to more recent and modern phenomena, e.g. *dame* 'peace' (from Indo. *dame*) or *intel* 'spy' (from Indo. *intel*).

8.6.1 Analysis of word tokens

The total number of word tokens in my oral and written corpora combined amounted to over 70,000 tokens, out of which 62% were of native origin and 38% were borrowings:

All Words	Native		All Borrowings	
Word Tokens	No.	%	No.	%
73982	46337	63%	27645	37%

Table 54 Ratio of native and borrowed tokens in my oral and written corpora combined

The majority of these borrowings were Portuguese loanwords, followed by Indonesian, Malay and English ones:

	Number	%
All Words	73982	100%
All Borrowings	27645	37%
Native	46337	63%
Portuguese Borrowings	25807	35%
English Borrowings	151	0,2%
Indonesian Borrowings	1122	1,5%
Malay Borrowings	588	0,8%

Table 55 Ratio of foreign borrowings per language

Based on previous studies (Williams-van Klinken 2002, Williams-van Klinken & Hajek 2009b, Williams-van Klinken & Hajek 2016), I expected to find different results for oral speech and written language. As shown in Table 56, the number of borrowings in the written corpus is indeed much higher than in the oral corpus (43% vs. 25%):

	All Words	Native		All Borrowings	
	Word Tokens		%	No.	%
Oral	23222	17432	75%	5790	25%
Written	50760	28905	57%	21855	43%

Table 56 Ratio of native and borrowed tokens in the oral and in the written corpus

I also understood there might be differences between varieties of Tetun and expected to find higher number of borrowings in Tetun Prasa than Tetun Terik. In Tetun Prasa, the ratio of borrowings per informant ranges from 22% to 40%, with a mean of 29%:

Variety	Speakers	All Words	Native		All Borrowings	
		Word Tokens	No.	%	No.	%
TP	tx@JMB	286	223	78%	63	22%
TP	tx@JBP	1330	1017	76%	313	24%
TP	tx@MAS	1060	808	76%	252	24%
TP	tx@LEG	200	150	75%	50	25%
TP	tx@ACB	786	586	75%	200	25%
TP	tx@CS	1634	1215	74%	419	26%
TP	tx@FDP	204	151	74%	53	26%
TP	tx@NSG	790	573	73%	217	27%
TP	tx@BE	237	171	72%	66	28%
TP	tx@DLB	1407	959	68%	448	32%
TP	tx@JLS	1543	1009	65%	534	35%
TP	tx@LAOB	1118	703	63%	415	37%
TP	tx@TAB	346	206	60%	140	40%

Table 57 Ratio of native and borrowed word tokens in Tetun Prasa speakers

The numbers were lower in the two Tetun Terik subvarieties: Tetun from Suai and Tetun from Viqueque, although the disparity of results was much greater. In Tetun Suai, we find speakers who employed only 5% of borrowings and at the same time speakers whose speech contained nearly 40% of loanwords, with the mean value of 22%:

Variety	Speakers	All Words	Native		All Borrowings	
		Word Tokens	No.	%	No.	%
TS	tx@MA	149	141	95%	8	5%
TS	tx@CSG	447	413	92%	34	8%
TS	tx@JA	157	140	89%	17	11%
TS	tx@LLM	156	139	89%	17	11%
TS	tx@RASS	259	221	85%	38	15%
TS	tx@FJC	123	104	85%	19	15%
TS	tx@BB	170	136	80%	34	20%
TS	tx@MCA	199	159	80%	40	20%
TS	tx@MN	96	75	78%	21	22%
TS	tx@AGJ	523	394	75%	129	25%
TS	tx@NA	499	375	75%	124	25%
TS	tx@OC	434	320	74%	114	26%

TS	tx@ATB	348	255	73%	93	27%
TS	tx@OA	440	311	71%	129	29%
TS	tx@GJX	352	215	61%	137	39%

Table 58 Ratio of native and borrowed word tokens in Tetun Suai speakers

The numbers were similar in the Tetun Terik subvariety from Viqueque, although the majority of speakers showed less than 20% of borrowings, with the mean value of 21%:

		All Words	Native		All Borrowings	
Variety	Speakers	Word Tokens	No.	%	No.	%
TV	tx@AFA	406	383	94%	23	6%
TV	tx@MS	909	798	88%	111	12%
TV	tx@LP	482	412	85%	70	15%
TV	tx@JCV	2990	2552	85%	438	15%
TV	tx@JCP	519	423	82%	96	18%
TV	tx@AP	175	142	81%	33	19%
TV	tx@MVJ	2448	1553	63%	895	37%

Table 59 Ratio of native and borrowed word tokens in Tetun Viqueque speakers

As documented in Figure 35, the percentage of Portuguese borrowings is lower in the two subvarieties of Tetun Terik:
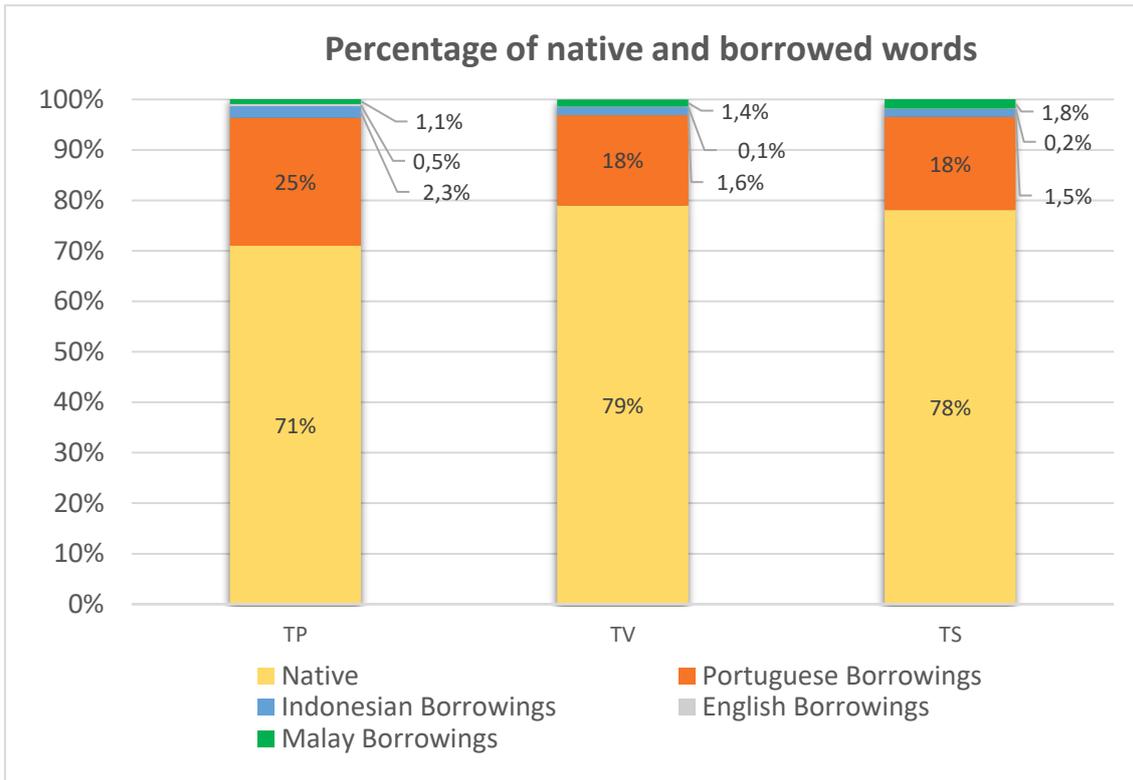
Figure 35 Percentage of native and borrowed words in Tetun varieties

However, when it comes to loanwords from other languages, the numbers are very similar. Based on Williams-van Klinken (2010), I expected to find a higher number of Indonesian loanwords in Tetun Terik from Suai, as this region borders Indonesia and Tetun Fehan presents rather elevated number of Indonesian borrowings: 12% in van Klinken's (1999) dictionary file. According to my analysis, the ratio of these loanwords is very low in Tetun Suai speakers (see Table 60) and, surprisingly, lower than in the other two varieties, as shown in Figure 35.

		Indonesian Borrowings	
Variety	Speakers	No.	%
TS	tx@LLM	1	1%
TS	tx@MA	1	1%
TS	tx@AGJ	4	1%
TS	tx@MN	1	1%
TS	tx@CSG	5	1%
TS	tx@BB	2	1%
TS	tx@NA	6	1%
TS	tx@ATB	11	3%
TS	tx@FJC	5	4%
TS	tx@OC	17	4%
TS	tx@MCA	13	7%
TS	tx@JA	0	
TS	tx@RASS	0	
TS	tx@OA	0	
TS	tx@GJX	0	

Table 60 Ratio of Indonesian borrowings in Tetun Suai speakers

8.6.2 Analysis based on sociolinguistic variables

Now let us look at sociolinguistic variables that could influence the number of borrowings used by each speaker. As I mentioned above, I have been working with three hypotheses:

- speakers who were educated during the Portuguese times and whose language of instruction in school was Portuguese show a higher number of Portuguese borrowings;
- speakers proficient in Portuguese show a higher number of Portuguese borrowings;
- speakers with higher education show a higher number of Portuguese borrowings.

According to the first set of analyses, the language of instruction in school and the period in which Tetun Prasa speakers were educated showed to have little influence on the number of Portuguese borrowings they used (see Figure 36). Age does not seem to be an important variable in this case, although it is interesting to see lower numbers of borrowings in the younger speakers.
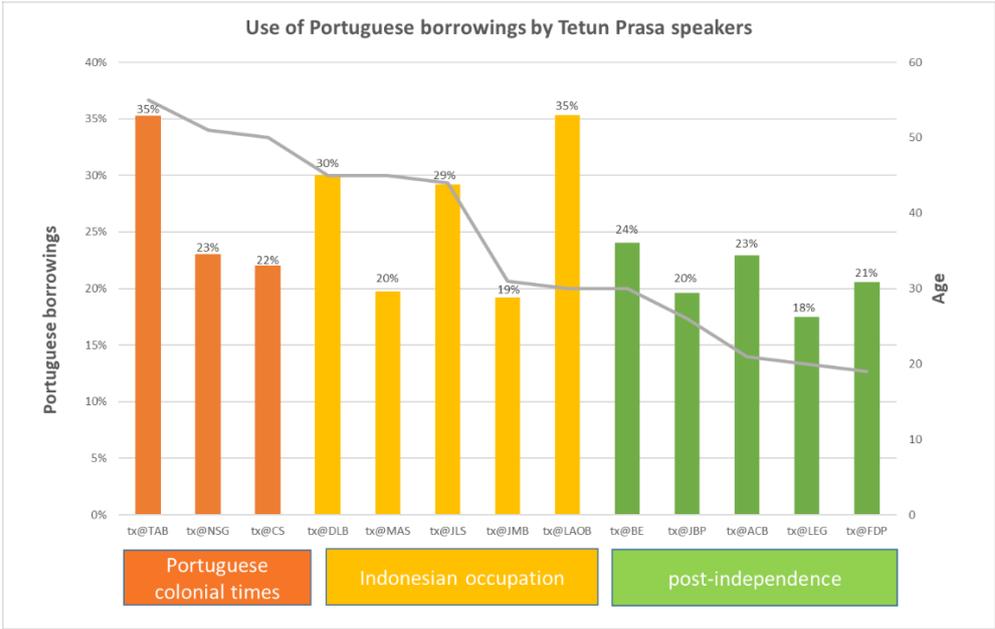
Figure 36 Ratio of Portuguese borrowings based on the period in which Tetun Prasa speakers were educated

However, what seems to play a role is the combination of three social factors: proficiency in Portuguese, level of education and sex. Female speakers (tx@JMB, tx@LEG, tx@FDP and tx@CS) with primary/secondary education and some/no understanding of Portuguese show the lowest numbers of borrowings, regardless of their age, as shown in Figure 37, Figure 38 and Figure 39.
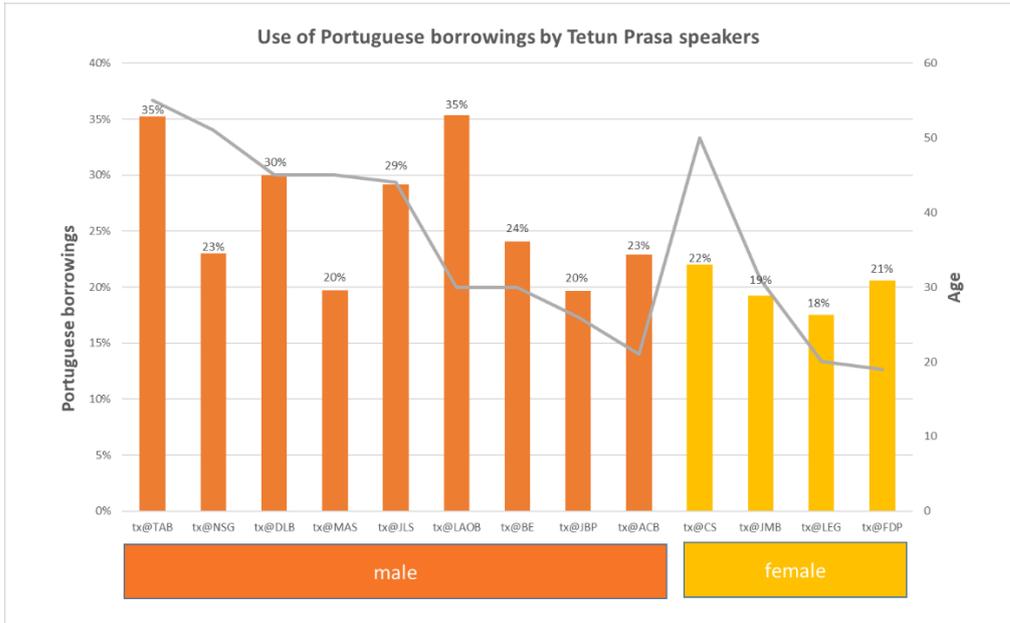
Figure 37 Ratio of Portuguese borrowings based on sex (Tetun Prasa speakers)
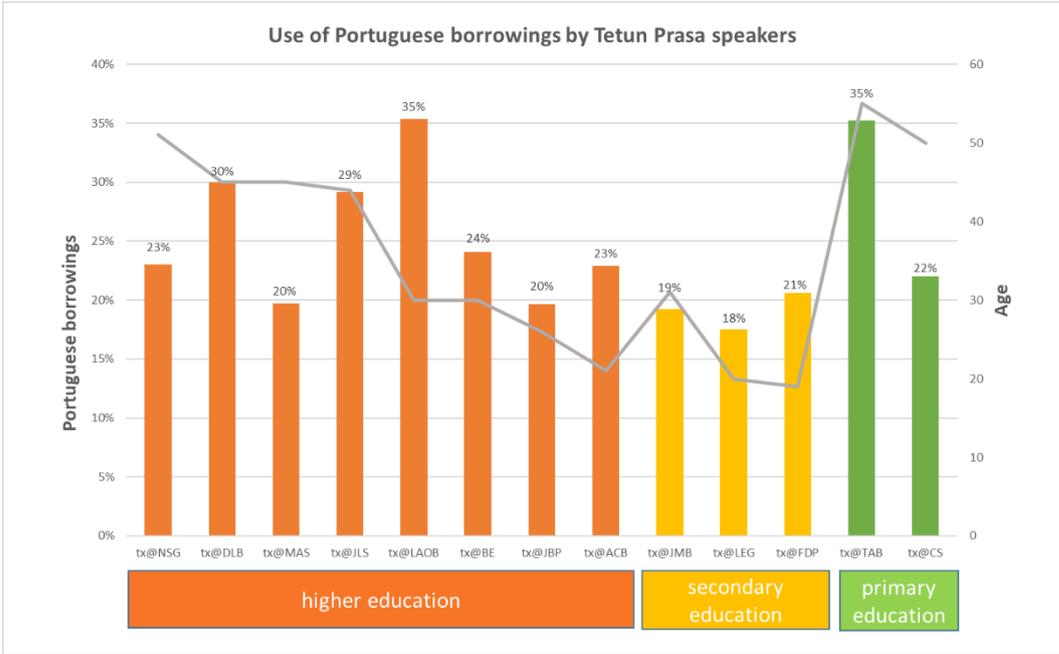
Figure 38 Ratio of Portuguese borrowings based on level of education (Tetun Prasa speakers)
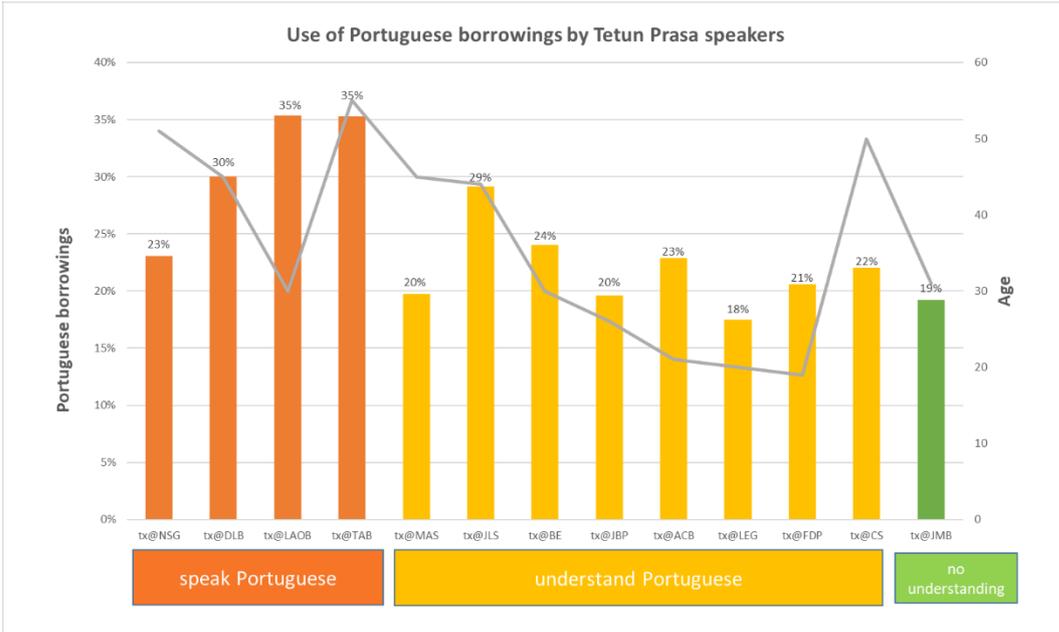
Figure 39 Ratio of Portuguese borrowings based on proficiency in Portuguese (Tetun Prasa speakers)

Tetun Terik data are much more difficult to interpret as the ratio of Portuguese borrowings ranges from 3% to 38% in case of Tetun Suai speakers and 4% to 33% in Tetun Terik speakers. For this analysis, I will treat the two subvarieties separately.

Compared with Tetun Prasa, data from the Tetun Suai-speaking area seem to show the inverse tendency. Younger speakers, regardless of age, with lower education and limited proficiency in Portuguese demonstrate more than average percentage of

382

Portuguese borrowings (see Figure 40, Figure 41, Figure 42 and Figure 43). We can only hypothesize that younger speakers have been influenced more by Tetun Prasa and that proficiency in Portuguese does not affect the number of borrowings in this variety.
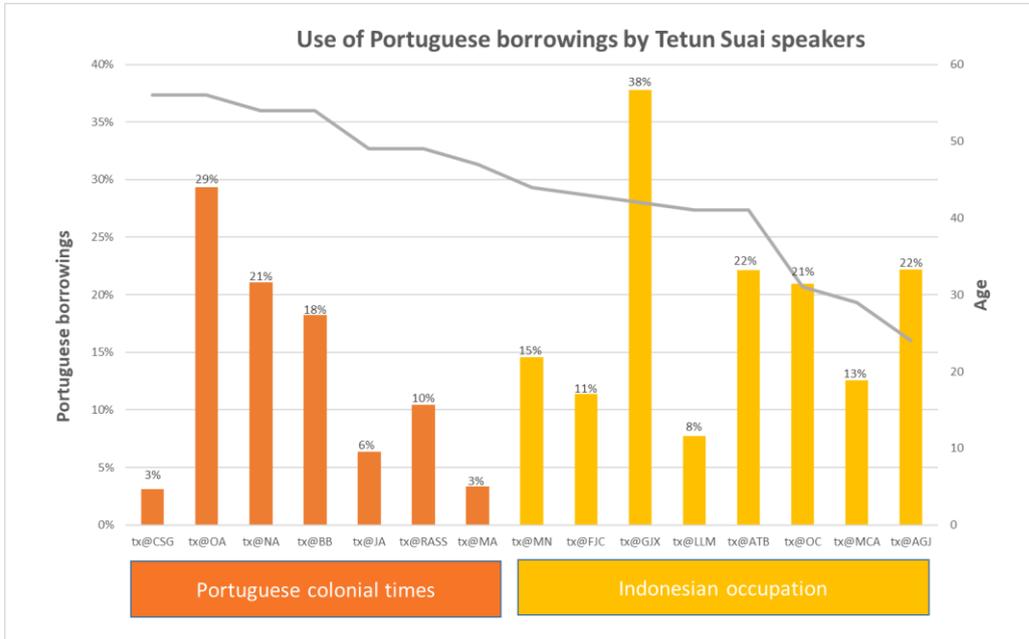
Figure 40 Ratio of Portuguese borrowings based on the period in which Tetun Suai speakers were educated
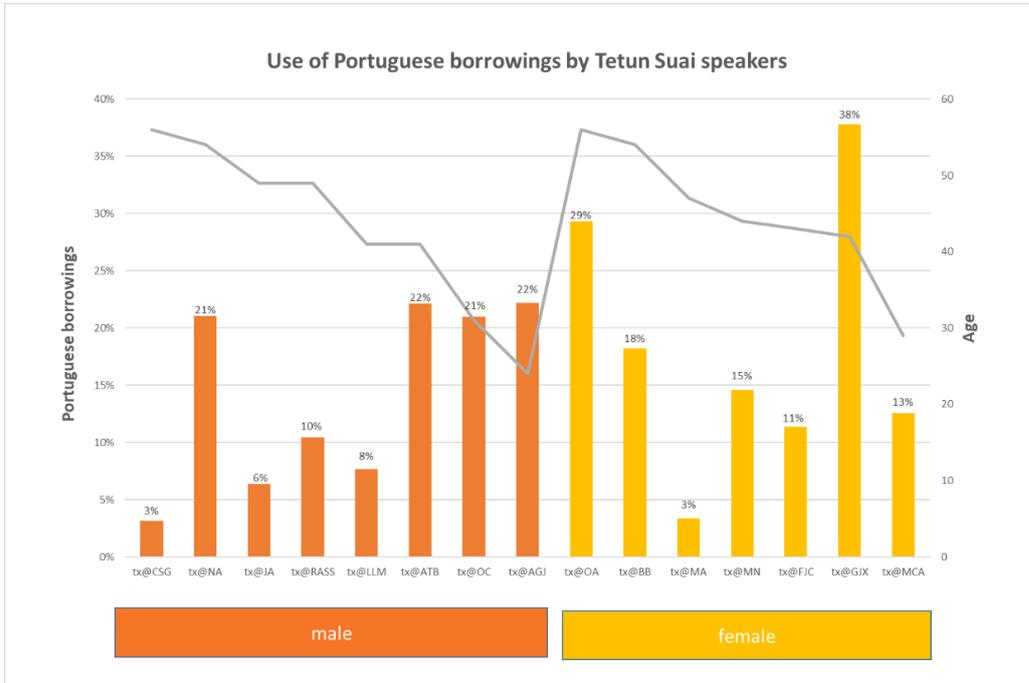
Figure 41 Ratio of Portuguese borrowings based on sex (Tetun Suai speakers)
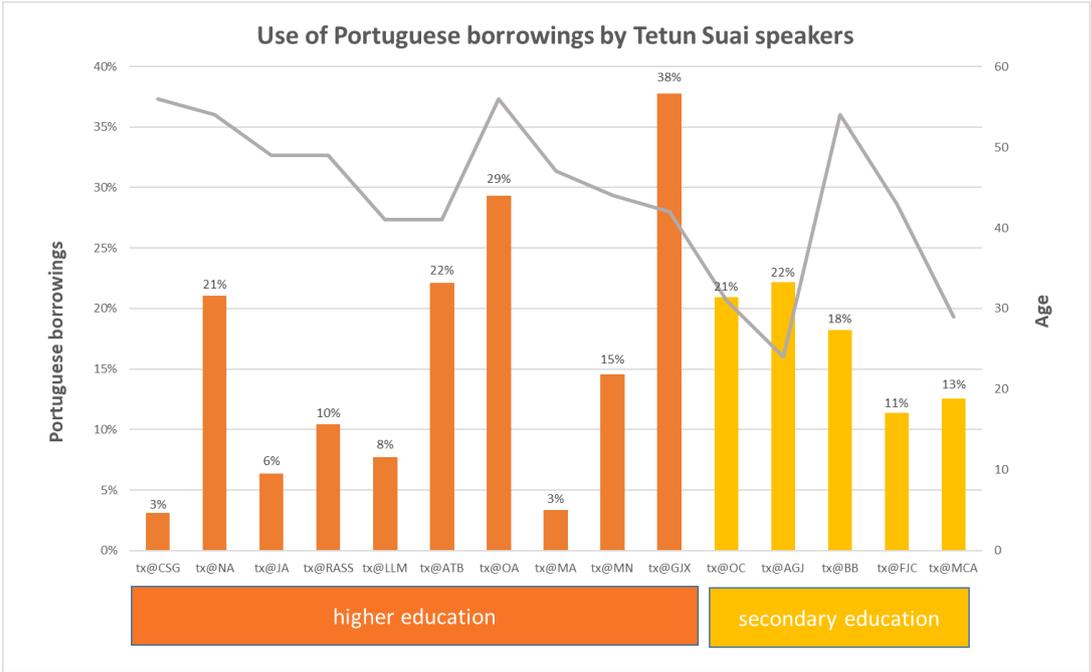
Figure 42 Ratio of Portuguese borrowings based on level of education (Tetun Suai speakers)
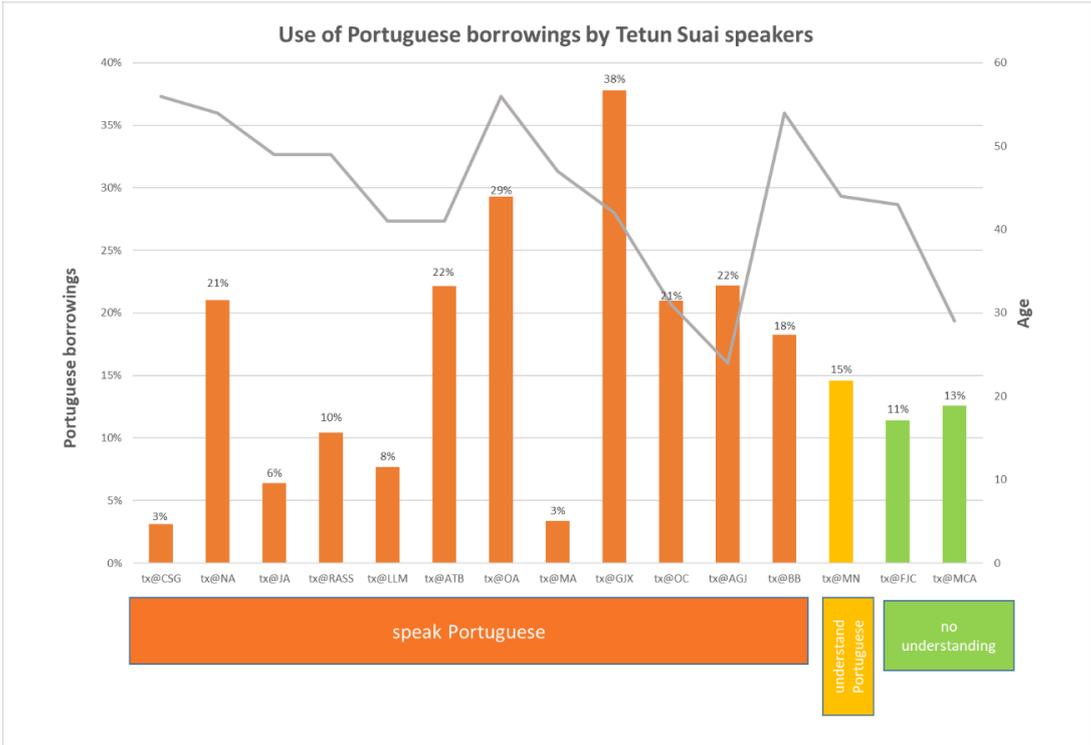
Figure 43 Ratio of Portuguese borrowings based on proficiency in Portuguese (Tetun Suai speakers)

In the case of Tetun from Viqueque, only one speaker (tx@MVJ) showed high percentage of borrowings (33%), which can be explained by the fact that he spent some time living and studying at a university in Dili and that he uses Portuguese on a daily basis in his work environment. All the other speakers show much lower percentages,

although no conclusions can be made based on the three social variables: the period in which they were educated (see Figure 44), the level of education (see Figure 45) and their proficiency in Portuguese (see Figure 46).
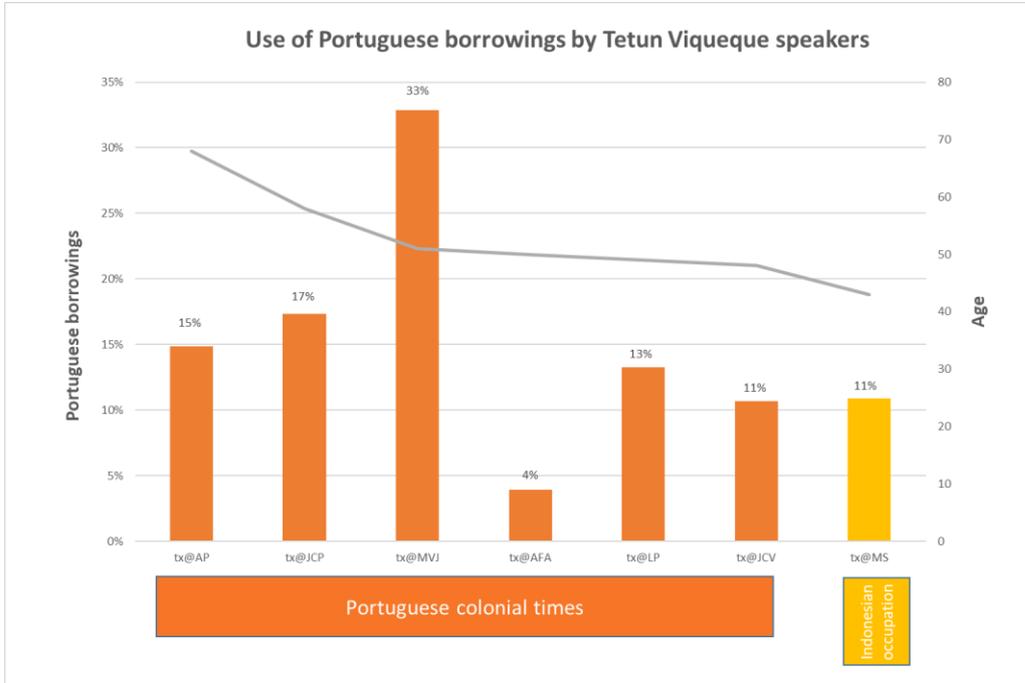
Figure 44 Ratio of Portuguese borrowings based on the period in which Tetun Viqueque speakers were educated
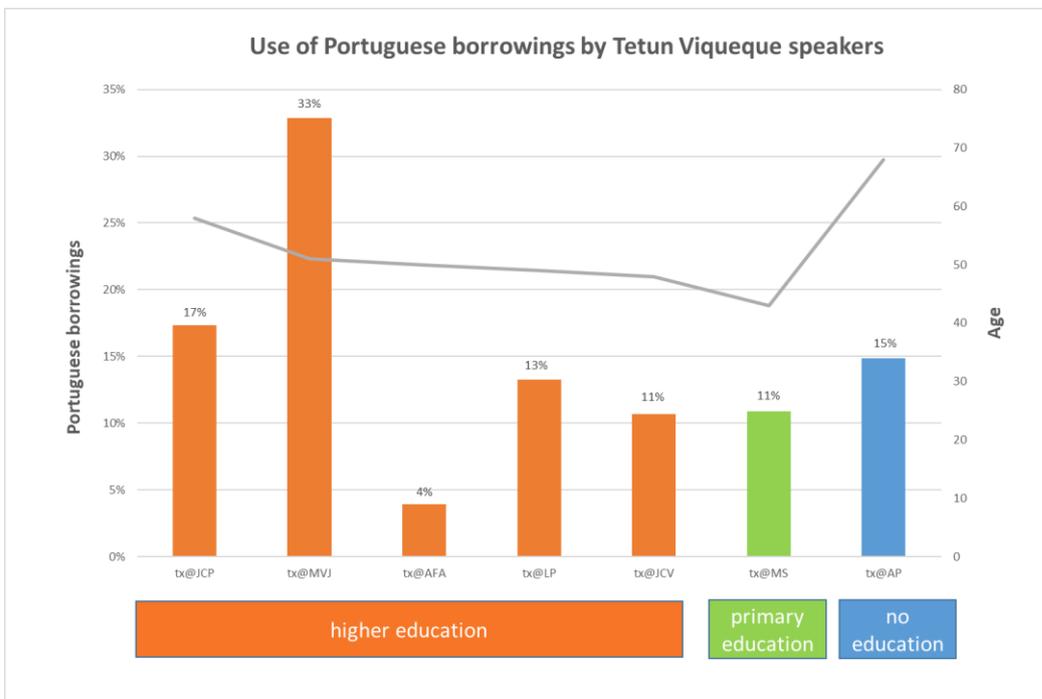
Figure 45 Ratio of Portuguese borrowings based on level of education (Tetun Viqueque speakers)
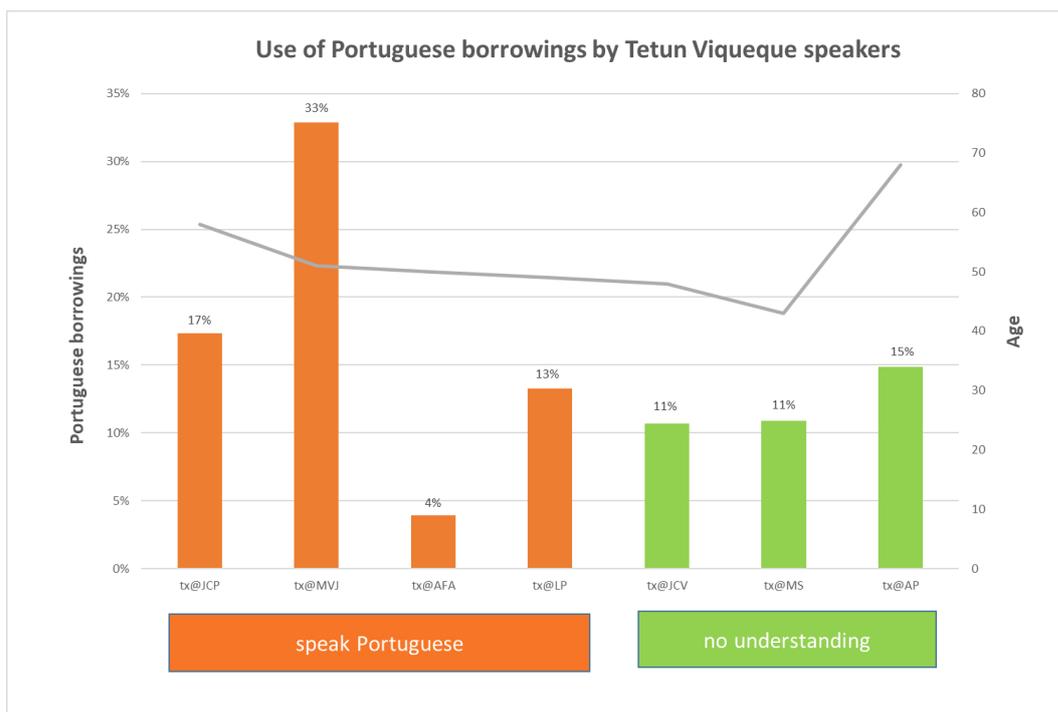
Figure 46 Ratio of Portuguese borrowings based on proficiency in Portuguese (Tetun Viqueque speakers)

The graphical interpretation of results discussed in the section above illustrate that for single variables, such as Portuguese proficiency or age, we can see correlations between the number of borrowings and a given variable. Using a single variable regression analysis[206], we can represent these graphical results in terms of the coefficient of determination denoted by R-squared which is "the proportion of the variance in the dependent variable that is predictable from the independent variable" ("Coefficient of determination", 2017). Undertaking regression analysis provides an indication of how conclusions based on the relative 'influence' of multiple independent variables (e.g. age, sex or Portuguese proficiency) affect the percentage of Portuguese borrowings used by each speaker.

Table 74 in Annex 3 shows the outputs of the borrowings analysis described in the above section. For purposes of this section of analysis, only Portuguese borrowings, which were shown to be predominant, will be further analysed.

---

[206] Regression analysis is a statistical process for estimating the relationships among variables. Regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed.

From the 5 main sociolinguistic variables and their 16 subcategories[207], which have been used for this analysis, the first step was to identify the subcategories with the most statistical significance (relative to the other subcategories). For this I have analysed Tetun and its dialects (Tetun Prasa and Tetun Terik) and subdialects (Tetun Suai and Tetun Viqueque) against the 16 subcategories. Table 61 shows the results of this analysis. All values that are above 18% (highlighted in yellow) have been identified as more statistically significant than other results. For example, in the case of male vs. female speakers of Tetun Prasa, this gives us an R-squared of 28.8%, compared to R-squared of 4.3% for Tetun Suai speakers, which means that 'sex' is a more significant variable in Tetun Prasa. Linking this to the observations above, we can see that the relationship observed in Figure 37 is supported by an R-squared value which is relatively significant compared to the other Tetun variates.

Tetun	No. Speakers	Sex		Age	Age groups		
		Male	Female	Age	1-23	23-47	47+
Tetun (All)	35	0,02%	0,02%	2,05%	0,19%	4,57%	5,76%
Tetun Prasa	13	28,82%	28,82%	19,29%	26,20%	8,48%	4,67%
Tetun Terik	22	3,48%	3,48%	0,83%	Insufficient data	3,72%	3,72%
Tetun Suai	15	4,32%	4,32%	2,30%	Insufficient data	8,56%	8,56%
Tetun Viqueque	7	Insufficient data	Insufficient data	2,94%	Insufficient data	3,75%	3,75%

Tetun	No. Speakers	Level of education				Language of instruction		
		No Education	Primary	Secondary	Higher Education	Portuguese	Indonesian	Tetun
Tetun (All)	35	0,60%	1,61%	0,45%	0,01%	2,91%	0,08%	7,73%
Tetun Prasa	13	Insufficient data	9,37%	26,05%	4,62%	0,01%	0,56%	12,45%
Tetun Terik	22	0,04%	1,32%	0,68%	0,02%	4,65%	0,23%	23,57%
Tetun Suai	15	Insufficient data	Insufficient data	0,55%	0,55%	0,03%	0,22%	14,39%
Tetun Viqueque	7	0,00%	3,75%	Insufficient data	2,23%	60,80%	0,00%	60,80%

Tetun	No. Speakers	Proficiency in Portuguese		
		Speaks	Understands	No understanding
Tetun (All)	35	0,85%	1,97%	8,08%
Tetun Prasa	13	54,63%	31,05%	6,90%
Tetun Terik	22	4,58%	0,07%	4,60%
Tetun Suai	15	2,94%	0,18%	2,91%
Tetun Viqueque	7	7,83%	Insufficient data	7,83%

Table 61 Coefficient of determination for fixed variable 'Tetun' and its varieties and subvarieties against independent variables (insufficient data entries denote where the analysis did not return a value)

There are several further observations we can make on the data. Firstly, given the low number of Tetun Viqueque speakers interviewed (7 informants), any data analysis of this variety will not yield meaningful results and is therefore not discussed further. Secondly, there is a trend where Tetun Prasa has consistently the highest R-squared for 7 out of 8 significant (over 18 %) subcatagories, which represents 87,5% of results. And even though only one subcategory achieved R-squared value higher than 50%, we can conclude that relative to the other results for different varieties of Tetun, Tetun Prasa values show significant relative correlations for sex, age, and Portuguese proficiency.

---

[207] Sex (2 subcategories), Age (1 subcategory), Age groups (3 subcategories), Level of education (4 subcategories), Language of instruction in school (3 subcategories), Proficiency in Portuguese (3 subcategories), as shown in Table 61.

Figure 47 graphically represents these three independent sociolinguistic variables and we can see a clear trend showing higher % of borrowings in male speakers (top), older speakers (middle) and full proficiency in Portuguese (bottom).

	Tetun Prasa and Sex		
		Sex	
	Tetun variety	Male/Female	
1	Tetun (All)	0,02%	
	Tetun Prasa	28,82%	
	Tetun Terik	3,48%	
	Tetun Suai	4,32%	
	Tetun Viqueque	Insufficient data	

	Tetun Prasa and Age		
		Age	
	Tetun variety	All	
2	Tetun (All)	2,05%	
	Tetun Prasa	19,29%	
	Tetun Terik	0,83%	
	Tetun Suai	2,30%	
	Tetun Viqueque	2,94%	

	Tetun Prasa and proficiency in Portuguese		
		Portuguese	
	Tetun variety	Speaks (R2 value)	
3	Tetun (All)	0,85%	
	Tetun Prasa	54,63%	
	Tetun Terik	4,58%	
	Tetun Suai	2,94%	
	Tetun Viqueque	7,83%	

Figure 47 Regression analysis with 2 independent variables: TP and sex (top), TP and age (middle) and TP and proficiency in Portuguese (bottom)

In the next stage of the analysis I wanted to see if an increase in the overall R-squared value would be observed if combining the three independent variables: Tetun Prasa, sex and age. Figure 48 shows an $R^2 = 31.2\%$ which is higher than that of the two respective '2 independent variable' analysis. We can conclude that considering a multivariational analysis shows increase in correlation with the % of Portuguese borrowings used by Tetun Prasa speakers.
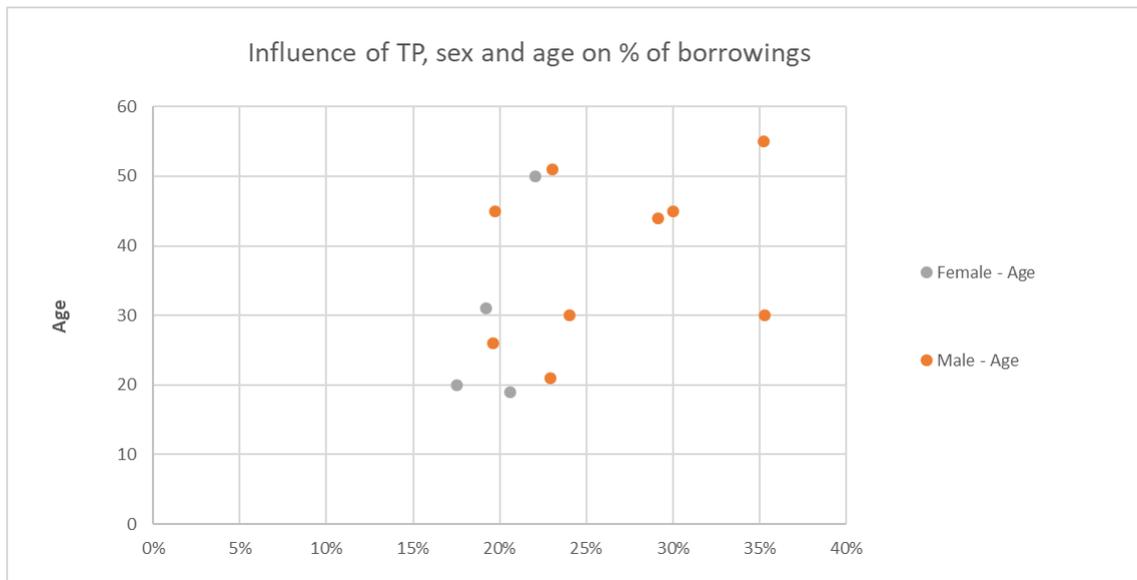
Figure 48 Regression analysis with 3 independent variables

In the final analysis I tried to find out if an increase in the overall R^2 value would be observed by combining the four independent variables: Tetun Prasa, sex, age, Portuguese proficiency. Figure 49 shows an $R^2 = 51.8\%$ which is higher than that of the '3 independent variable' analysis. Therefore, based on this multivariational analysis, we can conclude that the highest percentage of Portuguese borrowings is observed in older male speakers of Tetun Prasa who are proficient in Portuguese.
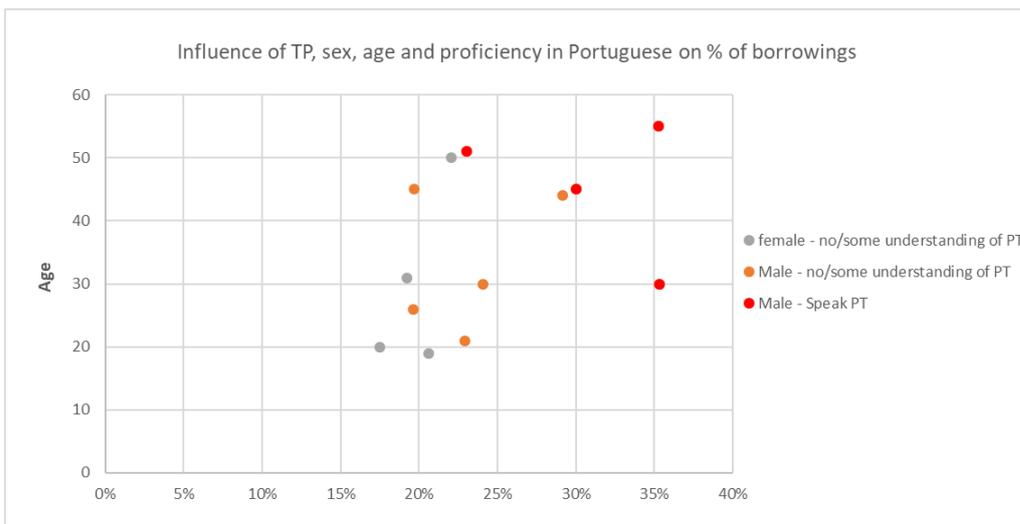
Figure 49 Regression analysis with 4 independent variables

In conclusion, the regression analysis was used to see how the percentage of borrowings (dependent variable) was related to the sociolinguistic data (independent variables like sex, age, level of education, language of instruction and proficiency in Portuguese) in relation to Tetun and its varieties and subvarieties. Due to the limited data set and the low R^2 values observed, we cannot state with significant confidence that the relations indicated may apply to a large sample size. However, there are some clear patterns which would warrant further investigation with a larger survey. For the purposes of this research, undertaking a larger survey given the work involved in data collection, processing and analysis was not practical. I also believe that other variables, such as language attitudes and a preference for certain languages need to be considered in further studies. In many cases, high proficiency in Portuguese was not conditioned by advanced age or Portuguese-based education, but rather by family tradition and positive attitude towards the language.

8.6.3 Analysis of word types

To better understand the extent of native and borrowed lexicon as well as frequency of core vocabulary, I also conducted an analysis of word types. Most of the native words are thought to belong to the core vocabulary and their frequency is higher than that of foreign loans. More than a thousand Portuguese loanwords occurred only

once in my corpus and some might argue whether they have been fully integrated into the Tetun Prasa lexicon or should be considered nonce borrowings.

This analysis yielded rather different results and a staggering number of borrowings in general:

All Words	Native		All Borrowings	
Word Types		%	No.	%
4118	1092	26,5%	3026	73,5%

Table 62 Ratio of native and borrowed word types in the oral and written corpora combined

The numbers got even more interesting when I split the total number of word types into two subcategories: oral speech and written languages:

	All Words	Native		All Borrowings	
	Word Types		%	No.	%
Oral	1724	762	44%	962	56%
Written	3021	547	18%	2474	82%

Table 63 Ratio of native and borrowed types in oral and written corpus

Most of the borrowings in the oral corpus are of Portuguese origin, but there is also a considerable number of Indonesian borrowings which are almost absent in the written language, as documented by Figure 50 and Figure 51.
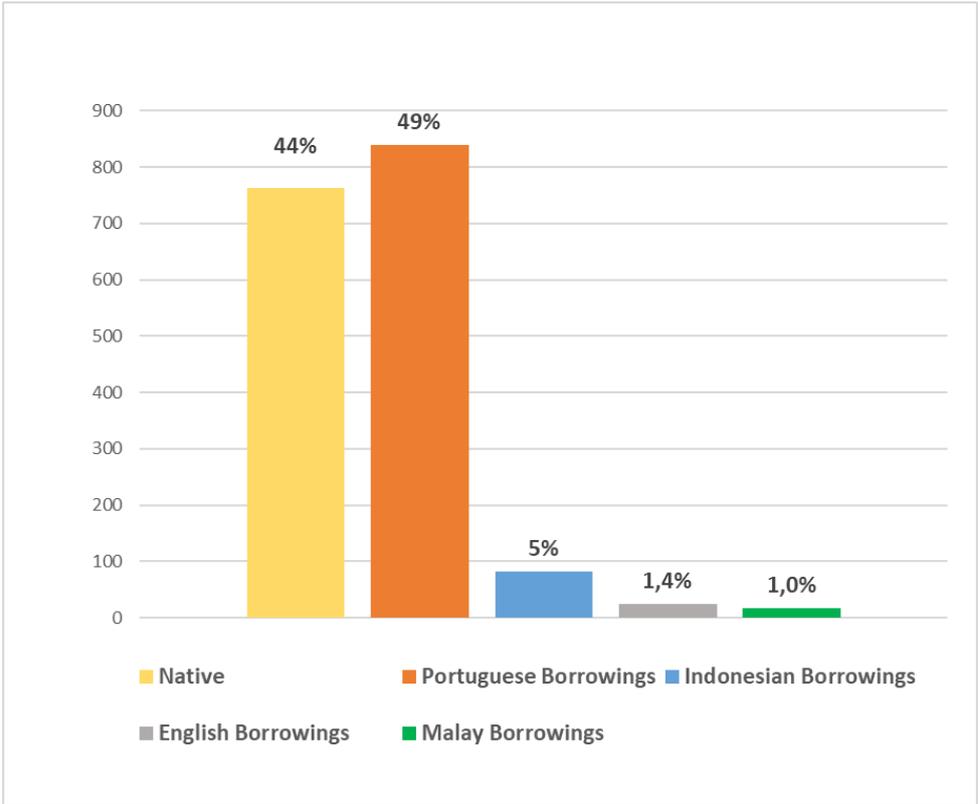
Figure 50 Ratio of borrowings in the oral corpus per language
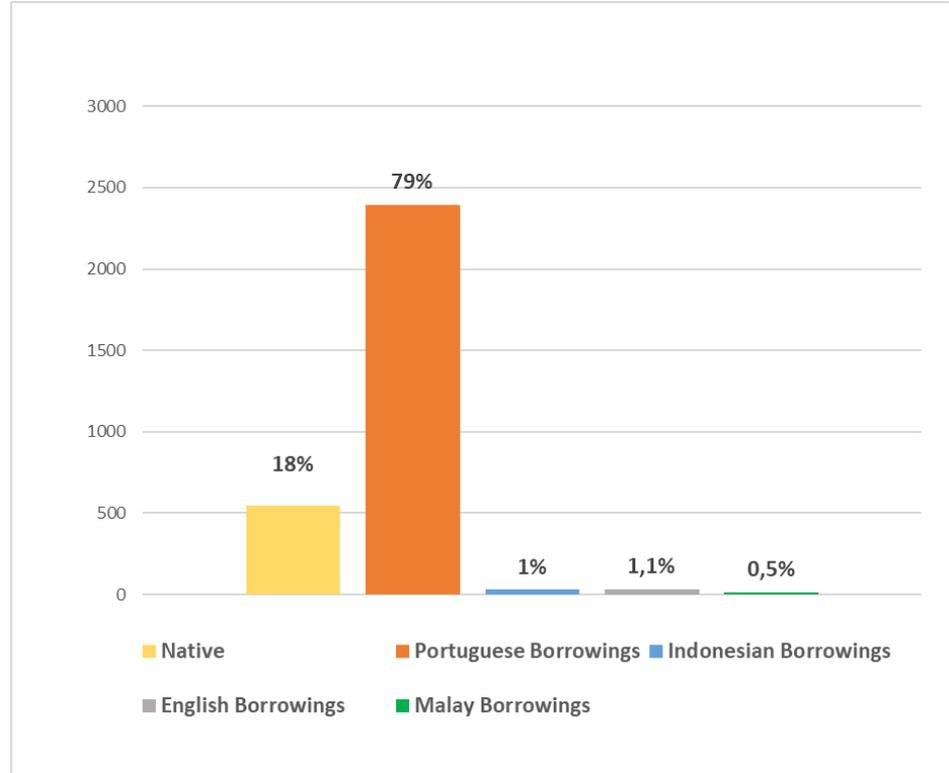
Figure 51 Ratio of borrowings in the written corpus per language

8.6.4 Analysis of word classes

Putting native words aside, in my further analysis I subdivided borrowings into two larger categories: content words (nouns, verbs, adjectives, adverbs and numerals) and function words (pronouns, conjunctions, preposition, auxiliary verbs etc.). There is a clear empirical evidence that content words are more likely to be borrowed than function words. Some of these function words are discussed in the Chapter 6, where I looked at 'lexical pairing', as described in Hajek (2006b:170), in which native and borrowed function words and grammatical forms appear together, e.g. purposive conjunction *atu hodi, para atu, atu para, para hodi* 'in order to'.

As you can see in Table 64, the number of borrowed function words is much lower than that of content words, which is nothing surprising as languages normally contain more content words than function ones. However, if we make a comparison with native words, we see that the number of borrowed content words is almost three times as high as that of native ones (2714 versus 932). Most of these words are nouns, verbs and adjectives, which is documented by Table 65.

	Borrowings	Native
Content words	2714	932
Function words	62	132

Table 64 Content and function lexical borrowings in Tetun

This supports the claims of many authors (Silva 1885, Dores 1907) that, originally, the number of abstract nouns in Tetun was limited and that Portuguese words were borrowed to fill this gap. These were then complemented by nouns, verbs and adjectives referring to modern concepts, as is further discussed in section 8.6.5.

Functional borrowings were also numerous, especially in categories like conjunctions, prepositions and interjections, and they amounted to 32% (62 out of 194 function words). The majority of them were so-called core borrowings, duplicating and coexisting with native words. It was rather unexpected to find more functional loans in my oral corpus, as shown in Table 65. Since function words are more difficult to borrow, I would assume they are introduced via written language which presents much higher number of borrowings in general but also which makes greater use of borrowed syntactic

structures. No borrowings were found in categories like TAM markers, focus markers, possessive markers or classifiers.

Borrowings	All	Oral	Written	Word type
Noun	1660	537	1407	Content words
Verb	437	158	401	Content words
Adjective	515	111	465	Content words
Numeral	42	41	10	Content words
Adverb	60	40	38	Content words
	2714	887	2321	
Conjunction	24	23	15	Function words
Preposition	16	14	11	Function words
Pronoun	4	2	3	Function words
Interjection	11	10	4	Function words
Auxiliary verb	4	4	1	Function words
TAM	0	0	0	Function words
Determiner	3	1	3	Function words
Focus marker	0	0	0	Function words
Classifier	0	0	0	Function words
Possessive marker	0	0	0	Function words
	62	54	37	

Table 65 Lexical borrowings in Tetun per word classes and per corpus

Borrowings, mainly of Portuguese origin, have enriched the three biggest open word classes: nouns, verbs, and adjectives. In Tetun, many words are pre-categorical and their word class and function can only be determined by their syntactic position (e.g. *moris* 'to live; life; alive').[208] On the other hand, Portuguese has quite clear borders when it comes to basic word classes and this distinction is usually preserved also when the word is borrowed (Hajek 2006b:169-170). However, words can also undergo conversion.

Portuguese nouns are borrowed without articles and mostly in the singular form, occasionally plural. It is also very common to borrow whole NPs with grammatical agreement that are considered collocations also in the source language:

(364) a. *primeiru ministru* 'prime minister' (from Port. *primeiro ministro*)

b. *eskola sekundária* 'secondary school' (from Port. *escola secundária*)

c. *primeira gerra mundiál* 'First World War' (from Port. *primeira guerra mundial*)

---

[208] To classify the native words, I consulted my oral and written corpora to check for the syntactic contexts.

Portuguese verbs are borrowed in 3S present indicative tense, as in (365), with a few exceptions, as illustrated by (366). They are not conjugated for person, number or tense. Occasionally, verbs are derived from nouns via zero derivation, as shown (367), also discussed in section 5.3.8.

(365)   a. *kontinua* 'continue' (from Port. *continuar* 'continue.INF' vs. *continua* 'continue. 3S.PRS.IND'')

b. *tenke* 'must' (from Port. *ter que* 'must.INF' vs. *tem que* 'must.3S.PRS.IND')

(366)   a. *fiar* 'trust, believe' (from Port. *fiar* 'trust.INF')

b. *pasiar* 'stroll' (from Port. *passear* 'stroll.INF')

(367)   a. *servisu* v. 'to work' from T n. *servisu* 'work' (from Port. n. *serviço* 'work')

b. *eskola* v. 'go to school' from T n. *eskola* 'school' (from Port. n. *escola* 'school')

One of the word classes that was especially enriched by borrowings was that of adjectives. Van Klinken (1999:52-53) classified native Tetun adjectives as verbs because they share various features with intransitive verbs. Although adjectives are often found as modifiers in NPs, they are even more often found used predicatively. In my corpus, I have collected more than 500 adjectives, a considerable part of which corresponded to Portuguese past participles. I decided to group them together with other adjectives, since they cannot function as verbs on their own, they do not combine with TAM markers, and Tetun Prasa does not productively derive these participles from verbs. They made their way into Tetun Prasa via the written language, which often blindly translates newspaper articles from Portuguese into Tetun Prasa. In the process of translation, the copula is lost and the past participle occupies the slot of a predicative adjective, although often the syntax does not undergo any change:

(368)   a. Tetun Prasa:

[...]	*preokupasaun*	*no*	*dezafiu*	*sira*	*ne'ebé*	***partilladu***
	concern	and	challenge	PL	REL	shared

<table>
<tr><td></td><td><i>hosi estadu</i></td><td><i>frajil</i></td><td><i>sira</i></td><td>[...]</td></tr>
<tr><td></td><td>by state</td><td>fragile</td><td>PL</td><td></td></tr>
</table>

'[...] concerns and challenges (that are) shared by fragile states [...]'
(GovernuTL_2010)

b. Portuguese:

<table>
<tr><td>[...] <i>preocupações</i></td><td><i>e</i></td><td><i>desafios</i></td><td><i>que</i></td><td>são</td></tr>
<tr><td>concern.PL</td><td>and</td><td>challenge.PL</td><td>REL</td><td>COP</td></tr>
</table>

<table>
<tr><td>partilhados</td><td><i>por</i></td><td><i>estados</i></td><td><i>frágeis</i></td><td>[...]</td></tr>
<tr><td>share.PST.PTCP</td><td>by</td><td>state.PL</td><td>fragile.PL</td><td></td></tr>
</table>

'[...] concerns and challenges that are shared by fragile states [...]'

I looked at these three word classes to see which were more likely to be borrowed. According to Tadmor (2009:61), nouns are, in general more likely to be borrowed than verbs (2:1). Originally, it was thought that this was because of the high complexity of verbs, but we get similar results in languages that are highly isolating and their verbs do not undergo any morphosyntactic modification (Tadmor 2009:61-63).[209]

Tetun Prasa is a highly isolating language and borrowed verbs do not need to be specially adapted or conjugated to fit into the morphosyntactic structure of the language, which is probably why the ratio of borrowed nouns is similar to that of borrowed verbs: 58 % of borrowed nouns (1660 out of 1995 nouns) versus 57% of borrowed verbs (437 out of 768 verbs), as shown in Table 66.

	All	Borrowings	%	Native	%
Nouns	1995	1660	58%	335	42%
Verbs	768	437	57%	331	43%
Adjectives	636	515	81%	121	19%

Table 66 Ratio of borrowed and native nouns, verbs and adjectives

It might also come as a surprise that the number of borrowed adjectives (81%) is higher than that of verbs (57%) (see Table 66). This could be caused by the fact that, considering the composition of the native lexicon, nouns and verbs are the most numerous

---

[209] However, the author also added that the borrowability of nouns and verbs did not depend only on grammatical factors but also on social ones: nouns are more likely to be borrowed because the things and concepts they denote can be easily adopted by different cultures (Tadmor 2009:63).

word classes, and that many of these adjectives were derived from past participles that were part of a VP in the source language but came to be used as predicative adjectives in Tetun Prasa[210].

Adverbs are also easily borrowed, especially those of manner and time. Adverbs of manner are easily recognizable by the Portuguese derivational suffix *-mente*:

(369)   a. *ilegalmente* 'illegally' (from Port. *illegalmente*)

b. *diretamente* 'directly' (from Port. *diretamente*)

c. *formalmente* 'formally' (from Port. *formalmente*)

Temporal adverbs (e.g. *depois* 'later' from Port. *depois)* are often used in temporal adverbial phrases as conjunctions, e.g. *depois* 'after' from Port. *depois de* (see section 6.3.4.1).

8.6.5   Analysis of semantic fields

Understanding which word classes are more prevalent bring us to the next questions: what semantic fields are most affected? According to Tadmor (2009:64-65), some semantic fields are more likely to be affected by borrowing than others, and they tend to be the same across all languages. These are especially religion and belief, clothing and grooming, and the house. Technical vocabulary was not included in the World Loanword Project but we can be sure that it would yield a high number of borrowings as well, as technical terms also scored high in my analysis of semantic fields (see Table 67). Tadmor (2009) further argued that the categories with the lowest ratio of borrowings were: sense perception, spatial relations, the body and kinship terms, since these are universal concepts shared by most human societies. I used some of the semantic fields used in the World Loanword Project but added also some other ones to better reflect my oral and written corpora, like politics, economy and technology. According to Albuquerque (2011:75), new Portuguese borrowings cover mainly these semantic fields:

---

[210] In Portuguese, past participles can either be part of a VP (COP + PPP) or act as adjectives (Mateus et al. 2003:374).

technical and scientific terminology, and juridical and administrative terminology. Apart from semantic fields, I also divided loanwords into two bigger categories: cultural borrowings and core borrowings. As Figure 52 illustrates, the majority of loanwords (86%) belong to cultural borrowings and represent concepts that were foreign to the receiving culture. At the same time, a considerable number of core borrowings (14%) were also borrowed into Tetun Prasa, either replacing or co-existing with native words.

Semantic domain	All
Politics and law	437
Economics	169
Army and police	153
Technology	134
Toponymy	124
Speech and language	90
Education	85
Religion and belief	75
Time and date	71
Emotions and values	57
Numbers	46
Agriculture and vegetatio	40
Body	29
Building	27
Kinship	25
Cognition	24
Spatial relations	16
Household	13
Food and drink	12
Clothing and grooming	10
Motion	7
Animals	5
Nature	0
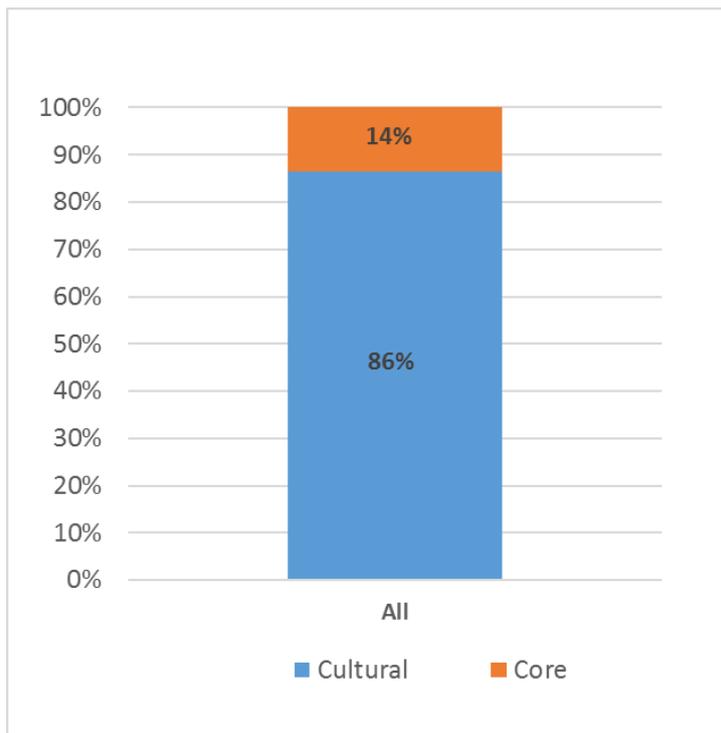
Table 67 Analysis of semantic domains based on word types

Figure 52 Borrowings belonging to core and cultural vocabulary

Interestingly, clothing and grooming and household scored quite low, which was probably caused by the topics covered. On the other hand, domains which are supposedly less likely to welcome borrowings – such as body and kinship terms – contained around 25 different loanwords, including core vocabulary:

(370)   a. *grávida* 'pregnant' (from Port. *grávida*) ↔ T. *isin-rua,* lit. 'body-two'

   b. *korasaun* 'heart' (from Port. *coração*) ↔ T. *fuan*

   c. *pai* 'father' (from Port. *pai*) ↔ T. *aman*

   d. *fillu* 'son' (from Port. *filho*) ↔ T. *oan-mane*, lit. 'child-male'

It is not surprising that the most numerous semantic fields in oral and written corpora differ substantially, as shown in Table 68 and Table 69. While the written corpus contains exclusively newspaper articles dealing with national and international politics, economy, technology and current issues, the oral corpus consists of recordings with several underlying topics, like family background, everyday life, history, current issues and traditional stories, as detailed in section 3.2.

Semantic domain	Oral		Semantic domain	Written
Politics and law	102		Politics and law	399
Education	55		Economics	161
Speech and language	47		Army and police	134
Numbers	43		Technology	120
Religion and belief	40		Toponymy	102
Time and date	32		Speech and language	62
Technology	29		Education	57
Economics	28		Time and date	56
Toponymy	26		Emotions and values	51
Army and police	25		Religion and belief	50
Kinship	21		Agriculture and vegetation	35
Emotions and values	18		Body	28
Building	10		Cognition	22
Agriculture and vegetation	9		Building	20
Cognition	8		Spatial relations	15
Food and drink	5		Numbers	13
Motion	5		Household	13
Spatial relations	4		Food and drink	10
Clothing and grooming	4		Kinship	7
Animals	4		Motion	6
Body	3		Clothing and grooming	6
Household	3		Animals	4
Nature	0		Nature	0

Table 68 Semantic domains in the oral corpus   *Table 69 Semantic domains in the written corpus*

However, it is interesting to see that more core borrowings are to be found in the oral speech than written language (see Table 70). As discussed in 8.6.7, the press register makes bigger use of borrowings in general and journalists often import foreign loans that have a native counterpart in Tetun Prasa. The fact that many times, speakers opt for a borrowing instead of a native word can be explained by the fact that certain borrowings have already been fully established and are not viewed as foreign elements by speakers.

Further research is needed to understand whether their choices are intentional and whether they are motivated by some social and sociolinguistic factors.
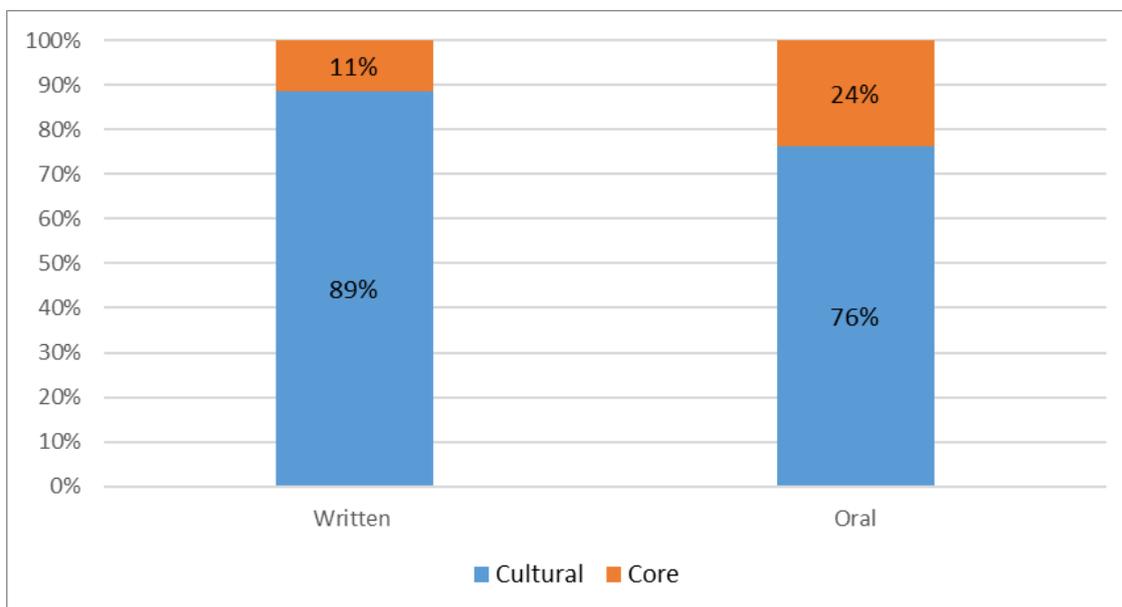
Table 70 Core versus cultural borrowings in the written and oral corpora

8.6.6 Semantic change

With respect to semantics, another interesting issue to address is semantic change, the process through which words acquire new meaning, for example, the original, more general meaning becomes more specific – in this case, we can talk about semantic narrowing:

(371)   a. *komarka* 'prison' (from Port. *comarca* 'judicial territory')

b. *reza* 'pray to God or the Saints' (from Port. *rezar* 'pray')[211]

Many Portuguese words were borrowed back in the colonial times and although they are not widely used in Portugal anymore or are considered uncommon/old-fashioned, they have preserved their original meaning in Tetun Prasa:

(372)   a. *remata* 'finish' (from Port. *rematar* 'finish', now more frequent *acabar, finalizar* 'finish' in European Portuguese)

---

[211] As opposed to *hahoran* 'pray for somebody' and *hamulak* 'pray to spirits' in animist religion.

b. *dotór* 'doctor' (from Port. *doutor* 'doctor', now mainly refers to anyone with a higher education degree)

c. *matabixu* 'breakfast' (from Port. *mata-bicho* 'breakfast')[212]

In other cases, it was not only the meaning that changed, but also the word class:

(373) a. *manha* 'spoiled' (from Port. *fazer manha* 'act spoiled' while *manha* on its own means 'malice')

b. *baixa* 'hospitalized' (from Port. *ter baixa* 'to be on sick leave')

c. *custa* 'difficult' (from Port. *custar* 'be difficult')

d. *basta* 'as long as' (from Port. *basta* 'enough')[213]

Then there are other words that seem to be of Portuguese origin, but their meaning has not been derived from an existing meaning in the source language:

(374) *perigu* 'danger' but also 'amazing' (from Port. *perigo* 'danger')

A special category has been described by Godinho (2013:190-192) as loanwords borrowed from one source language but having the meaning and/or form and derivational morphology borrowed from a different language:

(375) a. words borrowed from Portuguese but whose meaning was acquired from Indonesian (Godinho 2013:190):

- *sosializasaun* 'promotion, dissemination of ideas' (from Port. *socialização* 'socializing' and Indo. *sosializasi* 'promotion')

b. Indonesian borrowings of English origin with Portuguese pronunciation and/or morphology (Godinho 2013:191):
- *evaluasaun* 'evaluation' (from Indo. *evaluasi* and Eng. *evaluation* and Port. *-ção*)
- *komitmentu* 'commitment' (from Indo. *komitmen* and Eng. *commitment* and Port. *-mento*) (GovernuTL_2010)

---

[212] Now mostly used in the former Portuguese colonies (Brazil, Mozambique, Angola) and in rural European Portuguese, as opposed to more common *pequeno-almoço* (Portugal) and *café da manhã* (Brazil).
[213] Probably via contexts like Port. *Tu consegues, basta quereres.* 'You can do it, as long as you want it.' (Hugo Cardoso, p.c.)

c. English borrowings with Portuguese pronunciation and/or morphology (Godinho 2013:192):

- *environmentu* 'environment' (from Eng. *environment* and Port. -*mento*)
- *involvimento* 'involvement' (from Eng. *involvement* and Port. -*mento* or *envolvimento*) (STL_March_2016)

Godinho (2013) does not mention the number of occurrences for each of these words so we do not know if they are stable loans or just nonce borrowings. As for the examples from my corpus, all of them occurred just once.

8.6.7 Analysis of the press register

As I have discussed in Chapter 1, apart from different varieties, there are also different registers of Tetun: ritual, church and press register. I did not collect data on ritual Tetun Terik as I believe that as a highly traditional register spoken by a very restricted number of people and used only in formal ritual contexts (Williams-van Klinken 2002:2), there has been very little contact with other languages. The ritual language often uses special vocabulary not found in everyday speech and many people find it difficult to understand (Williams-van Klinken 2002:2). As for Church Tetun, I have discussed the influence of Portuguese missionaries on the lexicon in section 8.4.

However, what I was really interested in was the press register. There have been many studies written about the massive amount of lexical borrowings in media language and how this new vocabulary, together with new morphosyntactic structures, is being spread across other language domains. It is not unusual for language innovations to first occur in written or media language. In Timor-Leste, this is possibly because newspapers still use, alongside Tetun Prasa, also Portuguese, Indonesian and English. Articles dealing with domestic issues are written in Tetun Prasa, many others are just simply translated from Portuguese or Indonesian, importing new words (or full NPs with grammatical agreement) and structures (like Portuguese plurals). According to Williams-van Klinken

(2002:5), the staggering number of Portuguese loans is caused by four factors: 1. all journalists were educated either in Portuguese and/or in Indonesian and are used to technical vocabulary and abstract nouns, which are usually loanwords; 2. by using Portuguese loans they can refrain from using Indonesian borrowings which are to be avoided in formal contexts; 3. Portuguese has a more positive association than Indonesian; and 4. there is a serious shortage of qualified translators, so the translations into Tetun Prasa tend to be very literal.

Just by reading the headlines in Timorese journals, one can notice the astonishing number of Portuguese loanwords, as highlighted below:

(376) a. **Rezignasaun** *Xanana*   *Fó* **Impaktu**   *ba*
          resignation   Xanana   give impact   to

          Ezekusaun **Orsamentu** **Jerál**   **Estadu**   *2014*
          execution   budget   general   state   2014

          'Resignation of Xanana will impact the execution of the General State Budget 2014'

       b. *Cabral*   **Informa**   **Situasaun**   *Giné Bisau*
          Cabral   inform   situation   Guinea Bissau

          ba   **Prezidente**   **Repúblika**
          to   president   republic

          'Cabral informs the President of the Republic about the situation in Guinea-Bissau'

       c. **Esbosu**   **Lei**   **Imprensa**   **Reprezenta**   **Aspirasaun**
          draft   law   press   represent   aspiration

          Komunidade **Média**
          community   media

          'Draft of the Press Laws represents an aspiration of the media community'

       d. **Komandu**   **Jerál**   **PNTL** **Lansa** **Planu** **Estratéjiku**
          command   general   PNTL   launch   plan   strategic

PNTL
PNTL

'PNTL General Command launches a strategic plan for PNTL'
(examples from Suara Timor Lorosae, February 11, 2014, 1-3)

According to Williams-van Klinken & Hajek (2009b), 30-40% of word tokens in the press register are Portuguese loans compared to only 10 to 20% of Portuguese loanwords in informal urban speech, which corresponds to my analysis of word tokens in the written corpus:

	All Words	Native		Portuguese Borrowings	
	Word Tokens		%	No.	%
Written	50760	28905	57%	20814	41%

Table 71 Ratio of native and Portuguese word tokens in the written corpus

These figures are much higher for open word classes. Many of these borrowings are not understood by the general public, especially by people who do not speak Portuguese. The authors looked at the verbs of speaking used in the press register and divided them into three categories (Williams-van Klinken & Hajek 2009b):

- widespread in the community: *dehan* 'say', *fó sai* 'reveal', *hateten* 'say', *hato'o* 'pass on', *hatutan* 'add', *husu* 'ask', but also *esplika* 'explain' (from Port. *explicar*), *informa* 'inform' (from Port. *informar*), *deklara* 'declare' (from Port. *declarar*);

- Tetun Terik-based media terms: *haktuir* 'tell, recount', *katak* 'say', *tenik* 'say';

- Portuguese-based media terms: *afirma* 'affirm, state' (from Port. *afirmar*), *akrexenta* 'add' (from Port. *acrescentar*), *alega* 'allege' (from Port. *alegar*), *apela* 'appeal' (from Port. *apelar*), *esklarese* 'explain' (from Port. *esclarecer*), *komenta* 'make a comment' (from Port. *comentar*), *lamenta* 'lament' (from Port. *lamentar*), *mensiona* 'mention' (from Port. *mencionar*), *reforsa* 'stress' (from Port. *reforçar*), *salienta* 'stress, point out' (from Port. *salientar*), *subliña* 'underline, stress' (from Port. *sublinhar*), *sujere* 'suggest' (from Port. *sugerir*), *kestiona* 'question' (from Port. *questionar*), *anunsia* 'announce' (from Port. *anunciar*).

As we can see, there are native Tetun verbs of speaking complemented with Tetun Terik and Portuguese words. It is questionable whether these Portuguese loanwords were

borrowed out of necessity, due to the lack of quality translation, or whether they present certain sociolinguistic values (e.g. higher prestige) that make them preferable over native words. According to Williams-van Klinken (2002), even university-educated speakers have problems understanding many of these words and reading newspapers in general, let alone rural population with low literacy rates and Tetun Prasa as their L2.

Williams-van Klinken and Hajek (2016), but also many Timorese, refer to the Tetun overflowing with Portuguese loans as Tetungés (see Table 72). The authors had this text read by a group of people and the majority could not understand the Tetun translation, including university graduates.

Enkontru Komisaun Dezenvolvimentu Distritu [3]	The second District Development Commission (KDD) meeting
(KDD) ba dala rua (II) mak enkontru ida ne'ebé	
organiza hosi **Koordenador KDD**	was chaired by KDD coordinator
ho **partisipasaun membru KDD**	with participation from KDD members
hodi **diskuti** no **analiza**	to discuss and analyze
Relatóriu no **Sumáriu Proposta Projetu**	report and summary of project proposal
ne'ebé mak **prepara** hosi **Sekretariadu KDD**.	which was prepared by KDD secretariat
--	and to make a decision.
Objetivu final husi **enkontru KDD** mak	The final objective of KDD meeting is
aprova sumáriu proposta projetu	approval of a multiproject proposal summary
hodi **prodúz ezbosu Planu Investimentu Distritál (PID)**.	for producing a draft of District Investment Plan (PID)
--	for the following year.

Table 72 Example of so-called Tetungés (Williams-van Klinken & Hajek 2016:32)[214]

On the other hand, the ratio of Indonesian borrowings in the written context is very low. According to Williams-van Klinken and Hajek (2009b), avoidance of Indonesian borrowings is typical for all formal registers and the low numbers just prove it:

	All Words	English Borrowings		Indonesian Borrowings		Malay Borrowings	
	Word Tokens	No.	%	No.	%	No.	%
Written	50760	87	0,2%	678	1,3%	278	0,5%

Table 73 Ratio of English, Indonesian and Malay word tokens in the written corpus

One last issue that needs to be addressed is the social attitude towards borrowings. According to Haspelmath (2009:43), "[i]nnovating speakers may face criticism by older speakers for using a loanword, and this contributes to the general awareness of the degree to which a word is an accepted and established part of the

---

[214] Words in bold are Portuguese borrowings.

language." Although I did not specifically ask my speakers about their attitudes towards borrowings, I got the impression that many speakers do not make a conscious decision to use Portuguese and other loans in lieu of native words. Tetun Prasa has been in an intensive contact with Portuguese since the 18th century and many of the words have become accepted as Tetun words. In fact, we can find many lexical doubles or triplets coexisting together, e.g. three words with the meaning of 'bridge': T. *lalete* 'bridge'*, ponte* 'bridge' (from Pt. *ponte*), and *jambatan* 'bridge' (from M. *jambatan*) (Hajek 2006b:169).

8.7 Discourse markers

Discourse markers are "words that indicate the discourse role (e.g. topic) of the associated noun phrase" (Schachter & Shopen 2007:35). They operate at the level of discourse and are relatively syntax-independent. In Tetun Prasa, I have found a couple of examples of discourse markers borrowed from Portuguese:

(377)   a. *porezemplu* 'for example' (from Port. *por exemplo*)

b. *kerdizér* 'I mean' (from Port. *quer dizer*)

c. *pois* 'indeed' (from Port. *pois*)

d. *entaun* 'so' (from Port. *então*)

These discourse markers were found almost exclusively in oral speech and in Tetun Prasa speakers with higher education and proficiency in Portuguese, except from *entaun* 'so', which is already widespread among speakers of different varieties and backgrounds:

(378)   *Tempu barak liu   ko'alia inglés,   **pois**.*
        Time   many very   speak English   indeed

        'Indeed, a lot of time they speak English.'
        (0055TD_MAS_45_M_Uni_Mid)

(379)	**_Kerdizér,_**	_lian_	_materna_	_seluk._
	I.mean	language	mother	other

'I mean, other mother tongues.' (0051TD_DLB_45_M_Uni_Kem)

8.8 Conclusions

The lexicon is undeniably the component of Tetun Prasa most affected by language contact and the results of my analyses were hardly surprising. On the contrary, the statistical analyses only supported the general claims about the high number of loanwords in Tetun Prasa and offered more detailed quantitative analyses of different Tetun varieties and registers. The borrowings, especially from Portuguese, made it into the written as well as the oral language, often replacing or complementing their native counterparts. The numbers are especially high in the press register, which makes use of literal translation from foreign languages, and the language of the media has been recently referred to as 'Tetungés'. As documented by my data from three different Tetun-speaking regions, Tetun Prasa contains the highest number of borrowings (29% of word tokens), compared with more conservative varieties of Tetun Suai (22%) and Tetun Viqueque (21%). It appears that male speakers of Tetun Prasa who are proficient in Portuguese and have acquired higher education tend to use more borrowings in their speech than female speakers of the same variety with limited education and proficiency in Portuguese. Although the number of borrowings in Tetun Terik is lower than in Tetun Prasa, we can see that speakers of this variety are also very familiar with many, especially older Portuguese borrowings. However, to draw any general conclusions about the factors influencing the use of loanwords in this variety, more data are needed to explain the wide range of percentages found in my corpus. One hypothesis that turned out not to be true was the amount of Indonesian borrowings found in Tetun spoken in Suai. Although the region borders Indonesia, Tetun Suai speakers do not show high numbers of these borrowings.

Tetun Prasa has borrowed a considerable amount of modern vocabulary, expanding semantic fields like technology, politics and law, economics, and army and

police. It has also acquired numerous function words whose use has also affected the syntax and morphology of Tetun Prasa.

9. CONCLUSIONS

The aim of this thesis has been to investigate the role of language contact in the continuing development of Tetun Prasa in Timor-Leste from a socio-historical perspective. The main focus was on Tetun Prasa's contact with Portuguese, Indonesian, Malay and Asian varieties of restructured Portuguese and Malay, and I have tried to document it in sociolinguistic terms as well as in terms of lexis, phonology, morphology and syntax.

Chapter 1 offered an introduction to the Austronesian and Timor-Leste languages in particular, as well as a discussion on different classifications of Tetun varieties. As stated in this chapter, I adopted my own classification that was useful for the purpose of this thesis. In general, I focused on two main varieties, Tetun Prasa and Tetun Terik, which I further subdivided into two subvarieties, Tetun Suai and Tetun Viqueque, since I have noticed certain syntactical differences between these two. I also relied on van Klinken's data on Tetun Fehan, as it turned out this variety spoken in western and eastern part of the island is the most distinct one from Tetun Prasa.

Because of the general focus on language contact and on how sociolinguistic history helped to shape the development of Tetun Prasa, it was important to offer a detailed description of different historical periods in Timor-Leste's history: from the early Chinese and Malay traders, the Timorese society before the arrival of Europeans, the Portuguese presence in Southeast Asia and their settlement of Timor to Indonesian occupation and hard-won independence. The long-lasting presence of diverse foreign ethnolinguistic groups did not only impact the Timorese genetics and way of life but has dramatically influenced the local languages and Tetun Prasa in particular. So much so, that researchers began referring to Tetun Prasa as a creole (Grimes et al. 1997, Hull 1999c, Avram 2008), pidgin (Avram 2008) or a hybrid language (Hull 1999c). One of the main research question was thus to look at Tetun Prasa from the theoretical and structural point of view to see if any of these claims could hold. In Chapter 3, I provided a discussion on theory of creoles, pidgins and mixed languages and discussed the possibility of Tetun Prasa being a result of a specific type of language contact. It became clear from the discussion that the sociolinguistic context in which creoles and pidgins develop, e.g. need

for communication between communities that lacked a common language, has not been attested in Timor, since Malay and later Tetun served as a vernacular language between the outsiders and the local peoples. Neither from the linguistic point of view can Tetun Prasa be considered a creole language because it lacks one of the most salient characteristic of creoles: most of its vocabulary is still native and not derived from a superstrate language, as was later illustrated in Chapter 8. Tetun Prasa cannot be considered a mixed language either, at least not in the traditional sense. If we look at other mixed languages, we usually see a clear split, either between lexicon and grammar or verbal and nominal systems. However, Tetun Prasa has not reached that level of mixing (yet) and both lexicon and grammar, as well as verbal and nominal systems, are still predominantly native. Instead, I argue that, based on sociohistorical and linguistic data, it would be more fitting to consider the possibility of Tetun Prasa being a koine with heavy Portuguese lexical influence, especially when compared with the more conservative varieties of Tetun Terik and Tetun Fehan. Tetun was brought to Dili as a vehicular language, replacing the local Mambae, and it eventually became the mother tongue of children born in the capital. However, before it became the first language in the area, it has undergone certain reduction in forms like subject marking and plural marking and at the same time it has developed new forms (e.g. possessives, passive voice). And although Tetun Prasa has been spoken in Dili for centuries now and has been stabilized as a new variety of Tetun, it still presents a wide range of variation as it is spoken as L2 by most of the speakers outside of the capital.

I have conducted an extensive research into the literature available on Tetun and other Timorese languages and was especially interested in the older works on Tetun Prasa dating back to the end of the 19th century. These documents gave me an insight into the state of Tetun at the turn of the 20th century and were also proof that Tetun Prasa started to acquire more and more Portuguese borrowings with the arrival of Portuguese missionaries. In chapters on morphology, syntax and phonology, I also relied on the previously published and very detailed descriptive grammars by Williams-van Klinken et al. (2002a) and Hull and Eccles (2005), since my objective was not to offer another general overview of Tetun Prasa's structure but rather focus only on features that have been the result of language contact.

To be able to base my hypothesis and assumptions on real and up-to-date data, I conducted two field trips in Timor-Leste and subsequently created an oral corpus for the two varieties I compared: Tetun Prasa and Tetun Terik. The corpus consists of recordings of natural speech, in which participants were free to discuss a whole range of topics. The linguistic data extracted from the recordings were then matched with the sociolinguistic background of the speakers to produce qualitative, and in a more limited way also quantitative research, taking in consideration variables like age, sex, level of education, period in which speakers were educated and proficiency in Portuguese. At the same time, I also built a written corpus composed of newspaper articles and press releases taken from the archives of Timorese daily newspapers and the official government website. As documented by many examples in the structural chapters, sometimes the differences between these two registers – written formal and oral informal – were immense.

Having assessed all the language components of Tetun Prasa, it is clear that morphology has been least influenced by outsiders' languages. The basis for the analysis in Chapter 5 is the fact that, compared to other Central Malayo-Polynesian languages, Tetun Prasa is more analytical and isolating in nature. It does not show rich inflectional and derivational morphology and the few derivational affixes that can be found in it survived only in a fossilized form. When it comes to inflectional morphology, one of the most salient features that sets Tetun Terik apart is subject marking. The loss of this strategy in Tetun Prasa can be explained on the grounds of language contact and second-language acquisition. But according to my data, although this phenomenon is still present in both Suai and Viqueque region, it is not as pervasive as it might have been before. I have documented it almost exclusively with *h*-initial words in both subvarieties, contrary to Hull and Eccles's claims. Also, subject marking was used in less than 35% of cases with an overt personal pronoun or with another element/verb placed between a personal pronoun and an *h*-verb. Unfortunately, I was unable to record natural speech of younger speakers (less than 18 years old) to see if this generation uses subject marking on daily bases, although when instructed, they were able to produce these structures effortlessly.

Some derivational morphology is still rather productive, especially derivation of causative verbs and actor nouns. I compared prefixes found in Tetun Fehan (van Klinken 1999) and Tetun Terik with those described in works on Tetun Prasa: Williams-van Klinken et al. (2002a), Hull & Eccles (2005) and Albuquerque (2011), but most of the

affixes found in more conservative varieties exist in Tetun Prasa in a fossilized form. There is also one affix that has been borrowed from Portuguese and become fully productive deriving actor nouns even from native basis: *-dór*, coexisting with other similar affixes of native origin *-teen, -na'in* and *mak-(-k)*. As documented by Williams-van Klinken and Hajek (2003), no other Austronesian language which has been in contact with Portuguese or Castilian uses this or any other Romance suffix in a productive way. Other Portuguese affixes like *-mente* that derives adverbs and *eis-/vise-* that derive nouns have only been documented with borrowed basis in my corpus although I believe that in the future they might come to be used also with native roots.

Other word-formation strategies include zero derivation (conversion), compounding and reduplication. Since word classes in Tetun are fuzzier than the ones in Portuguese, zero derivation is often used with Portuguese borrowings, e.g. *eskola* n. 'school' and v. 'to attend school' (from Port. *escola* n. 'school'). Some instances can be also found with native words where 'affix dropping' is becoming more common in the younger generation (Williams-van Klinken et al. 2002a:18), a widespread phenomenon also in colloquial/informal Indonesian.

Compounding as a word-formation process is relatively common in Tetun Prasa. There is a large number of native compounds in Tetun Prasa and many new are emerging, either using Tetun Terik words, foreign borrowings or calques. An interesting case are so-called loanblends which consist of one native and one borrowed element, e.g. *uma-andár* 'multi-storey building' (from T. *uma* 'house' and Port. *andar* 'storey'). Other compounds, that were borrowed from a foreign language, only act as compounds in Tetun Prasa, e.g. a. *armáriu-livro* 'bookshelf' (from Port. *armário de/para livros* lit. 'cupboard for books').

Reduplication is also very common in Austronesian languages, although this strategy is more restricted in Tetun Prasa, considering it is only found with four word classes: nouns, adjective, adverbs and numerals and has limited range of semantic properties that include habituality, intensification, distributivity, and manner adjectives. Unlike Tetun Fehan, there is no reduplication of verbs in Tetun Prasa and pluralization by reduplication exists only in a fossilized form and can be assigned to Malay/Indonesian influence. Different reduplication strategies also exist in creole languages, for example

Malacca Creole Portuguese, and this fact led Avram (2008) to claim that Tetun Prasa could be a creole or an extended pidgin. However, since reduplication is wide-spread phenomenon in a more conservative dialect of Tetun Fehan, it is unlikely that this process was borrowed from other Portuguese-based creoles spoken in the region, nor has it been acquired during its development.

The influence of other languages is more visible at the level of syntax. For example, Tetun Prasa uses a whole range of pluralization strategies, such as the unmarked plural, the native plural marker *sira* 'they', the Portuguese plural marker *-s/-es* and limited reduplication. However, just like subject marking, plural prefixes *-r/-n* found in Tetun Terik are no longer used in this variety.

There is one syntactic structure that sets Tetun Prasa apart not only from its more conservative varieties but from other Austronesian and Papuan languages spoken in Timor-Leste: possession. In Tetun, it is formed using a possessive marker *(-)nia/(-)nian*, which always follows the possessor, and allows two possible orders – preposed and postposed – based on different factors, for example alienability/inalienability of a noun. Apart from this strategy, Tetun Terik and Tetun Fehan use simple juxtaposition 'possessor + possessum' or they employ a range of genitive suffixes, like *-n/-k* in Tetun Terik, attached to a personal pronoun. Comparing my data from two Tetun Terik-speaking regions, I found that in the southwestern subvariety from Suai, this construction (personal pronoun + genitive suffix) appears in a preposed position (e.g. *ita<u>k</u> buat* '1PI.GEN thing') while in the subvariety from the Viqueque region, the possessive determiner is always postposed (e.g. *rai ita<u>k</u>* 'country 1PI.GEN').

The origin and development of the possessive marker *nia* is quite interesting and so is the form of adnominal possession with pronominal possessors in Tetun Prasa. Besides acting as a possessive marker, *nia* is also the 3S personal pronoun. Hull and Eccles (2005) believed *nia* originated in Malay *punya* 'to own' and then assimilated to the 3S pronoun. Similar strategies with a possessive marker following a nominal possessor in adnominal possession can be also found in other local Timorese language, like Mambae, Baikeno, Makasae and others. However, the story is quite different in the case of adnominal possession with pronominal possessors. While other Timorese languages use simple juxtaposition of a personal/possessive pronoun and possessum, Tetun Prasa uses the following form: *personal pronoun + nia + possessum*. Baxter (1990)

and Baxter and Bastos (2012) suggested that parallel possessive structures existed in restructured varieties of Portuguese and according to Paauw (2008) also in Ambonese Malay. It is very possible that Tetun Prasa started to mark possession on personal pronouns indirectly by way of using a possessive marker *nia* after being in contact with these varieties.

When it comes to gender marking and agreement, prototypically, Tetun Prasa nouns are not marked for gender but followed by classifiers for sex distinction, although many Portuguese nouns describing humans (e.g. kin terms, professions) have been adopted in both masculine and feminine forms, like *primu* 'male cousin' (from Port. *primo),* and *prima* 'female cousin' (from Port. *prima).* Borrowed adjectives, regardless of whether they follow a native noun or a Portuguese loanword, are used in the default masculine form, e.g. *empreza públiku* 'public company' (from Port. *empresa* 'company.FEM' and *público* 'public.MASC'). However, many were borrowed as a part of fixed multiword expressions and have preserved the Portuguese gender agreement. According to my analysis, gender agreement was more common in written language and when the modifier (adjective) directly followed the noun phrase. Gender agreement is also more common in speakers proficient in Portuguese and with a higher level of education. In other cases, when the adjectives occur in relative clauses and predicative constructions, gender agreement rarely travels across the phrase boundary.

In terms of the verb phrase, Tetun Prasa shows a number of new emerging features, like copula, passive voice or borrowed structures that allow VS order. In all cases, the main 'culprit' has been the contact with the Portuguese language. Tetun Prasa, as most of the languages in the Pacific region, uses a zero copula strategy, although copula can be found in the dialect of Tetun Fehan. The slot that is normally occupied by a copula in many languages can be filled with a focus marker *mak/maka* in the case of referential nominal predicates and with *ne'e* in the case of affirmative equational clauses. But, according to Williams-van Klinken (2010), new copula *hanesan* 'like, as' has been developing and can now be found mostly in written texts. I have also noticed another preposition with a similar meaning, *nu'udar* 'like, as', being used as a copula in my written corpus. However, when I elicited these examples with my informants, they thought they 'sounded strange'.

Historically, Tetun Prasa did not possess passive voice either, but presently, there are various possibilities to express passive meaning, including calques from other languages. One of the basic passive strategies is to use the basic active form of the verb following the object. In this case, the subject of the active voice is simply eliminated and the verb is not distinctively morphologically marked (Keenan & Dryer 2007). Another strategy is the use of the emphatic construction with the OSV order, which is also found in Tetun Terik. For passive meaning, it is also possible to omit the agent of the transitive verb and shift the patient before the verb. Tetun also possesses so called 'strict morphological passive' (Keenan & Dryer 2007), using a prefix *nak-* which turns transitive verbs into intransitive verbs with a passive meaning, though this strategy is very little productive in Tetun Prasa. But other constructions, like the one with *hetan* 'get' + active verb, resemble structures in Portuguese, English and Indonesian. An even more glaring example of calquing is a passive that has no morphologically marked verb or an auxiliary, and the actor can be introduced by a preposition *hosi* 'from, by'. The source of this calque is most likely literal translations from Portuguese, English and Indonesian done by Timorese journalists.

Tetun Prasa's clause structure has been also altered due to language contact. Originally, complex sentences were mostly dominated by coordination, but borrowed subordinating conjunctions made way for a wider variety of syntactic relations. In most cases, Portuguese and Indonesian conjunctions complement the native ones, and in some cases, they are the most widely used ones; in others, they are considered to be part of the written language mainly. An interesting case is that of copulative conjunctions, where the innovation came from the more conservative dialect of Tetun Terik rather than an outsiders' language. In Tetun Prasa, different copulative conjunctions coordinate different types of structures, e.g. *ho* coordinates NPs, *mós* coordinates clauses and *i* coordinates clauses and sentences, but there is one conjunction that coordinates all types of constituents: *no*, which is possibly the 3S form of *ho* used in Tetun Terik (Williams van Klinken et al. 2002a). It is surprising to find a case in which various competing forms were not overpowered by Portuguese. Rather, a conjunction inherited from the conservative variety Tetun Terik became the norm in Tetun Prasa, as opposed to the Portuguese borrowing *i*, which is still confined to the coordination of only clauses and sentences and whose frequency in oral speech and written texts continues to be low.

The fact that borrowed conjunctions were often more common in oral speech than written text – which was against my initial hypothesis – was proven by other examples. Native adversative conjunction *maibé,* competing with its borrowed counterpart *mas/mais* (from Port. *mas),* occurred almost 400 times in my corpus, while the Portuguese conjunction occurred only 64 times, out of which 56 cases were documented in the oral speech.

Although many of the subordinating conjunctions are of native origin, this does not mean that they have always fulfilled this function. Take, for example, complementizers. In the dialect of Tetun Fehan, complementizers are few and used rather unfrequently. On the other hand, due to the close contact with Portuguese clause structure, Tetun Prasa employs a range of native (*katak, atu*) and borrowed (*para* from Port. *para*) complementizers. But the native *katak* was originally used as a verb meaning 'say' and the main function of *atu* is that of an irrealis TAM marker.

Another innovation of Tetun Prasa is complex conjunctions, like the complex complementizer *para atu*, the complex relativizer *ne'ebé ke* (from native *ne'ebé* and Port. *que*), or the complex adverbial conjunctions of time (*bainhira kuandu* 'when'), purpose (*para atu/para hodi* 'in order to') and condition (*se karik* 'if').

I decided to take a closer look at adverbial clauses of time, purpose, reason, condition and concession to find instances of borrowed conjunction, their frequency in oral and written language and semantic differences. In case of adverbial clauses of time, purpose and reason, the native conjunctions *bainhira* 'when', *hodi* 'in order to' and *tanba* 'because', respectively, were far more preferred than their borrowed counterparts, *para* 'in order to', *porké* 'because' and *komu* 'since' in both written texts and oral speech, except for *kuandu* 'when', which seemed to be more popular in spoken language. When it comes to conditional clauses, the borrowed conjunction *se* 'if' seemed to be more common in Tetun Prasa than the native *karik* 'maybe', although the two conjunctions are often found in the same clause, either together or disconnected. Williams-van Klinken et al. (2002a) and Hull and Eccles (2005) also discussed the possibility of temporal conjunctions *bainhira* and *kuandu* 'when' being used in conditional clauses. In my corpus, I have found examples of both temporal conjunctions also with conditional reading. The last group of adverbial clauses are those of concession. It seems that no

native conjunction exists to introduce this kind of subordinate clause, although sometimes *bele* 'can' is used to express this relation. Two borrowed conjunctions, Indonesian *biar* 'although' and Portuguese *mezmu* 'although' are also rather marginal, with *maski* 'although' being the preferred choice especially in written texts. This preference for *maski* might have to do with the fact that it was introduced into Tetun a long time ago. Originally derived from Port. *mas que*, the meaning of 'although' has disappeared from modern Portuguese. However, the same conjunction is found in Malay and its varieties as well as in Portuguese creoles of Malacca and Macau, which are the most probable sources of this borrowing.

As we can see, the influence of Portuguese and Malay/Indonesian has been intensive and long-lasting enough to affect the syntax of Tetun Prasa. New structures, not found in Tetun Terik and/or Tetun Fehan have emerged and other, borrowed ones have been competing with their native counterparts. Only time will tell if loaned conjunctions will eventually replace the native ones.

The phonology of Tetun Prasa, discussed in Chapter 7, was the component of language with the highest ratio of interspeaker variation and it was at this point that I realized that there must me something more besides the basic sociolinguistic characteristics that I collected that motivates the choices my speakers made. There were speakers of similar age and of similar background but whose phonemic inventories were completely different. Based on my findings, I strongly believe that future research should also look at language attitudes, which could be very indicative of speakers' choices. Originally, Tetun contained only 11 native phonemes but later adopted new segments from Malay and Portuguese, amounting nowadays up to 22 native and foreign segments. One might argue whether these borrowed segments, especially the palatals and palate-alveolar fricatives, are actual phonemes or just allophones to native phonemes, a question that has remained unanswered and would require much deeper analysis. Instead, I decided to attempt quantitative analysis in order to determine whether certain social variables like age, proficiency in Portuguese, level of education and language of instruction at school had any influence on the pronunciation of borrowed phonemes. Since it was impossible to list all possible pronunciations for every word I discussed, I decided to use the educated Portuguese-influenced pronunciation for my IPA transcriptions.

Variation has been also observed with native phonemes /w/ and /'/, which are not part of the Tetun Prasa phonemic inventory. The glottal stop is still used in the Tetun Terik regions, regardless of the age of the speakers, although not all Tetun Terik speakers possess this phoneme. What proved to be the decisive factor favouring the absence of this segment was the mother tongue of Tetun Terik L2 speakers (in our case, Bunak) and an extended period of living/working in a Tetun Prasa environment. The situation was rather different in Dili, where Tetun Prasa speakers rarely pronounce this sound. It is believed that it has disappeared after the contact with Mambae, a language originally spoken in Dili that also does not contain this phoneme (Williams-van Klinken 2010b, Albuquerque 2011). I analysed my Tetun Prasa recordings and contrary to the general belief, I came across occasional instances of glottal stop also in some of my Tetun Prasa speakers. When I tried to determine the reason, I found that the glottal stop was only pronounced by speakers whose mother tongue contained this phoneme, such as Waima'a and Midiki.

The reason behind the alleged disappearance of [w] from Tetun Prasa and its substitution with [b] also has to do with the influence of Mambae, which does not contain this phone. However, my data prove otherwise, since, according to my analysis, some speakers freely alternate between the two allophones. Again, this might have to do with the fact that they also speak other Timorese languages and do not find [w] foreign to their inventory.

The limited number of speakers and occurrences proved to be problematic when doing quantitative analyses of certain underrepresented phonemes, like the Malay/Indonesian /ŋ/ or the Portuguese palatals /ɲ/ and /ʎ/. I have thus decided to focus on the palato-alveolar fricative /ʃ/ and its allophones [ʃ], [s], and marginal [z], and run a couple of different analyses. Looking at Tetun Prasa data, I was able to confirm that speakers who received higher education and show proficiency in Portuguese are more likely to opt for the Portuguese pronunciation. However, I also came across cases which could not have been explained on the basis of these variables, such as a speaker proficient in Portuguese and educated in the Portuguese times with zero instances of Portuguese-like pronunciation in Tetun Prasa.

Having done the same analysis for Tetun Terik, the results were quite surprising, as it seemed that Tetun Terik speakers were more likely to choose the Portuguese-like

pronunciation than Tetun Prasa speakers. This was against my initial hypothesis that Tetun Terik speakers would opt for more Tetun-like pronunciation since this variety has not been in such intensive contact with Portuguese. I tried to find an explanation for this and focused on two factors: number of occurrences and the composition of the vocabulary. In many cases, the number of occurrences was less than 20, which proved to be a limitation to my quantitative analysis. As for the vocabulary, the majority of the 20 most frequent words containing /ʃ/ were content words connected to the school environment, where Portuguese is one of the languages of instruction. To confirm if these two factors were decisive, I analysed the same fourteen words found in both varieties. Again, the results were unexpected. While speakers of Tetun Prasa opted for Portuguese-like pronunciation only in 36% of cases, in Tetun Terik this figure was much higher: 61%. Although I was not able to find a reasonable explanation, one thing became clear – the gap between these two varieties is narrowing, at least when it comes to phonology.

My next analysis had to do with word classes. As mentioned above, Tetun Prasa often employs zero derivation, or conversion, when borrowing new words. This is how the Portuguese noun *escola* 'school' came to be used in Tetun Prasa as a noun but also as a verb *eskola* 'to attend school'. I was interested to see whether different parts of speech would yield different pronunciation, but my analysis showed that the percentages were exactly the same for both word classes: 78% of occurrences as a verb and 78% of occurrences as a noun were pronounced as [eʃ.ˈko.la].

The last analysis led me to look at the phonological context. In European Portuguese, final <s> is pronounced as [ʃ] when the following word starts with a consonant. In the case of vowel-initial words, the final <s> is realized as [z]. I decided to test this phonological law on one of the most frequent words in my oral corpus: *depois* 'later; after'. In the majority of cases, the final <s> was pronounced as [s], but the percentage was higher when the following word was V-initial (79%) than in the case of C-initial words (67%). Also, [s] was pronounced in almost all cases (92%) when the following word was /s/-initial. Although this proves that context plays role in the pronunciation of [ʃ], it is questionable whether the alternation has been modelled on the Portuguese rule.

Compared to the rich vowel phonology of Portuguese, Tetun Prasa has a simple five vowel system. Phonemes like /a/, /i/, /u/ are relatively fixed and do not undergo much variation, although they can be weakened when unstressed. The two vowels that are affected in the pre-tonic or final/post-tonic position are /o/ and /e/, respectively. Just like in Brazilian and African varieties of Portuguese and in the Portuguese-based creoles and some dialects of European Portuguese, /e/ can be raised to [i] in a post-tonic position. I looked closer at the pronunciation of /e/ in Portuguese borrowings in the post-tonic position in order to see whether age had any influence on the pronunciation in different varieties. The analysis of Tetun Prasa showed that, although not prevalent, [i] is pronounced by speakers across all age groups. The number of occurrences of [i] were much more limited in Tetun Terik, which might be, again, explained by the very low number of post-tonic /e/ occurrences in the first place.

An interesting category is that of diphthongs. Originally, there were no diphthongs in Tetun. In native words, two different vowels in a sequence are considered belonging to two different syllables, although in fast speech they are often pronounced as diphthongs (*ain* 'leg' [ˈa.in] ~ [ˈajn]). Tetun Prasa adopted diphthongs with Portuguese borrowings, although speakers sometimes simplify them to single vowels.

To summarize this chapter, I decided to look at the overall preference for Portuguese-like pronunciation in Tetun Prasa and Tetun Terik, based on different variables: historical period, level of education and proficiency in Portuguese. According to my analysis, speakers who were educated during the Portuguese times were more likely to adopt Portuguese pronunciation (70% of occurrences) than speakers educated during the Indonesian occupation (53%) and in the post-independence times (56%). Again, speakers with higher education (64% of occurrences) were more likely to opt for the Portuguese pronunciation then speakers with secondary (53%) and primary (49%) education. Only one speaker claimed to have no education but pronounced Portuguese sounds in 78% of cases. Surprisingly, proficiency in Portuguese seemed to be the least decisive factor in my analysis but this might have to do with the fact that the speakers self-assessed their knowledge of Portuguese and thus the data might be a bit misleading in some cases. Speakers who claimed to understand Portuguese pronounced Portuguese phonemes in 68% of cases, while speakers who can understand Portuguese only in 50% of cases. Speakers with no knowledge of Portuguese, especially in the Tetun Terik region,

scored quite high (61%) which makes us wonder if they did not underestimate their Portuguese skills.

The last Chapter 8 was dedicated to lexicon, the component of Tetun Prasa most affected by language contact. One does not need to be a linguist to see the abundance of Portuguese borrowings, especially in the written language. Based on the historical period, it was possible to identify older Malay and Portuguese borrowings as well as more recent Indonesian, Portuguese and English borrowings. But besides the origin, I was also interested to see what kind of words were most likely to be borrowed, based on their word class, content/function distinction or semantic field.

First, I ran two basic tests to evaluate the borrowability of basic vocabulary. The most famous list of basic vocabulary is so-called Swadesh list, named after its author, American linguist Morris Swadesh. I filled the abridged 100-item list and failed to find a single borrowing, except for 'green' which is translated as *verde* (from Port. *verde*) in certain contexts. A more recent list of basic vocabulary has been developed as a part of the Loanword Typology project and is referred to as the Leipzig-Jakarta list. The two 100-item lists share 62 words. Again, I was unable to find a single borrowing, which proves that the native Austronesian nucleus is almost intact. Together with the sociolinguistic history described in Chapter 2, this is another proof that Tetun Prasa should not be considered a Portuguese-based creole.

Second, I analysed my oral and written corpus, working with a list of initial hypotheses based on previous studies on borrowability and lexical borrowings in Tetun Prasa. I conducted two separate analyses, one based on work tokens, which gave me more insight into the frequency of native and borrowed words, and one based on word types which offered data on the extent of both lexicons. The analysis of word tokens confirmed the claims previously made by Williams-van Klinken (2002), Williams-van Klinken and Hajek (2009b) and Williams-van Klinken and Hajek (2016), that the ratio of borrowings is much higher in written texts (43% in my corpus) than in oral speech (25% in my corpus). It was also expected to find more borrowings in the speech of Tetun Prasa speakers (29%), compared to Tetun Suai (22%) and Tetun Viqueque (21%) speakers. One of the initial hypothesis that proved wrong was the ratio of Indonesian borrowings. Various authors have commented on the fact that the subvariety of Tetun Terik spoken in

the southwest shows higher number of these borrowings because of the proximity of Indonesian border. However, my data from Suai show otherwise: the ratio of Indonesian loanwords ranges from 0% to 7%, with an average of 1.5%.

Just like in the case of phonetic realizations, I attempted a quantitative analysis of my oral corpus based on four social variables: sex, historical period in which the speakers were educated, level of education, and proficiency in Portuguese. The language of instruction in school and the period in which Tetun Prasa speakers were educated showed to have little influence on the number of Portuguese borrowings they used, although it is interesting to see slightly lower numbers of borrowings in the younger speakers. What seems to play a role is the combination of three social factors: proficiency in Portuguese, level of education and sex. Female speakers with primary/secondary education and some/no understanding of Portuguese show the lowest numbers of borrowings, regardless of their age. It proved much more difficult to interpret Tetun Terik data because of its large disparity: the ratio of Portuguese loanwords ranged from 3% to 38%. Compared with Tetun Prasa, data from Tetun Suai-speaking area seem to show the inverse tendency. Younger speakers, regardless of age, with lower education and limited proficiency in Portuguese demonstrate more than average percentage of Portuguese borrowings. In the case of Tetun Viqueque, only one speaker shows a really high percentage of Portuguese borrowings (33%), which can be explained by the fact that he spent some time living and studying at a university in Dili and that he uses Portuguese on daily basis in his work environment. The rest of the speakers show 12% of Portuguese borrowings in average.

Analysing the press register and my written corpus yielded very different results. According to the previous research of Williams-van Klinken and Hajek (2009b), 30-40% of word tokens in the press register were Portuguese loans compared to only 10 to 20% of Portuguese loanwords in informal urban speech, which corresponds to my analysis of word tokens in the written corpus (38%). The numbers were much higher when I analysed word types: 79% of all word types in my written corpus were Portuguese borrowings, compared to 18% of native words, 1% of Indonesian borrowings, 1% of English borrowings and 0.5% of Malay borrowings. The dominance of Portuguese loanwords in the press register was expected: according to the INL policy, Indonesian borrowings are to be avoided. In the case of my oral corpus, the difference was not as big, but Portuguese

borrowings (49%) still outnumbered native words (44%). Indonesian loanwords are much more common in oral speech and amount to 5% of all word tokens.

There is clear empirical evidence that content words are more likely to be borrowed than function words and this proved to be true also in the case of Tetun. The number of borrowed content words was almost three times as high as that of the native words; on the other hand, functional borrowings amounted to 32% of all function words. Interestingly, more functional borrowings were found in my oral corpus, although one would expect to find them rather in written texts which make greater use of borrowed syntactic structures.

The three classes most likely to be enriched by borrowings were nouns, verbs and adjectives. In general, nouns are thought to be easier to borrow than verbs, because they do not have to undergo complex morphosyntactic modifications. However, in isolating languages like Tetun Prasa, in which verbs do not need to be specially adapted or conjugated to fit into the morphosyntactic structure of the language, the ratio of borrowed nouns and verbs tend to be similar, just like in the case of Tetun: 58% and 57%, respectively. Somehow surprising might seem the high number of borrowed adjectives (81%), but most of these were actually derived from past participles that were part of a VP in the source language but came to be used as predicative adjectives in Tetun Prasa.

And what can be said about semantic fields? According to Tadmor (2009), some semantic fields are more likely to be affected by borrowing than others, and they tend to be the same across all languages. These are especially technical terms referring to modern and abstract concepts, words related to religion, household and clothing. We can also distinguish between cultural borrowings, which describe new concepts introduced by the source language and culture, and core borrowings, which duplicate or replace native words that already exist in the recipient language. Most of the loanwords in Tetun are cultural borrowings (89% in written language and 76% in oral language), although the number of core borrowings is quite significant in oral speech (24%). When it comes to the most numerous semantic domains, politics and law, economy, army and police, technology and toponyms ranked among the top 5 in the written corpus. Oral language also contained many words referring to speech and language, numbers, religion and belief and education.

One of the by-products of lexical borrowings has been semantic change. We can find examples of semantic narrowing, archaisms, changes in word classes and words whose form/derivational morphology was borrowed from one source language and the meaning from a different language, as described by Godinho (2013).

All in all, Tetun Prasa, especially the press register, borrowed a considerable amount of new vocabulary and for its high ratio of loanwords it is nowadays also referred to as Tetungés. The functional borrowings subsequently affected and altered the syntax and morphology of Tetun Prasa, although, as I have argued above, some of these innovations have also spread to the more conservative subvarieties of Tetun Terik.

Unfortunately, because of the high inter- and intraspeaker variation, and because of rather low number of speakers, it proved difficult to conduct a more in-depth quantitative and variationist analysis. The fact that the Timorese society is extremely diverse and multilingual and that there are many variables influencing the speech of individuals was already noted by Williams-van Klinken at al. (2002a) 15 years ago. In many cases, my results contradicted those documented in older works but we need to be careful about attributing it solely to language change. We also have to factor in the composition of my focus group, which was still big enough to conduct a qualitative study of the present-day Tetun Prasa.

To summarize, the main outcome of this dissertation is a detailed overview of all borrowed structures at the level of morphology, syntax, phonology and lexicon. The general descriptions of Tetun Prasa were based on previous works on this language, especially the grammars of Williams-van Klinken et al. (2002a) and Hull & Eccles (2005) and complemented with my own observations of the synchronic and diachronic variation in Tetun Prasa based on the data from my oral and written corpora.

Since this research has been developed within the area of Portuguese linguistics, the main focus was the language contact between Tetun Prasa and Portuguese, as well as its restructured varieties spoken in South, East and Southeast Asia. Considering no other Austronesian language has been in such an intensive and long-lasting contact with Portuguese, it was interesting to observe all kinds of different dimensions of the Portuguese influence on Tetun Prasa, not only from purely linguistic point view but also

from sociohistorical perspective. One of the outcomes of the lexicon analysis is a compilation of a list of all Portuguese borrowings found in my oral and written corpora of Tetun Prasa and Tetun Terik with their respective word classes. The next step, which remains to be done, would be to include, when possible, information about their specific source language variety and the period when they entered Tetun Prasa, as well as any changes in their phonological form or semantic range.

One of the main questions – not only in my thesis but among linguists working on Tetun – is the status of Tetun Prasa as a contact language. I decided to tackle this issue at the very beginning with the intention to refute the idea that Tetun Prasa could or should be considered a creole. Based on the theory of language contact and looking at the sociohistorical context in which Tetun Prasa emerged, it is now clear that there were no conditions for a creole language to arise and that Tetun Prasa has not reached the necessary level of mixing for it to be considered a mixed language. I have proposed the possibility of Tetun Prasa being an immigrant koiné with heavy Portuguese influence which is something that Thomaz (2002) already implied when comparing Tetun Prasa to Língua Geral (Nheengatu). In this process of koineization, Tetun language spread from the place where it originated and became the language of the new capital as well as a lingua franca of the eastern part of the island. Because it was acquired by its new speakers as L2, some of its structures were reduced or simplified. However, once the language became an L1 of the Dili population, Tetun Prasa has seen linguistic expansion and new innovations that were not originally found in Tetun Terik.

Although a lot has been written on the differences between Tetun Prasa and Tetun Terik, no quantitative analysis has ever been done in this regard, possibly due to the limited data available about the more conservative variety. Many authors have commented on the 'unintelligibility' between the two varieties with some of them suggesting that the gap has been narrowing. Based on my analysis, we can tell that Tetun Terik dialects have been undergoing changes, possibly due to the contact with Tetun Prasa, and several speakers have presented relatively high numbers of borrowings. It would be interesting to conduct a similar analysis in the future to map the possible convergence between Tetun Prasa and Tetun Terik and/or other local languages. More research on this variety would also provide a better insight into the use of certain forms

and structures, e.g. derivational morphology and SVCs. Studies should also focus on younger speakers of Tetun Terik to see whether the contact with Tetun Prasa through social and mass media has influenced their mother tongue.

The oral corpus that I have created resulted from my fieldwork in Timor-Leste and will be made freely accessible for future research. The variationist analysis was the first ever done on Tetun Prasa although with the disclaimer that much more data will need to be collected and analysed to get results representative of Timorese society. What I have managed to show is that some sociolinguistic variables are more important than others while admitting that language attitudes also need to be considered in the future. During my three visits to Timor-Leste, it was apparent that the attitudes of younger generations were stronger towards certain foreign languages than others, in contrast with often-present nostalgia of elderly speakers. It was also clear from the quantitative and qualitative analysis of Tetun Prasa that often speakers from very similar backgrounds showed opposite results, which, again comes down to a deeper divide between language attitudes rooted in the history of their families. Unfortunately, because issues like this only emerged after I had collected my data and analysed my corpus, some questions have remained unanswered.

Since multilingualism is a norm in Timor-Leste, the setting in which Tetun Prasa emerged is rather specific from other contact situations. To get a better understanding of the sociolinguistic situation in which Tetun Prasa has developed, more research needs to be done on other Timorese (especially Austronesian) languages. A detailed mapping of the linguistic landscape in Timor-Leste is necessary to determine how many languages and their respective dialects are found on the island. This could offer us greater insight into the variation caused by the influence of various mother tongue of L2 speakers of Tetun Prasa but also help in the area of language planning and mother tongue-based multilingual education.

ANNEX 1 – QUESTIONNAIRE

(TETUN ORIGINAL AND ENGLISH TRANSLATION)

Kestionáriu

Naran: _____

Tinan: _____

Seksu: <u>mane – feto</u>

Servisu iha ne'ebé: _____

Estuda iha ne'ebé: _____

Moris iha ne'ebé (vila/distritu): _____

Hela iha ne'ebé (vila/distritu): _____

Fatin seluk ne'ebé uluk hela: _____

Aman-nia lia materna: _____

Inan-nia lia materna: _____

Ita-nia lia materna: _____

Lia seluk ne'ebé ko'alia: _____

Lia seluk ne'ebé persebe: _____

Lia (lia sira) iha iskola primária: <u>tetun – portugés – bahasa</u>

Lia (lia sira) iha iskola sekundária: <u>tetun – portugés – bahasa</u>

Lia (lia sira) iha universidade: <u>tetun – portugés – bahasa</u>

Lia (lia sira) ne'ebé ko'alia iha uma: _____

Lia (lia sira) ne'ebé ko'alia ho belun sira loro-loron: _____

Lia (lia sira) ne'ebé uza iha sirvisu: _____

Oinsá konsente: kestionáriu ne'e hala'o tuir peskiza iha liur konabá knaar dotoramentu (S3) nian iha Fakuldade Letras Universidade Coimbra nian, Portugál, ho títulu *Tetun iha Timor-Leste: knaar hosi lia-nia kontaktu atu haburas Tetun-Prasa*, ne'ebé Zuzana Greksáková hala'o. Ha'u konkorda katak dadu privadu sira hanesan mós gravasaun hotu ne'e ha'u halo ona bele uza ba knaar ne'e no/ka ba knaar seluk/artigu/aprezentasaun ruma.

Ha'u konkorda katak: □ **bele halo vídeo**

□ **la bele halo vídeo**

Asinatura: **Data:**

Questionnaire

Name: _____

Age: _____

Sex: <u>male – female</u>

Place of work: _____

Education: _____

I was born in (town/district): _____

I live in (town/district): _____

I previously also lived in: _____

Father's mother tongue: _____

Mother's mother tongue: _____

My mother tongue: _____

Other languages that I speak: _____

Other languages that I understand: _____

Language(s) of instruction in primary school: <u>Tetun – Portuguese – Indonesian</u>

Language(s) of instruction in secondary school: <u>Tetun – Portuguese – Indonesian</u>

Language(s) of instruction in higher education: <u>Tetun – Portuguese – Indonesian</u>

Language(s) I speak at home: _____

Language(s) I speak with my friends on a daily basis: _____

Language(s) I use at work: _____

Consent agreement: this questionnaire is part of the fieldwork carried out by Zuzana Greksakova for her doctoral thesis entitled Tetun in Timor-Leste: the Role of Language Contact in its Development within the doctoral program Portuguese Linguistics at Faculty of Letter of University of Coimbra in Portugal. I agree that my personal details and all the recordings can be used for the purpose of this thesis and/or other papers/presentations.

I agree that: □ video can be shot

□ video cannot be shot

Signature: Date:

ANNEX 2 – SPEAKERS AND THEIR SOCIOLINGUISTIC PROFILE

Recordings	Variety	Initials	Age	Sex	Profession/place of work	Education	Place of birth	Currently living in	Previously also lived in	Father's mother tongue	Mother's mother tongue	My mother tongue	Other languages that I speak	Other languages that I understand	Language of instruction in primary school	Language of instruction in secondary school	Language of instruction in university	Languages I speak at home	Languages I speak with my friends on a daily basis	Languages I use at work
0005TD, 0006TD	TP	BE	30	male	agricultural group	higher education	Haekoni, Bagia, Baucau	Dili	N/A	Makasae and Tetun	Makasae	Tetun and Makasae	Tetun Prasa, Indonesian, English, Portuguese	N/A	Indonesian, Tetun	Indonesian, Tetun, Portuguese	Indonesian, Tetun, English, Portuguese	Makasae, Tetun	Tetun	Tetun
0005TD, 0006TD, 0015 TD	TP	JBP	26	male	university student	higher education	Haekoni, Bagia, Baucau	Dili	N/A	Makasae	Makasae	Makasae	Tetun Prasa, Indonesian, English, Naueti	Portuguese	Portuguese, Tetun, Indonesian	Portuguese, Tetun, Indonesian	Portuguese, Tetun, Indonesian	Tetun Prasa, Makasae, Naueti	Tetun Prasa	Tetun Prasa
0008TD	TP	TAB	55	male	staff in local municipality office	pre-secondary	Lekidoe, Aileu	Dili	N/A	Mambae	Mambae	Mambae	Tetun, Portuguese, Malay	N/A	Portuguese	Portuguese	N/A	Tetun, Mambae	Tetun	Tetun
0014TD	TP	LEG	20	female	English course student	secondary	Bagia, Baucau	Dili	N/A	Makasae	Makasae	Makasae	Tetun, English, Indonesian	Portuguese	Portuguese, Tetun	Portuguese, English	N/A	Tetun	Tetun	English, Tetun, Portuguese, Indonesian
0014TD	TP	FDP	19	female	English course student	secondary	Bagia, Baucau	Dili	N/A	Makasae	Makasae	Makasae	Tetun, Portuguese, English	N/A	Tetun, Portuguese	Tetun, Portuguese, Indonesian, English	N/A	Makasae, Tetun	Tetun, Portuguese, English	Tetun, Portuguese, English
0014TD	TP	JMB	21	female	English course student	secondary	Bagia, Baucau	Dili	N/A	Makasae	Makasae	Makasae	Tetun, English, Indonesian	N/A	Tetun, Indonesian	Tetun, English, Indonesian	N/A	Tetun, Makasae	Tetun	Tetun, English, Indonesian
0014TD	TP	ACB	21	male	university student	higher education	Bagia, Baucau	Dili	N/A	Makasae	Makasae	Makasae	Naueti, Tetun, Portuguese, English	N/A	Portuguese, Tetun	Portuguese, English, Tetun	Indonesian, Portuguese, English	Tetun	Tetun	Portuguese, Tetun
0051TD	TP	DLB	45	male	professor at UNTL	higher education	Kailaku, Bobonaro	Dili	N/A	Kemak	Kemak	Kemak	Tetun, Portuguese, English	Bunak, Mambae	Tetun	Indonesian	Indonesian	Tetun	Tetun	Tetun, English, Portuguese, Indonesian
0053TD	TP	JLS	44	male	professor at UNTL	higher education	Baucau	Dili	N/A	Waima'a	Waima'a	Waima'a	Makasae, Malay, English, Kairui, Midiki	Tetun Terik, Mambae, Portuguese	Tetun, Portuguese, Indonesian, English	Tetun, Portuguese, Indonesian, English	Tetun, Portuguese, Indonesian, English	Tetun, Waima'a	Tetun	Tetun, Portuguese, English, Malay
0055TD	TP	MAS	45	male	professor at UNTL	higher education	Vemasi, Baucau	Dili	Baucau, Lospalos	Midiki	Midiki	Midiki	Fataluku, Makasae, Galolen, Malay, English	Portuguese, Spanish	Indonesian	Indonesian	Indonesian	Tetun	Tetun	Tetun
0056TD	TP	CS	50	female	maid	primary	Dili	Dili	N/A	Mambae	Mambae	Tetun	Indonesian	Portuguese	N/A	N/A	N/A	Tetun	Tetun	Tetun
0067TD	TP	LAOB	30	male	staff at GIZ	higher education	Dili	Dili	Baucau, Natarbora, Maliana	Makasae	Tetun	Tetun	Indonesian, English, Portuguese, Spanish, German	Makasae	Indonesian	Indonesian	Portuguese	Tetun, Portuguese	English, Portuguese, Indonesian, Tetun	Tetun, English
0054TD	TP/TS	NSG	51	male	professor at UNTL	higher education	Fohoren, Kovalima	Dili	Suai	Tetun Terik	Tetun Terik	Tetun Terik	Tetun Prasa, Indonesian, Portuguese	English, Bunak	Indonesian	Indonesian	Portuguese	Tetun Prasa	Tetun Prasa	Tetun Prasa, Portuguese

Recordings	Variety	Initials	Age	Sex	Profession/place of work	Education	Place of birth	Currently living in	Previously also lived in	Father's mother tongue	Mother's mother tongue	My mother tongue	Other languages that I speak	Other languages that I understand	Language of instruction in primary school	Language of instruction in secondary school	Language of instruction in university	Languages I speak at home	Languages I speak with my friends on a daily basis	Languages I use at work
0017TS	TS	MN	44	female	teacher	secondary	Suai Loro	Suai Loro	N/A	Tetun	Tetun	Tetun Terik	Malay	Bunak, Portuguese	Malay	Malay	N/A	Tetun	Tetun	Tetun, Malay
0017TS	TS	MCA	29	female	teacher	secondary	Suai Loro	Suai Loro	N/A	Tetun	Tetun	Tetun Terik	Malay	N/A	Malay	Malay	N/A	Tetun Terik	Tetun Terik	Tetun, Malay
0018TS	TS	JA	49	male	teacher	higher education	Kamanasa	Kamanasa, Suai	Dili	Tetun	Tetun	Tetun	Indonesian, Portuguese	Bunak	Indonesian	Indonesian	Portuguese	Tetun	Tetun	Tetun, Portuguese
0018TS, 0020TS	TS	RASS	49	male	cultural staff	higher education	Fatumea	Debus, Suai	N/A	N/A	N/A	Tetun Terik	Portuguese, Malay	English, Bunak	Tetun, Portuguese, English	Tetun, Portuguese, English	Tetun, Portuguese, English, Malay	Tetun	Tetun, Portuguese	Tetun, Portuguese
0019TS	TS	BB	54	female	teacher	secondary	Kamenasa	Kamenasa	N/A	Tetun Terik	Tetun Terik	Tetun Terik	Malay, Portuguese, English, Tetun Prasa	N/A	Portuguese, Tetun	Tetun, Portuguese, English, Malay	N/A	Tetun Terik	Tetun Terik, Malay, Portuguese, Tetun Prasa	Tetun and Portuguese
0019TS	TS	NA	54	male	teacher	higher education	Kamenasa	Kamenasa	N/A	Tetun Terik	Tetun Terik	Tetun Terik	Malay, Portuguese, Tetun Prasa	N/A	Portuguese, Tetun	Tetun, Portuguese, Malay, English	Tetun, Portuguese, Malay, English	Tetun Terik	Tetun Terik, Malay, Portuguese, Tetun Prasa	Tetun Terik, Portuguese
0020TS	TS	CSG	56	male	teacher	higher education	Fohoren, Kovalima	Kovalima	Suai	Tetun Terik	Tetun Terik	Tetun Terik	Malay, Portuguese, Tetun Prasa	Bunak	Tetun Terik, Portuguese, Bunak	Tetun, Portuguese	Tetun, Malay, Portuguese	Tetun, Bunak	Tetun Terik, Bunak	Tetun Terik
0023TS	TS	LLM	41	male	teacher	higher education	Matai, Kovalima	Matai, Kovalima	N/A	Tetun	Tetun	Tetun	Bunak, Portuguese, Malay	English	Tetun, Portuguese	Tetun, Portuguese, English, Malay	Tetun, Portuguese, English, Malay	Tetun	Tetun	Tetun, Portuguese
0023TS	TS	MA	47	female	teacher	higher education	Matai, Kovalima	Matai, Kovalima	N/A	Tetun	Tetun	Tetun	Bunak, Portuguese, Malay	English	Tetun, Portuguese	Tetun, Portuguese, Malay	Tetun, Portuguese, English	Tetun	Tetun, Portuguese	Tetun, Portuguese
0026TS	TS	FJC	43	female	teacher	secondary	Fohorua	Fohorua, Kovalima	Dili	Tetun	Tetun	Tetun Terik	N/A	N/A	Indonesian	Indonesian	N/A	Tetun Terik	Tetun	Tetun, Portuguese
0026TS	TS	ATB	41	male	teacher	higher education	Fatumea	Debus, Suai	Dili	Tetun	Tetun	Tetun Terik	Indonesian, Portuguese, English	N/A	Indonesian	Indonesian	Indonesian	Tetun	Tetun	Tetun
0032TS	TS	GJX	42	female	teacher	higher education	Matai, Kovalima	Kovalima	N/A	Tetun Terik	Tetun Terik	Tetun Terik	Tetun Prasa	Kemak, Portuguese	Tetun, Portuguese	Tetun, Portuguese, Malay, English	Tetun, Portuguese, Malay, English	Tetun Terik	Tetun Ofisiál/Prasa	Tetun Ofisiál
0032TS	TS	OA	56	female	teacher	higher education	Matai, Kovalima	Kovalima	N/A	Tetun Terik	Tetun Terik	Tetun Terik	Tetun Prasa, Galolen, Tetun Loro-loron	N/A	Tetun, Portuguese	Tetun, Portuguese	Tetun, Portuguese	Tetun Terik	Tetun Terik, Tetun Prasa	Tetun Prasa
0033TS	TS	OC	31	male	staff in school	secondary	illegible	Kovalima	Maukatar, Kovalima	Bunak, Tetun	Bunak, Tetun	Bunak	Tetun, Indonesian, Portuguese	N/A	Indonesian	Indonesian, Tetun, Portuguese	N/A	Bunak, Tetun	Tetun, Bunak	Portuguese, Tetun
0033TS	TS	AGJ	24	male	staff in school	secondary	Bobonaro	Kasabauk, Kovalima	N/A	Bunak	Bunak	Bunak	Tetun, Indonesian, Portuguese	N/A	Tetun, Portuguese	Tetun, Portuguese	N/A	Tetun, Bunak	Tetun, Bunak	Tetun, Portuguese

Recordings	Variety	Initials	Age	Sex	Profession/place of work	Education	Place of birth	Currently living in	Previously also lived in	Father's mother tongue	Mother's mother tongue	My mother tongue	Other languages that I speak	Other languages that I understand	Language of instruction in primary school	Language of instruction in secondary school	Language of instruction in university	Languages I speak at home	Languages I speak with my friends on a daily basis	Languages I use at work
0039TV	TV	MVJ	51	male	staff at land registry	higher education	Viqueque	Viqueque	Dili	Naueti	Tetun Terik	Tetun Terik	Portuguese, Indonesian	Naueti, Makasae	Portuguese	Portuguese, Indonesian	Tetun	Tetun Terik, Tetun Prasa	Tetun, Portuguese	Portuguese, Tetun
0042TV, 0049TV	TV	MS	43	male	farmer	secondary	Bahalarauain, Viqueque	Aisahe, Bahalarauain, Viqueque	illigible	Tetun Terik	Tetun Terik	Tetun Terik	N/A	N/A	Indonesian	Indonesian	N/A	Tetun Terik	Tetun Terik	Tetun
0043TV	TV	AP	69	male	farmer	no education	Viqueque vila	Viqueque vila	Welaku, Bahalarauain, Viqueque	Tetun Terik	Tetun Terik	Tetun Terik	Tetun Prasa	N/A	N/A	N/A	N/A	Tetun Terik	Tetun Terik	Tetun
0044TV	TV	JCP	58	male	teacher	higher education	Viqueque	Viqueque	N/A	Tetun	Tetun	Tetun	Portuguese	Indonesian	Tetun, Portuguese	Portuguese	Indonesian	Tetun	Tetun	Tetun, Portuguese
0045TV	TV	LP	49	male	teacher	higher education	Viqueque	Bahalarauain, Viqueque	Dili	Tetun Terik	Tetun Terik	Tetun Terik	Tetun, Portuguese	N/A	Indonesian	Indonesian	Indonesian	Tetun Terik	Tetun	Tetun, Portuguese
0046TV, 0050TV	TV	JCV	48	male	teacher	higher education	Viqueque	Viqueque	Bahalarauain, Viqueque, Baucau	Tetun Terik	Tetun Terik	Tetun Terik	N/A	Naueti, Makasae	Indonesian	Indonesian	Indonesian	Tetun Terik	Tetun Prasa	Tetun, Portuguese
0047TV	TV	AFA	50	male	teacher	higher education	Viqueque	Viqueque	Dili	Tetun Terik	Tetun Terik	Tetun Terik	Tetun, Portuguese	Naueti	Indonesian	Indonesian	Indonesian	Tetun Terik	Tetun Terik	Tetun, Portuguese

431

ANNEX 3 – ANALYSIS OF LEXICAL BORROWINGS FOR EACH SPEAKER BASED ON INDEPENDENT VARIABLES

Variety	Speakers	All Words Word Totals	Native No.	%	All Borrowings No.	%	Portuguese Borrowings No.	%	English Borrowings No.	%	Indonesian Borrowings No.	%	Malay Borrowings No.	%
TP	tx@NSG	790	573	73%	217	27%	182	23%			32	4%	3	0%
TP	tx@DLB	1407	959	68%	448	32%	422	30%	5	0%	11	1%	10	1%
TP	tx@LAOB	1118	703	63%	415	37%	395	35%			13	1%	7	1%
TP	tx@TAB	346	206	60%	140	40%	122	35%	4	1%	11	3%	4	1%
TP	tx@MAS	1060	808	76%	252	24%	209	20%			30	3%	13	1%
TP	tx@JLS	1543	1009	65%	534	35%	450	29%	8	1%	62	4%	15	1%
TP	tx@BE	237	171	72%	66	28%	57	24%	6	3%	7	3%	2	1%
TP	tx@JBP	1330	1017	76%	313	24%	261	20%	13	1%	25	2%	26	2%
TP	tx@ACB	786	586	75%	200	25%	180	23%	3	0%	10	1%	7	1%
TP	tx@LEG	200	150	75%	50	25%	35	18%	4	2%	10	5%	1	1%
TP	tx@FDP	204	151	74%	53	26%	42	21%	2	1%	4	2%	5	2%
TP	tx@CS	1634	1215	74%	419	26%	360	22%	3	0%	35	2%	21	1%
TP	tx@JMB	286	223	78%	63	22%	55	19%	2	1%	3	1%	3	1%
TS	tx@CSG	447	413	92%	34	8%	14	3%			5	1%	15	3%
TS	tx@NA	499	375	75%	124	25%	105	21%			6	1%	13	3%
TS	tx@JA	157	140	89%	17	11%	10	6%					7	4%
TS	tx@RASS	259	221	85%	38	15%	27	10%	2	1%			9	3%
TS	tx@LLM	156	139	89%	17	11%	12	8%			1	1%	4	3%
TS	tx@ATB	348	255	73%	93	27%	77	22%	5	1%	11	3%	1	0%
TS	tx@OA	440	311	71%	129	29%	129	29%						
TS	tx@MA	149	141	95%	8	5%	5	3%			1	1%	2	1%
TS	tx@GJX	352	215	61%	137	39%	133	38%					4	1%
TS	tx@OC	434	320	74%	114	26%	91	21%			17	4%	6	1%
TS	tx@AGJ	523	394	75%	129	25%	116	22%			4	1%	9	2%
TS	tx@BB	170	136	80%	34	20%	31	18%			2	1%	1	1%
TS	tx@MN	96	75	78%	21	22%	14	15%			1	1%	6	6%
TS	tx@FJC	123	104	85%	19	15%	14	11%			5	4%		
TS	tx@MCA	199	159	80%	40	20%	25	13%			13	7%	2	1%
TV	tx@JCP	519	423	82%	96	18%	90	17%			2	0%	4	1%
TV	tx@MVJ	2448	1553	63%	895	37%	805	33%	5	0%	71	3%	14	1%
TV	tx@AFA	406	383	94%	23	6%	16	4%	1	0%			6	1%
TV	tx@LP	482	412	85%	70	15%	64	13%			5	1%	1	0%
TV	tx@JCV	2990	2552	85%	438	15%	320	11%			43	1%	75	3%
TV	tx@MS	909	798	88%	111	12%	99	11%	1	0%	4	0%	7	1%
TV	tx@AP	175	142	81%	33	19%	26	15%					7	4%

Table 74 Analysis of borrowings from each language per speaker and Tetun variety

Variety	Speakers	Sex		Age				Level of education				Language of instruction			Portuguese		
		Male	Female	Age	1-23	23-47	47+	No Education	Primary	Secondary	Higher Education	Portuguese	Indonesian	Tetun	Speaks	Understands	No understanding
TP	tx@NSG	1	0	51	0	0	1	0	0	0	1	1	1	0	1	0	0
TP	tx@DLB	1	0	45	0	1	0	0	0	0	1	0	1	1	1	0	0
TP	tx@LAOB	1	0	30	0	1	0	0	0	0	1	1	1	0	1	0	0
TP	tx@TAB	1	0	55	0	0	1	0	1	0	0	1	0	0	1	0	0
TP	tx@MAS	1	0	45	0	1	0	0	0	0	1	0	1	0	0	1	0
TP	tx@JLS	1	0	44	0	1	0	0	0	0	1	0	1	1	0	1	0
TP	tx@BE	1	0	30	0	1	0	0	0	0	1	0	1	1	0	1	0
TP	tx@JBP	1	0	26	0	1	0	0	0	0	1	1	1	1	0	1	0
TP	tx@ACB	1	0	21	1	0	0	0	0	0	1	1	1	1	0	1	0
TP	tx@LEG	0	1	20	1	0	0	0	0	1	0	1	0	1	0	1	0
TP	tx@FDP	0	1	19	1	0	0	0	0	1	0	1	0	1	0	1	0
TP	tx@CS	0	1	50	0	0	1	0	1	0	0	1	0	0	0	1	0
TP	tx@JMB	0	1	31	1	0	0	0	0	1	0	0	1	1	0	0	1
TS	tx@CSG	1	0	56	0	0	1	0	0	0	1	1	1	0	1	0	0
TS	tx@NA	1	0	54	0	0	1	0	0	0	1	1	1	1	1	0	0
TS	tx@JA	1	0	49	0	0	1	0	0	0	1	1	1	0	1	0	0
TS	tx@RASS	1	0	49	0	0	1	0	0	0	1	1	0	1	1	0	0
TS	tx@LLM	1	0	41	0	1	0	0	0	0	1	1	1	1	1	0	0
TS	tx@ATB	1	0	41	0	1	0	0	0	0	1	0	1	0	1	0	0
TS	tx@OA	0	1	56	0	0	1	0	0	0	1	1	0	1	1	0	0
TS	tx@MA	0	1	47	0	0	1	0	0	0	1	1	0	1	1	0	0
TS	tx@GJX	0	1	42	0	1	0	0	0	0	1	1	1	1	1	0	0
TS	tx@OC	1	0	31	0	1	0	0	0	1	0	0	1	1	1	0	0
TS	tx@AGJ	1	0	24	0	1	0	0	0	1	0	1	0	1	1	0	0
TS	tx@BB	0	1	54	0	0	1	0	0	1	0	1	0	1	1	0	0
TS	tx@MN	0	1	44	0	1	0	0	0	0	1	0	1	0	0	1	0
TS	tx@FJC	0	1	43	0	1	0	0	0	1	0	0	1	0	0	0	1
TS	tx@MCA	0	1	29	0	1	0	0	0	1	0	0	1	0	0	0	1
TV	tx@JCP	1	0	58	0	0	1	0	0	0	1	1	1	1	1	0	0
TV	tx@MVJ	1	0	51	0	0	1	0	0	0	1	1	1	1	1	0	0
TV	tx@AFA	1	0	50	0	0	1	0	0	0	1	0	1	0	1	0	0
TV	tx@LP	1	0	49	0	0	1	0	0	0	1	0	1	0	1	0	0
TV	tx@JCV	1	0	48	0	0	1	0	0	0	1	0	1	0	0	0	1
TV	tx@MS	1	0	43	0	1	0	0	1	0	0	0	1	0	0	0	1
TV	tx@AP	1	0	68	0	0	1	1	0	0	0	0	0	0	0	0	1

Table 75 Independent variables analyzed against the percentage of borrowings from each language

ANNEX 4 – LIST OF PORTUGUESE BORROWINGS IN TETUN

abandona	v.	afetivu	adj.
abastese	v.	afirma	v.
abastesimentu	n.	afirmasaun	n.
abertura	n.	afrikanu	adj.
abilidade	n.	agora	adv.
abituál	adj.	agostu	n.
abolisaun	n.	agradese	v.
aborda	v.	agradesimentu	n.
abortu	n.	agráriu	adj.
abranjente	adj.	agravadu	adj.
abrigu	n.	agresaun	n.
abríl	n.	agresividade	n.
absolve	v.	agríkola	adj.
absolvisaun	n.	agrikultór	n.
abstensaun	n.	agrikultura	n.
abstratu	adj.	água	n.
abuzu	n.	agudu	adj.
adapta	v.	ajenda	n.
adekuadu	adj.	ajénsia	n.
adere	v.	ajente	n.
adeus	interj.	ajuda	n.
adisionál	adj.	ajuda	v.
adjuntu	adj.	akaba	v.
administradór	n.	akadémia	n.
administrasaun	n.	akadémiku	adj.
administrativu	adj.	akizisaun	n.
admite	v.	aklamasaun	n.
admoestasaun	n.	akomoda	v.
adora	v.	akompaña	v.
adosaun	n.	akompañamentu	n.
adota	v.	akontese	v.
adubu	n.	akontesimentu	n.
advogadu	n.	akórdaun	n.
advokásia	n.	akordu	n.
aéreu	adj.	akreditadu	adj.
aero-	affix	akrexenta	v.
aeronave	n.	akta	n.
aeroportu	n.	aktu	n.
afavór	adv.	akumula	v.
afegaun	adj.	akuza	v.
afeta	v.	akuzasaun	n.
afetadu	adj.	aldeia	n.

alegasaun	n.	animadora	n.
aleijadu	adj.	animál	n.
alemaun	adj.	animizmu	n.
alende	prep.	aniversáriu	n.
alende	conj.	ankorajen	n.
alerta	n.	antagóniku	adj.
alerta	v.	anteriór	adj.
alfabetizasaun	n.	antes	adv.
alfabetu	n.	antesipa	v.
alfándega	n.	antesipadu	adj.
aliansa	n.	antesipasaun	n.
aliás	adv.	antigu	adj.
alimentár	adj.	anti-	affix
alimentasaun	n.	anu	n.
alkansa	v.	anuál	adj.
alkanse	n.	anula	v.
alma	n.	anunsia	v.
almosa	v.	anúnsiu	n.
almosu	n.	apá	n.
alojamentu	n.	aparellu	n.
aloka	v.	aparénsia	n.
altera	v.	apartamentu	n.
alterasaun	n.	apela	v.
altu	adj.	apelu	n.
altura	n.	aperta	v.
aluga	v.	aplika	v.
alunu	n.	aplikasaun	n.
ambientál	adj.	apoia	v.
ambiente	n.	apoiante	n.
ambisaun	n.	apoiu	n.
ambisiozu	adj.	aposta	v.
ámbitu	n.	aprende	v.
ameasa	n.	aprendizajen	n.
ameasa	v.	apresia	v.
amerikanu	adj.	apresiasaun	n.
amigu	n.	aprezenta	v.
amizade	n.	aprezentasaun	n.
amu	n.	aprofunda	v.
amu-bispu	n.	apropriasaun	n.
amu-papa	n.	aprosima	v.
analiza	v.	aprosimasaun	n.
análize	n.	aprova	v.
andár	n.	aprovadu	adj.
angolanu	adj.	aprovasaun	n.

aproveita	v.	asina	v.
aprovizionamentu	n.	asinadu	adj.
arbitrajen	n.	asinante	n.
área	n.	asinatura	n.
areia	n.	asinke	conj.
arguidu	n.	asiste	v.
arguidu	adj.	asisténsia	n.
argumenta	v.	asistente	n.
argumentu	n.	asosiasaun	n.
arkipélagu	n.	aspetu	n.
arkitetóniku	adj.	aspirasaun	n.
arkiva	v.	asume	v.
arkivu	n.	asumidu	adj.
arma	n.	asuntu	n.
armadu	adj.	asvezes	adv.
armazén	n.	atake	n.
armonia	n.	até	prep.
armonizasaun	n.	até	adv.
arogánsia	n.	atende	v.
arranka	v.	atendimentu	n.
arrasta	v.	atensaun	n.
arrendamentu	n.	atentadu	n.
arrepende	v.	atinje	v.
arte	n.	atitude	n.
artesanatu	n.	ativa	v.
artezaun	n.	atividade	n.
artifisiál	adj.	ativista	n.
artigu	n.	ativu	adj.
artikulasaun	n.	atlántiku	adj.
artista	n.	atlas	n.
asaltu	n.	atletizmu	n.
asaun	n.	atrai	v.
asegura	v.	através	prep.
aseita	v.	atravesa	v.
aseitasaun	n.	atraza	v.
asembleia	n.	atrazu	n.
asentu	n.	atribui	v.
asentuasaun	n.	atribuidu	adj.
aserta	v.	atua	v.
asesór	n.	atuál	adj.
asesoria	n.	atualiza	v.
asesóriu	n.	atualizasaun	n.
asesu	n.	atuasaun	n.
asidente	n.	audiénsia	n.

auditóriu	n.	bengala	n.
aula	n.	bensaun	n.
aumenta	v.	benvindu	interj.
aumentu	n.	berlinde	n.
ausiliár	adj.	BI	n.
australianu	adj.	biblioteka	n.
autodeterminasaun	n.	bilaterál	adj.
autojestaun	n.	billete	n.
automatizasaun	n.	biogás	n.
autónomu	adj.	biolójiku	adj.
autór	n.	birmanés	adj.
autoridade	n.	bispu	n.
autorizasaun	n.	bloku	n.
avalia	v.	boa	adj.
avaliasaun	n.	boas-vindas	n.
avansa	v.	bola	n.
avansadu	adj.	bolsa	n.
avaria	n.	bolsu	n.
avenida	n.	bondia	interj.
aviasaun	n.	bonoite	interj.
aviaun	n.	botarde	interj.
aviza	v.	brankeamentu	n.
avizu	n.	brasu	n.
avó	n.	brazileiru	adj.
aziátiku	adj.	brigadeiru	n.
azilu	n.	britániku	adj.
bairru	n.	broxura	n.
baixa	adj.	brutál	adj.
balansu	n.	brutalidade	n.
baliza	n.	da	prep.
bankada	n.	dadu	n.
bankáriu	adj.	dama	n.
banku	n.	danu	n.
barríl	n.	data	n.
baxarelatu	n.	de	prep.
baze	n.	debate	n.
bazeadu	adj.	debate	v.
bazeia	v.	defaktu	adv.
báziku	adj.	defende	v.
bebé	n.	defensór	n.
bein-estár	n.	defensoria	n.
benefisia	v.	defeza	n.
benefisiáriu	n.	define	v.
benefísiu	n.	definisaun	n.

definitivu	adj.	despaxa	v.
dékada	n.	despaxu	n.
deklara	v.	despeza	n.
deklarasaun	n.	desportu	n.
dekretu	n.	destaka	v.
dekretu-lei	n.	destinu	n.
delega	v.	destrosu	adj.
delegadu	n.	destrui	v.
delegasaun	n.	destruidór	n.
deliberadu	adj.	desvantajen	n.
delimitasaun	n.	detalle	n.
demais	adv.	detein	v.
demarkasaun	n.	determina	v.
demokrasia	n.	determinadu	adj.
demokrata	n.	determinasaun	n.
demokrátiku	adj.	detesaun	n.
demonstrasaun	n.	deteta	v.
demora	v.	devagár	adv.
departamentu	n.	deve	v.
depende	v.	devér	n.
dependénsia	n.	devolve	v.
dependente	adj.	dezafiu	n.
depois	adv.	dezastre	n.
depois	conj.	dezde	prep.
depozita	v.	dezde	conj.
deputada	n.	dezembarku	n.
deputadu	n.	dezembru	n.
derrepente	adv.	dezempeñu	n.
dés	num.	dezempregu	n.
deseñadu	adj.	dezena	n.
desentralizasaun	n.	dezeña	v.
deside	v.	dezenkorajadór	adj.
desizaun	n.	dezentendimentu	n.
desizivu	adj.	dezeñu	n.
deskansa	v.	dezenvolve	v.
deskoberta	n.	dezenvolvidu	adj.
deskobre	v.	dezenvolvimentu	n.
deskola	v.	dezigualdade	n.
deskoñesidu	adj.	dezisténsia	n.
deskonfia	v.	dezliga	v.
deskontu	n.	dezloka	v.
deskulpa	n.	dezlokasaun	n.
deskulpa	v.	dezmobiliza	v.
desmobilizasaun	n.	dezmobilizasaun	n.

dezvantajen	n.	distritu	n.
dezvia	v.	distúrbiu	n.
dezviu	n.	diversifika	v.
dia	n.	divertimentu	n.
dialetu	n.	divizaun	n.
diálogu	n.	divulgasaun	n.
diáriu	n.	dixiplina	n.
diferensa	n.	dixiplinár	adj.
diferente	adj.	doadór	n.
difikuldade	n.	dobru	n.
difikulta	v.	dois	num.
difisil	adj.	dokumentu	n.
dignidade	n.	dolar	n.
dignu	adj.	doméstiku	adj.
dimensaun	n.	domina	v.
dinámiku	adj.	dominante	adj.
dinamiza	v.	domingu	n.
dioseze	n.	domíniu	n.
diploma	n.	don	n.
diplomasia	n.	dosente	n.
diplomata	n.	dosié	n.
diplomátiku	adj.	dotrina	n.
direitu	adj.	doutór	n.
direitu	n.	doze	num.
diresaun	n.	drama	n.
diretiva	n.	droga	n.
diretór	n.	du	prep.
diretóriu	n.	duké	conj.
dirije	v.	duplika	v.
dirijente	n.	dura	v.
disemina	v.	duradoru	adj.
diseminasaun	n.	durante	prep.
disionáriu	n.	dúvida	n.
diskursu	n.	edifísiu	n.
diskusaun	n.	edisaun	n.
diskute	v.	edita	v.
disponibilidade	n.	eduka	v.
disponibiliza	v.	edukasaun	n.
disponivel	adj.	edukasionál	adj.
disputa	n.	efetivu	adj.
distánsia	n.	efikás	adj.
distintivu	adj.	efikásia	n.
distribuisaun	n.	efisiente	adj.
distritál	adj.	egu	n.

eis-	affix	enfatiza	v.
ekilibradu	adj.	enfrenta	v.
ekilíbriu	n.	enklave	n.
ekipa	n.	enkontru	n.
ekipamentu	n.	enkuantu	conj.
ekolójiku	adj.	enrelasaun	adv.
ekonomia	n.	enseramentu	n.
ekonómiku	adj.	ensinu	n.
éktare	n.	entaun	adv.
elabora	v.	entende	v.
elaborasaun	n.	entermus (de)	conj.
eleisaun	n.	entidade	n.
eleitorál	adj.	entrada	n.
eleitu	adj.	entre	prep.
elementu	n.	entrega	n.
elétriku	adj.	entrega	v.
eletrisidade	n.	entretantu	adv.
elojia	v.	entrevista	n.
ematolojia	n.	entuziazmu	n.
embaixada	n.	enviadu	adj.
embaixadór	n.	envigór	adv.
embaixadora	n.	envolve	v.
embargu	n.	envolvimentu	n.
embarkasaun	n.	episentru	n.
emenda	v.	episkopál	adj.
emerjénsia	n.	erói	n.
emisáriu	n.	esbosu	n.
emisór	n.	esensiál	adj.
emite	v.	esforsa	v.
emosaun	n.	esforsu	n.
emosionál	adj.	eskala	n.
empata	v.	eskalaun	n.
empeñadu	adj.	esklarese	v.
empeñu	n.	eskola	n.
emprega	v.	eskola	v.
empregu	n.	eskolár	adj.
empresta	v.	eskrita	n.
empréstimu	n.	eskritu	adj.
empreza	n.	eskuadra	n.
emprezáriu	n.	eskudu	n
enaltese	v.	espalla	v.
enderesu	n.	españól	adj.
enerjétiku	adj.	espasiál	adj.
enerjia	n.	espasu	n.

440

espekulasaun	n.	estimula	v.
espekulativu	adj.	estímulu	n.
espera	v.	estinsaun	n.
esperansa	n.	estrada	n.
esperiénsia	n.	estraga	v.
espertu	adj.	estragu	n.
espesiál	adj.	estranjeiru	n.
espesialidade	n.	estranjeiru	adj.
espesialista	n.	estraordináriu	adj.
espesializadu	adj.	estrasaun	n.
espesialmente	adv.	estratéjia	n.
espesifikamente	adv.	estratéjiku	adj.
espesifikasaun	n.	estrativu	adj.
espesífiku	adj.	estreitu	adj.
espirasaun	n.	estremista	n.
espíritu	n.	estremista	adj.
esplika	v.	estremu	n.
esplikasaun	n.	estrofe	n.
esplísitu	adj.	estrutura	n.
esplora	v.	estruturál	adj.
esplorasaun	n.	estuda	v.
espoza	n.	estudante	n.
espozisaun	n.	estudu	n.
espresa	v.	etapa	n.
espresaun	n.	étika	n.
estabelese	v.	etnia	n.
estabelesimentu	n.	europeu	adj.
estabilidade	n.	euru	n.
estabiliza	v.	evakua	v.
estabilizadór	n.	eventu	n.
estabilizasaun	n.	evidénsia	n.
estadu	n.	evita	v.
estadu-membru	n.	evolusaun	n.
estájiu	n.	exelénsia	n.
estandarte	n.	exesu	n.
estasaun	n.	ezame	n.
estatál	adj.	ezamina	v.
estatístika	n.	ezaminasaun	n.
estátua	n.	ezekusaun	n.
estatutu	n.	ezekuta	v.
estavel	adj.	ezekutivu	adj.
estensaun	n.	ezekutivu	n.
esteriór	n.	ezemplu	n.
esternu	adj.	ezerse	v.

441

ezérsitu	n.	finál	n.
ezije	v.	finaliza	v.
ezijénsia	n.	finansa	n.
eziste	v.	finansas	n.
ezisténsia	n.	finanseiru	adj.
ezonera	v.	finansia	v.
ezonerasaun	n.	firme	adj.
faktu	n.	fiskál	adj.
falla	v.	fiskalizasaun	n.
fallansu	n.	fiu	n.
falsifika	v.	fizikamente	adv.
falsifikasaun	n.	fíziku	adj.
falsu	adj.	fogete	n.
falta	n.	foka	v.
falta	v.	fokál	adj.
família	n.	fonte	n.
familiár	adj.	forma	n.
farmásia	n.	forma	v.
fásil	adj.	formál	adj.
fasilidade	n.	formalidade	n.
fasilita	v.	formaliza	v.
fatór	n.	formalmente	adv.
favorese	v.	formandu	n.
faze	n.	formasaun	n.
federál	adj.	fórmula	n.
feira	n.	formuláriu	n.
felisita	v.	fornesimentu	n.
fenómenu	n.	forsa	n.
feriadu	n.	fortalesimentu	n.
ferramenta	n.	fortaleza	n.
fertilizante	n.	forte	adj.
festa	n.	fortifika	v.
fevereiru	n.	forum	n.
fiar	v.	foto	n.
fibra	n.	fotografia	n.
figura	n.	fragmentu	n.
fiksa	v.	frajil	adj.
fiksu	adj.	frajilidade	n.
filla	n.	frakeza	n.
fillu	n.	fraku	adj.
filme	n.	franka	adj.
filozofia	n.	fransés	adj.
finadu	n.	fraude	n.
finál	adj.	frei	n.

442

frente	n.	grava	v.
friza	v.	grave	adj.
fronteira	n.	grávida	adj.
fuma	v.	gravidade	n.
fundadór	n.	gregu	adj.
fundamentál	adj.	grupu	n.
fundamentu	n.	guarda	n.
fundasaun	n.	HIV-SIDA	n.
fundu	n.	i	conj.
funerál	n.	íbridu	adj.
funsaun	n.	idade	n.
funsiona	v.	idealizmu	n.
funsionamentu	n.	ideia	n.
funsionáriu	n.	identidade	n.
futuru	adj.	identifika	v.
futuru	n.	identifikadu	adj.
gabinete	n.	identifikasaun	n.
galardoadu	adj.	ideolojia	n.
garante	v.	ignora	v.
garantia	n.	igreja	n.
gás	n.	iguál	adj.
gasta	v.	igualdade	n.
gazóleu	n.	ilegál	adj.
gerra	n.	ilegalmente	adv.
gerrilla	n.	illa	n.
gerrilleiru	n.	imajen	n.
gia	n.	imediatu	adj.
gineense	adj.	imigrasaun	n.
globál	adj.	imobiliáriu	n.
golpe	n.	impaktu	n.
gosta	v.	impase	n.
governa	v.	impede	v.
governadór	n.	impedimentu	n.
governadór-jerál	n.	imperadór	n.
governamentál	adj.	imperatríz	n.
governante	n.	imperiál	adj.
governasaun	n.	imperializmu	n.
governativu	adj.	implementa	v.
governu	n.	implementadór	n.
goza	v.	implementasaun	n.
graduasaun	n.	implika	v.
gramátika	n.	implikasaun	n.
grande	adj.	implísitu	adj.
gratuitu	adj.	importánsia	n.

importante	adj.	inkonstitusionál	adj.
importasaun	n.	inkrementu	n.
imposível	adj.	inkumprimentu	n.
impostu	n.	inovasaun	n.
imprensa	n.	insensu	n.
imprevizibilidade	n.	insidente	n.
imprevizivel	adj.	insiste	v.
imputa	v.	insituisaun	n.
inaugurasaun	n.	inspesaun	n.
indemnizasaun	n.	inspesiona	v.
independénsia	n.	inspetór	n.
independente	adj.	instabilidade	n.
indignasaun	n.	instala	v.
indíjena	adj.	instaladór	n.
indikadór	n.	instalasaun	n.
indikasaun	n.	instánsia	n.
índiku	adj.	instituisaun	n.
indireitu	adj.	institusionál	adj.
indispensavel	adj.	institutu	n.
individuál	adj.	instrumentu	n.
individualmente	adv.	insulta	v.
indivíduu	m.	integra	v.
indonéziu	adj.	integradu	adj.
indultu	n.	integrál	adj.
indústria	n.	integralmente	adv.
industriál	adj.	integrasaun	n.
infánsia	n.	integrativu	adj.
infesaun	n.	integridade	n.
influénsia	n.	íntegru	adj.
influensiadu	adj.	intelektuál	adj-
informa	v.	intelijénsia	n.
informasaun	n.	intensaun	n.
informativu	adj.	intensivu	adj.
infraestrutura	n.	intensu	adj.
inglés	adj.	inter-	affix
inglés	n.	interasaun	n.
inimigu	n.	interesante	adj.
inisia	v.	interese	n.
inisiál	adj.	interinu	adj.
inisiativa	n.	interiór	n.
inísiu	n.	interkámbiu	n.
injustisa	n.	interligasaun	n.
inklui	v.	intermédiu	adj.
inkluidu	adj.	internasionál	adj.

internasionalizasaun	n.	jestaun	n.
internasionalmente	adv.	jezuita	n.
internet	n.	Jezus	n.
internu	adj.	jihadista	n.
interpreta	v.	joga	v.
interpretasaun	n.	jornál	n.
interrompe	v.	jornalista	n.
intervensaun	n.	joven	n.
introdús	v.	joven	adj.
introdusaun	n.	judisiáriu	adj.
intruzaun	n.	juís	n.
inu	n.	juiza	n.
inundasaun	n.	julga	v.
invade	v.	julgamentu	n.
invazaun	n.	jullu	n.
inventa	v.	junta	n.
investe	v.	juñu	n.
investidór	n.	jura	v.
investiga	v.	juramentu	n.
investigadór	n.	jurídika	adj.
investigasaun	n.	jurisdisaun	n.
investimentu	n.	jurista	n.
inxadu	adj.	justifika	v.
irakianu	adj.	justifikasaun	n.
irmán	n.	justisa	n.
irmaun	n.	justu	adj.
irresponsabilidade	n.	juventude	n.
irrigasaun	n.	kabe	v.
istória	n.	kada	pron.
istóriku	adj.	kadeia	n.
italianu	adj.	kadeira	n.
izlámiku	adj.	kadete	n.
izoladu	adj.	kaduka	v.
janeiru	n.	kafé	n.
japonés	adj.	kaixa	n.
jenerál	n.	kaixaun	n.
jenerozu	adj.	kalendáriu	n.
jéneru	n.	kalifadu	n.
jeolojia	n.	kalmu	adj.
jeolójiku	adj.	kalsa	n.
jeradór	n.	kama	n.
jerál	adj.	kámara	n.
jerasaun	n.	kamaradajen	n.
jere	v.	kamioneta	n.

kamiza	n.
kampaña	n.
kampu	n.
kanál	n.
kandidatu	n.
kandidatura	n.
kanta	v.
kapa	n.
kapás	adj.
kapasidade	n.
kapasita	v.
kapasitasaun	n.
kapela	n.
kapitál	n.
kapitál	adj.
kapitalizmu	n.
kapítulu	n.
kapta	v.
kaptura	n.
kapturasaun	n.
karater	n.
karaterístika	n.
kardiolojia	n.
kargu	n.
karismátiku	adj.
karpinteiru	n.
karreira	n.
karreta	n.
karta	n.
kartás	n.
kartaun	n.
kasa	v.
kastigu	n.
katana	n.
katástrofe	n.
katekista	n.
katóliku	adj.
katorze	num.
kautelár	adj.
kauza	n.
kaza	n.
kazu	n.
ke	conj.
keixa	n.

kerdizér	
kestaun	n.
kestiona	v.
kilómetru	n.
kilu	n.
kímiku	adj.
kiñentus	num.
kinta	n.
kinta-feira	n.
kintu	num.
kinze	num.
klandestinu	adj.
klaru	adj.
klaru	adv.
klase	n.
klasifika	v.
kliente	n.
klínika	n.
kobertura	n.
kobre	n.
kobre	v.
kódigu	n.
koezaun	n.
kolabora	v.
kolaborasaun	n.
kolár	n.
kolega	n.
koletivu	n.
koletivu	adj.
koligasaun	n.
kolleita	n.
koloka	v.
kom	prep.
koma	n.
komanda	v.
komandante	n.
komandu	n.
komarka	n.
kombate	v.
kombate	n.
kombatente	n.
komboiu	n.
komemora	v.
komemorasaun	n.

446

komenta	v.	kondenasaun	n.
komentáriu	n.	kondisaun	n.
komérsiu	n.	kondisiona	v.
komesa	v.	kondús	v.
komete	v.	kondusaun	n.
kometimentu	n.	koñese	v.
komisariadu	n.	koñesidu	adj.
komisáriu	n.	koñesimentu	n.
komisáriu-jerál	n.	konferénsia	n.
komisaun	n.	konferensista	n.
komité	n.	konfesa	v.
komitiva	n.	konfiansa	n.
kompañeiru	n.	konfidensiál	adj.
kompañia	n.	konfirma	v.
kompara	v.	konfirmasaun	n.
komparasaun	n.	konflitu	n.
komparativu	adj.	konforme	adj.
kompatriota	n.	konforme	prep.
kompensa	v.	konfortu	n.
kompeténsia	n.	konfronta	v.
kompetente	adj.	konfrontasaun	n.
kompleksu	adj.	konfuzaun	n.
kompletu	adj.	kongratula	v.
komplikadu	adj.	kongregasaun	n.
kompoin	v.	kongresu	n.
komponente	n.	konjénitu	adj.
komportamentu	n.	konjugál	adj.
kompostu	adj.	konjuntu	n.
komprende	v.	konkista	n.
komprimentu	n.	konkista	v.
kompromete	v.	konklui	v.
komprometidu	adj.	konkluzaun	n.
kompromisu	n.	konkorda	v.
komputadór	n.	konkretiza	v.
komu	conj.	konkretizasaun	n.
komún	adj.	konkretu	adj.
komunidade	n.	konkursu	n.
komunikadu	n.	konsagra	v.
komunikasaun	n.	konsede	v.
komunitáriu	adj.	konsege	v.
kondekora	v.	konseitu	n.
kondekorasaun	n.	konsekuénsia	n.
kondena	v.	konsellu	n.
kondenadu	adj.	konsensu	n

447

konsentra	v.	kontrola	v.
konsentrasaun	n.	kontrolu	n.
konsertasaun	n.	konvensaun	n.
konserteza	adv.	konvense	v.
konservadór	adj.	konverjénsia	n.
konsidera	v.	konversa	n.
konsideradu	adj.	konversa	v.
konsiderasaun	n.	konversasaun	n.
konsolida	v.	konversaun	n.
konsolidasaun	n.	konvida	v.
konstituisaun	n.	konvidadu	n.
konstitusionál	adj.	konviksaun	n.
konstitusionalidade	n.	konvite	n.
konstrusaun	n.	konxiénsia	n.
konstrutivu	adj.	konxientemente	adv.
konsuladu	n.	koopera	v.
konsulta	n.	kooperasaun	n.
konsulta	v.	kooperativa	n.
konsultasaun	n.	koordena	v.
konsultivu	adj.	koordenadór	n.
konsultór	n.	koordenasaun	n.
konsumadu	adj.	koorganizadór	n.
konsume	v.	koorganizasaun	n.
konsumu	n.	kopia	v.
konta	n.	kopu	n.
konta	v.	kór	n.
kontajiozu	adj.	korajen	n.
kontakta	v.	korasaun	n.
kontaktu	n.	koronél	n.
kontensiozu	adj.	korporál	adj.
kontente	adj.	korpu	n.
kontentór	n.	korresponde	v.
kontestu	n.	korrije	v.
konteúdu	n.	korrupsaun	n.
kontinentál	adj.	kortezia	n.
kontinjénsia	n.	kosta	n.
kontinua	v.	kostume	n.
kontinuasaun	n.	kotasaun	n.
kontra	prep.	kredibilidade	n.
kontra-golpe	n.	kréditu	n.
kontratu	n.	kreximentu	n.
kontribui	v.	kria	v.
kontribuisaun	n.	kriasaun	n.
kontributu	n.	kriatura	n.

krime	n.		kursu	n.
kriminál	adj.		kurtu	adj.
kriminozu	adj.		kusta	v.
kriolizasaun	n.		kustu	n.
kristus	n.		kuziñeiru	n.
kritériu	n.		laboratóriu	n.
kritika	v.		lamenta	v.
krítika	n.		lamentavel	adj.
krítiku	n.		lansa	v.
krize	n.		lansamentu	n.
kruél	adj.		largu	n.
krús	n.		lasu	n.
kuadradu	adj.		lee	v.
kuadru	n.		legál	adj.
kualidade	n.		legalidade	n.
kualifikasaun	n.		legalista	n.
kualkér	pron.		legaliza	v.
kuandu	conj.		lei	n.
kuantidade	n.		leitór	n.
kuantu	conj.		leitura	n.
kuarenta	num.		lejizladór	n.
kuarezma	n.		lejizlativu	adj.
kuarta	n.		lejizlatura	n.
kuarta-feira	n.		lema	n.
kuartu	n.		lembra	v.
kuartu	num.		leste	adj.
kuatru	num.		levantamentu	n.
kuatrusentus	num.		lezadu	n.
kuaze	adv.		liberdade	n.
kubanu	adj.		liberta	v.
kuidadu	n.		libertasaun	n.
kulpa	n.		lidera	v.
kulpadu	adj.		lideradu	adj.
kultu	n.		lideransa	n.
kultura	n.		liga	v.
kulturál	adj.		liga	n.
kulturalizmu	n.		ligadu	adj.
kulturalmente	adv.		ligasaun	n.
kumpre	v.		limitadu	adj.
kumprimenta	v.		limitasaun	n.
kumprimentu	n.		limite	n.
kuorum	n.		liña	n.
kurrikulár	adj.		língua	n.
kurríkulu	n.		linguajen	n.

linguista	n.	manganés	n.
linguística	n.	manifestasaun	n.
linguistikamente	adj.	manobra	n.
lisaun	n.	mantén	v.
lisensiar	v.	manuál	n.
lisensiatura	n.	maremotu	n.
lista	n.	mariña	n.
livre	adj.	mariñu	adj.
livru	n.	marítimu	adj.
lixa	n.	marka	v.
lojikamente	adv.	markasaun	n.
lojístiku	adj.	marsiál	adj.
lokál	n.	marsu	n.
lokál	adj.	martir	n.
lokaliza	v.	marxa	n.
longu	adj.	mas	conj.
lugár	n.	másimu	adj.
lukru	n.	matabixu	n.
luta	n.	matemátika	n.
luta	v.	matéria	n.
lutadór	n.	materiál	n.
madre	n.	materna	adj.
magnitude	n.	maturidade	n.
maiór	n.	mau	adj.
maioria	n.	medalla	n.
maioritáriu	adj.	média	n.
mais	conj.	medianu	adj.
mais	adv.	mediasaun	n.
maiu	n.	medida	n.
maizumenus	adv.	medikamentu	n.
majór	n.	médiku	n.
mákina	n.	mediterráneu	adj.
makroekonómiku	adj.	médiu	adj.
mal	n.	meiu	n.
maladministrasaun	n.	meiudia	n.
malaiu	adj.	mekanizmu	n.
malaiu	n.	membru	n.
malandru	n.	memória	n.
malformasaun	n.	menina	n.
mamá	n.	menór	adj.
mana	n.	mensajen	n.
manda	v.	mentál	adj.
mandatu	n.	menus	adv.
maneira	n.	merese	v.

méritu	n.		**mistura**	n.
merkadoria	n.		**mizerikórdia**	n.
merkadu	n.		**mobiliár**	n.
més	n.		**mobilizasaun**	n.
mestra	n.		**moda**	n.
mestradu	n.		**modelu**	n.
mestre	n.		**moderadu**	adj.
meta	n.		**moderniza**	v.
metade	n.		**modernizasaun**	n.
metáliku	adj.		**modernu**	adj.
mete	v.		**modestu**	adj.
métodu	n.		**moeda**	n.
metru	n.		**momentu**	n.
meza	n.		**monitorizasaun**	n.
mezmu	conj.		**montante**	n.
mezmu	pron.		**morál**	adj.
migrante	n.		**morál**	n.
migrasaun	n.		**moratória**	n.
migratóriu	adj.		**morte**	n.
mikroempreza	n.		**mostra**	v.
mikrofinansas	n.		**motivasaun**	n.
mikrosefália	n.		**motivu**	n.
míl	num.		**motór**	n.
miléniu	n.		**motorizada**	n.
milísia	n.		**movimentasaun**	n.
militante	n.		**movimentu**	n.
militár	adj.		**muda**	v.
militár	n.		**mudansa**	n.
militariza	v.		**mullér**	n.
militarizasaun	n.		**multa**	n.
millaun	n.		**multikulturál**	adj.
minerál	n.		**multilaterál**	adj.
minerál	adj.		**multilateralizmu**	n.
mínimu	adj.		**multimédia**	n.
ministériu	n.		**multinasionál**	adj.
ministra	n.		**multisetoriál**	adj.
ministru	n.		**mundiál**	adj.
minoridade	n.		**mundu**	n.
minutu	n.		**munisipál**	adj.
misa	n.		**munisipalidade**	n.
misaun	n.		**munisípiu**	n.
misil	n.		**mutuál**	adj.
misionáriu	n.		**nada**	pron.
misteriozu	adj.		**namora**	v.

nasaun	n.	nove	num.
nasionál	adj.	novembru	n.
nasionalidade	n.	noventa	num.
nasionalizmu	n.	novesentus	num.
naturál	adj.	novu	adj.
natureza	n.	nukazu	conj.
naufrájiu	n.	nukleár	adj.
naun-	affix	numerál	n.
navál	adj.	número	n.
navegasaun	n.	nunka	adv.
nega	v.	obedese	v.
negativu	adj.	objetivu	n.
neglijénsia	n.	objetu	n.
negosia	v.	obra	n.
negosiasaun	n.	obriga	v.
negósiu	n.	obrigadu	interj.
negru	adj.	obrigasaun	n.
neozelandés	adj.	obrigatoriamente	adv.
nervozu	adj.	obrigatóriu	adj.
nesesáriu	adj.	observa	v.
nesesidade	n.	observadór	n.
neurolójiku	adj.	observasaun	n.
nivel	n.	observatóriu	n.
nome	n.	obstákulu	n.
nomeadu	adj.	ódiu	n.
nomeasaun	n.	oeste	n.
nomeia	v.	ofende	v.
nonu	num.	ofensa	n.
nordeste	adj.	oferese	v.
norma	n.	oferta	n.
normál	adj.	ofisiál	n.
normaliza	v.	ofisiál	adj.
norte	n.	ofisialmente	adv.
norte-amerikanu	adj.	ofísiu	n.
norueste	n.	oitavu	num.
nosaun	n.	oitu	num.
nota	n.	okaziaun	n.
nota	v.	okupa	v.
notavel	adj.	okupasaun	n.
notifika	v.	olandés	adj.
notifikasaun	n.	omenajen	n.
notísia	n.	omisídiu	n.
notisiozu	adj.	omologa	v.
noturnu	adj.	onestidade	n.

452

onestu	adj.	ospitál	n.
ONG	n.	otél	n.
onra	n.	otelaria	n.
ONU	n.	ótiku	adj.
onze	num.	otimista	adj.
opera	v.	ou	conj.
operasaun	n.	ouru	n.
operasionál	adj.	outubru	n.
opiniaun	n.	padraun	n.
oportunidade	n.	padre	n.
oportunu	adj.	padronizadu	adj.
opozisaun	n.	pagamentu	n.
opsaun	n.	pai	n.
oradór	n.	painél	n.
oralidade	n.	país	n.
oráriu	n.	pájina	n.
oras	n.	pakote	n.
orasaun	n.	palásiu	n.
orden	n.	palavra	n.
ordena	v.	palku	n.
ordenasaun	n.	pániku	n.
orgániku	adj.	pankada	n.
organiza	v.	papá	n.
organizadór	adj.	papél	n.
organizadu	adj.	para	v.
organizasaun	n.	para	conj.
órgaun	n.	paradigma	n.
orgullozu	adj.	parede	n.
orgullu	n.	parese	v.
orienta	v.	paresér	n.
orientál	adj.	parlamentár	adj.
orientasaun	n.	parlamentáriu	adj.
orijen	n.	parlamentu	n.
orijinál	adj.	parókia	n.
orijinalidade	n.	parseiru	n.
orizontál	adj.	parselár	adj.
orsamentál	adj.	parsialmente	adv.
orsamentu	n.	parte	n.
ortikultura	n.	partidu	n.
ortografia	n.	partikulár	adj.
ortopédiku	adj.	partilladu	adj.
oseániku	adj.	partisipa	v.
oseanu	n.	partisipante	n.
osidentál	adj.	partisipasaun	n.

partu	n.		peskiza	n.
pás	n.		peskizadór	n.
pasa	v.		pesoa	n.
pasadu	n.		pesoál	adj.
pasadu	adj.		pesoál	n.
pasajeiru	n.		pesoalmente	adv.
pasaporte	n.		petisaun	n.
pasiénsia	n.		petróleu	n.
pasiente	n.		petrolíferu	adj.
pasifikasaun	n.		petrolojia	n.
pasífiku	adj.		pezadu	adj.
páskua	n.		pilár	n.
pasu	n.		pilotu	adj.
paternál	adj.		pinta	v.
pátria	n.		piór	adj.
patrimóniu	n.		pista	n.
patrulla	n.		plaka	n.
pavillaun	n.		planeadu	adj.
pedagójiku	adj.		planeamentu	n.
pedasu	n.		planeia	v.
pedidu	n.		planifikasaun	n.
pedreiru	n.		planu	n.
pekuária	n.		plataforma	n.
pekulatu	n.		plenária	n.
pelumenus	adv.		plenária	adj.
pena	n.		podér	n.
penál	adj.		poema	n.
pendente	adj.		pois	conj.
pensaun	n.		pois	adv.
perdua	v.		polémika	n.
perfíl	n.		polísia	n.
pergunta	n.		polisiál	adj.
perigozu	adj.		polisiamentu	n.
perigu	n.		política	n.
período	n.		polítiku	adj.
peritu	n.		politiza	v.
perkursu	n.		pontu	n.
permanente	adj.		populár	adj.
permite	v.		populasaun	n.
persegisaun	n.		porezemplu	
persentajen	n.		porfavór	interj.
pertense	v.		porsentu	n.
pesa	n.		portál	n.
peska	n.		portantu	conj.

portavós	n.	presiza	v.
portfólio	n.	presizamente	adv.
portu	n.	presta	v.
portugés	adj.	prestasaun	n.
portugés	n.	prestíjiu	n.
pose	n.	presu	n.
posibilidade	n.	prevalese	v.
posibilita	v.	prevee	v.
pós-	affix	prevene	v.
postu	n.	prevensaun	n.
posu	n.	preventivu	adj.
poténsia	n.	previstu	adj.
potensiál	n.	previzaun	n.
potensiál	adj.	prezensa	n.
potensialidade	n.	prezente	n.
poupansa	n.	prezente	adj.
povu	n.	prezide	v.
pozisaun	n.	prezidénsia	n.
pozitivu	adj.	prezidensiál	adj.
pragmátiku	adj.	prezidente	n.
prasa	n.	prienxe	v.
prátika	n.	prima	n.
prátiku	adj.	primáriu	adj.
prazér	n.	primeiru	num.
prazu	n.	primu	n.
predesesór	n.	prinsipál	adj.
prefere	v.	prinsipalmente	adv.
prega	v.	prinsípiu	n.
prejudika	v.	prioridade	n.
prejuizu	n.	prioritáriu	adj.
prekondisaun	n.	privadu	adj.
preliminár	adj.	priviléjiu	n.
prematuru	adj.	prizaun	n.
premiadu	adj.	prizioneiru	n.
prémiu	n.	pró	n.
prende	v.	problema	n.
preokupa	v.	prodús	v.
preokupante	adj.	produsaun	n.
preokupasaun	n.	produtivu	adj.
prepara	v.	produtu	n.
preparadu	adj.	profesór	n.
preparasaun	n.	profesora	n.
pré-	affix	profisaun	n.
presaun	n.	profisionál	adj.

profundidade	n.	protokolu	n.
profundu	adj.	prova	n.
programa	n.	provadu	adj.
programátiku	adj.	provedór	n.
progresaun	n.	provedoria	n.
progresu	n.	providénsia	n.
proibe	v.	provínsia	n.
projesaun	n.	provizór	n.
projetu	n.	provizóriu	adj.
proklama	v.	provoka	v.
prokuradór	n.	provokadór	adj.
prokuradora	n.	psikotékniku	adj.
prolonga	v.	publika	v.
promesa	n.	publikasaun	n.
promete	v.	públiku	n.
promosaun	n.	públiku	adj.
promosionál	adj.	pulmaun	n.
promotór	n.	punidu	adj.
promove	v.	pur	prep.
promovidu	adj.	radar	n.
promulga	v.	radikál	adj.
promulgadu	adj.	raiña	n.
prontu	adj.	ramu	n.
prontu	interj.	ranxu	n.
pronunsia	v.	rasionál	adj.
propoin	v.	ratifika	v.
proponente	n.	ratifikativu	adj.
proposta	n.	razaun	n.
propriedade	n.	reabilitasaun	n.
própriu	adj.	reál	adj.
prorrogasaun	n.	realidade	n.
prosedimentu	n.	realiza	v.
prosesa	v.	realizasaun	n.
prosesu	n.	realsa	v.
prosesuál	adj.	reasaun	n.
prosimidade	n.	redasaun	n.
prósimu	adj.	rede	n.
prosperidade	n.	redús	v.
prósperu	adj.	reeleisaun	n.
proteje	v.	reentrada	n.
protejidu	adj.	referál	adj.
protesaun	n.	refere	v.
protesta	v.	referendu	n.
protestu	n.	referénsia	n.

456

reflesaun	n.	relasionamentu	n.
reflete	v.	relativamente	adv.
reforma	n.	relatór	n.
reforsa	v.	relatóriu	n.
refujiadu	n.	relevante	adj.
regra	n.	relijiaun	n.
regula	v.	relijiozu	adj.
reguladór	adj.	remata	v.
regulamentu	n.	remodela	v.
regulár	adj.	remodelasaun	n.
regulariza	v.	rende	v.
regularmente	adv.	rendimentu	n.
regulasaun	n.	rentabiliza	v.
reinu	n.	reorganiza	v.
reitera	v.	reorganizasaun	n.
rejeita	v.	reparasaun	n.
rejiaun	n.	repete	v.
rejime	n.	réplika	n.
rejimentu	n.	reprezália	n.
rejionál	adj.	reprezenta	v.
rejista	v.	reprezentadu	adj.
rejistu	n.	reprezentante	n.
rekerente	n.	reprezentasaun	n.
rekizitu	n.	repúblika	n.
reklama	v.	reseita	n.
reklamasaun	n.	resensiamentu	n.
rekomendasaun	n.	resesu	n.
rekoñese	v.	resife	n.
rekoñesidu	adj.	respeita	v.
rekoñesimentu	n.	respeitadu	adj.
rekonsiliasaun	n.	respeitozu	adj.
rekonstrusaun	n.	respeitu	n.
rekorda	v.	responde	v.
rekorre	v.	responsabilidade	n.
rekruta	v.	responsabiliza	v.
rekrutamentu	n.	responsável	n.
rekupera	v.	responsável	adj.
rekuperasaun	n.	resposta	n.
rekursu	n.	restaurasaun	n.
rekuza	v.	restrisaun	n.
relasaun	n.	restu	n.
relasiona	v.	retiru	n.
relasionadu	adj.	retornu	n.
relasionál	adj.	retratu	n.

457

reune	v.	
reuniaun	n.	
revista	n.	
revizaun	n.	
revolusaun	n.	
revolve	v.	
reza	v.	
rezerva	n.	
rezidénsia	n.	
rezidente	n.	
rezigna	v.	
rezisténsia	n.	
rezolusaun	n.	
rezolve	v.	
rezulta	v.	
rezultadu	n.	
rezume	v.	
rezumu	n.	
riku	adj.	
risku	n.	
rituál	n.	
romanu	adj.	
rombu	n.	
ropa	n.	
rotativu	adj.	
rurál	adj.	
rusu	adj.	
sábadu	n.	
sai	v.	
saka	n.	
sakraliza	v.	
sakrifísiu	n.	
saku	n.	
saláriu	n.	
salaun	n.	
salezianu	adj.	
salienta	v.	
salva	v.	
sansaun	n.	
santu	adj.	
satélite	n.	
satisfás	v.	
satisfeitu	adj.	
saudasaun	n.	

saúde	n.	
saun	adj.	
se	conj.	
sede	n.	
segimentu	n.	
segredu	n.	
segunda	n.	
segunda-feira	n.	
segundu	num.	
segundu	prep.	
seguransa	n.	
seguru	adj.	
sein	prep.	
seis	num.	
sekretária	n.	
sekretariadu	n.	
sekretáriu	n.	
sekretáriu-jerál	n.	
sekretu	adj.	
seksaun	n.	
seksuál	adj.	
sektór	n.	
sékulu	n.	
sekundáriu	adj.	
selebra	v.	
selebrasaun	n.	
selesaun	n.	
selesionadu	adj.	
sélula	n.	
semana	n.	
semellansa	n.	
semi-	affix	
semináriu	n.	
sempre	adv.	
senadór	n.	
senadu	n.	
senáriu	n.	
señór	n.	
señora	n.	
sensibilidade	n.	
sensibilizasaun	n.	
sensór	n.	
sensura	n.	
sente	v.	

sentensa	n.		simétriku	adj.
sentidu	n.		simples	adj.
sentimentu	n.		simplezmente	adv.
sentímetru	n.		sin	adv.
sentradu	adj.		sinál	n.
sentrál	adj.		síndrome	n.
sentru	n.		singulár	adj.
sentu	n.		sinku	num.
separa	v.		sinkuenta	num.
seráke			sinseridade	n.
serimónia	n.		sintoma	n.
sériu	adj.		sintu	n.
serra	n.		síriu	adj.
sertidaun	n.		sirkula	v.
sertifikadu	n.		sirkulasaun	n.
serve	v.		sírkulu	n.
serveja	n.		sirkunstánsia	n.
servidór	n.		sistema	n.
servisu	n.		sita	v.
servisu	v.		situa	v.
sesaun	n.		situasaun	n.
sesenta	num.		síviku	adj.
sesta	n.		sivíl	adj.
sesta-feira	n.		sízmiku	adj.
sestu	num.		só	adv.
sete	num.		soberania	n.
setembru	n.		soberanu	adj.
setenta	num.		sobre	prep.
sétimu	num.		sobrevivénsia	n.
setór	n.		sofistikadu	adj.
SIDA	n.		sofre	v.
sidadania	n.		sofrimentu	n.
sidadaun	n.		soldadu	n.
sidade	n.		solene	adj.
siénsia	n.		solidariedade	n.
sientífiku	adj.		sólidu	adj.
sientista	n.		solisitasaun	n.
sigarru	n.		solusaun	n.
sigla	n.		solusiona	v.
signifika	v.		sombriñu	n.
signifikadu	n.		son	n.
simbóliku	adj.		sorte	n.
símbolu	n.		sosiál	adj.
simeira	n.		sosializasaun	n.

sosiedade	n.	taxa	n.
sosioekonómiku	adj.	téknika	n.
sósiu	n.	tékniku	adj.
sotake	n.	teknolójiku	adj.
sua	det.	telefone	n.
subdistritu	n.	telekomunikasaun	n.
subliña	v.	televizaun	n.
submarinu	adj.	tema	n.
submete	v.	temperatura	n.
subordinadu	adj.	tempestade	n.
subsidia	v.	temporáriu	adj.
subsídiu	n.	tempu	n.
substánsia	n.	tende	v.
substitui	v.	tendénsia	n.
substituisaun	n.	tenente-jenerál	n.
sudeste	adj.	tenke	v.
sudeste	n.	tenkesér	v.
sudueste	n.	tensaun	n.
sufiksu	n.	tenta	v.
sufisiente	adj.	tentativa	n.
suisida	v.	termina	v.
sujestaun	n.	terminál	adj.
súl	n.	termu	n.
súl	adj.	terra	n.
súl-amerikanu	adj.	terra-ár	n.
súl-koreanu	adj.	terrenu	n.
superintendente	n.	terrestre	adj.
superiór	adj.	territoriál	adj.
supervizór	n.	territóriu	n.
suplementa	v.	terrorista	n.
suporta	v.	terrorizmu	n.
supremu	adj.	tersa	
susesu	n.	tersa-feira	n.
suspeitu	n.	terseiru	num.
suspende	v.	tersu	n.
suspensaun	n.	teste	n.
sustenta	v.	testemuña	n.
sustentasaun	n.	testu	n.
sustentavel	adj.	tia	n.
tabaku	n.	timorense	adj.
tál	pron.	tipu	n.
tankedegerra	n.	tiru	n.
tantu	adv.	titularidade	n.
tarde	adv.	títulu	n.

460

tiu	n.		tribunál	n.
tolera	v.		trimestre	n.
toleránsia	n.		triste	adj.
toma	v.		troka	v.
tomada	n.		tropa	n.
tonelada	n.		tsunami	n.
tópiku	n.		turista	n.
topografia	n.		turístiku	adj.
totál	n.		turizmu	n.
totál	adj.		turku	adj.
traballadór	n.		turma	n.
traballista	adj.		tutela	n.
traballu	n.		tv	n.
tradisaun	n.		ultrapasa	v.
tradisionál	adj.		um	num.
tradisionalmente	adv.		umanidade	n.
tradús	v.		umanitáriu	adj.
trafikante	n.		umanu	adj.
tráfiku	n.		umór	n.
trajédia	n.		unanimidade	n.
trajetu	n.		uniaun	n.
trámite	n.		unidade	n.
transfere	v.		unidu	adj.
transferénsia	n.		unifikadór	adj.
transnasionál	adj.		úniku	adj.
transparénsia	n.		unilateralidade	n.
transporte	n.		unilateralmente	adv.
transtornu	n.		universál	adj.
tranzasaun	n.		universidade	n.
tranzisaun	n.		universitáriu	adj.
tranzitóriu	adj.		urbanizasaun	n.
tranzmisaun	n.		urbanu	adj.
tranzmite	v.		urjente	adj.
trasa	v.		útil	adj.
trata	v.		utiliza	v.
tratadu	n.		utilizadór	n.
tratamentu	n.		uza	v.
tratór	n.		uzu	n.
tratu	n.		vaga	n.
trauma	n.		vaka	n.
treinamentu	n.		vákuu	n.
treinu	n.		valeta	n.
trés	num.		válidu	adj.
treze	num.		valór	n.

valoriza	v.	virtuál	adj.
valorizasaun	n.	vírus	n.
vandalizmu	n.	vise-	affix
vantajen	n.	vistu	n.
varanda	n.	vitál	adj.
vazu	n.	vitalísiu	adj.
venda	n.	vítima	n.
veraun	n.	vitória	n.
verbu	n.	vizaun	n.
verifika	v.	viziña	n.
verifikasaun	n.	viziñansa	n.
versaun	n.	viziñu	n.
vertente	n.	vizita	n.
vés	n.	vizita	v.
vestimentu	n.	vizitante	n.
vestuáriu	n.	vokabuláriu	n.
veta	v.	vokasionadu	adj.
veteranu	n.	vokasionál	adj.
vetu	n.	volta	n.
via	n.	volume	n.
via	prep.	voluntariamente	adv.
viajen	n.	voluntáriu	n.
viatura	n.	vontade	n.
vibrante	adj.	vosa	det.
vida	n.	vota	v.
vídeo	n.	votasaun	n.
vigáriu	n.	votu	n.
vigór	n.	vou	n.
vigora	v.	vulkániku	adj.
vijilánsia	n.	xave	n.
vila	n.	xefe	n.
vingansa	n.	xefia	n.
vinte	num.	xinés	adj.
viola	v.	xokante	adj.
violasaun	n.	zeladór	n.
violénsia	n.	zeru	n.
vírgula	n.	zona	n.

ANNEX 5 – TRANSCRIPTIONS OF THE ORAL CORPUS

Tetun Prasa

0005TD

tx@JBP   B..., Ita nu'udar grupu dezenvolvimentu agrikultura lokál saida mak prinsipál liu ba Ita atu halo iha fulan ida-ne'e?

tx@BE   Di'ak, obrigadu ba tempu. Buat importante mak ita halo no planu iha futuru mak primeiru, ita survei ita-nia rain, rai ne'ebé mak atu hala'o agrikultura nian. Segundu, ita halo relasaun, husu apoiu ruma ba ajénsia ne'ebé de'it mak relevante, atu nune'e bele ajuda ita-nia grupu. Terseiru, ita tama ba kampu de servisu, ida-ne'e mak importante liu.

tx@JBP   Ita ohin dehan katak Ita halo survei maibé ha'u seidauk komprende loos saida de'it mak Ita bele hala'o iha survei ne'e nia laran?

tx@BE   Survei katak ita haree buat saida mak ita presiza atu halo. Primeiru mak rai, segundu irrigasaun, terseiru mak fasilidade ne'ebé mak atu suporta ba ita-nia atividade agrikultura nian.

tx@JBP   Wainhira ha'u rona ema hateten dehan katak survei ne'e iha kualidade no survei ida-ne'ebé mak ladún iha kualidade. Survei ida kualidade mak hanesan ne'ebé no oinsá ita bele hatutan ita ninia programa liu husi dalan survei?

tx@BE   Di'ak, ko'alia kona-ba survei ida ne'ebé mak kualidade ita tenke haree ba fasilidade ne'ebé mak iha relevante kona-ba ita-nia grupu agrikultura nian. Depois iha survei ne'ebé mak la iha kualidade mak ita, bainhira ita survei tiha, ita la fó prioridade ba ita-nia survei ida-ne'e.

tx@JBP   Ha'u mós hakarak hatutan tan de'it kolega ninia esplikasaun ne'ebé ohin dehan katak survei ne'e loos dunik katak iha Timor-Leste iha sektór agrikultura ne'e wainhira kuandu ita uza métodu ka tékniku ne'e primeiru mak liu husi dalan survei. Tanba iha dalan survei ne'e ita bele observa no ita bele define didi'ak rai ne'ebé protejidu atu utiliza ba agrikultores ba agrikultura. No liu husi survei ida-ne'e mós ita bele toman no hatene rai ninia kualidade, ninia topografia no mós rai ninia bokur ne'ebé mak iha. No liu husi survei ida-ne'e mós ita bele hatene bee ne'ebé produtivu no rai ne'e ita bele hatene ninia kualidade no ninia bokur ne'e to'o iha ne'ebé. Entaun liu husi dalan survei ida-ne'e, loke dalan ba ema estudante tékniku agrikultura sira atu nune'e sira bele dezenvolve sektór agrikultura hodi hasa'e kualidade produsaun hodi tane aas sektór ekonomia iha Timor-Leste.

0006TD

tx@JBP   Halo señór B..., tuir Ita-nia hanoin kona-ba ita ninia programa ohin loron nian no ita hahú hala'o ita-nia servisu husi dadeer saan to'o iha loraik ida-ne'e oinsá Ita-nia sentimentu.

tx@BE   Di'ak, ita-nia atividade ba ohin loron ninia ha'u sente di'ak tanba ita hetan benefísiu ne'ebé barak liu. Pois ha'u mós kontente tanba ita mós hasoru ita-nia belun na'in-rua husi li'ur ne'ebé mak akompaña ita iha loron kmanek ida-ne'e.

tx@JBP   Benefísiu saida mak ita hetan ohin loron B...? No mós Ita-Boot bele introdús uitoan kona-ba kolega na'in-rua ne'ebé Ita-boot rasik hateten dehan katak mai husi rai li'ur.

tx@BE   Ha'u hatene benefísiu ohin nian mak ita hetan orsamentu ruma kona-ba ita-nia atividade agrikultura nian. Segundu ita-nia belun na'in-rua ne'ebé mak mai husi li'ur, ida ha'u hatene nia naran maibé ida ha'u la hatene. Ida ha'u hatene mak kolega Alan. No mak belun, ita-nia belun feto ida husi Eropa nian mak naran Zuzana.

tx@JBP   Entaun Ita kontente loos tanba ita hetan kolega ida-ne'ebé mak mai husi rai li'ur no ita bele halo ka halo kolega di'ak liután. No B..., buat ida importante liuliu mak ba ita-ninia knaar ohin ne'e. Karik ohin ne'e ita sente kole ka ita sente hanesan ladún di'ak karik ita halo ita-nia servisu ne'e.

tx@BE   Ha'u sente servisu ne'e normál hela. Ha'u sente di'ak hela tanba ita-nia servisu mak hanesan ne'e. Ha'u mós satisfeitu kona-ba ita-nia servisu ne'ebé ohin loron ninia, di'ak.

tx@JBP   Ha'u hatutan tan ha'u-nia kolega nia hanoin kona-ba ami-nia programa ohin loron ida-ne'e. Buat ne'ebé mak ami halo iha loron ida-ne'e. Buat hirak ne'ebé mak ami halo iha loron ida-ne'e maka primeiru, ami hakbesik an ba ema doadores sira, kompañia balun atu bele fó apoiu ba ami-ninia grupu iha sektór agrikultura nian. No iha loron ida-ne'e mós ami hakbesik dunik ba sira no ami husu sira nia apoiu balun. No maski ami, ami hetan uitoan de'it maibé ami agradese tanba ohin loron ne'e ami servisu mós la saugati. I buat ida-ne'ebé mak ami sente ba hanesan todan uitoan ba ami mak, ami hanesan transporte mak sai difikuldade ba ami nu'udar ami la'o de'it. Hanesan ne'e ami sente ba kole maibé ami sempre kontente ba ami ninia programa ohin loron nian. No mós ami haksolok nafatin tanba ami hetan saúde di'ak nafatin. No ami sei hala'o nafatin ami-nia programa ne'e iha tempu tuirmai. Obrigadu.

0008TD

tx@TAB   Di'ak, hanesan problema ne'ebé ke mosu entre Austrália ho Timor-Leste sobre dokumentus ne'ebé ke intel Australia foti ne'e sente ha'u hanesan ema povu Timor-Leste ida, sente ladún kontente tanba ne'e dignidade governu estadu Timor-Leste ninian. Tuir

loos Austrália tenkesér, naran sá, tenkesér akompaña e tenkesér simu tanba segunda gerra mundiál Timor mós fó fatin para Austrália hela hodi halo funu hasoru Japaun hodi defende Austrália bele "aman". I nune'e mós tempu fali noventa e nove INTERFET forsa Austrália mak mai salva Timor. Tuir loloos sira rua ne'e tuur hamutuk atu fó ideia ba malu hanesan nasaun viziñu, nasaun ida ke ajuda malu e depois simu mós. Austrália mós simu refujiadus Timor nian bá iha Austrália, ne'ebé tuir loos ne'e, Austrália hamutuk ho Timor hodi rezolve problema oan sira hanesan ne'e. No mós baliza marítima tasi nian para hodi define loos Austrália ninian ne'e, ninia baliza marítima to'o iha ne'ebé, Timor nian to'o iha ne'ebé, ne'e para atu bele kontrola rikusoin tasi ninia.

tx@TAB Di'ak, iha ne'e iha subdistritu Na'in-Feto, iha ne'e iha administradór ida, CDO ida, ODL ida no mós staf na'in-rua, cleaner ida, animadora ida no mós seguransa na'in-neen. Portantu ha'u uluk iha noventa e nove mai to'o fali dois míl e treze ha'u mak asumi kargu hanesan ODL. Maibé tanba ha'u-nia idade mós boot e depois ha'u mós sempre saúde ladún permite entaun ha'u tensér rezigna an husi ODL. I fó fali ba ha'u-nia kolega mak lori fali kargu hanesan ODL. I ha'u hanesan staf apoiu administrasi de'it para ha'u bele haree mós iha ha'u-nia saúde, saúde tanba idade ona. Be ha'u tensér servisu maibé ha'u mós tenke haree ha'u-nia saúde, kondisaun saúde ninian.

tx@TAB   Portantu estranjeirus barak mak iha subdistritu Na'in-Feto, ne'ebé iha subdistritu Na'in-Feto iha suku neen. Suku Lahane mak estranjeiru la hela tanba la iha otél, la iha apartamentu, maibé suku seluseluk ne'e ne'e iha ema estranjeiru barak mak hela. I sira mós loke otél barabarak tanba iha suku lima ne'e ne'e iha uma di'ak, kondisaun di'ak, iha apartamentu mós di'ak, ne'ebé estranjeiru barak. I ami iha subdistritu Na'in-Feto só simu sira hodi atende sira mai trata sira-nia prorrogasaun viza.

0014TD

tx@JBP Hotu, imi diak ka lae? Di'ak! Ha'u kontente loos hasoru imi. No iha loraik ida-ne'e ita aprezenta uitoan kona-ba ita ninia moris, liuliu ba ita-nia atividades loroloron ne'ebé bele regula ita-nia moris. Tanba ne'e importante liu ita define no identifika atividades balun ne'ebé ita presiza hala'o iha ita-nia moris atu nune'e bele fó valores ka benefísiu ba ita-nia moris. Señora A..., oinsá Ita-ninia hanoin kona-ba Ita-ninia atividade ohin loron, Ita sente nu'usá, di'ak ka lae?

tx@LEG   Ya atividades ohin loron ha'u sente kontente. Tanba ohin dadeer saan jam tujuh ha'u hadeer mai. Ha'u kose tiha nehan. Haro'us tiha depois jam tujuh tiga puluh ha'u halo matabixu. Matabixu hotu tiha, matabixu hotu tiha ha'u ajuda ha'u-nia biin, ajuda ha'u-nia biin hamoos uma laran. Hamoos uma laran hotu tiha ami silu batar, silu batar atu nune'e ami da'an. Hamoos uma laran hotu tiha ami silu batar, silu batar atu nune'e ami da'an. Da'an hotu tiha ami tuur deskansa uitoan. Deskansa uitoan hodi ami hein batar ne'e tasak. Tasak hotu tiha ami tuur hein. Ami kontente uitoan tanba ami kolega na'in-rua mai, ne'ebé hanesan ami hamutuk atu tuur, komentár. Depois ya halo buat seluk

ne'ebé mak hanesan kontente. I depois loraik uitoan bele fahe malu. Fahe malu, ha'u kaer fali te'in fali kalan nian hodi nune'e ami hamutuk ho família sira ami han. Han hotu tiha ami bele deskansa kalan. Oké, obrigadu.

tx@JBP   Obrigadu ba Señora A... ne'ebé fó komenta uitoan. Agora ba iha Señora T.... Oinsá Ita-ninia atividades ohin loron nian?

tx@JMB   Oké, obrigadu ba tempu ne'ebé fó mai ha'u. Ha'u-nia atividade loroloron ohin dadeer nian hanesan ohin dadeer ha'u hadeer. Ha'u bá sikat tiha nehan. Depois hotu ha'u prepara matabixu ba maun sira no ba ami hotu. Hotu tiha, matabixu hotu tiha, hamoos uma laran no dasa fo'er iha li'ur. Depois ohin dadeer ami haree hanesan fo'er ne'e kan nakonu tiha ona iha be fatin be fo'er nian. Depois ami lori ba fakar tiha mak fila fali ami bá silu batar. Silu batar hodi nune'e ami da'an atu prepara ba meiudia nian. Depois ami tuur hamutuk, ami rua be alin Lili. Ami rua ko'alia dehan orasida ita rua nia kolega mai. Ne'e dunik agora ita rua keta toba demais tanba ita rua tuur hein ita rua nia kolega sira, se maun J... telefone mai. Ita rua tuur iha ne'ebá para hein ita-nia kolega ne'ebé mak mai atu ita halimar. Depois kolega ida telefone mai atu hasoru mas la jadi hasoru, tanba ha'u hatene katak kolega sira atu mai halimar iha ne'e. Depois ami hein iha ne'e, kolega na'in-tolu mai ho maun Janu, mai ita tuur halimar no komentár. Ami lori batar mai ita han rame-rame hamutuk no hemu kafé, no tuur hamutuk konta istória. Só ha'u-nia liafuan mak ne'e de'it.

tx@JBP   Oinsá mana A..., Ita-nia hanoin, Ita-nia atividades ohin loron?

tx@FDP   Oké, di'ak, obrigadu ba tempu. Ha'u-nia atividade ohin loron nian mak ohin dadeer saan ha'u hadeer tuku neen. Hadeer hotu tiha, kose nehan. Kose nehan hotu, ha'u fase bikan. Fase bikan hotu ha'u ho alin sira bá silu batar. Silu batar hotu ami da'an, ami prepara ba meiudia nian. No ohin loron ne'e ha'u la bá fatin ida. Ha'u halimar de'it iha uma. No tuku rua kolega na'in-rua mós mai no ami kontente hasoru sira. Tanba ami mós practice ami-nia inglés ho sira. Iha loraik ida-ne'e ami kontente tanba hasoru kolega ne'ebé di'ak husi nasaun seluk ne'ebé mak ohin loron vizita ami to'o iha ami-nia uma. Ha'u-nia liafuan mak ne'e de'it. Obrigadu barak.

tx@JBP   Ó lalika hamanas karik di'ak. Be entaun loos ona. Ha'u perguntas ba segundu nian kontinua ho.   Maibé karik imi na'in-rua hakarak mós atu ko'alia buat ruma? Oké, ha'u fó tempu ba alin Amakai atu ko'alia uitoan.

tx@ACB   Oké, di'ak, obrigadu ba tempu ne'ebé fó mai ha'u. Antes de ha'u atu konta uitoan kona-ba ha'u-nia atividade indivíduu nian, uluknana'in ha'u la haluha mós ha'u-nia botarde ba maun Alan no mós ba kolega biin, mana Zuzana ho tan maun ne'ebé ke ohin brother ne'ebé ke ohin mai vizita ami-nia fatin ida-ne'e no mós vizita ami. I iha biban ida-ne'e mós ohin ha'u la iha ne'e maibé ha'u iha tarde uitoan. Tanba ohin dadeer saan, ha'u-nia kolega ida mai konvida ha'u-nia kolega ida Markus konvida ha'u ami bá tiha fatin ida, i ha'u hakarak konta de'it kona-ba atividade iha diáriu nian ne'e. Iha atividade ida-ne'e prinsipalmente ba ita hotu tanba atividade ida-ne'e mak sai hanesan

eventu ida-ne'ebé ke di'ak ba ita hodi regula ita-nia moris loroloron nian. Tanba iha atividade ida-ne'e ita nunka mais iha atu disponivel. Iha loroloron kona-ba atividade, kona-ba ha'u-nian, ohin dadeer, dadeer saan iha tuku lima hanesan ne'e ha'u hadeer tiha ona tanba iha mós tugas barak kona-ba ami-nia eskola nian. Tanba agora daudaun ha'u mós halo hela ha'u-nia estudu iha universidade ne'ebé mak agora daudauk la'o hela, tanba iha ne'ebá mós dosen sira fó tugas ba ha'u. Depois ha'u mós halo. Depois iha be dadeer saan tuku hitu hanesan ne'e ramata tiha ha'u-nia tugas ne'ebé ke ha'u halo ha'u ramata hotu tiha. Depois ha'u-nia biin sira, ha'u-nia alin feto sira halo kafé. Depois ne'e ha'u-nia alin sira bolu ha'u dehan Amu mai hemu kafé. Depois ha'u bá hemu, ha'u hotu tiha ha'u haree ba be, haree ba li'ur be iha varanda ne'e. Ha'u-nia alin feto depois ha'u-nia rian sira bá silu batar. Depois sira tau hamutuk hamoos nia kulit, depois sira da'an. Depois ne'e hotu tiha la kleur tan ha'u-nia maun nia oan naran Ebi, iha leten hela ho nia avó depois be nia tiun ida Markus ne'e fó hatene dehan: Ó tun ba fó hatene maun Amu. Depois sa'e mai para ami na'in-rua bá atividade ida iha universidade UNPAZ nian ne'ebé ó fó hatene nia para mai ami rua bá asiste atividade ne'e. Depois alin ne'e mai bolu ha'u. Depois ha'u sa'e ba, ha'u sa'e ba Markus dehan: Ó mai lailais para ita rua bá. Depois ha'u to'o tiha iha ne'ebá, nia dehan: Ai…ha'u sei falta uitoan kona-ba sigarru nian, ha'u hola tiha sigarru ida para orasida to'o iha ne'ebá karik, naran sá, karik ita fuma entaun loos ona. Depois to'o tiha iha ne'ebá, nia hola hotu tiha sigarru. Depois mai ami rua bá. Bá to'o tiha iha ne'ebá ha'u mós kontente tebetebes tanba haree ha'u-nia kolega universidade sira ne'ebé ke prepara fatin depois sira mós simu oradór husi universidade seluk. Tanba eventu iha ne'ebá mós sira halo kona-ba graduasaun e nune'e mós sira halo fó pose ba malu kona-ba be HMJ nian no mós iha senadu nian. To'o tiha iha ne'ebá, iha acara ne'e hotu tiha depois sira kontinua ho divertimentu. Divertimentu hotu tiha ha'u sente ba ha'u kole uitoan. Depois hodi kalan mós ha'u hadeer de'it. Depois ha'u fó hatene ba ha'u-nia ria: "Markus, di'ak liu ita rua fila ona." Depois ami na'in-rua fila, ami na'in mai tesik mai Komoro nian, nia dehan tan: "Amu, ha'u hola tan sigarru ida lai." Entaun loos ona, nia hola tiha sigarru hotu tiha depois ami rua tun mai. Tun mai to'o iha ne'e, ha'u-nia Markus para tiha, ha'u tun mai. Ha'u mós kontente loos tanba haree bainaka na'in-rua ne'ebé ke mai ho ha'u-nia maun J.... Tanba maun J... mós iha interese tebetebes tanba liga ho ita-nia bainaka na'in-rua tanba iha bainaka ida-ne'e mós sai nu'udar ita-nia belun di'ak ne'ebé ke hamutuk ho ita. Bainhira ita iha difikuldade kona-ba língua ruma, difikuldade kona-ba nesesidade ruma ka, naran sá, sira mós, sira hanesan ho ita, maibé sira mós iha kapasidade uitoan kona-ba língua nian mós iha, karik iha limitadu ba nesesidade karik ya. Sira mós neineik-neineik sei la'o no buka hodi ita bele suporta malu. Ne'ebé atividade, ko'alia kona-ba atividade, atividade ne'e prinsipál sempre la'o hamutuk ho ita. Ita nunka mais atu disponivel kona-ba ita-nia atividade ne'e. Tanba ita nu'udar ema umanu ita sempre iha atividade. Tanba iha atividade ida-ne'e, ita presiza uza liubá ita-nia nesesidade. Tanba bainhira ita presiza nesesidade ne'ebé ke di'ak, ita-nia korpu ne'ebé ke di'ak ita presiza iha atividade. Porezemplu, ha'u hakarak iha eskola, entaun bainhira hakarak iha eskola presizamente iha atividade. Atividade ne'ebé ke ita halo hanesan ekonomi. Ekonomi sebagai atividade ne'ebé ke ita halo. Bainhira ita hala'o atividade

ekonomi ida-ne'e ita bele hetan rendimentu uitoan kona-ba ekonomia, rendimentu uitoan kona-ba ekonomia ne'ebé ke... Bainhira ita hetan rendimentu ida-ne'e ita mós bele sustenta ba ita-nia moris. I nune'e mós ita bele suporta mós ba ita-nia alin ruma ka, ita-nia família ruma ne'ebé ke sira limitasaun ba nesesidade. Di'ak, ida-ne'e mak hanesan ha'u-nia komentáriu e ha'u-nia congratulation mós ba ami-nia bainaka na'in-rua ne'ebé ke ohin mai vizita no mós ha'u-nia maun e obrigadu barak.

tx@JBP   Ai…obrigadu barak ba imi na'in hira ne'ebé esplikasaun di'ak no furak tebetebes. Ita sei iha tan perguntas kontinuasaun ba segundu nian. Ne'e dunik Señora A..., Ita-Boot prontu ka lae? Pergunta segundu. Karik iha Ita ninia hanoin karik iha atividades balun ne'ebé mak Ita-Boot konsidera hanesan prinsipál atu Ita-Boot halo? Tanba iha atividades barak ne'ebé Ita-Boot halo. Maibé karik Ita-Boot iha interesante ba atividade seluk ruma ka programa seluk ruma ne'ebé mak hanesan prinsipál ba Ita-Boot atu halo. Ita-Boot gosta mais programa ida-ne'ebé mak Ita-Boot hakarak halo?

tx@LEG   Ha'u-nia moris kona-ba program ida-ne'ebé ke ha'u gosta loos mak hanesan program kona-ba komputer. No mós ha'u hakarak hanesan atu hatene liután kona-ba língua. Depois ha'u mós hakarak atu be aprende língua kona-ba inglés, portugés, bahasa maibé ha'u sei estuda neineik atu nune'e ha'u bele aprende. Oké, obrigadu.

tx@JBP   Ba iha Señora T...., porfavór!

tx@JMB   Programa ne'ebé mak prinsipál liu ba ha'u mak hanesan ha'u hakarak loos atu aprende komputadór no inglés, portugés no bahasa. Tanba bainhira ita atu hakarak atu realiza buat ruma iha ita-nia moris ita presiza loos komputer, inglés no língua hotu-hotu. Tanba ita servisu, se bainhira ita hakarak atu servisu iha office ruma ita presiza loos língua sira ne'e. Se bainhira ita la hatene línguas ne'e mak sei difikuldade ba ita atu hetan buat di'ak ita-nia moris. Ida-ne'e mak ha'u-nia prinsipál ba ida-ne'e de'it.

tx@JBP   Oinsá mana A...?

tx@FDP   Ba ha'u programa importante liu mak ha'u hakarak aprende língua inglés, portugés, bahasa no aprende komputer. Tanba iha ita-nia nasaun ema presiza liu mak hatene iha servisu ruma ema presiza liu ema ne'ebé hatene komputer, hatene ko'alia inglés no portugés. Ne'e dunik ha'u hakarak aprende línguas sira ne'e. Tanba ita hakarak realiza buat saida de'it mak ita hakarak iha ita-nia futuru, ita presiza hatene línguas sira ne'e. Karik ita hakarak realiza buat ne'ebé ita hakarak iha ita-nia futuru maibé ita la hatene língua sira ne'e sei difísil ba ita atu realiza buat ne'e. Oké, obrigadu.

tx@JBP   Obrigadu barak ba imi no imi-nia esplikasaun ne'e di'ak no imi halo sakrifísiu ida no imi oferese an respostas no pergunta rua ne'ebé ha'u fó ba imi. Loos dunik katak ita ema moris ita sempre la sees husi atividades no ita sempre iha órgaun, ita iha liman, ain, iha ibun atu halo ita ninia knaar loroloron no atividades loroloron nian. Hanesan ita hotu hatene katak iha atividades hirak ne'ebé ita halo ne'e prinsipál dunik ba ita-nia moris. Maibé ha'u hakarak ko'alia ne'e hanesan imi hanesan alin sira loos ka lae? Ha'u hakarak

468

ko'lia uitoan kona-ba ita-nia moris ne'e saida mak prinsipál liu ita halo. Ita haree husi ita, primeiru ita tenke haree husi ita-nia an. Ita ne'e, ita agora iha ne'ebé, ita-nia posisi ne'e saida no ita atu la'o ba ne'ebé. Ha'u hakarak ko'alia liu kona-ba ema estudante. Ema estudante ne'e nia knaar mak nia tenke lee di'ak, hakerek di'ak. Nia tenke rona ne'e di'ak. Nia tenke esplika ne'e di'ak. Atu hetan buat sira hanesan ne'e ne'e, primeiru, ne'e iha intensaun ida tenke haka'as an no esforsu an. Agora, atu, iha liu husi dalan haka'as an no esforsu an mak primeiru, ita tenkesér halo oráriu hanesan imi baibain ohin iha ne'e esplika. Ita tenke halo oráriu katak ha'u sei hadeer oras hanesan ne'e no sei estuda oras hanesan ne'e hanesan ne'e no ha'u sei bá eskola oras hanesan ne'e. Tanba liu husi dalan oráriu mós bele muda ita ninia moris no bele halo, bele ajuda ita atu susesu no hetan buat ne'ebé mak ita hakarak. Iha mós dalan seluk mak ita hanesan iha interesante ka iha dunik kemauan atu aprende dunik. Katak ita aprende ne'e tenke ho konsentrasi ka ho serius ita aprende dunik. La'ós aprende be ne'e ne'e dua menit, tiga menit hamriik la'o tiha ba halimar barak tiha depois, hanesan ne'e ne'e, ne'e dehan ita la iha intensaun di'ak atu be muda ita-nia moris. Ha'u hanoin buat sira ne'e hotu prinsipál ba ita hanesan ohin imi ko'alia. Estudante ida prinsipál ba estudante mós linguajen. Tanba ita hakarak atu ko'alia porezemplu ho ha'u-nia kolega ida mana Zuzana ho Alain, ita tenke hatene sira-nia lian. Se ita la hatene sira-nia lian ita ko'alia saida. Ne'e ita la bele ko'alia, ita la bele halo komunikasaun se ita la hatene língua, se ita la komprende língua. Ne'e língua mais prinsipál dunik ba ita, agora iha ita-nia polítika iha Timor mós presiza língua para bele ajuda ita atu asesu ba servisu. No mós iha atividades ne'ebé prinsipál liu hanesan ita-nia programa prinsipál liu mós ba estudante mak ita tenkesér buka no rezolve ita ninia nesesidade. Estudante barak, ha'u uluk, ha'u-nia istória ne'e hanesan ne'e. Uluk ha'u tama mai iha Dili, fa'an, fa'an be manu-tolun ne'e. Fa'an manu-tolun ema dehan: Oi, ó-nia tolun hirak mak ne'e? Ne'e, ka ema dehan ha'u fa'an, ha'u fa'an ho be Xina sira-nia sasán. Fa'an Xina sira-nia sasán ne'e ne'e. Ema goza ha'u hanesan ne'e, hanesan ne'e. Maibé ida-ne'e hanesan dalan báziku ida-ne'ebé mak bele hatán ba ita-ninia nesesidade. Maibé importante liu ba ita mak ita oinsá ita bele muda ita-nia moris husi la di'ak ba di'ak. Ida-ne'e mak prinsipál liu ba ita halo. No neineik-neineik, neineik-neineik ita bele muda ita-nia moris. Entaun dalan importante ba estudante mak ne'e, wainhira ita la iha osan ita presiza buka bisnis ruma ka nu'usá para ita hetan osan. Porezemplu wainhira ita difikulta ba matéria balu ne'ebé mak hanesan ita ladún toman ka la hatene ita tenke buka para hatene. Ida-ne'e mak knaar prinsipál ne'ebé ita nu'udar joven ka estudante. Liuliu ba hanesan be ita joven sira be hakarak ba atu hola, hola mane ka hola feto ne'e, ida-ne'e tenke estuda didi'ak ema e, ita-nia knaar mak ne'e. Se ó la estuda didi'ak ema, ó hola sala mai hela naran de'it moris ida la di'ak. Loos ka lae? Entaun hanesan ne'e de'it, ne'e mak liafuan ikus ba ita hotu. No obrigadu barak ba imi no imi-nia ajudas. Obrigada!

0051TD

tx@DLB   Di'ak. Perguntas di'ak. Sin, ha'u-nia naran kompletu D.... Ha'u ema kaben-na'in. Ha'u-nia oan haat. Agora mane ida, fetu tolu. I ha'u-nia kaben mós naran F.... Ne'e ha'u-nia família. I restu sobre ha'u-nia perkursu estudu. Sin, ha'u hasai ha'u-nia kursu lisensiatura to'o loos ho tiha iha tempu Indonésia. UNTL sei ho naran UNTIM. Mais tanba problema polítiku, kestaun polítiku tempu ne'ebá ne'ebé no mai to'o noventa e nove ami-nia kursu la bele remata iha ne'ebá tanba situasaun iha ne'ebé ho referendum iha akontese Timor-Leste. Hamutuk halai to'o sa'e depois de referendum buat hotu la'o ladia'ak. I depois ha'u kontinua ha'u-nia kursu ne'e halo termina iha dois míl e trés. Dois míl e trés ha'u-nia licensiatura hotu ne'ebá tinan barak ba kotuk ona. Depois durante ha'u mós sei servisu antes ha'u mai servisu iha INL UNTL, ha'u mós antes ne'e ha'u servisu iha. Uluk ha'u profesór rasik, profesór iha ensinu báziku. Depois ha'u mai koliah, mai ha'u-nia estudu iha universidade Timor Loro-Sa'e, lisensiatura. Depois ne'e ha'u kontinua, depois ne'e ha'u halo ona graduasaun, lisensiar-nia iha dois míl e trés. Momentu ne'ebá ne'e ha'u sei servisu ho ajénsia internasionál iha misaun Nasoins Unidas. Depois ha'u servisu ho organizasaun internasionál de traballu. Kontinua mai to'o noventa, mai to'o iha dois míl e onze, dés, onze. Depois ha'u-nia kontratu termina iha ne'ebá ha'u servisu fali ho organizasaun internasionál ICRC, International Committee of Red Cross, dezde dois míl e onze mai to'o iha, dois míl e dés mai to'o dois míl e doze. Servisu iha ne'ebá hanesan mission field officer. Depois ha'u rasik husu hakarak kotu ha'u-nia kontratu iha ne'ebá tanba ha'u-nia misaun ICRC iha momentu ne'ebá la seguru ba ha'u-nia kontinuasaun servisu futuru, entaun ha'u deside hapara ha'u-nia servisu iha ne'ebá. Depois ha'u tuir entrevista iha UNTL, momentu ne'e ha'u liu. I dezde ne'ebá ha'u servisu mai to'o ohin loron iha UNTL. Liu servisu iha instituisaun nasionál linguístika INL. I agora daudaun departamentu Tetun mós tan daudaun servisu ho ha'u-nia ekipa servisu alende servisu nu'udar ha'u membru estrutura linguístika. Iha ne'e mós ami servisu mestre iha departamentu Tetun hodi hanorin estudante sira to'o ohin loron. Ha'u-nia pergunta primeira bele hatán hanesan de'it.

tx@DLB   Sin, ida-ne'e preokupasaun, kestaun ema barak ko'alia iha ne'ebá klaru ida-idak nia ko'alia ne'e, hanoin ne'e. Só ke atu ko'alia kona-ba oinsá atu dezenvolve lian Tetun ha'u hanoin dezde ha'u tama iha ne'e dois míl onze, ha'u tama iha ne'e, INL ninia hala'o nia knaar hodi dezenvolve Tetun ne'e. La'o alende hanorin lian Tetun iha eskola, iha instituisaun governu nian ne'ebé mai sira husu, ami mós fó formasaun Tetun ba membru polísia sira, ba instituisaun governu. Alende ne'e ami formasaun Tetun hala'o loos iha Dili de'it. Ami forma, fó kursu ba profesór sira hotu. Uluk-uluk sei foka liu ba profesór sira iha Timor e ba kursu iha distritu. Depois, agora to'o mai ohin loron ami fó formasaun ba iha distritu ne'e la'ós de'it ba profesór sira maibé fó hotu ba funsionáriu públiku sira ne'ebé mak servisu iha estadu no mós NGO sira. Ami fó formasaun kursu, liuliu kursu ba ortografia Tetun padronizada no gramátika Tetun, mak ami fó ida-ne'e parte ida hosi atu dezenvolve, parte ida dezenvolve lian Tetun, lian ofisiál.

tx@DLB   Kerdizer, lian materna seluk.

tx@DLB   Sin, ami hanorin Tetun, ne'e kerdizer, atu dezenvolve, atu dezenvolve lian Tetun, sai Tetun ida-ne'ebé di'ak, ema barak bele hatene. E hanesan ema bele utiliza, halo hatene ortografia ne'e. Ema hotu-hotu bele hatene, bele utiliza. Ba aban-bainrua bele hotu-hotu uza ortografia ho ida de'it, sistema ida de'it. Ema hotu-hotu hanesan.

tx@DLB   Mais ida-ne'e la signifika atu halakon fali lian seluk-seluk la iha, lian seluk-seluk ninia moris, ninia dezenvolvimentu ne'e iha ninia rasik, porezemplu. Lian seluk-seluk agora INL mosu halo produsaun ba lian lokál sira barak ne'ebé mak sira-nia balu. Isi ita bele haree iha administrasaun INL, iha lian lokál, dialetu sira ne'ebé mak kolega sira balu prodús sira-nia disionáriu monolinguál iha porezemplu Makasae sira. Iha livru iha ne'ebá bele haree. Só ke seidauk hotu-hotu iha livru hosi dialetu ida-idak maibé parte balu iha ona liu ne'e. Kerdizer, katak loos Tetun ne'e iha para tan hamate lian lokál, lae. Tetun la'o ninia, lian lokál sira-nia pozisaun loos iha nafatin. Sira, iha membru INL iha esforsu mós atu hamosu disionáriu ki'ik monolíngua ba lian lokál sira. Ne'ebé agora seidauk halo hotu mas iha ne'e knaar INL ne'e para bele aban-bainrua sei bele prodús hotu disionáriu monolíngua ba dialetu lokál ou lian lokál sira ne'ebé mak iha, eziste.

tx@DLB   Problema boot. Sin. Tetun komu, Tetun ema Timor ko hanoin katak lian Tetun ne'e ita-nia ami-nia rasik. Timor-oan hotu-hotu be moris mai ho lian Tetun keras, ko'alia rasik hatene hakerek. Só ke problema ne'ebé boot mak ami hasoru iha ne'e tanba ofisialmente não linguistikamente lian Tetun ne'e foin, foin mak hahú iha dois míl dois. Ho iha ne'ebé mak governu nomeia universidade Timor Loro-Sa'e liuliu husi INL bele nia hamriik iha ne'e para atu dezenvolve Tetun ne'ebé. Entaun, ema Timor barak liu, labarik, joven-oan sira maka agora iha eskola, sira seidauk hotu-hotu koñese ortografia Tetun, oinsá hakerek liafuan ida ne'e, tanba barak liafuan lian Tetun ne'e, barak liu maka adota hosi lian portugés no mós lian Tetun Terik. Problema boot ami hasoru labarik Timor-oan sira ne'ebé mak joven mós, Timor-oan sira seidauk koñese didi'ak oinsá atu hakerek lian Tetun ne'e ho di'ak, liafuan portugés sira ema empresta mai Tetun atu hakerek ka traskipt, halo ninia formasaun, mudansa ba hosi ortografia portugés ba Tetun ne'e, sira seluk hatene. Ida-ne'e problema to'o agora ita hanorin, hanorin sira Tetun ne'e, liafuan hanesan ne'e ho nia sistema hakerek hanesan ne'e, hanorin S no fim. Sira ezame hotu hetan notas, maibé sira hakerek, ita halo sira hakerek, sei sala hela de'it, sei sala. Pontu, pontu sira ne'ebé mak uza, iha sinais sira ne'ebé mak uza iha Tetun, Tetun ofisiál ne'e. Sira hatene tiha, aprende tiha mas depois sira ikus, sira haluha tanba sira sei hela ida-rua Tetun, ida be, sira ko'alia baibain lororon nia. Mas iha balu-balu de'it mak bele hakerek loos, ho loos, tuir matéria ne'ebé ami hanorin. Mas barak liu seidauk, seidauk komprende, seidauk komprende. Hotu mós haluha hakerek ona sira, ita haruka sira hakerek. Sira sempre haluha ninia pontu sira ne'ebé sinál, porezemplu, kapa-tatolan ho tan asentuasaun agudu. Sira ne'e sempre sira haluha koloka iha liafuan sira ne'ebé sira mak hakerek. Ne'e mak problema boot. Mas ida-ne'e la, tanba formasaun la kontinua, ita

hein katak ba aban-bainrua, sira mós bele utiliza, hakerek, bele hakerek lian Tetun ho di'ak. Se, kuanto formasaun, ita fó beibeik.

tx@DLB   Sin, iha pro e kontra. Estudante sira agora, agora estudante sira ne'ebé mak ita ko'alia asuntu kona-ba estudante sira ne'ebé estuda iha UNTL. Kerdizer, ne'e .... iha, sira ne'e iha kompeténsia iha, UNTL ninia sira ne'ebé sira obrigatoriamente tensér aprende Tetun bainhira tama iha ne'e tanba ita, hateten ba sira katak ita: Tetun ita-nia lian mas ita aprende hanorin ida-ne'ebé di'ak para aban-bainrua ema hotu-hotu uza mezmu. Tensér, sira tensér kompre ortografia Tetun ne'ebé ema mai hanorin atu ne'e ne'e sira rasik mak sei dezenvolve fali Tetun ne'e ba ema seluk. Aban-bainrua se sira sai hotu iha ne'e, sira bele hanorin fali iha fatin seluk. Entaun ema hotu-hotu bele hetan sira-nia família, iha sosiedade ne'ebé sira hela ba ou labarik sira ne'ebé sira hela ba sira tenke hanorin tutan tan fali atu ema seluk bele hatene hotu. Tanba ne'e ligasaun ho lian portugés sin, klaru, ema baluk dehan Tetun ofisiál ne'e, Tetun ne'ebé ema sira balu dehan Tetungés tanba liafuan barak, vokabuláriu barak Tetun nian ne'ebé empresta hosi lian portugés. Mas ne'e komentáriu ema hola Tetun, Tetungés tanba lian Tetun ne'e ita ko'alia ema portugés bele rona liafuan, liafuan verbu portugués sira ne'e iha rona rasik ne'ebé sira-sira. Ne'e mak sira dehan Tetungés mais ne'e ko'alia de'it. Se ita hakarak haree didi'ak atu hatene Tetun di'ak ne'e la'ós fásil ida. Ema tensér estuda, komprende, buka atu komprende para bele hakerek ho koa'lia Tetun di'ak. Ema sira komentár arbiru hanesan Tetun aprende, Tetun ofisiál no Tetungés ne'e sira komenta e no fim sira rasik ko'alia la loos, hakerek la loos, ko'alia mós kahur arbiru de'it. Tetun ne'e ninia, ninia saida, ninia regra ofisiál ne'e iha. Se bainhira ita halo tuir, ita bele ko'alia ho di'ak, hakerek ho di'ak. Ita, ita, ema ko'alia, ita rona mós di'ak. Mas ba ema sira ne'ebé maka komentáriu barak, sira hakerek la hatene, ko'alia arbiru de'it. Entaun dehan maski ita hakarak tau ida ne'e hanesan kestaun importante ita bele aprende di'ak. Entaun ita bele uza, ita bele hatene, ita bele komprende didi'ak, uza didi'ak, ita hatene, ita bele saida. Ha'u hanoin kestaun ida-ne'e hanesan ne'e ha'u-nia komentáriu. Depois profesora bele aumenta tan komentáriu barak hosi ha'u-nia kolega, señór sira.

tx@DLB   Nada.

0053TD

tx@JLS   Mana hatene hela ha'u-nia naran to? Primeiru ba tempu ne'e ha'u bele dehan hanesan ne'e ha'u naran J.... Ha'u husi Timor-oan dunik ne'ebé la'ós husi ema nasaun seluk mak mai tanba husi ne'ebé mak ha'u hatán hanesan ne'e. Maibé iha ne'e ha'u-nia istória pasadu ne'ebé uluk ha'u hahú eskola iha Universidade Timor Loro-Sa'e, iha tempu Indonézia nian. Ha'u hotu dunik eskola ne'e iha Indonézia nian maibé la wisuda. La to'o graduasaun. Depois tama fali iha ukun-an mau foin tuir graduasaun iha dois míl nove kuandu ha'u fila husi Fiji, Suva mai fali Timor-Leste. Ha'u ne'e mak ha'u hato'o. I agora ne'e daudauk ha'u servisu ho Universidade Nasionál, liuliu iha unidade ida ita bolu dehan

Instituisaun Nasionál Linguística. I halai liu to'o nia objetivu halai liu ba oinsá atu dezenvolve Tetun ho rasik iha Timor nu'udar indentidade nasaun ninian ho halai liu mós kona-ba iha kulturalismu nian iha Timor-Leste ba lian Tetun.

tx@JLS   Iha ne'e ha'u hanorin matéria lubun ida. Maibé loloos ne'e primeiru ha'u hanorin iha UNTL hahú husi dois míl oitu, ne'e ha'u sei kontratu. Ne'e ha'u hanorin iha departamentu lian inglés to'o dois míl treze. Depois ha'u tuir test ne'ebé pasa fila fali ba iha linguística. Entaun ha'u tama fila fali ba iha Instituisaun Nasionál Linguística ne'e. La haree loos ba lian ida de'it maibé haree liu ba lian haat ne'ebé hakerek ona no mós legalidade husi konstituisaun república RDTL ninian ne'ebé Timor-Leste ita uza lian haat. Nu'udar Instituisaun Linguística Nasionál, ita iha responsabilidade tomak kona-ba lian Tetun, portugés ne'e rua ba lian ofisiais ho nasionais, depois ba lian traballu rua, malaiu ka bahasa indonézia i depois ho lian inglés.

tx@JLS   Ha'u hakarak husu fali iha tempu ida-ne'ebé? Tempu ida malae-mutin portugés ou tempu indonézia, tempu japaun?

tx@JLS   Tempu portugés ne'e tanba ita hatene hela sistema-sistema europeia sira ne'e. Iha sistema buat ida ema bolu dehan imperializmu ho kapitalizmu. Entaun iha ne'e la fó fatin ba ema sidadaun sira atu eskola. Entaun eskola iha momentu ne'ebá ne'e só ema sira ne'ebé, maluk sira ne'ebé hanesan liurai, xefe suku, xefe aldeia nia oan sira maibé sira ne'ebé iha osan. Entaun iha tempu ne'ebá sira eskola. Iha tempu ne'ebá eskola, kuandu ha'u la salah, inklui ha'u-nia aman eskola to'o iha terseira klase. Ne'e komesa hahú ona iha míl novesentus vinte um mai leten. Ne'e iha tinan rihun ida atus sia ruanulu-resin-ida mai leten. Timor-oan sira foin hetan edukasaun husi malae-mutin portugál ne'e, liu husi misionáriu sira, la'ós liu husi estadu portugés rasik. Tanba ida-ne'e mak prioridade ne'e fó liu ba ema liurai nia oan sira. Ida-ne'e iha tempu ne'ebá ita bele sura de'it iha lista edukasaun ninian, se ha'u se lembra fali, momentu ita iha zeru vírgula sinku porsentu Timor-oan ne'ebé eskola iha malae-mutin nia tempu portugál ninian, ne'e ita ladún hetan to'o dés porsentu mós lae to'o, la to'o.

tx@JLS   Iha tempu indonézia ita bele ko'alia fifty-fifty, tanba ita hetan tiha ona eskola husi ensinu, aprendizajen ne'e. Ita mai husi kedas eskola ida ita bolu infánsia ka eskola labarik ki'ik sira ne'e, ita bolu di'ak kitchengarden hanesan ne'e. Depois mai fali to'o iha, naran sá, iha sekundáriu, mai universitáriu ne'e ema barak ona, ema barbarak ona aprende iha lian malaiu, aprende política, aprende buat hotu-hotu, iha inglés. I oinsá para prosesu Timor-oan sira ne'e edukasaun tempu tinan ruanulu-resin-haat ne'e atu husik sira susar hotu-hotu ne'ebé uluk Portugál sira husik hela iha Timor. Maibé ho tempu ida-ne'ebé la hanesan, tempu la hanesan iha edukasaun ne'e. Tempu ida edukasaun ne'e iha malae-mutin ninia ne'e la obrigatóriu. Agora Indonézia nia tempu ne'e tinan ruanulu-resin-haat ne'e obrigatóriu para halo lakon buat ida ema bolu dehan alfabetizasaun iha Timor-Leste. Ne'e ema dehan katak buta huruf, hanesan.

tx@JLS   Iha Timor-Leste lalika dehan barak. Maioria ita nafatin ho kuríkulu indonézia sei barak tanba ita uza lian haat, língua haat. Primeiru, bainhira kuandu estabelese edukasaun, restorasaun independénsia dois míl dois, hahú husi ne'e ba mai ita uza barak liu kuríkulum ho sistema indonézia nian. Kuando mosu fali mai agora daudauk ita hahú husi dois míl dés, dois míl dés, tinan rihun rua sanulu ne'e, ita komesa ona adapta kuríkulum la'ós indonézia de'it maibé kuríkulum ita adapta mós husi Portugál, Austrália. Adapta mós husi Amérika nia uitoan, ho Indonézia nia uitoan ne'ebé kuríkulum ne'e ita sei kahur hela. Seidauk determina loos dehan kuríkulum ne'e ema Timor-oan ne'e mak ne'e la iha.

tx@JLS   Kuandu halai fali ba perguntas ida hanesan ne'e ne'e iha podér politik. Podér politik ne'e ita la bele naran foti desizaun ida kuandu hadi'a nasaun nia parte ida. Kuandu iha kargu, autoridade, primeiru ministru ne'e, ita-nia kargu ne'e la haree ba iha departamentu ida. Ita la haree ba iha unidade ida, maibé ita haree iha jerál de'it, jerál komún, ne'e umum. Bainhira ita kuandu haree ba ida de'it, setór ida de'it entaun ita la bele dezenvolve setór ida. Entaun iha ne'e primeiru pasu mak oinsá kuandu sai nu'udar lideransa ne'ebé di'ak ne'e primeiru pasu maka ita tenke oinsá koloka orsamentu di'ak liubá iha fatór edukasaun para ema mestre sira ne'e la bele baruk, sira iha osan la'ós, sira iha osan hanorin, sira iha osan fasilidade, suporta sira bele hanorin estudante di'ak liu iha kualkér matéria saida, lian saida ou sistema política saida para sira bele aprende ba iha futuru nasaun ninia. Ida-ne'e mak importante maibé kuandu ita hela ho naran primeiru ministru sira ita haree Timor foin ukun an. Ita iha primeiru ministru na'in-tolu ona, na'in-lima, ne'ebé ita haree, ita ko'alia ne'e Timor-Leste la dezenvolve an iha setór jerál. Orasida iha dependente política ne'e desizaun la'o ketaketak. Kuandu edukasaun orasida la'o, orasida la la'o, política edukasaun, entaun iha ne'e ita la fó vantajen ba malu. Entaun iha ne'e susar uitoan. Maka ba fali ha'u hanesan mehi ida karik perguntas ida ohin ne'e hanesan mehi ida. Karik aban-bainrua ha'u sai primeiru ministru ha'u haree ba fatór tolu mak importante liu ba nasaun. Primeiru, edukasaun, segundu, saúde, terseiru, agrikultura. Para oinsá futuru nasaun ne'e la'o, se nasaun ne'e la'o, maka edukasaun la la'o, agrikultura la la'o, saúde la iha entaun ema eskola ne'e, eskola sai saida. Eskola folin la iha, osan la iha, buat hotu la iha, ema inan-aman baruk, lakohi haruka sira-nia oan sira ba eskola, tan buat hotu-hotu la iha.

tx@JLS   Kuandu edukasaun, horibainhira iha dois míl katorze ne'e edukasaun hetan an doze porsentu ne'e jerál, duabelas persen. Ne'e doze porsentu bá jerál, edukasaun jerál, husi ensinu báziku to'o ensinu superiór. Entaun mai fali iha dois míl kinze redús fila fali ba dés porsentu, dés porsentu. Entaun mai iha ensinu superiór hetan de'it sinku vírgula seis porsentu bá orsamentu jerál do estadu. Tanba orsamentu ne'ebé governu atu aloka ba iha edukasaun ne'e sei menus liu fali, programa edukasaun ne'ebé la'o husi ensinu báziku to'o iha ensinu superiór. Ida-ne'e mak nia persentajen ne'e ladún la'o di'ak. Ida fila fali kuandu edukasaun ninia persentajen ne'e dala ruma ema dehan katak ami Timor-oan rasik servisu mak la hamutuk entre ulun-boot sira to'o mai iha kraik ne'e ami servisu la hamutuk hanesan ne'e. Entaun fahe malu. Ida-idak mai tau ninian, ida ne'ebá di'ak liu,

ida ne'bá di'ak liu, ida-ne'e di'ak liu, entaun mosu tendénsia política antagónika ho non-antagónika, fahe malu iha klaran. Ita komprende, to? Ne'e mak problema ida iha ne'e.

tx@JLS   Razaun ida de'it. Uluk malae-mutin hanorin portugés ne'e só zeru vírgula sinku, zeru ne'e koma tiha depois lima ba tan ne'e, ema uitoan de'it. Ema dehan uneducated people mak barak liu. Uneducated people ne'e barak liu tanba ema ne'ebé la eskola, la bele ko'alia lian portugés iha tempu ne'ebá. Maibé só ema ne'ebé eskola de'it mak bele ko'alia lian portugés iha ne'e. Agora mai fali iha Tetun, ha'u bele dehan hanesan ne'e. Tetun ne'e mesmu Timor-oan sira la hatene lee, hakerek. Iha sékulu sanulu-resin-neen, iha sékulu sanulu-resin-neen, míl kiñentus doze, tinan rihun ida atus lima sanulu-resin-rua, bainhira Portugál sira haksoit ain mai iha, mai iha Goa, Malaka, antes sira atu tama mai iha Oekusi ne'e, Timor-oan sira ko'alia tiha ona Tetun. Tetun ne'ebé ita la'ós bolu dehan Tetun Prasa maibé Tetun Terik, ne'ebé sira ko'alia ona iha Webiku Wehali. Webiku Wehali iha ne'ebá sira ko'alia mai iha Suai, Suai bá iha Natarbora, Soibada, Welaluhu, sira halai fila fali ba iha, naran sá, Viqueque, entaun to'o iha ne'ebá ne'e, iha Luka, iha Viqueque vila, iha Bikari sira ne'e ko'alia Tetun Terik kuaze atu hadulas tiha. Iha ne'e Tetun Terik, maibé la'ós iha parte Timor-Leste de'it, maibé Timor ida ema bolu dehan Timor ida husi osidentál mós ko'alia Tetun Terik. Entaun Tetun ne'e ema ko'alia nanis ona. Maibé komu la hatene hakerek, la hatene lee, entaun sira la bele hakerek hela istória lian Tetun ne'e iha tempu ida ne'ebá. Entaun komesa ona iha ohin ha'u dehan iha míl novesentus vinte mai ohin komu ema Timor-oan sira metenek ona. Iha fali iha míl novesentus vinte komu misionárius sira hala'o ona sira-nia edukasaun to'o iha kuarta klase, SD kelas empat, hanesan ne'e. Entaun komesa misionárius sira ne'e dezenvolve ona sira-nia lian Tetun ida Timor-Leste nian ne'ebé tau ba iha Tetun Prasa hodi halo misa. Ka hala'o misionáriu sira-nia knaar iha igreja para hotu-hotu tenke aprende lian ida de'it, para sira bele fásil, fasilita sira para bele dezenvolve mós relijiaun iha Timor-Leste. Tanba lori tinan barak, lori sékulu barak ona, dékada hira hanesan ne'e. Tan Timor-oan sira seidauk bele hatene lee, hakerek entaun jesuita sira fó dalan ba Timor-oan. Ha'u-nia perguntas Tetun ladún barak. Politik de'it mak barak.

0054TD

tx@NSG   Hateten naran, né? Ha'u N.... Ha'u servisu iha Institutu Nasionál de Linguístika. I ha'u mós hanesan dosente hanorin Tetun iha departamentu lian Tetun nia.

tx@NSG   Ha'u husi Suai, Covalima.

tx@NSG   Ha'u lian língua materna Tetun Terik.

tx@NSG   INL ninia esforsu halo ona dezde tinan rihun rua rua liubá asinke Timor-oan sira deside atu uza Tetun no portugés nu'udar lian ofisiál ba nasaun ne'e. INL hetan kedas servisu todan ne'e hodi dezenvolve, buka hatene oinsá Timor-oan sira bele uza Tetun ne'e la'ós Tetun nu'udar lian ida hori ko'alia maibé tenke tama ona ba hanesan lian sira iha

mundu ne'ebé lori hodi hakerek, halo komunikasaun ofisiál sira. Nune'e INL halo ona padraun ida para ema hotu-hotu atu alende hakerek no uza termu sira ne'ebé Tetun nia, katak ka liafuan barak ne'ebé foti hosi portugés i balu mai hosi Tetun Terik i nune'e ema barak atu ko'alia sira tenke entende malu. Ne'e INL halo ona esforsu hodi disemina Tetun ne'e ba rai laran tomak liuliu ba profesór sira atu nune'e Timor-oan sira bele aban-bainrua ko'alia Tetun oin ida de'it. Sira kuandu ko'alia bele entende malu, selae, tanba liafuan sira hosi Tetun ofisiál ne'e ninia fonte ne'e mai hosi lian rua, ho portugés ho Tetun Terik. Ne'e ba parte balu ne'ebé estuda portugés nia iha vantajen nian boot liu para atu hatene lian sira mai hosi portugés. Nune'e balu ne'ebé la hatene portugés, sira koko ona atu estuda, atu hatene. Nune'e mós ho Tetun Terik. Ne'ebé ba ema mak dalen foun sira, sira konsidera lian sira ne'e foun ba sira.

tx@NSG   Termu Tetun Prasa ne'e tuir Geoffrey Hull ninia estuda ne'e foin moris iha tinan atus ida, atus rua liubá bainhira Portugál sira harii sira-nia sidade iha Dili. Maibé Tetun ida Terik ne'e mak tanba uza vokabuláriu sira mai husi portugés, Tetun ida Prasa ne'e uluk ema ko'alia iha Dili, iha kapitál. Ne'e, Tetun Prasa ne'e liafuan barak ne'ebé invade ba Tetun. Ne'e tuir Geoffrey Hull, nia dehan kriolizasaun daruak. Iha ne'ebé iha dehan kriolizasaun ida uluk. Sira deve sira-nia lian tama uluk mai malaiu, sira-nian tama ho lian lokál sira seluk tau ba Tetun. Ne'e tuir Geoffrey Hull, kriolizasaun daruak nia ne'e mak portugés ninia vokabuláriu barbarak tama. Nune'e Tetun Prasa ne'e dezenvolve hosi kedas Tetun Terik nian. Ne'ebé, porezemplu, Tetun Terik ninia verbu sira mai iha Tetun Prasa, sira uza de'it jerál. Porezemplu, iha parte balu sira la, hanesan verbu sira, sira ignora tiha ema Tetun Terik sira-nian. Porezemplu, iha Tetun Terik ne'e dehan ha'u ká, ó má, nia ná, ita há, sira rá. I kuandu tama ona ba Tetun Prasa, sira uza de'it mak ida ita, ita han. Entaun tama iha ne'e han de'it. Iha liafuan balu, porezemplu tama iha Tetun Prasa hanesan babeer, crista do galo, manu nia babeer. Liafuan ne'e mai husi Tetun Terik hanesan to'o ohin loron ema la hatene ida-ne'e la'ós hosi Tetun Terik. Tetun Terik nian beer, beer mak hanesan gosta. Mas ohin loron tama iha Tetun Prasa ne'e, beer la iha, nia babeer hotu kedas. I ninia diferensa mak ne'e Tetun Prasa. Portugés ninia vokabuláriu tama barak enkuantu Tetun Terik ne'e sei mantén ninia orijen.

tx@NSG   Ne'e kala tama uluk kedas. porezemplu, iha liafuan malaiu ou Tetun komu sira mai hosi Austronézia i liafuan barak hanesan nanis. Porezemplu kuda, liafuan sira, lima, ne'e tama nanis, ida uluk. Mas liafuan foun sira oras ne'e ema barak sei uza mak numerál sira, porezemplu ida, rua, tolu. Sira asvezes konta satu, dua, tiga. Ne'e sira ne'e tama mais liafuan sira hanesan uluk Tetun Prasa nian iha ona. Ne'e Indonézia nian alat tama. Porezemplu, liafuan sira hanesan bondia, bonoite. Ne'e, oras ne'e ema la uza selamat pagi, ema la uza ona.

tx@NSG   Iha Tetun Terik iha sufiksu ida -n. Ida-ne'e mak tama mai Tetun Prasa. Kleur ona sira muda tiha. Porezemplu, ha'u-nia inan iha Tetun Prasa, mas iha Tetun Terik sei ha'u inan. Ha'u inan. Ou liu rona ha'u-nia livru, iha Tetun Terik sei dehan livru ne'e ha'un. Mais Tetun Prasa ohin loron nia tama ona ba Tetun Terik sira. Tetun Terik sira

mós oras ne'e hatene ona uza Tetun Prasa, tanba Tetun Prasa ne'e mak ninia hun. Ohin loron sai Tetun nasionál, depois sai entaun Tetun ofisiál.

tx@NSG   Sin, sin.

tx@NSG   Ó má.

tx@NSG   Nia ná, sin. Sira la uza?

tx@NSG   Sira la uza. Kuandu sira enfrenta ema estranjeiru, sira koko atu uza Tetun Prasa.

tx@NSG   Ita se bá halo peskiza iha Fohorén, iha Suai, sira sei uza.

tx@NSG   Sin.

tx@NSG   Suai Loro.

tx@NSG   Suai Loro. Sira ne'e sei uza. Sira sei uza. Mas sira mós kuandu ko'alia baibain ho ema ne'e ne'e, ho ema estranjeiru sira la uza sira-nia Tetun lokál. Sira uza ona Tetun Prasa. Nune'e sira asvezes dehan mai ita han, ha'u han ona. Ne'e Tetun Prasa nian sira uza. Mas Tetun ida lokál ne'e só sira uza entre sira, entre komunidade, ho komunidade sira iha suku ne'e.

tx@NSG   Pois é. Língua lokál, em termos de abranjente territóriu ne'e nia fatin oan ne'ebá de'it sira ko'alia. Mas Tetun Prasa, ohin loron sai Tetun ofisiál ne'e, ninia abranjente territoriál ne'e, kobre tomak territóriu Timor. Entaun sira-nia statement bele, bele loos.

0055TD

tx@MAS   Di'ak. Ha'u M.... Ha'u servisu iha Institutu Linguístika Nasionál nu'udar mós dosente iha departamentu inglés hanorin dalen inglés ba estudante sira. Ha'u mai husi Vemasse. Ha'u-nia lian inan Midiki. Ha'u mós bele ko'alia lian lokál seluseluk hanesan Fataluku, Galolen, Makasae, Naueti, Kairui, bele ko'alia bahasa indonézia.

tx@MAS   Ha'u hanoin nia ne'e, lian Tetun nu'udar lian ofisiál ne'e, la bele halo lakon lian lokál sira seluk. Tanba lian Tetun ne'e, ema lokál sira kuandu ko'alia, bainhira sira la hatene, ou ema ne'ebé sira ko'alia ho ne'e la hanesan, sira la hatene sira-nia lian inan. Entaun sira uza lian Tetun ne'e nu'udar lian hodi ko'alia ba malu. Ha'u hanoin nia la bele halo lakon lian lokál sira seluk.

tx@MAS   Ha'u hanoin ha'u hakerek ona disionáriu badak kona-ba lian Midiki ho lian Waima'a. Iha liuliu Waima'a no Midiki ne'ebé ko'alia iha zona Vemasse, Baucau nian. Disionáriu ne'e hakerek ho disionáriu Midiki-Tetun, Waima'a-Tetun. Depois ha'u mós hakerek iha Inglés, Midiki-Inglés, Waima'a-Inglés. Ha'u la halo peskiza formalmente maibé ha'u hakerek ona disionáriu badak kona-ba lian.

tx@MAS   Tanba lian rua ne'e iha ona Institutu Linguística Nasionál. Iha ona ninia perfíl kona-ba lian sira iha Timor laran ne'ebé ha'u uza ida-ne'e ne'e atu hodi hakerek lian Midiki ka Waima'a ne'ebé tanba lian rua ne'e, sira ne'e rua haree husi vokabuláriu sira-sira atu hanesan. Iha sotake dala ruma la hanesan maibé liafuan barak liu hanesan. I depois ninia perfíl sira iha ona. Entaun haree husi perfíl ne'e ha'u bele hakerek ha'u-nia lian, lian inan.

tx@MAS   Tanba lian sira ne'e, lian sira iha Timor ne'e, maioria hun ida de'it ne'ebé sira-nia maneira ko'alia atu hanesan ne'ebé. La asesu husi alfabetu romanu maski iha son balu ne'e iha ha'u-nia lian Midiki ho mós Waima'a diferente uitoan ho son ne'ebé eziste iha Tetun. Porezemplu ami iha liafuan soe. Ne'e ha'u haree linguista sira fó símbolu s ho, s ho kapa tatolan, so'e. Ne'e ha'u hakerek ne'e loos tanba ha'u haree husi peskiza ne'ebé. Linguista sira iha INL hatuur tiha ona. Maibé karik ha'u husu ba ona malu sira ne'ebé la estuda ne'e ne'e sira-sira hakerek diferente. Maibé son loloos maka ida-ne'ebé hatuur tiha ona iha, ou hakerek tiha ona nia perfíl lian Tetun, ah, lian Midiki ho Waima'a.

tx@MAS   Ha'u-nia iha, iha, iha ha'u-nia subdistritu iha suku ida sei ko'alia ne'e ne'e. Maibé tanba agora ne'e ema tuur hamutuk de'it ona. Iha ema tuur hamutuk de'it ona. I lian sei kahur malu iha ha'u-nia suku rasik. Ha'u haree ema halai liu, labarik sira ikus mai ne'e, sira uza barak liu Waima'a. I lian Midiki ne'e, sira ne'ebé ida de'it boot ko'alia. Ha'u-nia inan mak kala hanesan uitoan, hanesan uitoan ba lian Midiki, tanba sira iha ha'u-nia suku, sira uza fali lian Waima'a ne'e hanesan língua franka ida. Tanba iha ha'u-nia subdistritu, liuliu ona suku ne'e, iha lian boot rua iha ne'e, iha Galolen ho, Galolen ho Waima'a. I iha mós Midiki. Maibé kuandu ema Galolen ho Midiki sira atu ko'alia sira uza Waima'a fali. Ida-ne'e haree hanesan ami sei uitoan Midiki.

tx@MAS   Iha ha'u-nia família, ami uza Tetun. Maibé se ha'u ko'alia ho ha'u-nia alin sira ne'ebé husi ne'ebá mai, ami ko'alia ho ami-nia lian materna, lian inan. Maibé ha'u ho ha'u-nia oan sira, ha'u ko'alia Tetun. Ho ha'u-nia kaben, tambén nia lian Fataluku entaun ami ko'alia Tetun de'it. Maski ha'u mós hatene Fataluku, dala ruma de'it mak ha'u ko'alia ona Fataluku. Maibé dala barak liu ami ko'alia Tetun. I ami-nia oan ko'alia Tetun de'it.

tx@MAS   Husi ha'u, ha'u la hatene ho nia problema saida. Maibé husi parte estudante, ha'u haree problema hatene ka koñesimentu ne'ebé. Estudante sira lori mai, ne'ebé halo ha'u atu hakat lalais iha hanorin ne'e susar. Ne'e mak ha'u hakarak avansa liu maibé saida mak sira hatene ne'e menus liu. Entaun halo ha'u tenke la'o neineik. Ne'ebé ida-ne'e mak ha'u hanoin ha'u-nia problema boot tanba ha'u hakarak ha'u-nia estudante sira hatene di'ak saida mak ha'u hato'o ba sira. Maibé ida-ne'e ha'u hatene liu husi teste ne'ebé ha'u fó ba sira. Barak la hetan nota di'ak. Ka lae ida-ne'e iha relasaun ho matenek, koñesimentu ne'ebé sira lori husi kraik mai.

tx@MAS   Iha ida-rua ko'alia di'ak, mais maioria sira tenke haka'as an. Tanba dala ruma sira mai iha ne'e, kala haree ema barak tama iha ne'e i sira mós hakarak hatene. Maibé

478

sira-nia formasaun, sira-nia preparasaun nian kraik ne'ebá ladún forte atu dudu sira mai iha ne'e. Tanba sira hanoin: oh, fatin ne'e di'ak, ka? Entaun sira mai maibé koñesimentu báziku. Ha'u haree husi sira-nia aparénsia. Sira, sira ladún hatudu iha ne'ebá. Ne'ebé ha'u-nia ida-ne'e problema boot iha ha'u-nia, ha'u-nia servisu knaar nu'udar manorin inglés nian.

tx@MAS   Tanba ida-ne'e departamentu inglés, ami profesór sira iha obrigasaun atu ko'alia ho inglés. Tempu barak liu ko'alia inglés, pois. Halo maklotak sira ka estudante sira rona oinsá ema ko'alia inglés ne'e. Maibé iha momentu ne'ebé kuandu ami esplika konseitu ruma ne'ebé susar ami bele uza Tetun i até lian lokál sira. Tanba ha'u hatene lian Midiki, Fataluku, Galolen, ha'u mós bele esplika ho Galolen ba sira ne'ebé hatene lian ne'e. Ou sira bele hatene sá ha'u esplika ho Tetun para ajuda sira komprende. Tanba se ita uza de'it inglés dala ruma vokabuláriu loos aas tan depois konseitu sira susar. Se ita la tulun sira ho lian sira ne'e hanesan ne'ebé sira hatene, ne'e hosi sira la komprende di'ak saida mak ita hato'o ba sira.

tx@MAS   Ha'u hanoin sei iha. Ne'e la'ós iha universidade de'it. Ha'u hanoin iha eskola sira tomak iha Timor laran tomak. Nia akontese tanba manorin sira ka profesór sira iha Timor laran tomak ne'e barak liu forma iha tempu indonézia. Tanba ne'e sira sei uza bahasa indonézia. Maibé formasaun língua portugeza ba profesór sira la'o nafatin iha hein katak ke lalais ka neineik profesór sira sei ko'alia di'ak portugés hodi hanorin. Instituisaun ne'e sei mós profesór Timor-oan sira, hasoru bainhira Indonézia sira tama mai iha ami-nia rain i UNTL ba profesór Timor-oan sira mós la hatene bahasa indonézia. Sira haree matéria ne'e bahasa indonézia maibé sira komprende ou portugés ou sira-nia lian. Sira esplika ba estudante, uza sira-nia lian ou Tetun ou bele mós portugés. I situasaun hanesan ne'e agora akontese. Tama uluk mestre barak hetan formasaun bahasa indonézia i siénsia ne'ebé sira hetan ho bahasa indonézia. I agora sira tenke transfere fali siénsia ne'e iha portugés. Ida-ne'e la'o mas neineik uitoan de'it. I presiza tempu para bele la'o, la'ós katak portugés sususar atu saida. Maibé tama formasaun ne'e nune'e hein katak sei di'ak iha futuru tanba uluk hanesan ha'u foin ko'alia tiha ona, Indonézia tama ami-nia rai mós hanesan ne'e. Profesór Timor-oan barak la hatene bahasa. Sira iha tiha formasaun barabarak, barabarak ikus mai. Sira hotu-hotu ko'alia ho bahasa indonézia. I situasaun ida-ne'e mak agora profesór Timor-oan sira ne'ebé uluk hetan formasaun ho bahasa indonézia hasoru.

0056TD

tx@CS   Mana, ha'u-nia família só ida de'it mak servisu ba kona-ba profesora, hanorin iha daudauk ne'e, iha edukasaun, maibé ha'u-nia oan, la'ós ema. Iha ha'u-nia biin hanorin profesora iha Amigo de Jesus. Só ha'u de'it mak servisu hanesan ne'e kona-ba cleaning. Maibé la iha seluk, la servisu tama, la iha, la iha forsa. La iha eskola ba iha matenek tanba eskola iha kraik de'it. Entaun sira la bele hetan servisu ba kantor. Sira tensér servisu ba

iha, narsá, estrada, halo ba estrada, halo ba buat ne'ebé, pinta, buat sira ne'e. La iha osan ba ha'u-nia oan hanesan hanorin maibé foin hanorin seidauk mós hanesan sai ba profesora, maibé hanesan provizór, provizores, hanesan ne'e de'it. Maibé moris família, la iha barak mak servisu iha kantor. Maibé sei servisu ba iha de'it hanesan bá iha to'os, bá iha estrada, bá sira hanesan hadi'a, bá buat sira ne'ebé iha estrada, sira ne'e, pinta buat sira ne'e. La iha servisu. Família la servisu tanba la iha eskola boot entaun sira hotu-hotu servisu la iha. Balun mak iha uma de'it. Balu hakarak ne'e buka badiu iha estrada, tuir estrada. Família la servisu. Ha'u-nia família ne'e tan barak la eskola. Entaun la, la, ne'e iha ida só, rua de'it mak servisu iha kantor, maibé rua ne'e mak sira profesora hotu. Ida bá profesora TK edukasaun nian. Ida profesora bá iha Amigo de Jesus. Señór padre Joao Felgueiras, amigo Jesus-nia, nia hanorin iha ne'ebá. Hanesan seluk la hetan ona. Só ha'u de'it mak servisu ba hanesan cleaning.

tx@CS Mana, ha'u eskola i kala iha segunda klase ida de'it. Ha'u la eskola boot ida, la eskola.

tx@CS Iha tempu portugés eskola segunda klase. Tanba iha ha'u-nia pai ho main ne'e la iha forsa para atu selu ha'u bá eskola. Entaun ami tensér sai para... Pai mós mate ona, main mós mate ona, ami sai eskola para bele kaben tiha de'it. La'en sira mós, saida, kaben maibé ha'u-nia kaben mós mate tan tiha. Agora ha'u mesak mak buka servisu hanesan ne'e, moris hanesan ne'e de'it. Malae hetan ona, mana, agora ha'u mesak maka servisu hanesan ne'e para haree sira na'in-haat. Ida sira agora mak atu bá universidade UNTL. Ida mak agora profesora. I rua maka sei eskola bá iha li'ur. Ida bá eskola iha Inglaterra. Ida mak agora iha be Indonézia.

tx@CS Indonézia iha Bali.

tx@CS Ida de'it mak agora tinan, hela tinan ida tan ne'e sura hotu ona, remata ona eskola para mai, mai Timor. Naran mós Isabel. Ha'u (...) tan ida mane ne'e. Agora mak la hatene se Maromak ajuda para nia bele liu UNTL. Se nia eskola, selae, ha'u haruka nia ba eskola iha Bali. Eskola tan hanesan nia biin bá foti turizmu iha Bali. Tanba biar ha'u servisu hanesan ne'e maibé ha'u hakarak halo sira bá eskola. Biar osan la barak. Maibé ha'u hakarak sira hanesan buat ida karik ha'u tensér dudu sira. Ha'u la eskola maibé ha'u hakarak halo sira bá eskola.

tx@CS Ha'u edukasaun importante liu ba ha'u. Tanba maibé ha'u, ha'u-nia pai i main ne'e fó dunik ha'u eskola maibé eskola uitoan de'it. Sira la iha forsa eskola, fó eskola ha'u. Entaun ha'u hakarak sira tensér bá eskola oioin, tensér iha futuru ne'ebé sira-nia. Se aban-bainrua se ha'u mate tiha karik, sira mós bele iha sira-nia be kapasidade para bele haree sira-nia aban-bainrua, sira-nia moris, sira-nia família, sira ne'e, bá iha sira-nia família aban-bainrua ninian. Se agora sei tuur, ha'u sei haruka, ha'u sei bele haforsa an, ha'u bele haruka sira sei eskola. Se ha'u hanesan la forsa ona, pasiénsia. Pasiénsia tanba la iha forsa. La iha forsa, tensér la haruka sira ba oin. Mas se iha forsa, tensér halo sira bá oin. Buat ne'ebé sira hakarak, ha'u mós tensér halo sira bá. Se sira lakohi mas ha'u tensér

obriga sira para sira bele bá oin. La bele sira mai fali kotuk. Ha'u lakohi tanba iha sira-nia tempu agora ne'e, tempu ba sira nian. Ne'ebé ita inan ita, aman ita tensér haforsa ita-nia an para ita-nia oan sira bá oin. Agora hanesan, ba ha'u, ho idade ona ha'u la bele servisu tan forsa. Ha'u forsa la iha ne'e. Ha'u servisu hanesan ne'e mas ha'u dehan di'ak, di'ak ba ha'u. Hemu de'it tua maibé hanesan, mana, hanesan, Maromak bele ajuda tan ha'u, bele fó fali ha'u-nia vida naruk, halo ha'u mós reza barak ba ha'u oan sira sei aban-bainrua eskola. Hanesan mós mana agora mai daudaun, hasai mana mós bele husi... Mana nia rain naran saida?

tx@CS   Slovakia. Bele mai Timor. Aban-bainrua ha'u mós hakarak sira bele eskola. Bele bá fali rai seluk, mana, bá Inglaterra ou atu bá Indonézia, ou atu bá Portugál. Bele halo buat ne'ebé hanesan mana mós agora mai la'o daudauk, hanesan mós mai ami Timor ne'e. Ida-ne'e mak ha'u-nia hakarak de'it.

tx@CS   Ha'u husi, ha'u-nia papá, ha'u-nia papá ema Ainaro. Ainaro. Ha'u-nia mamá ema Ermera. Maibé ami moris iha Dili, moris iha Dili, ami moris.

tx@CS   Tetun. Tetun. Ha'u-nia apá mós Tetun. Ha'u-nia apá maibé Tetun, Tetun maibé ho Mambae, Ainaro distritu. Ha'u-nia mamá mós hanesan, Mambae, Ermera distritu. Maibé hanesan ami mak ladún hatene. Hanesan rona, rona, rona maibé ami ladún. Komprende uitoan de'it maibé hanesan ha'u, hanesan ami hatene, tanba ami sempre bá iha ha'u-nia pai i ha'u-nia main distritu, sempre iha avó ko'alia mai ami, ami rona.

tx@CS   Família barak liu iha distritu. Família barak iha hanesan ha'u-nia mamá nia alin mós sei iha. Husi ha'u-nia papá nia mós iha maibé iha kala haat, haat de'it. Mate hotu. Indonézia tiru ba mate hotu. Balu bá Atambua, Indonézia mate. Maibé husi ha'u-nia mamá nia mak tomak liu, husi barak. Iha distritu Railaco, distritu Ermera, Bazartete. Ha'u-nia mane família de'it. Família tanba iha alin, hola feto bá distritu Bazartete, distritu Railaco, Ermera. Ha'u-nia mane família ne'e sei hotu. Hola, halo família barak, halo família barak, ne'ebé maibé to'o iha Dili, ami maibé iha Dili. Maibé ami, hanesan, karang-karang hakarak haree família ou sira mós mai ba vizita, vizita sira. Ami lori mós ami-nia oan sira bá haree sira, iha sira-nia distritu, haree família sira, dehan di'ak ka lae. I família sira hateten mai ami mós dehan katak, ami hanesan ami-nia, bele haka'as an nia oan sira eskola didi'ak. Sira ami mós hateten ba sira, sira mós tensér eskola didi'ak para aban-bainrua atu halo ami-nia rain ne'e halo di'ak liu tan, iha Timor ne'ebé di'ak liu tan. Hanesan mós la bele iha. Akontese iha buat ruma mai ikus fali. Iha mate barak. I depois iha mate barak tanba ami haree ba Indonézia. Iha ami joven barak mate, maibé mate Santa Cruz. Tanba ha'u iha ha'u-nia oan ida agora daudauk profesora ne'e, sei ki'ik uitoan, halo tinan haat, maibé ami na'in-rua bá merkadu iha daudauk, agora merkadu halo bá iha daudauk ne'e, merkadu lama ne'e. Mana hatene? Iha mikrolete, para, merkadu lama ne'e, uluk iha ne'e. Indonézia halai ne'e maibé ha'u ho ha'u-nia oan, mana, ha'u hola modo, hola modo fa'an, ha'u lori ba fa'an. Indonézia, ha'u tama iha klaran, Indonézia tiru. Mana, ha'u nakdedar, ha'u tama iha baleta kuak. Mais to'o agora ha'u la iha halo buat sira ne'e bá iha, halo hanesan, atu hateten ba mas ha'u lakohi halo. Tanba ha'u-nia rai, ha'u la

presiza atu halo buat sira hanesan ne'e, tanba ne'e ha'u-nia rai. Ha'u lakohi atu oras ida de'it bá halo hanesan hakarak osan tanba ha'u mate. Ha'u sei moris mas ha'u mate. Ha'u hakarak atu bá halo osan, atu simu osan veteranus. Ha'u mós ha'u sei mate maibé ha'u-nia rai ha'u haree de'it, ha'u haree tan ha'u-nia maluk mate barak. Tiru, ha'u ho ha'u-nia oan ha'u konta hela hanesan ne'e, ha'u tama iha baleta ne'e. Mana, baleta hanesan ne'e ne'e, ho fo'er, ho plastik. Ha'u tensér taka de'it. Ha'u lori ha'u-nia hena ne'e hodi taka metin ha'u-nia oan. Mane mak tiru, ha'u la bele sai. Kalan boot, tuku ida, ha'u foin sai ho ha'u-nia oan bá uma. Ema hateten ha'u mate. Mas ha'u la mate, maibé ha'u reza maka'as. Mana, ha'u akontese dala rua. Akontese ida iha ha'u, iha Indonézia ida tempu tiru iha ne'e. Akontese ida ami halai Indonézia sai ona. Ha'u mai ha'u-nia uma atu foti modo ho foos, ha'u-nia susubeen ba ha'u-nia oan. Bapa kesi metin hotu kilat, kompletu hotu. Mana, ha'u ho ha'u-nia oan feto ida, terseira filla, atu mai foti sasán. Mana, ha'u subar iha ai-tahan ida de'it. Ai-tahan ida de'it hodi subar ha'u ho ha'u-nia oan. Mas, ha'u, hanesan ne'e, bá ema liu hanesan ne'e. Ha'u mak ne'e, mas ema la haree ha'u. Mas tanbasá mana ha'u, ha'u reza barak. Ha'u tensér reza barak. Sebelum ha'u ema kiak maibé tensér reza barak. Tanba ha'u-nia rai, ha'u reza ba nia, sei rai ne'e sei taka nia matan, nia tensér liu de'it. Nia la bele haree ha'u. Mas, mana, liuliu tiha, depois ha'u, ha'u sai. Ha'u bá. Ne'e bolu hanesan ne'e, ha'u dehan ha'u foti, ha'u foti foos. Atu bá ne'ebé? Ha'u dehan atu bá Indonézia. Mas lae, ha'u la bá Indonézia. Ha'u sa'e. Ha'u bá Dare. Bá Indonézia? Ha'u dehan ha'u bá Indonézia, agora sa'e karreta bá Indonézia. Mas lae, mana, ha'u ho ha'u-nia oan sa'e, halai liu. Ah, mana, ema atu tama agora, hanesan ha'u ta'uk, ha'u trauma. Ha'u hanoin barak. Ha'u trauma. Ha'u dehan, ha'u reza barak tanba iha dala ruma akontese iha ikus ha'u ho ba ha'u-nia oan sira nia moris. Ha'u la hatene. Halai moris ou mate. Tanba akontese dala rua ona ema la haree ha'u. Maibé keta ikus mai fali, bele akontese. Ikus, ha'u sei mate, maibé ha'u reza Maromak. Maromak sei, sei haree ha'u, ho ha'u-nia oan sira, ha'u-nia família tomak. Balu iha Indonézia, balu Inglaterra, balu bá, balu iha Indonézia militár. Ha'u iha ha'u-nia main nia alin sira maibé ha'u mós reza ba sira bá halo funu karik. Reza ba sira para sira mós bele fila mai husi ne'e família halo di'ak. I sira mós dala ruma sei mai haree to'o ha'u. Ha'u reza barak. Halo buat sira ne'ebé ke di'ak iha Timor ne'e. Ita só, só ita mak haforsa ita-nia an. Se ita la haforsa ita-nia an, mana, buat hotu-hotu sei la di'ak ba ita-nia nasaun tanba ita mak husik ita-nia nasaun. Ita dunik mak halo ita-nia nasaun fo'er, halo buat ne'ebé ka la di'ak ba ita-nia nasaun. Kuandu ita halo di'ak, ita-nia nasaun sei di'ak. Kuandu ita halo la di'ak, ne'e nasaun mós sei sai la di'ak. Sai la di'ak mai ha'u.

tx@CS   Nada.

0057TD

tx@LAOB   Ha'u naran L.... Moris iha Dili, iha loron sanulu-resin-rua, fulan Juñu iha tinan rihun ida atus sia ualunulu-resin-lima. No ha'u mai husi família, husi pai, ha'u-nia aman mai husi Baucau. No ha'u-nia inan mai husi Dili. Ha'u eskola iha, eskola primária,

iha eskola primária bázika número kuatru ne'ebé besik ha'u-nia bairru, Bairro Pité. Iha tinan rihun ida atus sia sianulu to'o tinan atus, rihun ida atus sia sianulu-resin-neen iha eskola primária número kuatru Bairro Pité. I depois husi tinan rihun ida atus sia sianulu-resin-neen to'o sanulu rihun ida atus sia sianulu-resin-sia ha'u eskola iha pré-sekundária número dois iha Comoro, Dili. I depois iha ne'ebá iha tinan rihun ida, rihun ida atus sia sianulu-resin-sia ha'u kontinua bá eskola sekundária, bá iha eskola ida naran eskola sekundária número um Becora. Maibé la konsege remata tanba iha Setembru, loron Setembru, ya, Agustus ami tenke feriadu tanba atu hahú halo votasaun populár. I entaun ami eskola feriadu no la hotu tanba i depois iha Setembru iha funu, iha Timor. Entaun depois fila fali iha rihun, tinan rihun rua, ha'u kontinua eskola, ha'u kontinua iha Colégio Paulo VI. Maibé ami aprende de'it lian inglés no portugés ba fulan balu. Maibé depois ha'u muda fali ba eskola públika, hanaran eskola sekundária número dois UNAMET Dili iha Balide, to'o iha tinan rihun rua tolu. Depois de tinan rihun tolu ha'u finaliza ha'u-nia eskola sekundária, ha'u bá kontinua iha Universidade Nasionál Timor Loro-Sa'e, iha kooperasaun ho governu portugés. Sira iha implementa programa ida hanaran Fundasaun das Universidades portugueza. Ha'u foti ha'u-nian kursu siénsias agrares kona-ba agrikultura. No ha'u remata iha tinan rihun rua, tinan rihun rua, tinan rihun rua sanulu. Ha'u remata ha'u-nian, ha'u-nian lisensiatura iha ne'ebá. Presiza tinan neen ho balu hodi finaliza eskola tanba iha tinan rihun lima, to'o rihun lima, karik rihun lima neen, iha krize. Entaun ami iha falta aulas. Entaun ami tenke halo fali aula seluk hodi kompensa aulas ne'ebé ami lakon durante iha krize. Agora daudaun ha'u servisu iha programa ida, iha ne'ebé agora remata iha Dezembru ne'e, iha ajénsia internasionál husi Alemaña, Kooperasaun Internasionál Alemaña nian iha Timor-Leste. Iha programa ida hanaran dezenvolvimentu, programa dezenvolvimentu rurál ba dala haat ne'ebé servisu hamutuk ho Ministériu agrikultura no peskas, liu kona-ba oinsá mak atu haforsa servisu funsaun, servisu estensaun rurál nian. Maibé antes ha'u, antes ha'u servisu iha fatin seluk, iha ajénsia internasionál, ajénsia internasionál seluk hanesan ho programa USAID, ne'e ho servisu bá liu iha eskolas téknikas agríkolas iha Timor Loro-Sa'e tomak, iha Maliana, iha Fuiluru, Natarbora, no mós iha Oekusi. I depois ha'u hetan servisu ida iha Baucau. Maibé ha'u tenke bá servisu iha Viqueque, Ossoroa. Ne'e hamutuk koorganizasaun ida hanaran CRS, Catholic Relief Service, programa kona-ba seguransa alimentár nian. I depois ha'u mai fali Dili. Ha'u servisu iha nafatin ho organizasaun ida ho husi organizasaun amerikana ida, naran ACDI-VOCA ne'ebé servisu liu kona-ba hakiak i, hakiak kadiuk.

tx@LAOB   Tranzisaun husi noventa nove?

tx@LAOB   Tempu tranzisaun nian iha, depois iha tinan rihun ida atus sia sanulu-resin-sia. Iha tempu ne'ebá la fásil atu enfrenta, la fásil atu enfrenta moris. Tanba buat hotu-hotu difisil atu ita hetan. Iha Nasoins Unidas no sei iha ema barak no mai ho multikulturál no hanesan ema ne'ebé moris mós iha faze, iha períodu ne'ebá tenke mós aprende atu oinsá adapta ho ambiente foun. Ita la hanesan tanba husi liuliu edukasaun. Primeiru ho remata husi edukasaun husi tomak ho lian indonézia, ne'ebé ha'u rasik estuda durante husi eskola primária to'o iha, to'o iha eskola pré-sekundária, liuliu hamutuk ba tinan sia

ne'e ho lian indonézia. Maibé depois ida-ne'e, tranzisaun ida-ne'e, mós ami nafatin estuda lian indonézia tanba profesór sira, maioria sira eskola, finaliza sira-nia eskola mós husi ho lian indonézia, indonézia. Entaun maibé sira seidauk preparadu atu sai profesór, atu hanorin. Sira mós la iha, barak mak la iha esperiénsia, esperiénsia nu'udar profesór no ema hotu-hotu, profesór sira, momentu ne'ebá, sira sai profesór loloos también sira-nia karreira, tan loloos sira-nia profisaun maibé tanba de'it atu hetan osan, tanba de'it ekonomia, atu bele sustenta mós sira-nia moris. Entaun sira sei de'it bele sai profesór. Entaun iha momentu ne'ebá edukasaun, sistema edukasaun i eskola, liuliu eskola públika fraka tebetebes. Ami la iha kadeira i la iha livrus. Ami la iha biblioteka. No dala barak profesór sira falta. La iha justifikasaun. Ou balu falta maibé sira fó tuir livru. I depois fó responsabilidade ba estudante hodi hakerek iha kuadru. No estudante sira kopia saida de'it mak hakerek iha kuadru. Pois, nia mai tama, la iha esplikasaun, revista sein hakerek tuir ou lae. Entaun iha ne'ebá tanba ha'u rasik iha tempu indonézia ha'u eskola iha eskola ne'ebé di'ak. I entaun ha'u sente sofre tebetebes tanba ho edukasaun ne'ebá simu. Razaun seluk ha'u la bele muda ba eskola seluk tanba se bá eskola privada tanba ha'u komesa na eskola públika. I se ha'u bá eskola seluk ha'u tenke bá fali ha'u husi primeiru anu, tinan. Ha'u la bele bá fali kontinua hanesan iha anu ne'ebé ha'u eskola, ha'u iha, ha'u, hanesan eskola ida públika. Entaun ha'u ho kolega sira iha ne'ebá ida-ida tenke esforsa an, aprende rasik. No ami fahe malu ho kolega sira ne'ebé ami sente ne'ebé sira hatene liu, matenek liu. Entaun ami la moe atu simu malu, atu hanorin malu fali. I ida-ne'e mak tempu ne'ebé difisil tebetebes mai ami. No depois tanba, ami hotu-hotu hetan valór di'ak. Maibé kualifikasaun di'ak, maibé ha'u rasik la fiar ho ha'u-nia kualifikasaun mezmu ha'u-nia kualifikasaun di'ak. Tanba ha'u sente ha'u la aprende buat ruma husi profesór iha tempu ne'ebá. Entaun saida mak ami enfrenta wainhira tama iha universidade, iha matéria ne'ebé ke, iha universidade ha'u tama universidade, ha'u iha universidade ne'ebé ko'alia lian portugés. Mezmu ida-ne'e, ha'u hasoru problema barak tanba i depois iha universidade mak ami foin, ha'u foin hatene katak, saida mak ha'u estuda iha eskola sekundária, ne'e báziku tebetebes. Hanesan eskola iha pré-sekundária matéria ne'ebá ha'u simu iha sekundária la kompletu. La kompletu. Ha'u hatene husi ne'ebé tanba iha ami-nia turma laran iha kolega balu mai husi eskola, ami mai husi eskola diferente. Balun mai eskola, mai husi eskola ne'ebé privada. Balu mai husi eskola téknika. Balu mai eskola públika hanesan ha'u ho kolega seluk. No ami halo diferensa malu no husi eskola ne'ebé ami mai antes ne'e ami-nian koñesimentu la hanesan. No ami-nia, no matéria ne'ebé ami simu mós la hanesan iha eskola ne'ebé ami estuda antes iha eskola sekundária. Entaun saida maka ami enfrenta. Ami la bele aprende di'ak. La bele komprende lalais. I wainhira profesór ruma esplika buat ruma, loloos, ne'e loloos kona-ba lian. Maibé, porezemplu, iha matemátika, ne'e ita la presiza hatene lian maka lian inglés ka lian portugés ka lian indonézia ka lian Tetun mak ita hatene matemátika. Matemátika ne'e universál. Ne'e se ita hatene ida-ne'e entaun ita, prontu, bele komprende buat seluk. Matemátika, matemátika ne'ebé ha'u aprende iha eskola sekundária ne'ebé ba, mezmu profesór halo teste, ha'u la konsege hatene tanba ha'u la aprende iha eskola sekundária. Mezmu ne'e

iha ke tama iha kurríkulu husi eskola sekundária nian. I entaun ha'u tenke aprende fila fali saida mak ha'u aprende, saida mak seidauk aprende iha eskola sekundária.

Tetun Suai

0017TS

tx@MN   Alin Marcelina, be naran sá husi ne'ebé, mai husi ne'ebé?

tx@MCA   Ha'u kusi Suai Loro.

tx@MN   Ita-nia inan-aman sei moris ka mate ona?

tx@MCA   Ha'u ina sei moris, ha'u ama sei moris.

tx@MN   Ita-nia aman naran saida?

tx@MCA   Ha'u ama naran Luís Amaral, ha'u ina naran Anita Da Costa.

tx@MN   Ita oan pertama ka?

tx@MCA   Ha'u oan ikun.

tx@MN   Oan ikun.

tx@MCA   Ami iha, ami na'in haat, mane ida feto tolu.

tx@MN   Ita eskola, tama eskola saida?

tx@MCA   Ha'u tama eskola SMA.

tx@MN   Iha?

tx@MCA   Iha tinan rihun rua haat.

tx@MN   Agora ó manorin iha ne'ebé?

tx@MCA   Ha'u kanorin iha EBC Suai Loro, kanorin kelas empat.

tx@MN   Ó alunu sira iha na'in hira?

tx@MCA   Ha'u alunus iha tolunulu resin hitu.

tx@MN   Iha ita eskola ne'e iha mestre, mestre na'in hira?

tx@MCA   Eskola EBC Suai Loro ne'e tama loron, tama loron ida i... balu tama sei sawan, balu tama loron manas, sei sawan husi kelas satu to'o kelas lima, loron manas kelas enam to'o SMP.

tx@MN   Iha imi eskola ne'e, ninia be naran sá, kondisaun ne'e di'ak ka lae?

tx@MCA   Iha eskola ne'e kondisaun di'ak mas livru mak sei falta. Sei la to'o.

tx@MN   Kona-ba profesór, to'o ka lae?

tx@MCA   Profesór mós la to'o dauk tanba kelas ida balu sei to'o neennulu resin, ne'ebé profesór na'in-ida nanorin la bele tan ema wa'in liu.

tx@MN   Eskola ne'e nia be narsá sala ne'e to'o ka lae?

tx@MCA   Lae, eskola EBC Suai Loro ne'e sala la to'o ne'e mak be eskola balu sei sawa, balu loron manas ne'e tan sala la to'o. Ne'e tiha sala balu mós aat tiha, nakfera kala anin ne'e. Mestra sira ho alunus sira la iha laran, sai ba luan lae tembok monu.

tx@MN   Ne'e tiha imi sei sawan mai eskola, fila ba uma halo saida?

tx@MCA   Bá to'o iha, bá to'o iha uma ha'u be toba oan ida tiha, istirahat tiha oan ida tiha, ha'u prepara lisaun atu kodi mai kanorin ujian awan nia ka'i.

tx@MN   Depois iha uma imi ajuda imi inan-aman ka lae?

tx@MCA   Iha uma ha'u ajuda ha'u ina te'in ba han loron ho kalan niak, ajuda hotu tiha foin prepara be materiál atu hodi mai hanorin ujian awan nia ka'i.

0018TS

tx@RASS   Entaun señór mak be Diretór Interinu.

tx@JA   Sin ha'u mak Diretór Interinu nian.

tx@RASS   Ita mai husi ne'ebé?

tx@JA   Ha'u husi, ha'u mai kusi Kamanasa.

tx@RASS   Kamanasa, bele ko'alia uitoan kona-ba ita-nia tempu infánsia, tempu ki'ik ne'e ne'e halo nu'usá ita-nia moris loroloron ne'e.

tx@JA   Uluk ami sei ki'ik ne'e, ami halimak, bá to'os, bá hein karau, depois bá eskola iha Debos la'o rai de'it bá ne'e, dala ruma sa'e kuda ba to'os, bá haró kuda, bá ko'a hare ba kuda.

tx@RASS   Karik fila husi eskola, seluk ko'a du'ut ba kuda, iha servisu seluk atu halo tan, tuir be labarik ninian ne'e ne'e iha ka lae?

tx@JA   Ne'e ami, ami sei ki'ik ne'e, ami hi'it maluk, be lori malu ne'e ami bá halimak sa'e kuda ai los, de'ut malu ami le'u lixa, bá hariis mota.

tx@RASS   Kamanasa ne'e hela besik tasi ibun ne'e?

tx@JA   Ami besik tasi, ami hela besik tasi.

tx@RASS   Entaun uluk sei ki'ik ne'e ne'e iha tempu ruma bá halimar iha tasi ka lae?

tx@JA   Ami uluk sei ki'ik ne'e ami hein karau iha be tasi ibun ne'e ami bá hariis tasi hodi ai-notak tama tasi.

tx@RASS   Seluk sei tama tasi sei halo sá tan iha tasi ibun nu'udar labarik iha tempu ne'ebá?

tx@JA   Iha tasi ibun ne'e halimar buat wa'in, dalaruma ami be su'u rós oan iha tasi be, tuir be maun sira be hodi, bero oan be tama tasi, atu bukat.

tx@RASS   Kamanasa ne'e ita haree nuu barak, entaun prosesu be atu be atu sa'e nuu halo nu'usá.

tx@JA   Iha momentu ne'ebá. Ami uluk ne'e e sa'e nuu ne'e ami be kesi tali, kesi tali ne'e ami bolu be sá, ai-kenanak, tau ba ain-telok mak hodi sa'e nuu.

0019TS

tx@NA   Ó oras ne'e, ó malo tinan hira tiha?

tx@BB   Ha'u kalo tinan limanulu resin haat.

tx@NA   Limanulu resin haat. Be uluk ó sei ki'ik ne'e ne'e ó madiuk mak sé?

tx@BB   Ha'u uluk ha'u sei ki'ik ha'u kadiuk de'it ko maun alin sira.

tx@NA   Ho maun alin sira, kala ha'u ho kadiuk maka ho maluk ida maka oras ne'e mate tiha. Tuirmai, iha uluk liu, oras lokraik ha'u ko ha'u maluk oan sira ne'e bá haris mota, ne'e tian dalaruma ami bá haris tasi, halimar bola iha tasi ho kolega ho maluk sira iha ne'ebá. Kala B... kuandu uluk sei ki'ik ne'e ne'e imi hadiuk ne'e iha ne'ebé, iha uma de'it ka ou dalaruma keta mai imi tene malu bá su'u ros ka, buat sira ne'e ne'e.

tx@BB   Lae, ha'u uluk, ha'u mós sei ki'ik, ha'u ko ha'u maluk sira tiha tena malu ami bá hariis mota no tempu lokraik, kalau, sai tiha eskola foin ami bá…

tx@NA   Oras loraik foin bá.

tx@BB   Loraik ami bá, ami sai tiha eskola foin bá hariis mota.

tx@NA   Tuirmai B... eskola ne'e ne'e ho tinan hira?

tx@BB   Ha'u eskola ne'e iha pré-primária ho tinan lima.

tx@NA   Ho tinan lima.

tx@NA   Kala ha'u eskola iha hanesan mós pré-primária, hahú tinan hitu, ha'u tama ona eskola la hatene ho momentu ne'ebá ne'e iha tinan rihun ida atus sia neenulu resin ualu. Ha'u tama eskola, hahú husi uluk ne'e ne'e husi ABC. Iha uluk ne'ebá ne'e ami eskola

iha desa ida, desa Kamanasa iha Temukun Fatu-sik. Iha uluk ne'ebá ne'e ami eskola de'it iha uma tali, uma ne'e la iha, ami halo de'it eskola uma tali. E ami-nia mestra mak agora sei iha, naran be Olivia. I depois to'o tiha be primeira klase ami muda ba iha eskola iha Tebus Mata Air. Seterusnya, ami muda fali ba iha Misaun. Ami halo kuarta klase, ami bá eskola iha Maliana terus kuarta klase atu eskola nafatin maka iha uluk ne'e ne'e tanbá katuas sia eskola. Oan sia hodi de'it tabaku. Ne'ebé inan-aman komu forsa la to'o, ne'ebé iha loron ne'ebá ne'e iha momentu ne'ebá ne'e, ha'u to'o de'it kuarta klase i ha'u para husi eskola. Kalau B..., eskola primária hotu, kuarta klase hotu tiha sei kontinua ka la kontinua?

tx@BB   La kontinua.

tx@NA   Tuirmai iha tinan rihun ida atus sia hitunulu resin lima, ha'u sai hanesan profesór iha eskola primária Kamanasa to'o ohin loron. Señora B..., depois eskola kuarta klase hotu tiha, ohin la hahú, la para tiha husi kuarta klase ne'e ne'e señora B... servisu saida?

tx@BB   Ha'u para tiha iha kuarta klase. Ha'u sei servisu ou tu'an inan-aman halo servisu iha uma, iha to'os. To'o Indonesia tama mai iha tinan rihun ida atus sia hitunulu resin sia. Ha'u komesa hahú tama servisu iha admistrasaun, konsellu de admistrasaun, Suai, to'o tinan rihun ida atus sia sianulu resin sia komesa revolusaun, sá ne'e. Ha'u iha hahú tinan rihun ida atus sia sianulu resin sia fulan Dezembru, ha'u voluntáriu ba hanorin iha eskola ida ne'e rasik, eskola Audian to'o ohin loron.

tx@NA   Di'ak. Ne'ebé oras dadaun ne'e, señora B... hanesan profesór iha eskola ida ne'e, ne'ebé hanesan profesór ne'e ne'e Ita hanoin nune'e be, profesór ne'e ne'e buat ida di'ak ka o nune'e be?

tx@BB   Ba ha'u, foin hahú, ha'u hanorin, ha'u sente todan tebe-tebes ba ha'u, mais to'o ohin loron, ha'u sente ba ha'u ne'e, la iha buat ruma ona, hanesan baibain de'it.

tx@NA   Ha'u, buat ida profesór ne'e di'ak, tanba ita hanorin ema, halo sai ema, hanorin ema atu halo tuir buat ne'ebé maka loos, ne'ebé kala ha'u, buat ida profesór ne'e ne'e, di'ak. Tanba ita sei hanorin ema, ema ne'ebé maka la'o laloos ita sei haloos, i hanorin halo sai matenek, tanba ohin loron ita-nia nasaun ida maka atu hetan di'ak, ne'e liu husi profesór sia. Tanba kalau profesór sia la iha, maka ha'u kanoin nasaun ne'e ne'e susar atu bá oin, ne'ebé ho profesór ne'e ne'e buat ida mak di'ak. Tanba, ó bele atu sai dotór, bele sai prezidente, bele sai primeiru ministru maibé tenke liu husi profesór. Só prefesór mak hanorin sia, ne'e ita profesór ne'e buat ida maka di'ak tebetebes. Ita bele terus maibé ita-nia naran la lakon. Ita bele mate ba mais ita-nia naran la lakon. Bele nia sai prezidente, atu sai dotór fali maibé profesór nia naran nia la naluha, ne'e buat ida profesór ne'e buat ida maka di'ak tebes.

0020TS

tx@RASS   Maun naran sá?

tx@CSG   Ha'u naran C....

tx@RASS   C.... Maun bele ko'alia to'o be bele dale ..., uluk, uluk sei ki'ik ne'e ne'e halo sá? Imi seidauk tama eskola ne'e imi kuandu moris lororon ne'e nu'usá?

tx@CSG   Uluk ami sei ki'ik ami biasa halimar. Halimar ho maluk sia mak iha ita kan iha kota laran ne'e ami halimar. Halimar hana kaleik. Halimar sa'e nuu, han. Ami halimar taa ai, halo babiduk hodi tali bobar para soe, halo nalai. Ami hadiuk mak ida ne'e, haduir ka'ut ne'e uluk ami sei ki'ik.

tx@RASS   Halimar ne'e ne'e mane mesak de'it ka ou?

tx@CSG   Ami feto mane.

tx@RASS   Feto mane.

tx@CSG   Feto mane uluk halimar hamutuk de'it.

tx@RASS   Maun bele ko'alia dauk uluk be iha tempu ne'ebá, iha nato'on ne'e ema hotu mama, imi sei ki'ik ne'e ne'e mama ona ka seidauk?

tx@CSG   Ami sei ki'ik ne'e la mama, ami haree katuas no ferik sia mak mama. Ferik no katuas sia mak mama. Ne'ebé ami haree ba katuas sia ne'e mama, tula tabako, ami haree. Ne'e ami sei ki'ik haree.

tx@RASS   Tanbasá mak sia mama ne'e?

tx@CSG   Tanba sia mama ne'e, kala ita husi uluk ita bein sia ne'e kala mama nanis ona nune'e be. Sira mós kalau haree tuir ita bein sia moris uluk ne'e, ne'e mak sia mama.

tx@RASS   Maun bele esplika to'ok moris iha ne'ebé?

tx@CSG   Ha'u moris iha Fohoren, iha dato ida naran Ai-to'os, suku Dato Rua.

tx@RASS   Fohoren ne'e, Fohoren nune'e ema dale Tetun hotu ka?

tx@CSG   Fohoren ne'e ema dale Tetun ema suku rua de'it. Suku rua dale Bunak.

tx@RASS   Suku rua dale Tetun ne'e mak naran suku ne'ebé ne'e?

tx@CSG   Suku rua dale Tetun ne'e mak Dato Rua no Fohoren. Dale Bunak, Lato'os no Dato Tolu.

tx@RASS   Dale Bunak, Lato'os ho Dato Tolu. Tanbasá mak katuas sia ne'e ne'e mama tiha, tenki tula nok tabako ne'e?

tx@CSG   Tanba katuas sia mama no ferik sia mama tiha tulak nok tabako ne'e tanba dalaruma ita han ho ahu eh. Ita mama ne'e han ho ahu, ahu dalaruman han hodi tabako ne'e mak dudu. Dudu na'ak ahu ne'e la bele na ita.

tx@RASS   Halo mamar?

tx@CSG   Halo mamar, halo ita ibun ne'e mamar.

tx@RASS   Ha'u rona maun mós ema futu-manudór ida?

tx@CSG   Ha'u futu-manudór.

tx@RASS   Futu-manu ne'e halo nu'usá, hanesan toman ida ka ou hanesan be taru?

tx@CSG   Hanesan futu-manu ne'e ita hanesan taru. Ita taru manu, ita haree emak manu, manu ida ne'ebé mak hanesan noo, ita buka manu seluk ida atu hasoru nia. Ita hasoru nia ne'e ita taru, ne'e ita taru. Ne'e uluk katuas sia, ami moris mai ne'e, haree katuas sia nuu iha Fatumean. Fatumean iha manu meo ida. Fatumean ka, Tilomar ka, iha manu meo ida, katuas sia iha Fohoren rona: oh manu meo ida ne'e iha Fatumean. Ita hodi sá manu mak itak atu bele hasoru mak manu ne'e. Rona tuir ida ne'e, ita manu meo ne'e Fatumean ne'e ita buka manu iha Fohoren, ita buka, manu ida ne'ebé mak rahun di'ak atu hodi hasoru nia, ita bá ho katuas sia taru, futu.

tx@RASS   Ha'u rona iha liafuan ida naran dore au?

tx@CSG   Dore au?

tx@RASS   Bele…

tx@CSG   Dore au ne'e, hanesan ne'e.

tx@RASS   Bele esplika to'ok dore au ne'e oinsá?

tx@CSG   Iha tempu uluk katuas sia dore au ne'e, hanesan buat ida, itak buat ida lakon ne'e. Itak buat ida lakon iha to'os. Ita kolu lai ema ne'e, tapi ema ne'e ita haree tiha, hatene nia, ita la kaer nia. Ita la kaer nia.

tx@RASS   Mais nia nega?

tx@CSG   Nia nega, ita depois ita bá hasoru dato sia hanesan liurai sia ka, atau xefe suku sia ita bá, ema ne'e mak hoirseik kan nia bele soe jagung batar ka, i ha'u iha kan to'os ka, ha'u karee rasik nia. Ha'u karee duni nia mak tahan kan hudi ne'e. Depois kaer hudi mai ita nia nega? Nia la terik ne'e ha'u lae, tanba ida ne'e maka dore au. Dore au hodi futu-manu.

tx@RASS   Liu husi futu-manu hodi buka lia loos.

tx@CSG   Hodi buka lia loos.

tx@RASS   Ha'u hanoin to'o iha ne'e de'it.

0023TS

tx@MA   Iha uma ne'e mak pernah ami halo ne'e ne'e ami kuru wee, te'in, dasa rai, ne'e tiha ami bá to'os, bá to'os fila mai tiha, ami hodi modo ami te'in. Hotu tiha tuirmai te'in hotu ami fase bikan. Fase bikan ne'e tiha haree hahan be mane ho feto sira, ami-nia inan-aman, ho alin sira ho, naran sá, bei-tuak ho inan bei sira. Ne'e tiha to'o mai hotu tiha agora ami loroloron iha uma ne'e. Ami bá kuru bee, ami bá kuru bee, bá hola ai, bá lia, ne'e kan ami halo oioin de'it. Tuirmai, tama hotu. Ne'e iha ne'e teki...tuirmai ami sei, hotu tiha ami halimar, halimar ho ami nia kan halimar ho oan, alin sira, ho alin feto mane iha uma. Ne'e hotu tiha kan ami tuur hamutuk halimar. Ami hadiuk, ami hadiuk ne'e, hadiuk be, hadiuk halo servisu, servisu ne'e halo mane sira bá halo to'os, ami feto sia iha uma. Ne'e ha'u-nia liafuan mak ne'e de'it.

tx@LLM   Di'ak. Ha'u iha tempu uluk, ami sei ki'ik oan, ami hadiuk ai-tati'ik. Ami hadiuk bola. Ami hadiuk berlindus. Ami tilak badut kamii. Loroloron amik hadiuk mak ida ne'e. Loraik, ami bá to'os, selu kuda, selu karau. Sei sawan ami bá eskola. Ami hadiuk buat oioin. To'o kalan ami toba. Sei sawan hadeer mai, ami hariis i ami bá eskola. Ami mós ajuda inan-aman. Amik ina, amik ama sia, halo to'os de'it. Hafaho du'ut. Hodi fó han ami. To'o silu batar, to'o sohi batar, ami mós bá sohi hotu, ami lalin hodi mai uma. Lalin to'o hotu. Mai fai batar. Te'in etu ba ami, han etu batar de'it. Tanba ami iha foho, wee la iha ne'ebé ami han batar de'it. Sei atu han foos hare, hola iha pasar. To'o tempu hamlaha, ami ho amik ina no amik ama sia, fai akar. Fai akar, kuru wee hodi lele akar to'o hetan isin tiha, foin kadaka ami han. Obrigadu!

0026TS

tx@ATB   Ok, bondia kolega.

tx@FJC   Bondia.

tx@ATB   Ohin loron ita iha ne'e ita atu ko'alia uitoan kona-ba eskola ida ne'e. Oinsá ita hanorin ita-nia oan sira. No saida-saida mak ita hanorin. Ho durante ida ne'e mós ita hanorin prefere liubá língua Tetun? Ne'ebé ha'u hakarak husu ba mestra, durante ne'e, iha tinan hira ne'e mestra hanorin. Tópiku saida mak mestra hanorin, ho klase hira?

tx@FJC   Di'ak. Di'ak, durante ha'u hanorin iha tinan ida ne'e ha'u hodi língua Tetun ba ita-nia oan sira, ba ita-nia oan sira be husi kelas, kelas dua, IPA satu to'o iha IPS dua, ne'ebé ha'u hakarak. Ne'ebe iha, ha'u hakarak haka'as-an ba alunus, ha'u hakarak haka'as-an ba ita oan sira atu nune'e sira mós bele hatene Tetun ne'e tanba Tetun ne'e importante ba ita tanba iha ita-nia nasaun. Agora ne'e ita tenkesér ho ko'alia ho Tetun de'it. Tanba Tetun nu'udar ita-nia liafuan, ita-nia lian uluk kedas ho ita.

tx@ATB   Ok, mestra hateten kan Tetun ne'e hanesa ita-nia lian, lian kahorik ya lian orijinál hodi tempu uluk to'o ohin loron. Maizumenus, mestra nia esforsa-an uza métodu

halo nu'usá mak ita bele hanorin ita-nia oan sira di'ak, atu kompriende di'ak liu iha siénsia ida ne'e? Ok, karik mestra iha konfusaun uitoan, ha'u hakarak fó hetene ba média katak, distritu Kovalima ne'e distritu ida ne'ebé refere liu iha ne'e distritu Kovalima, distritu ne'ebé mak ko'alia duni lian Tetun. Ohin loron Tetun mós sai ba língua nasionál. Iha oportunidade, ita hotu mós tenke haka'as-an duni atu dezenvolve ita-nia lian ida ne'e, iha país ida ne'e, atu bele berlatih tuir língua sira seluk iha mundu rai klaran hanesan lian Inglés, Portugés, Mozambike, ho sira seluk tan. Ohin loron sai hanesan profesór Tetun, iha métodu buka hanoin oinsá bele hanorin labarik sira di'ak, atu bele la'o tuir língua sira seluk, gramátika língua sira seluk. Atu nune'e língua Tetun bele dezenvolve di'ak liu ba jerasaun foun ho sidadaun foun. Tetun ezite ona iha Timor, iha siénsia, iha siénsia Tetun ne'e la'o iha Timor durante... Timor hetan indenpendénsia to'o ohin loron maski la'o durante tinan sanulu nia laran ona Tetun sei hasoru masalah barak. Maibé nu'udar Timor-oan sei haka'as-an nafatin oinsá atu bele dezenvolve lian ida ne'e ho di'ak. Maizumenus ha'u-nia hanoin hanesan ne'e, kolega nia hanoin sira seluk tan, se iha, bele fó tan.

tx@FJC Ha'u-nia hanoin di'ak liu tanba ita hotu aprende Tetun ne'ebé ita hotu mai atu aprende Tetun tanba Tetun importante ba ita hotu atu hatene. Tanba, ha'u sente katak Tetun ne'e difísil ba ita maibé ita haka'as-an atu hatene Tetun.

tx@ATB Ok tanba kolega hateten daudauk katak Tetun ne'e mós difísil ba ita iha ita-nia nasaun ida ne'e, ne'e duni ha'u husu ba liu husi oportunidade ohin loron, ha'u husu ba ministériu edukasaun atu kria mekanizmu ida, kria sistema ida oinsá iha Timor-Leste ne'e mós, ita mós bele halo seminar ka kongresu lian Tetun ba profesores Tetun tomak mak iha Timor-Leste atu nune'e, atu bele marka ninia...

0032TS

tx@GJX Tempu livre saida mak Ita-boot gosta halo?

tx@OA Iha tempu livre ha'u gosta halimar ho ha'u-nia oan sira. Ha'u gosta te'in hahan ne'ebé ke ha'u gosta para ha'u han.

tx@GJX Entaun hahan saida mak Ita-boot gosta?

tx@OA Ha'u gosta han liu ne'e hanesan modo-tahan, ba ha'u na'an ha'u ladún gosta han.

tx@GJX Maizumenus be na'an Ita han, bele han semana ida dala ida loos ka lae?

tx@OA Maizumenus semana ida dala ida konforme, selae semana ida mós dala ruma la han na'an, han modo-tahan de'it.

tx@GJX Entaun hanoin be se ita karik kompara Ita-boot nia hela fatin ho fatin Ita-boot servisu ho Ita-boot hanorin ne'e ne'e, se Ita-boot kompara karik, distánsia entre Ita-boot

nia uma, husi uma mai iha fatin servisu, hosi, porezemplu Ita-boot Mota-ain mai iha Ogus ne'e besik ka dook, tuir Ita-boot nia hanoin?

tx@OA   Se karik halo komparasaun ba ha'u-nia hanorin bá mai entaun ha'u-nia hela fatin ne'e dook husi eskola ne'ebé ke ha'u hanorin.

tx@GJX   Susar Ita-boot be sente?

tx@OA   Ha'u sente iha difikuldade uitoan tanba ha'u mai ne'e, ha'u mai hanorin, ha'u hela fatin dook ha'u mai hanorin iha EBC Ogues ne'ebé karik dala ruma ha'u sente difikuldade tanba se karik orsamentu selu trasporte bá mai nian la iha entaun dala ruma ha'u bele falta iha semana ida dala ida tanba difikuldade ba transporte.

tx@GJX   Ita-boot bele iha difikuldade ne'e bainhira karik, iha, iha ita hetan difikuldade iha ne'e meius de transporte.

tx@OA   Sin.

tx@GJX   Porezemplu karik ita, menus iha osan, osan menus ona ita la bele atu selu ba motorizada entaun oinsá Ita-boot nia sentimentu karik bainhira husik hela sala?

tx@OA   Ha'u, ha'u nu'udar profesóra, se karik ha'u la iha osan atu selu trasporte tanba ha'u hela fatin dook, entaun ha'u iha uma ne'e, ha'u triste tebetebes.

tx@OA   Ha'u hanoin ha'u-nia alunus sira tanba ohin loron matéria ne'ebé ke ha'u fó, ha'u la fó tanba difikuldade ha'u la hetan trasporte atu hodi mai to'o iha eskola. Entaun, iha tempu ne'ebé ke ha'u la mai, ha'u hanoin tebetebes tanba alunus ne'ebé ke ha'u hanorin ha'u konsidera mós hanesan ha'u-nia oan maka hamutuk ho ha'u loroloron iha uma.

tx@GJX   Ita komprende karik iha difikuldade ne'e sai dezafiu, sai obstákulu iha Ita-nia knaar. Entaun oinsá Ita-boot hanoin ba profesora, oinsá Ita-boot nia hanoin atu rekupera lisaun ne'ebé Ita-boot, atu rekupera, oinsá Ita-boot atu rekupera, atu falta ne'ebé iha?

tx@OA   Se karik hanesan difikuldade ne'ebé ke ha'u iha, ha'u iha difikuldade ba transporte. Karik iha tempu ne'ebá ha'u la mai entaun se iha loron ida livre mai ha'u hanesan profesór ruma se karik falta, entaun tempu ne'ebé ke profesora falta ne'e, ha'u bele tama hanorin fila fali matéria ne'ebé ke iha loron ne'e ha'u la mai.

tx@GJX   Karik durante seis dias, karik durante loron neen ne'e iha be iha oráriu ne'e sempre la iha oras ne'ebé mamuk. Entaun oinsá ita buka, Ita-boot atu buka, meius saida maka Ita-boot bele uza?

tx@OA   Sei…lae, irmán, ha'u hanoin iha oráriu mamuk ne'e nia sempre iha de'it tanba karik iha loron ida ba loron ida ne'e profesór balun se karik nia moras nia la tama entaun ha'u aproveita tempu ida ne'e, ha'u lori tama fila fali hanorin matéria ne'ebé ke loron ne'ebé ha'u la fó.

tx@GJX   Entaun haree ba nivel da edukasaun. Haree ba nivel da edukasaun ne'ebé ita iha depois Ita-boot, Ita-boot nia espesialidade ne', espesialidade ne'e língua portugesa entaun nu'usá nu'udar profesora de língua portugesa, obstákulu saida mak iha tan?

tx@OA   Ha'u nu'udar profesora língua portugesa ha'u iha obstákulu boot ida tanba falta livrus ba alunus, hanesan falta materiál, hanesan língua portugesa, livrus hanesan ba matemátika tanba durante ne'e livru mak iha ne'e ba iha sétimu ano ho oitavu ano de'it. Nonu anu la iha. Ne'ebé ha'u hanesan iha obstákulu boot ida hasoru ida ne'e, falta de livrus.

tx@GJX   Entaun iha língua portugesa iha domíniu haat ne'ebé profesora nu'udar profesora iha língua portugesa iha domíniu haat ne'ebé, iha domíniu haat ne'ebé Ita presiza transmitir ba alunus. Porezemplu oralidade, leitura, eskrita i koñesimentu esplísitu. Ha'u foti ezemplu de'it karik iha leitura. Dala ruma iha leitura ne'e iha liafuan barak, liafuan balun ne'ebé alunus la komprende. Pois entaun palavras deskoñesidas. Entaun oinsá ita-boot, profesora buka maneira ida bele, responde, bele esplika liafuan ne'ebé balun alunus la komprende ne'e ne'e profesora prepara rasik ka ou haruka, ba maneira seluk bele haruka alunus rasik buka oinsá?

tx@OA   La iha, dala ruma ha'u haruka alunus buka maibé karik alunus sira buka ba la komprende entaun ha'u rasik buka meius hanesan ha'u, hanesan sira lee hotu, depois ha'u lee, sira lee tiha depois ha'u bele esplika fila fali ho Tetun atu nune'e alunus sira bele komprende sá mak iha ne'e.

tx@GJX   Entaun liafuan ne'ebé, alunus la komprende ou liafuan palavras difíseis ne'ebé iha ne'e ne'e profesora prepara ka, karik prepara profesra rasik prepara atu hodi esplika ba alunus sira entaun. Materiál, materiál saida mak Ita-boot uza hodi hasai signifikadu palavras deskoñesidas?

tx@OA   Se karik hanesan ha'u mós nu'udar profesora ida ha'u lee ba hanesan alunus la komprende husu mai ha'u karik ha'u mós la komprende maka ha'u tenke buka iha disionáriu. Ne'e para ha'u bele hodi hetan fila fali signifikadu saida mak ne'e.

0033TS

tx@OC   Ne'e ne'e nia maka komesa hakarak bá ona.

tx@AGJ   [...] kona-ba be saida? Ó hakarak kona-ba saida?

tx@OC   Kona-ba feto.

tx@AGJ   Ita kona-ba feto ka? Ko'alia mane, feto saida ne'e.

tx@OC   Ita kan agora ita Timor ita iha ne'e ne'e kan ko'alia Tetun.

tx@AGJ   Ne'e to'o, Tetun ita-nia, ita-nia mós diak.

tx@OC   Se mana, mana Zuzana tan mai hanesan nia, be, halo peskiza kona-ba língua entaun ita, ninia nakarak ita-nia língua Tetun ne'e atu naran sá grava.

tx@AGJ   Ne'e loos to'o ona. Entaun ita hanesan konta istória kona-ba buat ruma, istória saida?

tx@OC   Ko'alia kona-ba istoria saida nian fali.

tx@AGJ   Ita depende, depende Ita-boot istória kona-ba saida nian.

tx@OC   Ami bele ko'alia barak, mana? Bisa berbicara banyak? Hanya sedikit saja tidak ada masalah.

tx@AGJ   Ko'alia kona-ba saida? Ko'alia kona-ba saida?

tx@OC   Ko'alia ita-nia, agora ko'alia ita-nia kona-ba istória Timor, naran sá, de'it ona, ita-nia kona-ba língua, ita kan ko'alia Tetun.

tx@AGJ   La iha.

tx@OC   Entaun ko'alia kona-ba saida?

tx@AGJ   Hanesan ne'e, ita hanesan konta istória ida, konta istória ida para iha grava.

tx@OC   Nia kala komesa grava daudaun ona.

tx@AGJ   Daudaun mós ita konta istória ida mak di'ak.

tx@OC   Istória saida? Ha'u istória la hatene atu ko'alia istória saida.

tx@AGJ   Ou feto ka.

tx@OC   Feto [...]

tx@AGJ   Ko'alia saida?

tx@OC   Ko'alia ne'e kona bola de'it ona, ó kan biasanya hobi bola. Ha'u hobi bola hanesan ha'u koalia ho ó mós hanesan de'it ona.

tx@AGJ   Ó la hatene bola.

tx@OC   [...] la bele to'o. Entaun buat ne'ebé mak ó hatene ne'e mak bele, ne'e hau ko'alia ho ó di'ak. Entaun ita ko'alia kona-ba istória saida, istória la iha ona.

tx@AGJ   Política ka. Entaun política de'it ona.

tx@OC   Politik ida ne'e, inan naneer la hotu, ita naran atu diskusi ba kona-ba politik inan-nia buan naran sá ne'e mak ne'e.

tx@AGJ   Política, politik, kona política de'it ona. Agora ha'u husu, ha'u husu. Oinsá tuir ó-nia hanoin kona-ba be sira atu, ita-nia be Primeiru Ministru atu rezigna an.

tx@OC   Ya, ne'e defende ninia política, nia atu rezigna an, apakah ida ne'e tebes duni nia atu rezigna an ou atu nia provoka situasi iha Timor laran ne'e.

tx@AGJ   Lae, tuir ha'u-nia hanoin ne'e, ha'u sente nia rezigna an ne'e rezigna an duni tanba nia, ida ne'e ne'e fó sai iha públiku tiha ona ne'ebé ida ne'e ne'e ha'u sente justu ona. Mais ida ne'e ne'e la hatene mak política hotu ba saida mak ida ne'e ema boot sira nian, ita mós sei dúvidas uitoan, ita hein, halo ita hein be finál.

tx@OC   Nia dehan katak atu rezigna ona mais nia liafuan ne'e seidauk, nia sai ona, fó sai ona iha públiku i ema barak mós debate ona kona-ba ninia ida rezigna an ne'e. Maibé ita la hatene loos tanba ne'e ema boot nian tanbasá mak nia hakarak rezigna an bainhira nia mandatu seidauk hotu.

tx@AGJ   Ida ne'e maka ita la hatene, ida ne'e ne'e tanba nia fó sai derepente de'it, derepente ho kazu ida be señór Mauk Moruk nian ne'e. Ida ne'e ne'e mak la kleur de'it, nia fó sai tiha be isu-isu kona-ba be nia atu rezigna an ne'e. Hanesan ne'e problema ida ne'e, problema sira, iha buat ne'e uluk problema desde ke uluk sira sei iha rezisténsia nian dada mai to'o até agora.

tx@OC   Ne'e buat ida ne'e kan ita la hatene tuir ita-nia maun boot Xanana nian tanba ne'e ema ida be level, nivel ida altu nian.

tx@AGJ   Ita hanesan, ita hanesan tuir-tuir de'it ninian be situasaun ne'ebé mak sira agora daudaun be fiar tanba ne'e ita hein tuir-tuir de'it ona. Ita haree tuir de'it tanba saida mak nia atu, atu rezigna an. Tuir ita-nia, tuir ita-nia hatene tuir ita-nia interpreta mak ita hatene mak dehan katak tan problema ida hanesan mak nia atu rezigna an. Mais be politik sira-nia, ita la hatene ida.

tx@OC   Mak ne'e, ne'e mak ita agora ita povo Timor ita mak ki'ik oan, ita dúvida hela kona-ba nia rezigna an. Ninia mandatu seidauk hotu mais nia hakarak rezigna an. I depois governu ida subtitui nia ne'e, nia bele halo tuir buat ne'ebé mak nia halo ona iha durante tinan hira nia laran be nia bele halo tuir hotu ka lae ida ne'e. Ida ne'e problema boot.

tx@AGJ   Ida ne'e kan maibé nafatin ida partidu ne'ebé mak hanesan maioria hanesan sira ne'ebé mak maioria ho be abut barak hanesan CNRT. Sira ne'e mak delega fali nia ema sai troka fali nia, nia para bele hala'o fali sira-nia be, sira-nia misaun ou sira-nia promesa ne'ebé mak sira hato'o durante kampaña ne'e. Sira la bele hili fali ema seluk.

tx@OC   Ne'e mak foin lalais ha'u haree liu husi televizaun husi ita-nia deputadu balu i husi bankada Fretelin. Fó sai ona katak maun Xanana atu rezigna an maibé ema ne'ebé mak atu em nome ba hanesan atu subtitui ba nia kecuali maioria parlamentár ou bloku parlamentu, hanesan ita-nia CNRT ho PD ou Fretelin mudansa.

tx@AGJ   Sin, ida ne'e ne'e mais nia ne'e loos, maibé sira ida be ita dehan be sira ida bloku ne'e bele mais sira atu, sira bele hala'o be sira-nia be, hala'o fali be, sira ida CNRT

nia be programa ne'e ka lae, programa ne'ebé mak sira ida be durante kampaña sira fó sai, fó sai be ba públiku ne'e, sira hala'o ka lae.

tx@OC   Ida ne'e konforme lisaun politik, ita la hatene loos sira atu em nome se atu eleisaun ka, atu eleisaun antesipada ne'e ita la hatene tuir.

tx@AGJ   Ne'ebé eleisaun antesipada karik dala ruma la akontese mais tanba ida ne'e ne'e kan buat ne'ebé mak justu tiha ona, sira atu hein ita hatene dehan katak sira partidu be CNRT ne'e ke sira manán ho be votu ne'ebé mak masin, ne'ebé buat ida eleisaun antesipada ne'e atu naran akontese ne'e la iha tanba sira mós iha votu ne'ebé mak barak, ne'ebé mak sira, voto ne'ebé barak atu. Sira iha podér tomak atu hili fali sira-nian lideransa, sira ne'ebé mak atu serve sai duni be lideransa husi governu, Primeiru ministru nia, ida ne'ebé maka serve atu sai Primeiru Ministru.

tx@OC   Ne'e mak ne'e iha ikus liu tanba ita mesak ki'ik tanba ita haree tuir informasaun ne'ebé mak publika iha TVTL ne'ebé ita atu diskusi liután mas la di'ak ona ita haree fila fali hein, fila fali maun Xanana nian desizaun ne'ebé mak atu rezigna an.

tx@AGJ   Entaun loos ona, hotu ona.

Tetun Viqueque

0039TV

tx@MVJ   Bele, bele. Iha Tetun Terik? Di'ak. Dadeer kmanek ida ne'e. Ha'u hakarak ko'alia ho, dada lia uitoan ho menina Zuzana. Mai husi Xekozlovákia, ne'ebé nia estuda iha Universidade Coimbra iha Portugál ne'ebé oras ne'e daudauk sira mai halo peskiza kona-ba sira-nia estudu iha siénsia linguística iha Timor. Ne'ebé sira hakarak ko'alia ho ita, Timor-oan rasik, ho Tetun, ho maluk sira ne'ebé Tetun, orijen Tetun. Sira hakarak atu estuda kle'an uitoan kona-ba língua Tetun Terik ne'ebé oras ne'e daudauk Timor-oan sira ne'ebé agora parte Terik atu, sira atu aprende kona-ba lian Tetun Terik. Iha ne'e menina Zana hakarak dada lia uitoan ho ha'u, M..., ne'ebé Tetun Terik, ema Tetun Terik rasik iha Viqueque. Hakarak ko'alia uitoan kona-ba família. Ha'u-nia inan-aman moris. Ha'u-nia inan-aman naran, aman naran José, inan naran Carmelinda. Ha'u-nia aman mai husi Uatu-Karbau. Ha'u-nia inan ema Viqueque, Tetun Terik, Karaubalu. Sira na'in-rua hetan malu iha funu japonés nia tempu. Bainhira ha'u-nia aman atu hetan ha'u-nia inan bainhira iha prosesu, liu husi prosesu ida ne'ebé nia inan mai hola ema Karau-Balu ho ema Uma Ki'ik, liurai Uma Ki'ik, naran Miguel. Ne'e mak nia lori ho nia alin ida naran José de Araújo, ne'e mak ha'u-nia aman. Pois mai iha Viqueque. Mai hela ho nia biin. Nia sei, iha tempu ne'ebá, sei iha okupasaun olandesa iha Timor, olandesa iha Timor, sei karik la sala iha míl novesentus kuarenta i oitu. Ninia iha ne'e depois bá servisu iha kampu aviasaun. Kampu aviasaun hatene? Aeroplane.

tx@MVJ   Iha Viqueque. Servisu iha kampu aviasaun. Lori, lori fatuk. Iha tempu ne'ebá servisu obrigatóriu. Lori fatuk, fatuk rahun. I depois lori ba halo kampu. Bainhira kuandu ha'u-nia aman sira bá halo kampu, lori fatuk obrigatóriu. La iha ema ida, kuandu iha tempu servisu, la iha ema ida atu fuma sigarru. Fuma sigarru, deskansa la iha obrigatóriu. Sé mak fuma, ne'e, baibain ha'u-nia aman katuas sira dehan katak ne'e tuan-tuan, tuan-tuan mai, liurai baku. Entaun baku sira ne'ebé deskansa. Sira tenke servisu de'it. Tempu ne'ebá servisu, bolu dehan, servisu obrigatóriu. Entaun iha ne'ebá ha'u-nia inan mós idade sira sa'e hanesan, servisu hotu tiha kampu aviasaun. Tanba iha ne'ebá sira na'in-rua kala konkista malu, prontu, hola malu. Sira na'in-rua hola malu. Hela iha Viqueque, iha Tetun Terik. Komu iha tempu ne'ebá kulturalmente tuir lia tradisaun adat, husi família mane nian la ko'alia ho família feto nian. Entaun sira tenke hela metin iha Viqueque. Hela metin iha família feto nian. Entaun iha ne'e, oras sira na'in-rua hetan oan na'in-sanulu-resin-rua. Doze filhos. Na'in-saunulu-resin-rua, oan na'in-sanulu-resin-rua. Mane lima, mane hitu, feto lima. Mane hitu, feto lima. Iha ne'e mane na'in-lima fila ba rai, fila. Agora hela mane na'in, mane na'in-tolu, feto rua fila ba rai. Entaun hela mane na'in, hela mane na'in-lima, feto rua maka moris hela. Iha ne'e, prontu, iha ne'e ida boot eskola iha Colégio Ossú ho ida, primeiru fillu ne'e eskola iha Colégio Ossú, segundu fillu iha Colégio Ossú iha tempu portugés nia tempu. Ne'e sira na'in-rua konsege akaba sira-nia estudu iha kuarta klase iha língua portugeza, iha Ossú. I depois iha terseiru fillu mós eskola. Nia eskola iha kompañia, kompañia Betise, kompañia tropa mílitár da Indonézia, oh, portugeza. Iha kompañia Betise, mós akaba nia kuarta klase. Prontu, feto na'in-rua sei ki'ik, ha'u rasik mós sei ki'ik ne'ebé seidauk eskola. Pois kuandu to'o iha míl novesentus i sesenta i sinku, kuandu ha'u moris, to'o atinje sete anus, tinan hitu, tenke bá eskola iha primeira, pré-primária. Eskola pré-primária.

tx@MVJ   Iha Viqueque. Mós iha kompañia Betise nian, peska. Kompañia Betise. Até primeira klase, até segunda, depois da segunda klase, pasa ba iha, transfér fali. Komu iha setenta i kuatru, setenta i kuatru, kuandu prezidente ne'e, prezidente Portugál ne'e Marcelo Caetano ou António Spinola fó liberdade ba nasaun sira ne'ebé CPLP, para sira bele determina sira-nia, fó demokrasia ida ba sira atu determina sira-nia destinu. Entaun iha tempu ne'ebá, ami nu'udar estudante, ami sei ki'ik i espesialmente ami Timor tomak seidauk koñese saida maka dezenvolvimentu ba política ninia, seidauk koñese. Entaun iha ne'ebá iha baluk koñese mas Portugál ho ninia, ninia esperta ne'e, ho ninia esperta hatene karaterístika, aspetu de ema Timor, hatene. Entaun sira haree ema Timor ne'ebé matenek uitoan to'o ona, eskola to'o ona, sétimu anu, nonu anu, oitavu, oitavu anu, prontu. Koñesimentu kona-ba política ona, entaun sira tenke muda ba eskola iha Portugál. La bele iha Timor. Tanba sira tenke iha otak ho karater bé ema Timor kuandu koñese kona-ba ne'e sira bele, sira bele kontra. Tan sira haruka. Entaun prontu, hela maka Timor-oan sira ne'ebé eskola seidauk, seidauk atinje sira-nia estudu ba iha klase ida ne'ebé boot, edukasaun ida ne'ebé boot ba hanesan. Kintu, sestu, sétimu anu ne'e mak hela iha Timor. Kuandu iha kuartu, terseiru anu, kuartu anu, kintu anu, bele hela iha Timor mas kuandu a nível de universidade, la bele. Entaun prontu, iha ne'e, iha kuandu António Spinola ho

Marcelo Caetano fó liberdade ba Timor-oan, ba nasaun hotu-hotu ne'ebé CPLP i CPLP nia mós inkluidu mós Timor, ema Timor. Entaun prontu, iha ne'e, iha komesa iha mudansa ba edukasaun, liuliu iha, liuliu iha kompañia sira hanesan militár, iha ensinu militár. Sira komesa transfér, transfér ona ami husi Betise ba fali ona iha Olobai, kompañia de Olobai. Entaun iha ne'ebá to'o iha setenta i, iha vinte kuatru de, vinte sinku de Abríl, se não me engano, vinte sinku de Abríl, míl setenta sinku. Prontu, entaun iha ne'ebá ami komesa muda. Prontu, iha, iha setenta i seis, setenta i seis, não, setenta i kuatru, vinte sinku de Abríl setenta i sinku, prontu iha ne'e, iha setenta i sinku, ne'e iha setenta i sinku ne'e mós komesa mós ona fó ona liberdade. Entaun Timor-oan sira, sira ne'ebé matenek, komesa harii ona, idak-idak fó ona liberdade ba ema ida-idak atu harii partidu. Harii partidu, entaun iha Timor iha partidu lima. Iha, iha FRETILIN. FRETILIN mak até agora ho ninia ne'ebé defende nia ideolojia, to'o mak FRETILIN. Ida maka APODETI, Asosiasaun Populár Demokrátika de Timor-Leste, APODETI. Nia hakarak servisu hamutuk ho Indonézia, APODETI. Prontu, ida maka UDT, Uniaun demokrátika de Timor-Leste. Ida maka KOTA. Ida mak traballista. Tem cinco partidos políticos. Iha Timor kuandu António Spinola ho Marcelo Caetano fó liberdade ba nasaun CPLP ida-idak atu define ninia destinu. Entaun iha tempu ne'ebá, prontu to'o iha, iha setenta i sinku, setenta i sinku, nia laran kuaze Marsu, Abríl komesa mosu ona kontra-golpe iha Dili, militár kontra-golpe. Entaun prontu, iha Dili kuandu kontra-golpe, sira ne'ebé, apoiante sira ne'ebé hakarak defende ideolojia Timor nian, hakarak ukun an. Sira komesa militár fahe malu. Iha divizaun rua. Balu ho Portugál, balu ho Timor. Agora sira ne'ebé apoia Portugál, nafatin ho Portugál, ho militár de Portugál, sira ne'ebé Timor komesa halo, halo kontra-golpe, tiru malu hasoru Portugál iha tempu ne'ebá. Entaun Portugál sente la bele. Sira evakua ne'ebá iha Atauro para depois bele hetan meius ruma para bele liu bá iha sira-nia rain, bá fali Portugál. Entaun iha ne'e komu Indonézia mós, partidu APODETI mós iha ligasaun ho Indonézia, sira mós komesa halo tama mai ona iha Timor, hakarak mai invade Timor. Entaun iha setenta i sinku, setenta i sinku, dia sete de Dezembru komesa Indonézia mai ona ho aviaun. Sete de Dezembru Indonézia sira mai ho aviaun, ho ró, ho tankedegerra. Armas pezadas boboot mai iha Timor. Entaun forsa FALINTIL sira FRETILIN ninian ne'e tiru malu sira la bele. Evakua sa'e ona bá iha foho leten, sa'e foho leten. Entaun prontu, iha tempu ne'ebá, Timor-oan barak mak sa'e bá ai-laran, bá foho leten. Iha vila laran komesa okupa husi inimigu militár Indonézia ninian mak okupa. Komu ho sira-ninia política ne'ebé ohin intensaun atu ukun rai ida ne'e, entaun sira tenke ho meius oioin para depois bele sakraliza fali povu Timor lubuk ida ne'ebé. Balu família iha ai-laran. Balu la tuir sira-nia ideia. Entaun sira tenke oho ema sira ne'e. Sin, insentimente ho ema, sin. Entaun iha ne'e prontu, povu Timor kuantu mais hetan presaun de podér sira ne'e, sira kuantu mais hetan presaun, kuantu mais aumenta ideias kontra mais maka'as. Iha planu oioin para atu halo buat sira ne'e. Ne'ebé iha iha Timor rasik, nia iha Timor rasik, Timor-oan rasik, sira bele halo saida de'it, bele hasoru sira de'it. Mas sira, sira ninia baze fundamentál ba atu defende konseitu ruma, defende ideias ruma, é presiza sira ninia baze fundamentál maka sira ninia tradisaun, identidade kulturál ne'e. Sira tenke kaer metin, pois sira iha unidade para atu bele defende sira-nia prinsípiu ne'ebé

sira kaer metin. Ne'e bazeia ba sira-nia kulturalmente tradisaun, kultura ne'e sira kaer metin ninia maka bele la'o ba oin. Ida ne'ebé ha'u hanoin hanesan menina presiza atu hatene uitoan kona-ba kultura, liuliu kona-ba lian iha Timor. Ha'u-nia lian inan maka hanesan nia ho hateten lian iha Timor ne'e, lian liu oioin, barak. Kuaze iha, ha'u ladún fiksa ona nia be volume, número ne'e, mas lian ne'e iha, também iha, iha Timor, iha, iha, iha munisípiu sanulu-resin-tolu, iha sesenta i sinku subdistritus, postu administrativu. I iha mós kuatrusentus i kuarenta i dois, atus haat haatnulu-resin-rua, suku atus haat haatnulu-resin-rua. I iha mós aldeia. Iha aldeia ne'e, iha rihun rua atus rua ruanulu-resin-lima. Rihun rua atus rua ruanulu-resin-lima. Barak. Ne'e ninia bahasa mós ninia bé bahasa, língua mas em indoneziamente ida bahasa. Sira-nia lian mós barak ne'ebé iha, iha Mambae, iha Galolen, iha Idate, iha Kairui, iha Waima'a, iha Tetun Terik, iha Makasae, iha Bunak, iha Dagadá, iha Baikenu. Iha lian lubuk ida ne'ebé balu mós ha'u seidauk la bele hatún, ko'alia hotu, barak liu. Prontu, entaun ho sira-nia lian ne'ebé barak i sira ninia tradisaun moris loroloron mós la hanesan. Tradisaun moris loron-loron mós la hanesan. Tuir sira-nia uma-lisan uzus, kostumes, kada uma-lisan ida-idak ninia ne'e la hanesan. Ne'ebé iha Timor ne'e, iha ninia uma-lisan, mós lubuk ida ne'ebé nia uma-lisan ne'e la hanesan. Ida-idak ninia, ninia kultura, ninia cara atu halo ninia métodu, métodu de, como se chama, tradisionalmente atu hamulak atu ne'e la hanesan. La hanesan. Ne'ebé mak hanesan daudauk foin ita haree, ha'u hanoin menina sira akompaña, ka lae, inaugurasaun igreja Viqueque? Lae. Mas é, haree ninia, maizumenus ninia.

tx@MVJ   Ha'u-nia dia vinte dois de Agostu, vinte um, vinte dois de Agostu, ne'e selebrasaun.

tx@MVJ   Fulan kotuk, sin. Maun Goncalo, ami-nia koordinador bá, hasai foto hamutuk ho adat ne'ebé ami husu ho ha'u, ho ha'u-nia espoza, hako'ak malu ho señór Goncalo. Señór Goncalo iha klaran, ha'u-nia espoza ne'e bolu dehan ha'u-nia kaben husi sorin, ha'u husi sorin. Hasai foto ho tais. Ne'e mak iha, iha Timor ne'e, iha ninia kultura la hanesan, iha ninia adat, hatais, vestuáriu ne'e, kada kultura ida-idak ho ninia motif, nia tipu la hanesan. Diferente mas iha ninia tipu ou motif ne'ebé sira prodús ne'e, fó, fó ninia hanoin, fó ninia versaun, fó ninia versaun kulturalmente ba iha ninia tradisaun ne'e. Ne'ebé la, la, la hanesan vestimentu ida ne'ebé ita hakarak ne'e kór atu hatais. Mas nia iha nia signifika ida ne'ebé aprofunda tebes kona-ba kulturalmente iha tradisaun ida. Ne'e kada, kada Timor-oan, ida-idak nia kultura hanesan ne'e. Iha Viqueque ninia, Viqueque rasik Tetun Terik, espesialmente Tetun Terik. Tetun Terik ema ne'ebé mai ho karater moris mai iha ne'e. Tetun Terik rasik. Ne'e karater moris mai iha ne'e, ema ne'ebé kalmu, kalmu. Ko'alia, em termos de ko'alia, falam também com calma, devagár, depois rona maluk i respeitozu, mesa respeita ema, rona maluk. I depois Tetun Terik ne'e sira, iha diskusaun ruma, sira la hola desizaun imediatu. Mas sira tenke hanoin, hanoin ninia vantajen i desvantajen mak foin foti desizaun. Ne'e ema Tetun Terik. Mas Makasae outra coisa. Nia mai ho emosionál, ho dominante pelo emosionál. Entaun mai ho brutalidade. Primeiru, ema mak mai pur brutalidade.

tx@MVJ Makasae. Makasae. Makasae ninia karater hanesan. Ne'e todos é de Makasae. Tem, sin, iguál. Mas agora, ita atu aprosimasaun ba Makasae, nia tenke ita uza sistema, saida ne'e. Nia kuandu mai ho emosionál, ita lalika kontrola emosionál. Ita bá ho kalmu, kalmu. Kuandu nia ko'alia husik nia ko'alia, ita simu de'it. Saida mak nia ko'alia atu ko'alia ladi'ak, di'ak, ita simu de'it. Depois nia kalma tiha, nia tuur. Kalma tiha, ita tama neineik. Entaun iha ne'ebá ne'e ita konvense nia ideia ne'ebá, mas nia mai ho emosionál, ita mós bá ho emosionál, ho brutalidade, entaun prontu, mosu ona konfrontasaun. Fizikamente, é de crime. Mas kuandu nia mai ho buat ne'ebé maka'as, ita kalma. O que é que ele fala ou disse: Deixa. Não me diga como. Pois, vai com calma. Pois, nia kalma tiha, depois ita aprosima nia ho neineik. Depois iha ne'e, nia mós komesa iha ona. Kuandu nia kalma ona nia ne'ebé tensaun komesa tun ona. Tun ona, ita tama neineik hanesan ne'e, hanesan ne'e, bá, ai, vai com respeito maun, irmaun, pai, não fala assim porque é assim, não está bem, tem que assim assim, prontu. Ne'e ita bele mas iha, iha mós, iha ne'e iha Kairui, Waima'a. Iha, iha mós, iha ne'ebé ema lae, la iha, sira Liaruka, Naueti. Naueti kuaze iguál kom Tetun Terik, mas Tetun Terik ninia, iha Tetun Terik ninia unidade ne'e ladún, ladún iha. Mínimu, unidade Tetun Terik mínimu. Sira prefere individualmente, ne'e aas Tetun Terik. Mas kuandu halai ba kulturalmente, ne'e sira hamutuk fiar halo. Mas kuandu hanesan atu buka dezenvolve vida moris ne'e, sira ida-idak, individuál. Vai, bá ho ninia matenek, ho ninia buat ne'ebé nia iha, nia buka rasik. Agora, karater ida Tetun Terik ne'e kuaze hanesan ho, ho Naueti. Naueti mós hanesan, kuaze hanesan. Mas só ke Naueti, sira iha sistema kulturál ne'e maka'as, kulturál maka'as. I depois sira iha, iha unidade ne'e mós metin. Metin. Ne'ebé sira sempre ajuda malu, dada malu, ajuda malu. Ne'e, Naueti ninia, ninia karater. Ne'e kuaze atu hanesan ho Tetun Terik mas sira unidade boot liu fali Tetun Terik. Unidade, unanimidade ne'e boot liu fali Tetun Terik. Tetun Terik nia halai ba individuál maibé ida moris individuál ne'e maka'as. Ne'e maka iha, iha Viqueque ne'e agora de'it tanba situasaun okupasaun rua portugés depois de kuatrusentus sinkuenta anus, portugés destaka em Timor. Depois de, kuandu embargu, portugés dezembarku iha Timor iha em setenta i sinku. Indonézia tama fali, okupasaun Indonézia vinte kuatru anus, ne'e mak hakarak atu harahun Timor ninia, ninia orijinalidade, orijinalidade, hakarak mai hafahe, povu, i hakarak atu harahun orijinalidade Timor i hakarak mós atu aproveita natureza ne'ebé iha Timor iha. Maibé Timor nafatin ho ninia kulturál ne'ebé diferente, ninia, ninia dialetu ne'ebé la hanesan, ninia suku liu la hanesan. Mas sira kontinua mantén sira ninia pozisaun atu defende sira ninia identidade nu'udar sidadaun Timor. Ne'ebé sira nafatin ida-idak halo tuir ninian. Ha'u hanoin iha, iha Timor, hetan esperiénsia lubuk ida husi invazaun ba invazaun, ne'e sai hanesan referénsia lubuk ida para depois Timor bele halo análize ba futuru, bele halo mudansa ruma ke di'ak liu ba iha futuru. I ne'e mós la nunka tau buat ida ne'ebé liuliu ona, que passou, é passou, ne'ebé liu, liu ona ne'ebé sira hakarak atu dezenvolve rai ida ne'e husi, husi ema hotu ninia kooperasaun, husi ema hotu-hotu ninia, husi posibilidade, iha ninia disponibilidade para atu koopera ho povu Timor atu servisu iha, atu dezenvolve iha aspetu sira de'it ho Timor-oan nian, prontu. Atu bele koopera hamutuk hodi lori rai ida ne'e sai husi fatin ida ne'ebé uluk saida tiha ona. Ha'u hanoin señora presiza atu ko'alia tan buat ruma? La iha

ona. Ok. Entaun ita em relasaun ne'e dada lia uitoan ho menina Zana ne'ebé hakarak halo nia peskiza, mai iha Timor liuliu halai ba área linguística ne'ebé ha'u hanoin karik iha dada lia ne'e iha liafuan baluk ke ladún, tama iha ladún. Menina Zana la simu, bele orasida halo, bele sai tiha i ha'u ladún mós karik liafuan balu bele ofende karik ha'u deskulpa. Ha'u husu deskulpa tanba iha tempu ne'ebá mós situasaun hanesan ne'e duni maibé ita ko'alia buat ne'ebé lojikamente, buat hanesan ne'e mas tenke hanesan ne'e mak ita bele to'o iha nia rohan. Ne'ebé ha'u hanoin hela hanesan ne'e de'it. Di'ak i ha'u remata ha'u-nia dada lia ho menina Zana. Obrigadu.

0042TV

tx@MS   Ha'u pesoalmente ba ha'u-nia be moris, ha'u moris mai. Ha'u moris primeiru, ha'u hanesan ema kra'es ida. Ha'u ema dehan hanesan saida. Iha ne'e ha'u-nia inan, nia mós hanesan ema hanoin la to'o ida, hanesan bulak. Entaun iha ne'e ha'u moris halo be loron ida, ha'u-nia inan ne'e lori ha'u bá soe iha ai-laran de'it. Bá soe iha ai-laran. Durante be fulan tolu nia laran mak ha'u-nia tiun sira bá buka fali. Ha'u-nia avón, ha'u-nia avón, avón bá buka fali. Buka mak hetan. Hetan iha ne'e sira hodi, hodi tau mai uma. Mai ha'u-nia inan, bá buka tuir fali ha'u. Buka tuir fali ha'u. Lori halai fila fali ba ai-laran. Nia na'ok tiha ha'u, lori fali ba ai-laran. Iha ne'e mak to'o ha'u-nia tiun sira bá buka, la hetan. Sira bá kasa fali mak foin hetan. Sira lori asu ba kasa mak asu duni fali be rusa. Sira bá soro. Ha'u-nia tiun sira bá soro fali mak foin asu lori rusa ba, rusa halai liu mak asu bá, narii fali ha'u, ha'e foin sira hetan fali ha'u. Entaun sira hodi fali mai. Entaun sira mai hein fali ha'u-nia, ha'u na'in mai fali. Ha'u na'in mai, mai sira kaer tiha ha'u na'in, futu tiha ha'u na'in. Tama nia bulak. Se la futu nia, entaun orasida nia bele lori fila fali ona ha'u ba ai-laran. Ida ne'e mak ha'u-nia moris mai, entaun iha ne'e ha'u-nia bei-tuak sira ha'e rauа'ek tiha fali ha'u. Ha'u bei-tuak sira raua'ek tiha fali ha'u. Ha'u-nia ina, sira la fó ona ba, ne'ebá ha'u bei-tuak ha'e naua'ek ha'u, no na'i ferik ha'e raua'ek ha'u. Entaun iha ne'e, to'o ha'u boot, iha ne'e bá eskola. Bei-tuak sira ratama bá, iha ne'e, to'o bei-tuak no na'i ferik mate tiha. Entaun ha'u mesak de'it ona. Ha'u la eskola ona. Ha'u hanoin de'it atu be hakarak eskola maibé la iha ema ida atu saida. Entaun iha ne'e mak ha'u buka tun, buka sa'e. Entaun ha'u bá hela ho fali maun ida mak hatama ba eskola. Iha ne'e to'o ha'u tama tiha ba SD. Entaun ha'u atu kontinua eskola, la saida ona. Entaun iha ne'e ha'u hanesan buat ida halai badiu. Halai ba badiu ona iha Dili, bá mak hela fali ho ema ida, mak foin hodi be hatama fali ha'u bá eskola SMP. To'o bá SMP atu ujian, atu bá ujian akontese ona doze Novembru. Entaun iha ne'e ha'u hetan aleijadu. Ha'u hetan ona tiru. Entaun iha ne'e ha'u mós la, la kontinua ona eskola i to'o loron ohin ha'u la eskola. Ne'e hanesan ha'u-nia istória badak kona-ba ha'u-nia moris ne'e, ha'u-nia moris mak ne'e. Kona-ba buat seluk, entaun ha'u lian Tetun ne'e, ha'u ne'e ema moris Tetun ne'ebé dala wa'in tiha ona bá ha'u aprende tiha ba Tetun be Terik ho Tetun Dili ne'ebé la, la hanesan tiha ona. Ne'ebé ha'u hanesan ne'e de'it, orasida ha'u sei bele lia hanoin tan istória badak ruma kona-ba kultura ninian. Ha'u bele konta tan fali istória kona-ba kultura ninian. Ha'u-

nia liafuan mak ne'e. Ne'e ha'u orasida lai bele fó fali ba ha'u-nia aman sira seluk atu ne'e bele saida ha'u oras ida ha'u husi konta fali kona-ba kultura ninian. Kultura ninian. Lia adat kultura ninian.

0043TV

tx@AP   Naran A.... Maibé ha'u ko'alia, hanesan ne'e, tanba ha'u inan ho aman hamuruk mai ne'e ha'u mesak de'it. Ha'u mesak de'it. Maibé ha'u-nia aman ho ha'u-nia inan atu haruka ha'u ba eskola tanba ema ajuda ha'u la iha. Ne'ebé ha'u hein de'it karau, kesi de'it kuda, halo de'it to'os, halo de'it natar. Maibé ha'u-nia oan ne'e na'in-ualu, ha'u-nia oan na'in-ualu bá eskola ne'e. Ema ida la ajuda. Ha'u de'it maka buka rasik de'it. Ha'u-nia isin-lolon hodi hatama ha'u-nia oan sira bá eskola to'o loron ida be ohin ne'e ka sira hotu matenek hotu. Sira feto haat, mane haat. Ne'ebé fó han mós ha'u de'it, buat ne'ebé iha to'os ka natar ka ha'u mak halo hodi fó han de'it ha'u-nia oan sira. Ha'u-nia oan sira mós eskola hotu remata. Feto ho mane eskola remata. Matenek remata. Ne'ebé ha'u-nia, komesa ha'u-nia moris mai ne'e ha'u la iha konta istória boot ruma ba ema, hasoru buat boot ruma ba ema, ne'ebé ha'u moris mai hanesan ha'u mesak de'it. Ha'u la iha, halo istória ho ema ida. Ha'u moris ho uma laran de'it.

0044TV

tx@JCP   Naran J.... Ha'u-nia naran J.... Ha'u, dezde moris, ha'u-nia lia, ami-nia lian ne'e, lian Tetun de'it, Tetun Terik. Makasae mós la iha, Naueti mós ha'u la hatene. La katene. Ami hatene de'it lian Tetun. Be Tetun ha'u rona ne'e, Tetun, ho Tetun agora ne'e Tetun nasionál ninia ne'e la hanesan ho ami-nia Tetun. Ami-nia ne'e Tetun Terik. Tetun Terik maibé ninia lokat, ninia hahalok ne'e oin seluk. Hanesan porezemplu. Tetun Terik, ita dehan Tetun Timor agora ne'e ita dehan ema maksa'ek, ema ne'ebé sa'edór ai, ou sa'edór kuda. Ne'e agora ne'e dehan fali dehan maksa'ek. Aviaun ne'ebé semo kona leten ami dehan loer. Tetun amik ne'e dehan loer husi kalohan maibé agora Tetun, agora nasionál ne'e dehan maksemok. Ne'ebé Tetun la hanesan ho ami-nian iha ne'e. Agora ne'e ami Tetun uluk ne'e. Ami dehan loer, sira agora Tetun nasionál agora ne'e dehan katak maksemok. Sa'e emadór, sa'edór kuda, sa'edór ai, ne'e agora uluk ne'e ami bolu dehan ema ne'e sa'edór ai mas agora dehan ema maksa'ek. Tetun la hanesan uluk, uluk ninian. Tetun agora Tetun nasionál, ami, ami hotu-hotu halo agora hanesan buat ida foun, ami aprende hotu ho alunu sira. Ha'u iha, ha'u eskola hahú iha míl, tinan rihun ida atus sia neennulu-resin-neen. Ha'u bá iha Soibada, iha eskola iha ne'ebá, Colégio Nuno Álvares Pereira. Ha'u bá eskola segunda klase, terseira, kuarta. Ha'u sai kuarta klase iha míl novesentus setenta i dois. Ha'u fila fali husi ne'ebá mai, iha ne'e komu profesór ausiliár iha Buikari ne'e la iha. Ne'e ha'u hanorin iha ne'e, ha'u kanorin iha ne'e, kanorin iha tinan rua nia laran i depois funu tama. Ami halai tiha ba ai-laran. Fila fali husi ai-laran, mai iha dia loron ualu fulan Agostu, míl novesentus, tinan rihun ida atus sia

ualunulu Indonézia haruka ami mai hanorin fila fali iha ne'e to'o agora, to'o daudaun ne'e ha'u sei hanorin iha ne'e. Ha'u eskola, sai tiha, mai hanorin. Agora ne'e ha'u sei iha hanorin nafatin iha ne'e. Ha'u dezde moris, ha'u-nia lian Makasae mós la hatene, Naueti mós la hatene. Ami hatene mak Tetun Terik. Tetun Terik iha Viqueque ninian la hanesan ho Tetun nasionál agora hotu tanba Tetun uluk oin ketak Tetun ikus ne'e hanesan. Tau hotu ba ema hotu atu hateten ko'alia halo lian ida de'it ne'e mak bolu dehan katak manu, ami-nia bolu dehan manu semo, manu tete, maibé agora sira dehan halo lian Tetun nasionál ninian dehan katak maksemok. Ema sa'edór kuda. Sa'edór nuu, sa'edór ai. Ema ne'e ami dehan sa'edór maibé Tetun nasionál ninian dehan katak ema maksa'ek. Ne'ebé ami, agora ami hotu aprende hanesan lia husi nasionál ninian mai ami ho labarik sira aprende hamutuk. Tanba Tetun nasionál ho Tetun Terik, Tetun Terik ne'e só para distritu ida-idak ninian. Maibé ba hun, ba tomatomak tanba balu Makasae, balu Naueti, balu Kairui, balu Bunak, balu Waima'a, ne'ebé sira ladún la ratene ami lian ne'ebé tende uza lian ne'e ba ida-ida de'it para hotu-hotu bele tuir lian ne'e. Se nukazu sira mai ko'alia ho lilian Tetun Terik amik sira la hatene. La tada. Ne'ebé agora estadu nasaun tau lia ida dehan katak, língua ofisiál ou língua Tetun halo lian ida para hotu-hotu, ne'e bele hatene hotu. Ne'e mak agora ami hotu-hotu hatene Tetun ne'ebé nasionál ninia. Ne'e ha'u-nia liafuan mak ne'e de'it i obrigadu.

0045TV

tx@LP   Ha'u-nia naran, naran L.... Ha'u moris iha loron ruanulu-resin-ualu, fulan Outubru, tinan rihun ida atus sia neennulu-resin-lima. Ha'u moris iha rai ida naran, ami-nia rain tuur fatin iha Kulumaten. Ha'u-nia inan naran Bui Lera, aman naran Kai Seran. Ha'u moris iha ne'e. Iha tinan rihun ida atus sia hitunulu-resin-rua, ha'u-nia aman ida bé naran ikun ne'e, bolu dehan te'iku (ultimo), nia lori ha'u bá iha Viqueque. Nia iha ne'ebá, nia servisu ne'ebá hanesan be badain-ai, karpinteiru, badain-ai. Ha'u bá hela ho nia, eskola iha ne'ebá to'o iha tinan rihun ida atus sia hitunulu-resin-haat komesa ona be golpe ona, ha'u bá tiha be klase rua. Ami hanesan halai ona ba ai-laran. Bá iha ne'ebá, halai to'o Indonézia mai fali iha ne'e. Tinan rihun ida atus sia hitunulu-resin-lima ami mai fali. Ha'u komesa bá eskola fali iha tinan rihun ida atus sia hitunulu-resin-lima ne'e. Ami eskola iha ne'ebá, kontinua iha, iha klase rua. Eskola, eskola. Iha tinan rihun ida atus sia ualunulu-resin-rua eskola bá hanesan klase neen ne'e hotu ona. Ha'u bá fali SMP, SMP. Tinan tolu tan iha ne'ebá. Hotu iha be tinan rihun ida atus sia ualunulu-resin-lima ha'u bá ona eskola iha Dili. Dili mós dehan Dili be, iha tinan rihun ida atus sia ualunulu-resin-lima foin bá eskola iha Dili. Bá eskola iha ne'ebá, bá eskola kona-ba mestre ne'e. Mestre ne'e tinan tolu. Tinan tolu ne'e hotu. Hotu iha be tinan rihun ida atus sia ualunulu-resin-ualu. Hotu, hotu ne'e ha'u hela iha ne'ebá, halimar iha ne'ebá. Tiu sira iha ne'ebá. Ne'e uluk dehan, bapa nia tempu, xefe bolu dehan: Ita bá ajuda ha'u, ajuda ne'e servisu. Ha'u mai servisu iha be pedreiru, hanesan be fatuk nian. Mai iha Vemasse. Iha ne'ebá ne'e fulan tolu hanesan ne'e ha'u iha Vemasse. Sira bolu ona dehan: Ó, ida be hanesan tama

tiha ona uluk bapa nian dehan SPG ne'e, mestre sira. Ha'u mai ona, komesa ona, mai, mai ne'e bolu ona, dehan, imi atu bá ona hanorin. Ha'u hanorin, mai ona, no amu-papa mai iha, mai tiha depois ami mai ona fahe ami atu mai. Ha'u mai iha ne'e iha tinan rihun ida atus sia ualunulu-resin-sia ne'e, mai ne'e. Komesa hanorin ona. Hanorin iha Ratahur depois mai fali iha ne'e to'o agora ne'e ha'u iha, ha'u kaben-na'in ona. Oan na'in-tolu. Mane ida, feto na'in-rua, mane eskola iha Dili. Nia agora, daudaun ninia SMP klase tiga ona atu hotu. Feto SMP klase ida, ne'e eskola iha ne'e. Feto oan ida ki'ik liu nia SD klase tolu iha ne'e. Ha'u-nia istória mak badak ne'e de'it, ne'e. Ha'u hanorin iha ne'e, hanorin klase haat. Tama loraik, ami sei ezame, agora sei, ezame hanesan lokál nian. Manuál. Alunus barak. Ha'u-nia iha na'in-ruanulu-resin-hitu. Oan-feto, feto sanulu-resin-lima, mane sanulu-resin-rua. Hamutuk ruanulu-resin-hitu. Agora sei tuir ezame ne'e. Aprende lian Tetun ho Portugés hanesan be hanorin, depois ita oh bele tradús fali ba iha Portugés, sira hanesan neineik, neineik bele. Tetun. Hanesan be Tetun Terik agora ne'e ka ami kolega maluk, maluk ne'e mak tuir ne'e kona-ba kursu kona-ba Tetun iha Tetun Terik nian ne'e. Tetun Terik ne'e daudaun ne'e la hanesan uluk. Muda. Nia istória mak badak ona.

0046TV

tx@JCV   Ha'u kakarak atu konta istória Ué-Aifoun. Ué-Aifoun, uluk wee la iha. Ha'u bei-tuak naran Fahi Lekik, bá tuur heli iha wee karék ida. Foin wee karek ne'e, kalan, fahi bá roku. Bá roku iha ne'ebá, bei-tuak Fahi Lekik ne'e bá nameke heli. Foin fahi ida, la'ós fahi iha wee laran. Maibé malae-feto ida no mane ida sei rariis iha wee laran. Bei-tuak Fahi Lekik ne'e nia bá soro, nia nodi kilat, kilat fatuk ida, ai-huuk ida, diman ida, no rama ida. wainhira nia bá tuur heli, nia bá naree heli, fahi ne'ebé roku wee karék ne'e. La'ós fahi maibé malae-feto ida no mane ida ha'e sira rariis wee karék ne'e. La'ós wee karék maibé bee-lihun ida. Wainhira bei-tuak Fahi Lekik sei nameke malae-feto no malae-mane ne'e sei rariis, kilat fatuk ne'ebé bei-tuak nodi ne'e sadere ba kamii malae ida, kamii malae ne'e tohar, nia sanak maten ne'e tohar no kilat monu ba rai. Malae-feto no malae-mane ne'e foti ulun raree ba heli matan, retan kedas ha'u bei-tuak ida naran Fahi Lekik. Entaun malae ne'e bolu ona bei-tuak Fahi Lekik ne'e bá sira, atu bá iha sira-nia knua ida. Malae ne'e hodi ona bei-tuak Fahi Lekik ne'e, bá iha sira-nia knua. Entaun sira la'a, sira, malae ne'e haruka bei-tuak ne'e taka tiha matan. Ba ó hakarak, ó makarak saida. Bei-tuak na'ak ka ha'u kakarak buat ne'ebé Ita-boot sira hakarak. Entaun sira, entaun malae ne'e haruka bei-tuak Fahi Lehik ne'e, entaun ó makarak buat ne'ebé ami hakarak, entaun ó taka tiha ona nia matan. Wainhira ami haruka atu loke ha'a, ó loke ó-nia matan. Entaun malae ne'e kaer ba bei-tuak ne'e nia liman, malae haruka ne'e taka nia matan, taka tiha nia matan, la kleur malae haruka ba loke ona nia matan. Bei-tuak Fahi Lekik ne'e loke tiha nia matan, haree ba la'ós be ai-laran, la'ós wee karék. Maibé iha fali uma-andár iha sidade laran ida, sira tama ne'e iha sidade laran ida iha malae nia uman. Entaun malae ne'e nusu ba bei-touk Fahi Lekik ne'e. Buat hotu ne'ebé ami iha ne'e ó makarak ida-

ne'ebé. Entaun bei-tuak Fahi Lekik ne'e natán ba malae ne'e na'ak ka: ha'u kakarak de'it modo no ha'u kakarak de'it loron manas ruma, ha'u atu ba kariis wee namata, wee malirin, kodi buka modo ruma. Entaun ó makarak saida, makarak loos saida, makarak ami fó ikan ka fó tuna ka fó lafaek ka fó saida na'an ida, wee laran ida ba ita. Lae, ha'u kakarak fó na'an, ikan tasi laran nian maibé atu mosu fali iha wee, wee laran ida para ami tinan, tinan-tinan ami bele bá buka kmodo iha ne'ebá. Entaun malae ne'e halo ho fila fali nia bá mai fali iha rai wa'iwa'in, rai itak ne'e. Malae-feto no mane na'ak ka: entaun loos ona. Se ó makarak de'it ikan, makarak de'it modo ne'ebé iha wee laran entaun ita fila fali ba, ó bá tiha uma. Ah, rai kbesik tuku haat ó rona saida ha'e atu mosu iha fatin ida-ne'e. No ó bá, ó tenke fó matene ba ema sira ne'ebé te'in masin iha tasi, ne'ebé hela besik ba fatin ne'e karik, tenke sa'e tiha ba foho. Tanba aan rai-kmaus ne'e saida ha'e atu mosu ba fatin ne'ebé ami hariis ba ne'e. Entaun malae ne'e haruka bei-tuak ne'e dukur fali nia matan, entaun dukur tiha ó matan. Ita fila fali ba rai imi nian. Entaun bei-tuak ne'e taka tiha nia matan. derrepente haruka loke, nia loke fali nia matan, ne'e iha fali ona fatin ne'ebé ne'e hetan malae ba ne'e. Entaun malae haruka nia ó fila fali ba, aan rai-kmaus ó rona buat ruma tarutu iha ne'e karik. Ó lalika makfodak. Tanba buat ne'ebé ó presiza ne'e, ó makarak ne'e, ami fó naha mai ne'e. Entaun, fó matene ba la tuan sira, ferik oan sira ne'ebé te'in masin ne'e, sa'e fali tiha ba foho, bá fila fali tiha sira uman knua. To'o sira bá tiha uma knua, rai kbesik tuku tolu, tuku haat ne'e, udan tau ona iha wailoro bein ne'e iha keda fulan Outubru nia laran. Udan tau, udan tau la bá fatin seluk, iha de'it bá fatin ida be bei-tuak hetan malae-feto no malae-mane bá ne'e. Ne'ebé rai-tarutu, kakeu sira ne'e be lori tama remata ba tasi laran, foin wee fatin ida be kamii malae laran bee karek iha bá ne'e, lae naha wee karék mas sai fali tiha mota inan ida. Depois ikan mós wa'in loos. Foin rai-kmaus oan bei-tuak ne'e bá fali iha be fatin ne'ebá, bá netan ha'e malae ne'e. Netan hi'as malae ne'e iha fatin ne'ebá mas la'ós ona iha fatin be rai maran, sira bá retan fali malu iha bee-lihun laran. Tama tun tiha bee-lihun laran ne'ebá. Depois malae ne'e natudu tiha nia uman sira ne'e, fatin sira ne'e. Na'ak ka: wainhira imi atu mais suru ruma karik, ó tenki tama tun mai iha bee laran ne'e hodi sulan tiha kuak matan sira ne'e atu nune'e ikan sira ne'e la bele sa'e ba fatin ruma wainhira ema atu mai buka kmodo. Entaun bei-tuak ne'e fila tiha mai uma, nia mai fó natene ona liurai sira, dato sira, reinu sira, sira atu bá ona buka kmodo, bá suru. Wainhira sira to'o ba suru, molok atu bá suru, bei-tuak ne'e tenki bá tun uluk ba wee laran, bá nasoru uluk malae ne'e hodi taka tiha lai kuak matan, taka kuak matan ne'e iha rai, iha wee okos. Taka iha wee okos. Depois wainhira nia sa'e mai nia haruka bá liurai sira, na'i sira atu bá sa'u kosar karik, bele bá ona sa'u kosar ba, hisik kosar ba tanba kuak matan ha'u ksulan taha be. Abó sira la iha na'ak ka lafaek. Lafaek sira ne'e la iha taha, avó la iha ne'e, na'ak ka lafaek la iha taha ba. Ne'ebé bá ona hariis, bá ona sa'u kosar ba. Entaun liurai sira no dato sira haruka ona reinu sira ne'e bá ona suru, bá ona buka kmodo. Wainhira bá buka kmodo bei-tuak Fahi Lekik ne'e nia mai toba tiha iha nia kmamahan. Nia toba tiha nia kmamahan. wainhira sé mak tun ba suru wee, tenke bá foti tali lahan rua. Tali matak ne'e, tali mutin ne'e lahan rua, ida kesi ba ulun, ida kesi ba knotak, hodi hatudu na'ak ka ita ne'e mesak wee ne'e nia na'in. Ne'ebé wee nia na'in ruma, animál fuik ruma, lafaek ruma ka, ne'e la bele nasé,

la bele kamat ema ne'e. Ne'ebé molok sira atu tun ba suru, sira tenke bá foti tali ne'e lahan rua, ida kesi iha ulun, ida kesi iha knotak, hodi hatudu na'ak ka, ema ne'e mesak wee ne'e nia na'in. Ah, bei-tuak Fahi Lekik ne'e nia fila fali ba toba tiha nia kmamahan, nia la bá suru. Depois suru nahas tiha sa'e mai, ema sira ne'ebé tun ba suru ne'e, losu, losu na'an ikan ne'e, na'in rua-rua hatama ba liurai sira no dato sira. Wainhira hatama tiha ba dato no liurai, entaun liurai sira no dato sira ne'e foti fali ikan ne'e na'in-rua ema ida hatama ba liurai sira ne'e foti fali na'in ida-ida fó fali ba bei-tuak Fahi Lekik ne'e, hasa'e nu'udar nia, nia kmodo. Ne'ebé kada tinan-tinan sai hanesan ne'e, sai nafatin na'o ne'e, to'o oras ne'e no uhu nian ne'e. Ami bá iha suru wee ida naran Ué-Aifoun ne'e, komesa uhu nia, komesa aan, wainrua, waintolu, to'o Domingu tama kesak laran. Ne'ebé wee ida naran Ué-Aifoun ne'e kada tinan-tinan ami sempre bá buka kmodo iha ne'ebá, bá suru, na'ak ka ba sa'u kosar, hitis isin kosar, buka namata no malirin, hodi buka kmodo iha fatin ida naran Ué-Aifoun. Ne'ebé be Ué-Aifoun ne'e nia na'in, ne'e bei-tuak Fahi Lekik ne'ebé mate naha. Nia jerasaun mak hela ba aat, nia jerasaun ne'ebé daudaun ne'e norik bei-tuak Fahi Lekik ne'e oan la iha. Na'i-feen, ha'a na'i-feen ida naran no, naran Bui Rubik. Oan la iha. Nia nakiak fali ha'u na'in, naran Na'i Funuk, ha'e nela no sira no wee ne'e daudaun ba jerasaun ne'e ami sira ne'e. Ida-ne'e ha'u-nia duni. Ne'ebé ami daudaun ne'e ha'u fó fiar ba ha'u-nia primu sira ne'ebé la'ós servisu nu'udar funsionáriu, sira ha'e tau matan raree wee ida-ne'e kada tinan-tinan, sira ha'e prepara kesak, sira ha'e halo salaak ba liurai sira, fó hatene ba liurai sira suku sanulu ne'e atu mai suru. Ne'ebé istória badak ne'ebé ha'u konta kona-ba Ué-Aifoun, to'o naha ba ne'e no nahas na'a ba ne'e. Buat ruma ne'ebé ladi'ak karik, ita bele buka malu nafatin, hodi hadi'a no hodi halo kmanek liu tan atu istória ne'e sai furak ba ita-nia jerasaun, ba ita-nia oan, ita ne'ebé oan sira. Obrigadu.

0047TV

tx@AFA   Ha'u-nia naran, ha'u naran, naran A.... Ha'u mai husi ha'u tein ha'u na'in, mai husi ema ida be simples. Ema wa'iwa'in. Be ha'u tein ha'u na'in komesa husi ha'u na'in-ferik i ha'u na'in-tuak ema Tetun Terik tuur ba tanbasá terik, tanba tuur no ba tetuk no klobek ne'ebé nia bailoro leet nasoru malu, tan nia maluk, tan nia alin-maun, nia rian, nia sa'in, nia feton, nia inan. Sira la'a to'o malu, rasoru malu, so ratene de'it Tetun Terik ha'e rori lian ne'e rasoru malu. Lori lia, lia seluk, lian kona-ba iha kaletek kafotik sira la iha. Sira la ratene. Tanba tuur ba tetuk no klobek, ne'ebé sira kaer duni ba lia ne'ebé, lian Tetun Terik. Ne'ebé tuir mai oan ha'u, sira na'in oan ha'u. Wainhira basa ba hadak, ha'u tein to'o mai nusu, nusu ba ha'u na'in, nia bá tiha to'os, hakburas mai ha'u tein nusu na'ak ka warik oan tan ne'e nori feto eh mane. Ne'ebé ha'u na'in iha ahi-matan natán soru na'ak ka: ha'u oan-mane hodi loos an. Ha'u umakain, uma riin, lala'ok oan, leo la mara, ahi la mate tanba oan-mane iha nafatin ne'ebé nia moris ema ida sei lori la'a maibé Tetun Terik nia sei nodi nori lian ne'e, lian Terik nia sei la'a nafatin. Ne'ebé ha'u moris mai husi ha'u tein no ha'u na'in. Wailoro leet naruka ha'u bá kalo to'os, lori Tetun Terik

507

bá kala, bá nateten, ha'u krona ba, krona kusi lia ida-ne'e, naruka ha'u ba kesi kuda, lori lian ida-ne'e de'it ne'e ohin loron ha'u kateten tuir lian ida-ne'e de'it. Lian seluk ha'u la kateten tanba la'ós ha'u lian. Maibé iha tan lian seluk ha'u krona duni maibé tanba ranxu malu, la'a to'o malu ha'e hodi hateten. Lian ida ha'uk kontinua nanu'u ohin. Maluk sira ratein buat ne'e, lian ne'e mós mai iha nanis kedas ne'ebé tein no na'in ranorin kedas ida-ne'e, hateten ba ha'u na'ak: Ó atu bá mii lai karik, lori lian dehan mii lai, bá kuru wee karik lori lian na'ak ba kuru wee, bá, bá mai lailais, la bele ukun-laek, la bele malo tun malu sa'e, lae ó bá metan buat ladi'ak ruma iha dalan, susar ruma iha dalan. Ne'ebé ha'u moris mai komesa kedas iha loron ida de'it. Ha'u tein ha'u na'in nodi kedas lia ida-ne'e natama ba ha'u laran, ba ha'u to'o ne'e katak ke lian seluk hodi be lian Tetun Terik. Ne'ebé ha'u ida-ne'e de'it. Ha'u ka detein tanba ha'u moris kusi ema Terik nia oan mas se uhu lai bei-tuak, tein, nia lian mak ne'e. Obrigadu.

0049TV

tx@MS   Obrigadu ba tempu. Ha'u naran M.... Ha'u naran jentiu naran Rubi Fuluk tanba ha'u-nia ha'u tein naran Funu Kaik, ha'u nain naran Kasa Nahak. Ne'e duni ha'u ohin kateten liu tiha ba istória kona-ba ha'u-nia moris. Maibé ha'u-nia istória badak ida fali liu husi buat ne'ebé ha'u krona kusi ha'u tein no ha'u nain sira ne'ebé istória. Entaun ida-ne'e ha'u kaktuir katutan ida-ne'e atu ne'e bele sai istória ha'u tein naran Funu Kaik, ha'u nain naran Kasa Nahak. Ha'u bei-tuak ida naran Leki Lubik, ha'u nain nia, tein nia alik ne'ebé ha'u kela ko nia iha tempu kona-ba ha'u nain moras, moras no nodi husu lemo ne'e. Entaun ha'u kela fali ko bei-tuak. Ha'u bei-tuak ha'e nawa'i fali ha'u. Entaun bei-tuak konta ona istória ida. Nia bá to'os, nia bá to'os, nia loroloron nia bá to'os, bá naree lekirauk ran nahas tiha de'it hudi, oh lamak. Ran nahas tiha lamak. Kuda batar bó lekirauk ran nahas tiha. Tan loron ida nia bá natuur. Nia bá tuur, naree lekirauk ha'e mai ran, la'ós ema ha'e mai rola maibé lekirauk ha'e mai ran, nia bá ona nalo tatakak. Nia nalo tiha tatakak atu hodi taka lekirauk ne'e. Tuku haat rai-kbesik nia nalai ona bá atu tuur heli ne'e atu nodi taka lekirauk. Nia bá, lekirauk ida leok ona. Oh, imi na'in hira tomak mai be saida, ita bá han tiha aan atu bá hasoru oh Antony iha barasa, iha barasa ne'ebá. Ne'ebé imi mai be bá han tiha ita, bá hasoru ne'ebá. Ne'ebé latuan rona tiha ida-ne'e, nia bá atu taka lekirauk. Nia la konsege taka. Nia mai, nia bá ona Dili. Nia bá Dili, bá nein ona be saida, sá emboot Antony atu mai. Entaun nia bá iha Dili, bá iha ne'ebá, bá to'o rai-kmaus ona ró ne'e tun. Kapten ró ne'e tun, mai ida saida maka Antony entaun nia bá ona kaer liman. Oh Antony, ita horiseik tuir loloon ita daudaun, la sai, la mai iha ne'e, mai horiseik kuandu haktaka tiha ita karik ita la mai iha ne'e. Maibé saida, ne'e señór Antony nesuk ona kala la bele, la bele hateten sai ou fila fali ba iha ne'ebá ó to'os ha'u sei la, la kala aat tan. Ne'e duni istória ida-ne'e ne'e tanba animál sira fuik ne'e mós oh, sira mós oh sai ema nanu'u mós ita. Ne'e duni ho ida-ne'e animál fuik sira ne'e, ha'u foin ne'e, ha'u kohodór duni animál maibé to'o ba lekirauk ha'u kanoin, ha'u kanoin to'o ba ida-ne'e

entaun ha'u mós la, la koho animál lekirauk ho ida-ne'e. Ida-ne'e ne'e ha'u-nia istória badak.

0050TV

tx@JCV   Ho Tetun Terik? Di'ak, obrigadu ba tempu. Ha'u atu konta istória badak oan ruma kona-ba rai Aisusu Aisahe nu'udar rai ina, rai oa, rai husar, rai oar, iha Loro-Sa'e. Aisusu Aisahe, uluk bei'ala sira-nia tempu hela ona, hela naha ema lubuk ida ne'ebé moris iha organizadu ne'e laran. Iha ema sira-nia liurai iha, sira-nia xefe, sira-nia kbahen, lia-na'in sira, kompletu. Wainhira misionáriu sira iha, iha Portugál rona na'ak ka rai de Viqueque iha knua ida naran Aisahe, ema moris ho ninia liurai, ninia don, ninia koronél, ne'e dato no ne'e kbahen sira. Entaun iha Portugál padre na'in-rua hosi sira-nia misionáriu tomak mai sai husi Portugál. Padre na'in-rua mai iha Lifau, Larantuka, iha Larantuka. Sira mai atu mai liu Timor tanba Timor ema mesak animizmu. Mesak fiar ba lulik. Mesak jentiu. Wainhira sira rona iha Timor ema mesak jentiu, atu sira mós hakarak atu mai habelar, haklaken Na'i Maromak ninia dotrina ba ema sira iha rai ida naran Aisahe iha Loro-Sa'e. Entaun sira na'in-rua mai to'o iha Lifau, Oekusi. Malae-mutin seidauk tama iha Timor. Sira na'in-rua mai tama iha Oekusi, sira na'in-rua mesak mai ho sira-nia ekipa ida. Mai ekipa ne'e balu ema indonézia, balu ema arábia, balu ema índia, mai hamutuk ho padre na'in-rua ne'e. Mai balu hela iha Oekusi, Lifau. Iha ekipa ida mai ho padre ida naran Frei Joao Tavares. Sira mai iha, hela iha, sira na'in buka rai ida naran Aisahe. Iha Viqueque, Loro-Sa'e. Sira mai to'o buka rai Aisahe, sira mai hetan tiha rai naran Aisahe. Ema liurai sira, dato sira, simu ho kultura, tradisaun oioin hodi mean hodi osan mean, osan mutin, hodi fó ba na'i-lulik ne'e ho nia ekipa lubuk ida ne'e. Fó tiha bá hodi hatudu katak sira atu la bele fila ba sira-nia rain. Sira atu hela ho reinu sira iha rai Loro-Sa'e iha Aisusu Aisahe. Hela ho povu sira iha rai Aisahe. Wainhira sira mai hela hamutuk ho reinu sira, dato sira, liurai sira iha rai Aisahe, sira falta ai-han tanba sira-nia ekipamentus kona-ba hahan la iha. Sira nafatin hela ba ladún di'ak. Entaun iha madre balu fila tiha, frater balu fila tiha, sira hela de'it sei na'in-lima ka na'in-neen ida hanesan ne'e hamutuk ho na'i-lulik. Frei Joao hela iha rai ida Aisahe. Iha knua ida naran Aisahe. Sira hela iha ne'ebá. Na'i-lulik ne'e komesa haruka ninia, nia zeladór ou nia katekista naran Nahak Simao ho nia feen naran Bui Kiteria atu hanorin ema atu tuir dotrina, bele hatene atu sai sarani iha tempu ne'ebá. Wainhira Nahak Simao ho nia feen Buikiteria hanorin ema liurai ida iha Aisusu Aisahe ne'e nia sarani uluk tanba nia fiar na'ak ka Maromak ne'e loos liu fali ita-nia lulik ne'ebé ita adora durante ne'e. Tanba ema ne'ebé mai haklaken ne'e loos ita-nia maluk sira maibé ema mai husi rai nasaun seluk. Depois ema ne'e hanorin Maromak nia liafuan ne'ebé kona-ba hadomi, perdua no tulun malu. Wainhira sira hatene ita orasaun oan sira ne'e, sira komu na'i-lulik komesa sarani sira. Sarani uluk maka liurai ne'e, depois tuir mai nia dato sira, to'o ba nia reinu sira. Liurai ne'e iha ninia emar ne'e nia kuziñeiru, ne'e emar ne'ebé prepara ba hanesan zeladór ka katekista. Iha momentu ne'ebá ninia emar ne'ebé haree ba nia te'in naran Gaspár, prepara

na'i-lulik nia hahan, naran Gaspár. Wainhira nia prepara hahan tanba wee la iha, wee kdook, kuru wee kdook. Nia mesak de'it bá kuru wee, te'in, fase na'i-lulik nia ropa la bele entaun nia nalia tan na'e ema na'in-ida naran Goncalo husi sira Uma-Ki'ik. Ida Gaspár ne'e na'i-lulik lori, hodi no kedas husi, husi Larantuka, husi ema indonézia. Larantuka. Ida Nahak Simao ne'e ema Larantuka, ema Flores, ida ne'e feen naran Buikiteria ne'e ema husi Arábia. Ema padre mai hodi no kedan husi Arábia mai liu husi Índia. Wainhira to'o mai iha ne'e ida na'i-lulik ne'e emar ida prepara ai-han, fase nia ropa, prepara na'i-lulik ninia hariis, hatais, naran Gaspár ne'e, prepara ba la bele entaun nia bolu tan ona ida naran, nalia tan ida naran Gonsalo. Sira na'in-rua prepara hahan ba na'i-lulik, prepara hariis, fase ropa ba na'i-lulik. Entaun iha ne'e, ida na'i-lulik ne'e komesa hanorin, hanorin ema, ida bé Nahak Simao no Buikiteria ne'e hafee ona malu, hola ona, sai ona feen no la'en. Sira na'in-rua kaben. Depois sira-nia oan, jerasaun to'o mai ha'u ne'ebé dadauk hateten daudaun ne'e. Ida ninia emar na'in-rua ne'ebé haree hahan ne'e Gonsalo no Gaspár ne'e. Gonsalo ne'e husi Uma-Ki'ik, ida Gaspár ne'e husi ema indonézia maibé nia mai sai nu'udar ema Aisahe. Ida Gaspár ne'e oan la iha. Nia mai, la hola feto, to'o nia mate oan la iha. Ai, ida Buikiteria no Nahak Simao ne'e sira na'in-rua hola malu, sai feen no la'en, sira oan mane ida naran Nasu Nahak. Nasu Nahak nia oan naran Koli Nasuk. Koli Nasuk nia oan naran Naha Kolik. Naha Kolik nia oan Liu Nahak. Liu Nahak nia oan ne'e ha'u. Se ami lubuk ida kedas ami iha ha'u naran Koli Miuk. Joao Cristovao Viana. Ne'ebé sira moris iha ne'ebá. Depois padre ne'e promete wainhira loron karik ha'u deskansa, imi la bele fanun ha'u. Tanba ha'u deskansa ne'e dala ruma ha'u deskansa tebes, toba ka deskansa maibé dala ruma ha'u bá iha rai li'ur, hanesan ha'u-nia isin-na'in toba iha ne'e maibé ha'u-nia klamar ne'e sa'e ba fatin seluk, bá buka ba iha nasaun seluk ruma ne'ebé ha'u karik toba deskansa, imi la bele fanu ha'u. Maibé iha ne'e tanba bei-tuak sira ne'ebá la ratene, na'ak ka na'i-lulik ema ida kuandu nia toba liu naha loron rua, loron tolu sira hakfodak, sira tenki bá fanu. Entaun iha tempu ne'ebá bei-tuak ida naran Rubi Nahak bá naree na'i-lulik ne'e, toba la nadeer durante loron tolu nia laran, na'i-lulik ne'e la nadeer. Entaun nia lees tiha didin, ai-kele bikun ne'e, didin, lees tiha didin, nodi naree tama ba laran. Naree ba laran, bá na'i-lulik ne'e toba la nadeer, hanoin na'ak ka na'i-lulik ne'e mate naha. Entaun latuan ne'e bá ona, leot ona ba xefe sira no liurai sira: Oh, mai lalais, mai lalais, na'i-lulik mate naha toba la nadeer. Mai lalais! Entaun latuan sira ne'e ho liurai sira, dato sira ne'e bá lees tiha didin. Tanba koko ba na'i-lulik ne'e, nia la mate. Na'i-lulik nia isin manamanas nanis mais na'i-lulik ne'e la mate, iis la iha. Mas nia isin sei manas, isin sei mamar nanu'u ita wa'iwa'in ne'e. Maibé nia iis la iha. Entaun sira rakrake na'ak oh na'i-lulik mate tiha ona, na'i-lulik mate tiha ona. Ita bá ona, taman tiha. Sira tau de'it na'i-lulik ne'e ninia isin mate ne'e durante tinan haat nia laran iha sira-nia leet. Tanba sira radomi na'i-lulik ne'e. Sira rohi taman. Radomi, entaun sira rohi taman. To'o ikus mai sira hanoin be ladi'ak, sira tenke taman. Depois na'i-lulik ne'e nesuk na'ak ka. Wainhira ha'u, nia sei moris nesuk na'ak ka. Karik iha tempu ruma ne'ebé ha'u kuandu hakmatek karik, imi la bele taman ha'u. Se imi taman ha'u entaun taman tiha imi-nia sorte ne'ebé aan wainrua neineik-neineik imi oan, imi bei-oan sira sei, sei retan. Se karik imi taman tiha ha'u, kuda talin ida, daer iha karuik ida, ba

karuik ida, mós imi labele tuir sei nakotu tiha. Ai-tukan ida latan ba karuik ida ba, tau klalete ba mós imi haklete tuir sei tohar tiha. Tanba imi-nia sorte, imi taman tiha ona ne'e. Se imi la taman ha'u-nia isin maten karik, iha tempu ida ba tempu be ha'u-nia maun-alin sira, iha Vatikanu ratene, sira be, sira mai foti tiha ha'u-nia isin maten, rodi tiha ba no imi sei hetan buat di'ak ne'ebé imi la sente maibé nia sei mosu iha imi-nia leet. Maibé bei-tuak sira iha tempu ne'ebá la ratene. Sira taman tiha na'i-lulik ne'e nia isin maten. To'o ikus mai na'i-lulik ne'e ninia isin maten, sei tama nalik iha ami-nia knua ida naran Aisahe. No foin lalais iha horibainruak dia domingu, domingu horiwaintolu iha dia loron ruanulu-resin-rua liubá. Na'i-lulik ida husi vigáriu, husi parókia Viqueque bá naree tiha na'i-lulik ninia sasán sira ne'e. No na'i-lulik nia rate ne'ebé iha ami-nia knua ne'e taman hamutuk. Na'i-lulik ne'e iha estátua, taman iha ne'ebá no na'i-lulik nia rate ne'e sei kapás loos, sei di'ak, hada rate ne'e halo kabuar tiha, kabuar ida primeiru, halo klu'ak, maizumenus, metru sanulu. Depois tuir fali mai halo kabuar tan, metru lima. Halo kabuar. Depois iha ne'e tetek ida, depois iha fohon liu, iha klaran liu ne'e hada loos ona hanesan kaixaun sai nu'udar na'i-lulik nia rate. No to'o ohin nia loron ami sira jerasaun ba jerasaun ne'ebé família parte sira reprezenta família na'i-lulik ne'e niak. Wainhira na'i-lulik ne'e mate tiha mós, ami ne'ebé, ami-nia bei'ala sira mai ramutuk no na'i-lulik husi Portugál, husi Indonézia, mai iha ne'e, sai belar to'o ohin nia loron ne'e ami sira ne'e la husik na'i-lulik nia ne'e rate kada tinan Finadu ami sempre ba sunu ahi oan ba, kari ai-funan. I loron-boot ruma i tempu ne'ebé ami sente dehan, ami sente na'ak ka ami oan ruma atu bá iha semináriu karik, bá eskola atu buka matenek ruma. Ami sempre, ami bá sunu nafatin, ahi oan ba na'i-lulik ne'e ninia rate atu hodi husu nafatin matak malirin, atu ami-nia oan sira ne'e eskola bá ohin ho di'ak tanba na'i-lulik ninia akompañamentu, nia klamar kala akompaña nafatin ami-nia oan sira. Ha'u-nia istória to'o ba ida-ne'e. Se karik buat ruma ne'ebé ladún di'ak karik, iha tempu ruma ita bele buka malu, se bele hadi'a liu tan istória oan hirak ne'e. Maibé istória ne'e lo'os, lo'os dehan ami foti hanesan karan de'it ka ami naran tau de'it ba. Maibé na'i-lulik ninia rate sei iha ne'e no ninia sasán hirak ne'ebé nia mai hodi no bikan tuan sira ne'e, insensu fatin ho krús, estátua, ami tuur no iha, iha Aisahe ne'e. No foto retratu sira ne'ebé na'i-lulik niak ne'e dadaun ha'u kodi no iha ne'e. Na'i-lulik nia rate ha'u kasai kodi no, no bele, ha'u bele foti mai, señora bele hasai tiha mós bele. Tanba ha'u kodi no iha ne'e. Na'i-lulik ninia rate iha ne'e. Di'ak, hanoin kala ida-ne'e istória ida badak oan, keta buat ruma ladún di'ak karik ami husu deskulpa, buat ruma karik ladi'ak atu hadi'a karik, ita buka tempu ruma atu hadi'a. Obrigadu.

LIST OF REFERENCES

Adelaar, A. K. (2005). Structural diversity in the Malayic group. In A. Adelaar and N. P. Himmelmann (Eds.), *The Austronesian Languages of Asia and Madagascar* (pp. 202-226). London: Routledge.

Adelaar, A. K., & Prentice, D. J. (1996). Malay: its history, role and spread. In S. A. Wurm, P. Mühlhäusler, & D. T. Tyron (Eds.), *Atlas of Languages of Intercultural Communication in the Pacific, Asia and the Americas* (Vol 2., pp. 673-694). Berlin: Mouton de Gruyter.

Afonso, S., & Goglia, F. (2015). Portuguese in East Timor as a non-dominant variety in the making. In R. Muhr & D. Marle (Eds.), *Pluricentric Languages: New Perspectives in Theory and Description* (pp. 193-105). Frankfurt am Main: Peter Lang.

Aikhenvald, A. Y. (2006). Serial Verb Constructions in Typological Perspective. In A. Y. Aikhenvald & R. M. W. Dixon (Eds.), *Serial Verb Constructions: A Cross-linguistic Typology* (pp. 1-68). Oxford: Oxford University Press.

Albuquerque, D. B. de. (2011). *Esboço gramatical do Tetun Prasa: língua oficial de Timor-Leste* (Unpublished master's thesis). Universidade de Brasília, Brasília.

Albuquerque, D. B. de. (2014). *A língua portuguesa em Timor-Leste: uma abordgem ecolinguística* (Unpublished doctoral dissertation). Universidade de Brasília, Brasília.

Andrews, A. D. (2007). Relative clauses. In T. Shopen (Ed.), *Language Typology and Syntactic Description. Volume II: Complex Constructions* (pp. 206-236). Cambridge: Cambridge University Press.

Ansaldo, U., & Matthews S. (2001). Typical creoles and simple languages. The case of Sinitic. *Linguistic Typology 5*(2/3), 311-326.

Ansaldo, U., & Nordhoff S. (2009). Complexity and the age of languages. In E. O. Aboh & N. Smith (Eds.), *Complex processes in new languages* (pp. 345-363). Amsterdam: John Benjamins.

Avram, A. A. (2008). An Overview of Reduplication and Compounding in Tetun Dili. *Revue Roumaine de Linguisque* *53*(4), 427-448.

Baker, P. (1995). Motivation in creole genesis. In P. Baker (Ed.), *From Contact to Creole and Beyond* (pp. 3-15). London: University of Westminster Press.

Bakker, P. (1994). Pidgins. In J. Arends, P. Muysken & N. Smith (Eds.), *Pidgins and Creoles. An Introduction* (pp. 25-39). Amsterdam: John Benjamins.

Bakker, P. (2000). Convergence intertwining: an alternative way towards the genesis of mixed languages. In D. G. Gilbers, J. Nerbonne & J. Schaeken (Eds.), *Languages in contact. Studies in Slavic and General Linguistics* (Vol. 28, pp. 29-35). Amsterdam: Radopi.

Bakker, P. (2003). Mixed languages as an autonomous system. In Y. Matras & P. Bakker (Eds.), *The Mixed Language Debate: theoretical and empirical advances* (pp. 107-150). Berlin: Mouton de Gruyter.

Ballard, C. (2008). 'Oceanic negroes': British anthropology of Papuans, 1820-1869. In B. Douglas & C. Ballard (Eds.), *Foreign Bodies: Oceania and the Science of Race 1750-1940* (pp. 157-201). Canberra: ANU E Press.

Barbosa, D. (1992). *O livro de Duarte Barbosa*. Lisboa: Publicações Europa-América.

Barkhuizen, G. (2004). Social Influences on Language Learning. In A. Davies & C. Elder (Eds.), *The Handbook of Applied Linguistics* (pp. 552-575). Oxford: Blackwell Publishing.

Barreiros, D. L. (1943-44). *O dialecto português de Macau: Antologia*. Macau: Renascimento.

Batalha, G. N. (1959). Estado actual do dialecto macaense. *Separata da Revista portuguesa de filologia, 9,* 177-213. Coimbra: Casa do Castelo.

Batoréo, H. J. (2009). A língua portuguesa em Timor: de que forma deve o ensino de português adaptar-se às diferentes realidades nacionais? *Estudos Linguísticos/Linguistic Studies, 4,* 51-61.

Baxter, A. N. (1988). *A Grammar of Kristang (Malacca Creole Portuguese).* Canberra: Australian National University.

Baxter, A. N. (1990). Notes on the Creole Portuguese of Bidau, East Timor. *Journal of Pidgin and Creole Languages, 5*(1), 1-38.

Baxter, A. N. (1996). Portuguese and Creole Portuguese in the Pacific and Western Pacific rim. In S. A. Wurm, P. Mühlhäusler, & D. T. Tryon (Eds.), *Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas, Vol II: Texts* (pp. 299-338). Berlin: Mouton de Gruyter.

Baxter, A. N. (2009). Causative and facilitative serial verbs in Asian Ibero-romance Creoles – a convergence of substrate and superstrate system? *Journal of Portuguese Linguistics, 8*(2), 65-90.

Baxter, A. N., & Bastos, A. (2012). A closer look at the post-nominal genitive in Asian Creole Portuguese. In H. C. Cardoso, A. N. Baxter, & M. Pinharanda Nunes (Eds.), *Ibero-Asian Creoles: Comparative Perspectives* (pp. 47-80). Amsterdam: John Benjamins.

Baxter, A. N., & Cardoso, H. C. (2017). Early notices regarding Creole Portuguese in former Portuguese Timor. *Journal of Language Contact, 10*(2), 264-317.

Baxter, A. N., & Silva, P. de. (2004*). A dictionary of Kristang – English (with an English – Kristang finder list)*. Canberra: Pacific Linguistics/ANU.

Benu, N. N. (2016, May). *Possessive in Uab Meto*. Paper presented at Seminar Internasional Tahunan Linguistik (Annual International Seminar on Linguistics), Universitas Pendidikan Indonesia and Masyarakat linguistik Indonesia, Bandung.

Berkanis, A. (1996). *Fonologi Tetum Dili* (Unpublished master's thesis). Universitas Udayana, Denpasar.

Blust, R. (1993). Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics, 32*(2), 241-293.

Blust, R. (2013). *The Austronesian languages* (Rev. ed.). Canberra: Asia-Pacific Linguistics.

Booij, G. (2007). *The Grammar of Words: An Introduction to Linguistic Morphology* (2nd ed.). Oxford: Oxford University Press.

Bluteau, R. (1712-1728). *Vocabulario Portuguez & Latino: Aulico, Anatomico, Architectonico...,* (8 vols). Coimbra: Collegio das Artes da Companhia de Jesu.

Boxer, C. R. (1947). *The Topasses of Timor*. Amsterdam: Koninklijke Vereeniging Indisch Instituut.

Brito, R. H. P. de, & Corte-Real, B. de A e. (2006). Aspectos da política linguística de Timor-Leste. In M. de L. Martins, H. Sousa, & R. Cabecinhas (Eds.), *Desvendando contracorrentes. Comunicação e lusofonia: para uma abordagem crítica da cultura e dos media no espaço lusófono* (pp. 123-131). Porto: Campo das Letras.

Bybee, J. Perkins, R. & Pagliuca, W. (1994). *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World.* Chicago: The University of Chicago Press.

Cabral, E., & Martin-Jones, M. (2008). Writing the Resistance: Literacy in East Timor 1975-1999. *International Journal of Bilingual Education and Bilingualism, 11*(2), 149-169.

Capell, A. (1944a). Peoples and languages of Timor. *Oceania, 14*(3), 191-219.

Capell, A. (1944b). Peoples and languages of Timor. *Oceania, 14*(4), 311-337.

Capell, A. (1944c). Peoples and languages of Timor. *Oceania, 15*(1), 19-48.

Cardoso, H. C. (2009). *The Indo-Portuguese language of Diu*. Utrecht: LOT.

Cardoso, H. C. (2016a). O português em contacto na Ásia e no Pacífico. In A. M. Martins, & E. Carreira (Eds.), *Manual de Linguística Portuguesa* (pp. 68-97). Berlim: Mouton de Gruyter.

Cardoso, Hugo C. (2016b). Die Korrespondenz zwischen Tancredo Caldeira do Casal Ribeiro und Hugo Schuchardt. In Bernhard Hurch (Ed.) (2007-). *Hugo Schuchardt Archiv*. Retrieved October 31, 2017 from: http://schuchardt.uni-graz.at/id/letters/1273.

Cardoso, H. C. (2017). Descrição de línguas de Timor-Leste na transição dos séc. XIX e XX. *Moderna Språk*, *111*(1), 1-34.

Carneiro, A. S. R. (2010). As políticas linguísticas e de ensino de línguas em Timor-Leste: desafios de um contexto multilingue. *Filologia linguística portuguesa, 12*(1), 9-25.

Castro, A. de. (1867). *As possessões portuguezas na Oceania*. Lisboa: Imprensa Nacional.

Castro, A. O. de. (1943). *A ilha verde e vermelha de Timor*. Lisboa: Imprensa Nacional.

Clements, J.C. (2000). Evidência para a existência de um pidgin português asiático. In E. d'Andrade, M.A. Mota, & D. Pereira (Eds.), *Crioulos de Base Portuguesa* (pp. 185-200). Braga: Associação Portuguesa de Linguística.

Clyne, M. (1998). Multilingualism. In F. Coulmas (Ed.), *The Handbook of Sociolinguistics* (pp. 301-314). Oxford: Blackwell Publishers.

Coefficient of determination. (2017). Stat Trek. Retrieved December 2, 2017 from http://stattrek.com/statistics/dictionary.aspx?definition=coefficient_of_determination

Coelho, A. (1881). *Os dialectos romanicos neo-latinos na África, Ásia e América*. Lisboa: Casa da Sociedade de Geografia.

Collins, J. T. (1983). Syntactic change in Ambonese Malay: the possessive construction. In J. T. Collins (Ed.), *Studies in Malay dialects* (Vol. 17, pp. 28-41). Jakarta: Universitas Katolik Indonesia Atma Jaya.

Comissão Litúrgica da Diocese de Díli (1980). *Ordinário da Missa: Texto Oficial Tétum*. Díli.

Corrêa, A. A. M. (1944). *Timor Português: Contribuições para o seu estudo antropológico*. Lisboa: Imprensa Nacional.

Correia, A. J. G. et al. (2005). *Disionáriu Nasionál ba Tetun Ofisiál*. Lisboa: Imprensa Nacional.

Correia, A. P. (1934). *Gentio de Timor*. Lisboa: Imprensa Lucas & C.[a].

Corte-Real, B. de A. e. (1990). *A contrastive analysis between Tetun and English consonants – a preliminary study of some phonological features of both languages* (Unpublished bachelor's thesis). Universitas Kristen Satya Wacana, Salatiga.

Costa, L. (2000). *Dicionário de Tétum-Português*. Lisbon: Edições Colibri.

Costa, L. (2001). *Guia de Conversação Português-Tétum*. Lisbon: Edições Colibri.

Costa, L. (2015). *Língua Tétum – Contributos para uma Gramática*. Lisbon: Edições Colibri.

Costa, M. C. dos S., & Coelho, J. S. B. (2013). Derivação sufixal: funcionamento e sentidos do sufixo -dor[1] e -dor[2] no português arcaico. *Revista Crátilo, 6*(1), 66-71.

Crouch, M. A. (2012). Law and religion in Indonesia: The Constitutional Court and the Blasphemy Law. *Asian Journal of Comparative Law 7*(1), article 3.

Crowley, T. (1987). Serial verb constructions in Paamese. *Studies in Languages, 11,* 35-84.

Cunha, C., & Cintra, L. F. L. (1999). *Nova Gramática do Português Contemporâneo*. Lisbon: Edições João Sá da Costa.

Dalgado, S. R. (1919). *Glossário Luso-Asiático*. Coimbra: Imprensa da Universidade de Coimbra.

Dalgado, S. R. (1936). *Influence of Portuguese vocables in Asiatic languages*. Baroda (Vadodara): Oriental Institute.

Davies, A. (2004). The Native Speaker in Applied Linguistics. In A. Davies, & C. Elder (Eds.), *The Handbook of Applied Linguistics* (pp. 431-450). Oxford: Blackwell Publishing.

Democratic Republic of East Timor. (2002). *Constitution of the Democratic Republic of East Timor*. Retrieved October 31, 2017, from: http://www.icrc.org/ihl-nat.nsf/162d151af444ded44125673e00508141/127ced8944679aa5c12570910031f468/$FILE/Constitution%20Timor%20-%20EN.pdf.

Democratic Republic of East Timor. (2004). *O padrão ortográfico da língua Tétum*. [Decreto do Governo n º. 1/2004]. Retrieved October 31, 2017, from: http://mj.gov.tl/jornal/lawsTL/RDTL-Law/RDTL-Gov-Decrees-P/Decree-Law-2004-1.pdf.

Direcção Nacional de Estatística. (2006). *Censo da população e habitação: tabelas prioritárias nacionais.* Dili: DNE and FNUAP.

Direcção Nacional de Estatística. (2015). *Education.* Retrieved May 22, 2017, from: http://www.statistics.gov.tl/category/publications/census-publications/2015-census-publications/volume-3-social-and-economic-characteristics/1-education/.

Domingues, E. S. J. (1947). Línguas de Timor: nótulas bibliográficas. *Portugal em África – Revista da cultura missionária, 2*(4), 142-151.

Donohue, M., & Schapper, A. (2008). Whence the Austronesian Indirect Possession Construction? *Oceaninc Linguistics 47*(2), 316-327.

Dores, R. das. (1901, February 23). [Letter to José Leite de Vasconcelos]. Epistolário de José Leite de Vasconcelos (nr. 7026), Museu Nacional de Arqueologia, Lisbon.

Dores, R. das. (1907). *Diccionario teto-português*. Lisboa: Imprensa Nacional.

Dryer, M. S. (2007). Noun phrase structure. In Timothy Shopen (Ed.), *Language Typology and Syntactic Description. Volume II: Complex Constructions* (pp. 151-205). Cambridge: Cambridge University Press.

Duarte, T. (1930). *Timor (ante-câmara do inferno?!)*. Famalicão: Tip. Minerva.

Durand, F. (2010). *História de Timor-Leste da Pré-História à Actualidade/Istória Timor-Leste nian husi Pre-Istória to'o Atualidade*. Lisboa: Lidel.

Eccles, L. (1998). Perfective and Inchoative Verbal Markers in Tetum. *Studies of languages and cultures of East Timor, 1,* 39-51.

Eccles, L. (1999). The Verb Markers of Tetum-Praça and their Grammaticalization. *Studies of languages and cultures of East Timor, 2,* 22-37.

Eccles, L. (2001). A Tetum Reference Grammar: Some Problems in Grammatical Classification. *Studies of languages and cultures of East Timor, 4,* 37-47.

Eccles, L. (2004). Early Chinese accounts of Timor. *Studies of languages and cultures of East Timor, 6,* 178-187.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (2004). Individual Differences in Second Language Learning. In A. Davies, & Catherine Elder (Eds.), *The Handbook of Applied Linguistics* (pp. 525-551). Oxford: Blackwell Publishing.

Esperança, J. P. T. (2001). *Estudos de Linguística Timorense*. Aveiro: SUL – Associação de Cooperação para o Desenvolvimento.

Ewing, M. C. (2005). Colloquial Indonesian. In A. Adelaar, & N. P. Himmelmann (Eds.), *The Austronesian Languages of Asia and Madagascar* (pp. 227-258). London: Routledge Language Family Series.

Felgas, H. A. E. (1956). *Timor Português*. Lisboa: República Portuguesa, Ministério do Ultramar; Divisão de publicações e biblioteca, Agência Geral do Ultramar.

Ferguson, C. A. (1983). Language planning and language change. In J. Cobarrubias, & J. A. Fishman (Eds.), *Progress in Language Planning* (pp. i-xii). Berlin: Mouton.

Fernandes, A. J. (1931). *Esbôço histórico e do estado actual das missões de Timor e refutação dalgumas falsidades contra elas caluniosamente afirmadas por um ex-governador de Timor*. Macau: Mercantil de N. T. Fernandes & Filhos.

Fernandes, A. J. (1937). *Método prático para aprender o tétum*. Macau: Escola Tipográfica do Orfanato de Macau.

Fernandes, A. J. (1964). *Pequeno método prático para aprender o tétum.* [Dili?]: Comando Territorial Independente de Timor.

Fernandes, Pe. F. M. (2000). *D. António Joaquim de Medeiros (Bispo de Macau) e as Missões de Timor 1884-1897* (Unpublished Master's thesis). Universidade de Macau, Macau.

Ferreira, Fr. J. G. (1885, November 3). [Letter to Hugo Schuchardt]. Hugo Schuchardt Archiv (nr. 01-03853), University of Graz, Austria. Retrieved October 31, 2017, from: https://schuchardt.uni-graz.at/id/letter/1302.

Fiedler, H. (1929). *Die Insel Timor*. Friedrichssegen a. d. Lahn: Folkwang-Auriga-Verlag.

Fontoura, Á. E. N. de. (1940). *Colónia portuguesa de Timor (Álbum Fontoura).* Retrieved October 31, 2017, from: http://amrtimor.org/multimedia/fontoura/album/pag_inteiras/0.htm

França, B. da. (1897). *Macau e os seus habitantes: relações com Timor*. Lisboa: Imprensa Nacional.

Garcia, C. de P. (1901). *Algumas palavras sobre a colonização de Timor*. Lisboa: A Liberal – Officina Typographica.

Gil, D. (2013). Numeral Classifiers. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved October 31, 2017, from: http://wals.info/chapter/55.

Godinho, M. M. dos S. (2013). 'Environomentu kapaas?' Surgimentu de falsos amigos em tétum com ar a português. In H. Loney, A. B. da Silva, N. C. Mendes, A. da C. Ximenes, & C. Fernandes (Eds.), *Hatene kona ba/Compreender/Understanding/Mengerti Timor-Leste* (Vol.1, pp. 187-193). Hawthorn: Swinburne Press.

Government of Timor-Leste. (2011). *Timor-Leste Strategic Development Plan 2011-2030*. Retrieved October 31, 2017, from: http://timor-leste.gov.tl/wp-content/uploads/2011/07/Timor-Leste-Strategic-Plan-2011-20301.pdf.

Governu Timor-Leste. [news articles from 2010]. Retrieved January 15, 2016, from: http://timor-leste.gov.tl/?lang=tp

Grade, E. A. (1973). Timor 1973 – panorâmica do ensino. *Revista Militar, 25*, 211-252.

Greenfield, W. (1830). *A defence of the Surinam Negro-English version of the New Testament*. London: Samuel Bagster.

Grimes, C. E., Therik, T., Grimes, B. D., & Jacob, M. (1997). *A guide to the people and languages of Nusa Tenggara*. Kupang: Artha Wacana Press.

Hajek, J. (2000a). Towards language history of East Timor. *Quaderni del Dipartimento di Linguistica – Università di Firenze, 10,* 213-227.

Hajek, J. (2000b). Language planning and the Sociolinguistic environment in East Timor: Colonial Practice and Changing Language Ecologies. *Current Issues in Language Planning, 1*(3), 400-414.

Hajek, J. (2006a). Serial Verbs in Tetun Dili. In A. Y. Aikhenvald, & R. M. W. Dixon (Eds.), *Serial Verb Constructions: A Cross-Linguistic Typology* (pp. 239-253). Oxford: Oxford University Press.

Hajek, J. (2006b). Language Contact and Convergence in East Timor: The case of Tetun Dili. In A. Y. Aikhenvald, & R. M. W. Dixon (Eds.) *Grammars in contact: A Cross-Linguistic Typology* (pp. 163-178). Oxford: Oxford University Press.

Hajek, J., & Stevens, M. (2005). On the Acoustic Characterization of Ejective Stops in Waima'a. *Interspeech, 2005,* 2889-2892.

Hajek, J., & Williams-van Klinken, C. (2003). Um sufixo românico numa língua austronésia: *-dor* em Tetum. *Revue de linguistique romane, 67,* 55-65.

Hall, R. A. Jr. (1955). Sostrato e lingue creole. *Archivo glottologico italiano, 40,* 1-9.

Haspelmath, M. (2009). Lexical borrowing: Concept and issues. In M. Haspelmath, & U. Tadmor (Eds.), *Loanwords in the World's Languages: A Comparative Handbook* (pp. 35-54). Berlin: Mouton de Gruyter.

Hickey, R. (2010). Language Contact: Reconsideration and Reassessment. In R. Hickey (Ed.), *The Handbook of Language Contact* (pp. 1-28). Oxford: Wiley-Blackwell.

Himmelmann, N. P. (2005). The Austronesian languages of Asia and Madagascar: typological characteristics. In A. Adelaar, & N. P. Himmelmann (Eds.), *The Austronesian Languages of Asia and Madagascar* (pp. 110-181). London: Routledge Language Family Series.

Holm, J. A. (1989). *Pidgins and creoles: Volume II – Reference Survey*. Cambridge: Cambridge University Press.

Holm, J. A. (2004). *An introduction to pidgins and creoles*. Cambridge: Cambridge University Press.

Huber, J. (in press). Makalero and Makasae. In A. Schapper (Ed.), *The Papuan Languages of Timor, Alor and Pantar* (Vol. 2). Boston/Berlin: Mouton de Gruyter.

Hull, G. (1993). *Mai Kolia Tetun: a course in Tetum-Praça, the Lingua Franca of East Timor*. North Sydney: Australian Catholic Relief and Australian Catholic Social Justice Council

Hull, G. (1998a). The languages of Timor 1772-1997: A Literature review. *Studies of languages and cultures of East Timor, 1,* 1-38.

Hull, G. (1998b). The Basic Lexical Affinities of Timor's Austronesian Languages: A Preliminary Investigation (revised in 2001). *Studies of languages and cultures of East Timor, 1,* 97-202.

Hull, G. (1999a). Towards a Language Policy for an independent East Timor. *Studies of languages and cultures of East Timor, 2,* 1-7.

Hull, G. (1999b.) East Timor and Indonesia: The Cultural Facts of Incompatibility. *Studies of languages and cultures of East Timor, 2,* 55-67.

Hull, G. (1999c). *Standard Tetum – English Dictionary*. Sydney: Allen & Unwin.

Hull, G. (1999d). *Tetum Language Manual for East Timor*. North Ryde, NSW: University of Western Sydney Macarthur.

Hull, G. (2000a). O léxico tétum: Princípios de renovação. *Studies of languages and cultures of East Timor, 3,* 57-106.

Hull, G. (2000b). Historical Phonology of Tetum. *Studies of languages and cultures of East Timor, 3,* 158-212.

Hull, G. (2000c, March). *Current Language Issues in East Timor*. Paper presented during a public lecture at the University of Adelaide, Adelaide.

Hull, G. (2001a). A Morphological Overview of the Timoric Sprachbund. *Studies in Languages and Cultures of East Timor, 4,* 98-205.

Hull, G. (2001b). *Baikenu – Language Manual for Oecussi-Ambeno Enclave (East Timor).* Winston Hills, NSW: Sebastiao Aparicio da Silva Project.

Hull, G. (2003a). *Southern Mambai (Manbae-Ainaru nor Same).* Dili: Instituto Nacional de Linguística, UNTL.

Hull, G. (2003b). *Baikenu.* Dili: Instituto Nacional de Linguística, UNTL.

Hull, G. (2003c). *Galoli (Galolen).* Dili: Instituto Nacional de Linguística, UNTL.

Hull, G. (2005a). The Malay Lexical Element in Tetum. *Studies of languages and cultures of East Timor, 7,* 84-124.

Hull, G. (2005b). *Makasai (Makasae).* Dili: Instituto Nacional de Linguística, UNTL.

Hull, G. (2005c). *Fataluku (Fatalukunu).* Dili: Instituto Nacional de Linguística, UNTL.

Hull, G., & Correia, A. J. G. (2005). *Kursu gramátika Tetun: ba profesór, tradutór, jornalista no estudante-universidade sira.* Dili: Instituto Nacional de Linguística da Universidade Nacional Timor Loro'Sae.

Hull, G., & Eccles, L. (2001). *Tetum reference grammar*. Winston Hills, NSW: Sebastiao Aparicio da Silva Project.

Hull, G., & Eccles, L. (2005). *Gramática da Língua Tétum* [Translation of Hull & Eccles 2001]. Lisbon: Lidel.

Hull, G., & Pollard, T. (2005). *Disionário Malaiu-Tetun/Kamus Melayu-Tetun.* Dili: Instituto Nacional de Linguística da Universidade Nacional Timor Loro'Sae.

Instituto Nacional de Linguística. (2001). *Princípios de Ortografia Tétum: Sistema Fonémico*. Dili: Instituto Nacional de Linguística, Universidade Nacional Timor Lorosa'e.

Instituto Nacional de Linguística. (2002a). *Matadalan Ortográfiku ba Tetun Nasionál* [Orthographic Guide for National Tetun]. Dili: Direcção do Instituto Nacional de Linguística, RDTL.

Instituto Nacional de Linguística. (2002b). *Hakerek Tetun Tuir Banati: kursu ortografia padronizada nian* [Writing Tetun based on the Model: course of standardized orthography]. Dili: Direcção do Instituto Nacional de Linguística, RDTL.

Instituto Nacional de Linguística. (2004). *A Ortografia Pardonizada do Tétum: Os Seus 115 anos de Construção*. Dili: Direcção do Instituto Nacional de Linguística, RDTL.

Kartodirdjo, S. (1970). Religious and economic aspects of Portuguese-Indonesian relations. *Studia, 29,* 175-196.

Keenan, E. L., & Dryer, M. D. (2007). Passive in the world's languages. In T. Shopen (Ed.), *Language Typology and Syntactic Description. Volume I: Clause Structure* (pp. 325-361). Cambridge: Cambridge University Press.

Kerswill, P. (2002). Koineization and Accommodation. In J. K. Chambers, P. Trudgill, & N. Shilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 669-702). Oxford: Blackwell Publishing.

Klamer, M. (2002). Typical features of Austronesian languages in central/eastern Indonesia. *Oceanic Linguistics, 41,* 363–383.

Klamer, M., & Ewing, M. C. (2010). The languages of East Nusantara: an introduction. In M. C. Ewing, & M. Klamer (Eds.), *East Nusantara: typological and areal analyses* (pp. 1-24). Canberra: Pacific Linguistics.

Larangeira, M. M. (1932). *Cartilha-Tetun* (3rd ed.). Timor: Imprensa Nacional.

Leach, M. (2003). Privileged ties: Young people debating language, heritage and national identity in East Timor. *Portuguese Studies Review, 11,* 137-150.

Leach, M. (2017). *Nation-Building and National Identity in Timor-Leste*. Oxon: Routledge.

Lefebvre, C. (1998). *Creole Genesis and the Acquisition of Grammar: the Case of Haitian Creole*. Cambridge: Cambridge University Press.

Lencastre, J. G. de. (1929). Algumas regras gramaticais da língua tetum e vocabulário. *Boletim da Agência Geral das Colonias, 5*(54), 82-92.

Levis, J., & LeVelle, K. (2012). Social factors in pronunciation acquisition. In J. Levis, & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 1-10). Retrieved October 31, 2017, from: https://apling.engl.iastate.edu/alt-content/uploads/2016/06/Proceedings_3rd_PSLLT.pdf.

Lewis, M. P., Simons, G. F., & Fennig, C. D. (Eds.). (2013). *Ethnologue: Languages of the World, Seventeenth edition.* Dallas, Texas: SIL International.
Retrieved October 31, 2017, from: https://www.ethnologue.com/language/lva.

Lima-Salles, H. M. M., & Pilati, E. (2014). Correlações entre a ordem sujeito-verbo e as características das construções causais no Português Brasileiro (PB). *Veredas: Sintaxe das Línguas Brasileiras, 18*(1), 203-221.

Loch, A., & Tschanz, M. (2005). *Kleines Wörterbuch Tetum-Deutsch Deutsch-Tetum*. Hamburg: Helmuth Buske Verlag.

Lucchesi, D. (1994). Variação e norma: elementos para a caracterização sociolinguística do português do Brasil. *Revista Internacional de Língua Portuguesa, 12,* 435-482.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Magalhães, A. L. de. (1920). Subsídios para o estudo etnológico de Timor. *Trabalhos da Sociedade Portuguesa de Antropologia e Etnologia, 1*(2), 37-65.

Manhitu, Y. (2007). *Kamus Indonesia-Tetun Tetun-Indonesia*. Jakarta: Penerbit PT Gramedia Pustaka Utama.

Marcos, A. (1995). *Timor Timorense: Com suas línguas, literaturas, lusofonia….* Lisbon: Edições Colibri.

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three Misconceptions about Age and L2 Learning. *TESOL Quarterly, 34*(1), 9-24.

Martinho, J. S. (1943). *Timor: Quatro séculos de colonização portuguesa.* Porto: Livraria Progredior.

Mateus, M. H. M. et al. (2003). *Gramática da Língua Portuguesa*. Lisboa: Editorial Caminho.

Mathijsen, A. (1906). Tettum-Hollandsche woordenlijst, met beknopte spraakkunst. *Verhandelingen van het Bataviaasch Genootschap van kunsten en Wetenschappen* (Vol. 56, part 2). The Hague: Nijhoff.

Matos, A. T. de. (1974). *Timor Português 1515-1769: contribuição para a sua história*. Lisboa: Faculdade de Letras da Universidade de Lisboa.

Matras, Y. (2009). *Language Contact*. Cambridge: Cambridge University Press.

Maurer, P. (2011). *The former Portuguese Creole of Batavia and Tugu (Indonesia)*. London: Battlebridge.

Maurer, P. & the APiCS Consortium. (2013). 'Give' serial verb constructions. In S. M. Michaelis, P. Maurer, M. Haspelmath, & M. Huber (Eds.), *Atlas of Pidgin and Creole Language Structures Online*. Retrieved October 31, 2017, from: http://apics-online.info/parameters/86.

Maurer, P. & the APiCS Consortium. (2013). Nominal plural marker and third-person-plural pronoun. In S. M. Michaelis, P. Maurer, M. Haspelmath, & M. Huber (Eds.), *Atlas of Pidgin and Creole Language Structures Online*. Retrieved October 31, 2017, from: http://apics-online.info/parameters/25.

McWhorter, J. (2005). *Defining Creole*. Oxford: Oxford University Press.

McWhorter, J. (2007). *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars*. Oxford: Oxford University Press.

McWhorter, J. (2011). *Linguistic Simplicity and Complexity: Why do languages undress?* Berlin: Mouton de Gruyter.

Meakins, F. (2013). Mixed languages. In P. Bakker, & Y. Matras (Eds.), *Contact languages: a comprehensive guide* (pp. 159-228). Berlin: De Gruyter Mouton.

Mendes, M. P. (1935). *Dicionário tétum-português*. Macau: Tip. Mercantil de N.T. Fernandes & Filhos Ltda.

Menezes, F. X. de. (1992). Contacto de culturas no Timor Leste. *Estudos Orientais, 3,* 221-234.

Miranda, S. M., & Serafim, C. S. (2001). População e sociedade. In A. H. de O. Marques (Ed.), *História dos Portugueses no Extremo Oriente, vol. 2: Macau e Timor. O declínio do império* (229-257). Lisboa: Fundação Oriente.

Monteiro, F. (1985). *Kamus Tetun – Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Morais, A. F. de. (1944). *Sólor e Timor*. Lisboa: República Portuguesa, Ministério das Colónias; Divisão de publicações e biblioteca, Agência Geral das Colónias.

Morris, C. (1984). *Tetun-English Dictionary*. Pacific Linguistics, Series C, 83. Canberra: Australian National University.

Morris, C. (2003). *A Traveller's Dictionary in Tetun-English and English-Tetun* (originally published in 1992). Frankston: Baba Dook Books. Retrieved October 31, 2017, from: http://www.gnu.org/software/tetum/contributors/cliffMorris-xhtml/.

Mous, M. (2001). Paralexification in language intertwining. In N. Smith, & T. Veenstra (Eds.), *Creolization and Contact* (pp. 113-124). Amsterdam: John Benjamins.

Mufwene, S. (1996). Creole genesis: A population genetics perspective. In P. Christie (Ed.), *Caribbean Language Issues: Old and New* (pp. 163–196). Kingston: University of the West Indies Press.

Mufwene, S. S. (2001). *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.

Muysken, P. (1988). Are creoles a special type of language? In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey, Vol II. Linguistic Theory: Extensions and Implications* (pp. 285-301). Cambridge: Cambridge University Press.

Myers-Scotton, C. (2006). *Multiple Voices: An Introduction to Bilingualism*. Oxford: Blackwell.

Nicolai, S. (2004). *Learning independence: education in emergency and transition in Timor-Leste since 1999*. Paris: UNESCO-IIEP.

Noonan, M. (2007). Complementation. In Timothy Shopen (Ed.), *Language Typology and Syntactic Description. Volume II: Complex Constructions* (pp. 52-150). Cambridge: Cambridge University Press.

National Statistics Directorate and United Nations Population Fund. (2011). *Population and Housing Census of Timor-Leste, 2010. Volume 2: Population Distribution by Administrative Areas*. Dili: NSD.

Oliveira, L. de. (1950). *Timor na história de Portugal* (Vol. 2). Lisboa: Agência Geral do Ultramar.

Paauw, S. (2008). *The contact Malay varieties of Eastern Indonesia: A typological comparison* (Unpublished doctoral dissertation). University at Buffalo, Buffalo.

Paauw, S. (2013). Ambon Malay structure dataset. In S. M. Michaelis, P. Maurer, M. Haspelmath, & M. Huber (Eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved October 31, 2017, from: http://apics-online.info/contributions/68.

Payne, T. E. (2010). *Describing morphosyntax: a guide for field linguists*. New York: Cambridge University Press.

Pigeaud, T. G. Th. (1960). *Java in the 14th century: a study in cultural history: the Nāgara-Kĕrtāgama by Rakawi Prapañca of Majapahit, 1365 A.D. Vol. III Translations*. Hague: Martinus Nijhoff.

Pigeaud, T. G. Th. (1962). *Java in the 14th century: a study in cultural history: the Nāgara-Kĕrtāgama by Rakawi Prapañca of Majapahit, 1365 A.D. Volume IV Commentaries and Recapitulation*. Hague: Martinus Nijhoff.

Rêgo, A. da S. (1998). *Dialecto Poruguês de Malacca e outros escritos* (Reedition of the 1942 edition including re-edition of some of the author's articles). Lisboa: Comissão Nacional para as Comemorações dos Descobrimentos Portugueses.

Rêgo, F. J. H. (1968). Situação linguística de Timor. *Colóquios sobre as Províncias do Oriente, 1,* 59-69.

Ribeiro, T. C. do C. (1882, June 30). [Letter to Hugo Schuchardt] Hugo Schuchardt Archiv (nr. 01553), University of Graz, Austria. Retrieved October 31, 2017, from: https://schuchardt.uni-graz.at/id/letter/2960.

Ross, M. D. (1995). Some current issues in Austronesian linguistics. In D. T. Tryon (Ed.), *Comparative Austronesian Dictionary: An Introduction to Austronesian Studies* (Vol. 1, pp 45-120). Berlin: Mouton De Gruyter.

Rubino, C. (2012). Bilung in Zamboagueño Chavacano: the genericization of a substrate numeral classifier. In H. C. Cardoso, A. N. Baxter, & M. Pinharanda Nunes (Eds.), *Ibero-Asian Creoles: Comparative Perspectives* (pp. 239-262). Amsterdam: John Benjamins.

Sá, A. B. de. (1952a). Notas sobre Linguística Timorense: sistema de representação fonética. *Estudos Coloniais, 3,* 5-25.

Sá, A. B. de. (1952b). *Timor*. Lisboa: Sociedade de Geografia de Lisboa Semana do Ultramar.

Sá, A. B. de. (1955). *Documentação para a história das missões do padroado português do Oriente, Insulíndia vol. 2 (1550-1562).* Lisboa: Agência Geral do Ultramar, Divisão de Publicações e Bibliotecas.

Sá, A. B. de. (1956). *Documentação para a história das missões do padroado português do Oriente, Insulíndia vol. 4 (1568-1579).* Lisboa: Agência Geral do Ultramar, Divisão de Publicações e Bibliotecas.

Sá, A. B. de. (1961). *Textos em teto da literatura oral timorense*. Lisboa: Junta de Investigações do Ultramar.

Saliwangi, B., Muslich, M., Nurchasanah, & Kartomihardjo, S. (1991). *Sistem morfologi kata kerja (verba) bahasa Tetun*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Sampson, G. (2006). Does simply imply creole? In M. Suominen (Ed.), *A man of measure: Festschrift in honor of Fred Karlsson on his 60th birthday* (pp. 362-374). Turku: Linguistic Association of Finland.

Santa Maria, L. (1967). *I Prestiti Portoghesi nel Malese-Indonesiano*. Naples: Instituto Orientale di Napoli.

Sapo. [newspaper articles from March 2016]. Retrieved January 15, 2016, from: http://www.sapo.pt/.

Schachter, P., & Shopen, T. (2007). Parts-of-speech systems. In T. Shopen (Ed.), *Language Typology and Syntactic Description. Volume I: Clause Structure* (pp. 1-60). Cambridge: Cambridge University Press.

Schiffman. H. F. (1998). Diglossia as a Sociolinguistic Situation. In F. Coulmas (Ed.), *The Handbook of Sociolinguistics* (pp. 205-216). Oxford: Blackwell Publishers.

Schuchardt, H. (1891). Creolische Studien IX. Über das Malaioportugiesische von Batavia und Tugu. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften zu Wien, 122*, 1-258.

Serantes, P. J., & Doko, I. H. (1976a). *Pelajaran bahasa Indonesia / Tetun untuk rakyat Timor-Timur*. Bandung: Penerbit Ganaco NV.

Serantes, P. J., & Doko, I. H. (1976b). *Kamus kecil Indonesia – Tetun Belu – Tetun Dili*. Bandung: Penerbit Ganaco NV.

Serantes, P. J., & Doko, I. H. (1976c). *Pelajaran Bahasa Percakapan Sehari-hari Tetun-Indonesia*. Bandung: Penerbit Ganaco NV.

Siegel, J. (1985). Koines and koineization. *Language in Society, 14*(3), 357-378.

Siegel, J. (2001). Koine formation and creole genesis. In N. Smith & T. Veenstra (Eds.), *Creolization and Contact* (pp. 175-197). Amsterdam: Benjamins.

Silva, A. de M. (1789). *Diccionário da Lingua Portuguesa, Composto pelo Padre D. Rafael Bluteau, Reformado e Acrescentado por Antonio de Moraes Silva, Natural do Rio de Janeiro*. Lisboa: Officina de Thaddeo Ferreira.

Silva, M. M. A. da. (1900). *Noções da Grammatica Galoli: dialecto de Timor*. Macau: Typ. do Seminario de S. José.

Silva, M. M. A. da. (1903). *Catecismo da doutrina christã em portuguez e galoli*. Macau: Typ. Noronha & Ca.

Silva, M. M. A. da. (1905). *Diccionario portuguez-galoli*. Macau: Typ. Mercantil.

Silva, S. M. A. da. (1885). *Catecismo da Doutrina Christã em Tétum*. Macau: Typographia do Seminário.

Silva, S. M. A. da. (1889). *Diccionario de Portuguez-Tétum*. Macau: Typographia do Seminário.

Singler, J. V. (2008). The Sociohistorical Context of Creole Genesis. In S. Kouwenberg, & J. V. Singler (Eds.), *The Handbook of Pidgin and Creole Studies* (pp. 332-359). Oxford: Wiley-Blackwell.

Sneddon, J. N., Adelaar, A., Djenar, D. N., & Ewing, M. C. (2010). *Indonesian Reference Grammar* (2nd ed). St. Leonards, NSW: Allen & Unwin.

Soares, D. M. de D. (n.d). *Kamus bahasa Tetun – bahasa Indonesia – bahasa Portugis*. Jakarta: unknown.

Soedjiatno, S., Aminoedin N. A., & Djajengwasito, S. (1992). *Sistem morfologi kata tugas bahasa Tetum*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Sousa, S. M. de. (2013). Die Korrespondenz zwischen João Gomes Ferreira und Hugo Schuchardt. In Bernhard Hurch (Ed.) (2007-), *Hugo Schuchardt Archiv*. Retrieved October 31, 2017, from:
http://schuchardt.uni-graz.at/korrespondenz/briefe/korrespondenzpartner/alle/1705.

Souza, G. B. (1986). *The Survival of Empire: Portuguese Trade and Society in China and the South China Sea 1630-1754*. Cambridge: Cambridge University Press.

Stassen, L. (2013). Zero Copula for Predicate Nominals. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from: http://wals.info/chapter/120

Steele, S. (2000). The struggle for freedom: Predominantly Catholic East Timor slowly rises from the ashes after 24 years of occupation. *Florida Catholic Sept. 21, 2000*.

Retrieved October 31, 2017, from: http://www.etan.org/et2000c/september/17-23/21etslow.htm.

Steenbrink, K. (2007). *Catholics in Indonesia, 1808-1942: A documented history. Volume 2. The spectacular growth of a self-confident minority, 1903-1942*. Leiden: KITLV Press.

Suara Timor Lorosae (STL). [newspaper articles from July 2013 and March 2016]. Retrieved January 15, 2016, from: https://suara-timor-lorosae.com/.

Swadesh, M., & Sherzer, J. F. (Eds). (2017). *The origin and the diversification of language*. Oxon; New York: Routledge.

Tadmor, U. (2009). Loanwords in the world's languages: Findings and results. In M. Haspelmath, & U. Tadmor (Eds.), *Loanwords in the World's Languages: A Comparative Handbook* (pp. 55-75). Berlin: Mouton de Gruyter.

Taryono, A. R., Ibrahim, A. S., Rusmadji, O., & Moehnilabib, M. (1993). *Morfo-Sintaktis Bahasa Tetum*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Taylor-Leech, K. (2008). Language and identity in East Timor: The discourses of nation building. *Language Problems & Language Planning, 32*(2), 153-180.

Taylor-Leech, K., & Caet, A. (2012). Mother tongue-based multilingual education: A new direction for Timor-Leste. In M. Leach, N. C. Mendes, A. B. da Silva, B. Boughton, & A. da C. Ximenes (Eds.), *Peskiza foun kona ba / Novas investigações sobre / New Research on / Penelitian Baru mengenai Timor-Leste* (pp. 295-301). Hawthorn: Swinburne Press.

Teixeira, M. (1961a). *The Portuguese missions in Malacca and Singapore (1511-1958), vol. 1. – Malacca*. Lisbon: Agência Geral do Ultramar.

Teixeira, M. (1961b). *The Portuguese missions in Malacca and Singapore (1511-1958), vol. 2. – Malacca*. Lisbon: Agência Geral do Ultramar.

Teixeira, M. (1963). *The Portuguese missions in Malacca and Singapore (1511-1958), vol. 3. – Singapore*. Lisbon: Agência Geral do Ultramar.

The World Bank. (2015). *Youth literacy rate, population 15-24 years, both sexes (%).* Retrieved October 31, 2017, from:

http://data.worldbank.org/indicator/SE.ADT.1524.LT.ZS?locations=TL.

Thomason, S. G. (2001). *Language contact.* Edinburgh: Edinburgh University Press.

Thomason, S. G. (2003). Social factors and linguistic processes in the emergence of stable mixed languages. In Y. Matras, & P. Bakker (Eds.), *The Mixed Language Debate: theoretical and empirical advances* (pp. 21-40). Berlin: Mouton de Gruyter

Thomason, S. G. (2010). Contact Explanations in Linguistics. In R. Hickey (Ed.), *The Handbook of Language Contact* (pp. 31-47). Oxford: Wiley-Blackwell.

Thomason, S. G., & Kaufman, T. (1991). *Language Contact, Creolization, and Genetic Linguistics.* Berkeley, London: University of California Press.

Thomaz, L. F. F. R. (1974). Timor: notas histórico-linguísticas. *Portugaliae Historica, 2,* 167-300.

Thomaz, L. F. F. R. (1977). O afluxo ao meio urbano no Timor Português. *Revista da Faculdade de Letras de Lisboa, 4*(1), 459-553.

Thomaz, L. F. F. R. (1980). [Review of the book *Pelajaran Bahasa Indonesia-Tetun; Pelajaran Bahasa Percakapan Sehari-hari Tetun-Indonesia; Kamus Kecil Indonesia-Tetun Belu-Tetun Dili*, by P. J. Serantes, & I. H. Doko]. *Archipel, 20,* 348-351.

Thomaz, L. F. F. R. (1981). The formation of Tetun-Praça, vehicular language of East Timor. In N. Phillips, & A. Khaidir (Eds.), *Papers on Indonesian Languages and Literatures. Cahier d'Archipel 13* (pp. 1-30). Paris: Association Archipel.

Thomaz, L. F. F. R. (1982). Note sur le «Dicctionaire François et Timorien» de F. E. de Rosily. *Archipel, 23,* 105-108.

Thomaz, L. F. F. R. (1985). A língua portuguesa em Timor. *Congresso sobre a situação actual da língua portuguesa no mundo, 1,* 313-338.

Thomaz, L. F. F. R. (2002). *Babel Loro Sa'e: O Problema Linguístico de Timor-Leste.* Lisboa: Instituto Camões.

Thompson, S. A., Longacre, R. E., & Hwang, S. J. J. (2007). Adverbial clauses. In T. Shopen (Ed.), *Language Typology and Syntactic Description. Volume II: Complex Constructions* (pp. 237-300). Cambridge: Cambridge University Press.

Timberlake, A. (2007). Aspect, tense, mood. In Timothy Shopen (Ed.), *Language Typology and Syntactic Description. Volume III: Grammatical Categories and the Lexicon* (pp. 280-333). Cambridge: Cambridge University Press.

Tjia, J. (2004). The development of possessive and passive-like constructions in Ambonese language. In B. K. Purwo (Ed.), *Studies in Malay Dialects, part IV, NUSA, 53,* 51-62.

Tomás, M. I. (2009). The role of women in the cross-pollination process in the Asian-Portuguese varieties. *Journal of Portuguese Linguistics, 8*(2), 49-64.

Traugott, E. C., & König, E. (1991). The semantics-pragmatics of grammaticalization revisited. In E. C. Traugott, & B. Heine (Eds). *Approaches to grammaticalization* (pp. 189-218). Amsterdam: John Benjamins.

Trudgill, P. (1998). The chaos before the order: New Zealand English and the second stage of new dialect formation. In E. H. Jahr (Ed.), *Language Change: Advances in Historical Sociolinguistics* (pp. 197-207). Berlin: Mouton de Gruyter.

van Engelenhoven, A. (1995). *A description of the Leti Language (as spoken in Tutukei)* (Doctoral dissertation, University of Leiden). Ridderkerk: Offsetdrukkerij Ridderprint.

van Engelenhoven, A. (2002). Verb sequences in Melayu Tenggara Jauh: the interface of Malay and the indigenous languages of Southwest Maluku. In A. K. Adelaar, & R. Blust (Ed.), *Between Worlds: Linguistic Papers in memory of David John Prentice* (pp. 177-191). Canberra: Research School of Pacific and Asian Studies, Australian National University.

van Engelenhoven, A., & Williams-van Klinken, C. (2005). Tetun and Leti. In A. Adelaar, & N. P. Himmelmann (Eds.), *The Austronesian Languages of Asia and Madagascar* (pp. 735-768). London: Routledge Language Family Series.

van Klinken, C. L. (1999). *A grammar of the Fehan dialect of Tetun: an Austronesian language of West Timor*. Canberra: Pacific Linguistics. [see also Williams-van Klinken]

van Minde, D. (1990). The sound system of Ambonese Malay. *Cakalele, 1*(1/2), 61-98.

van Minde, D. (1997). *Malayu Ambong: Phonology, Morphology, Syntax*. Leiden: CNWS.

van Name, A. (1869-70). Contributions to creole grammar. *Transactions of the American Philological Association, 1,* 123-167.

Vasconcelos, L. de (1901). *Esquisse d'une dialectologie portugaise*. Lisboa: Centro de Estudos Filológicos.

Vaquinhas, J. dos S. (1883). Timor. *Boletim da Sociedade de Geographia de Lisboa* (pp. 277-328). Lisboa: Imprensa Nacional.

Vaquinhas, J. dos S. (1885, November 30). [Letter to Hugo Schuchardt]. Hugo Schuchardt Archiv (nr. 02-2615), University of Graz, Austria. Retrieved October 31, 2017, from: https://schuchardt.uni-graz.at/id/letter/4103.

Vaquinhas, J. dos S. (1886, July 14). [Letter to Hugo Schuchardt]. Hugo Schuchardt Archiv (nr. 03-2616), University of Graz, Austria. Retrieved October 31, 2017, from: https://schuchardt.uni-graz.at/id/letter/4104.

Veiga, N. V., & Fernández, M. A. (2012). Maskin, maski, masque in the Spanish and Portuguese creoles of Asia. In H. C. Cardoso, A. N. Baxter, & M. Pinharanda Nunes (Eds.), *Ibero-Asian Creoles: Comparative Perspectives* (pp. 181-203). Amsterdam: John Benjamins.

Velupillai, V. (2012). *An introduction to Linguistic Typology.* Amsterdam: John Benjamins Publishing Company.

Vennemann, T. (2012). Structural complexity of consonant clusters: A phonologist's view. IN P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer, & B. Kühnert (Eds.), *Consonant Clusters and Structural Complexity* (pp. 9-32). Berlin: De Gruyter Mouton.

Villiers, J. (1985). As derradeiras do mundo: the Dominican missions and the sandalwood trade in the lesser Sunda islands in the sixteenth and seventeenth centuries. *Separata de Actas do II Seminário Internacional de História Indo-Portuguesa*, 573-600. Lisboa: Instituto de Investigação Cientifica Tropical.

Wallace, A. R. (1869). *The Malay Archipelago: The Land of the Orang-utan, and the Bird of Paradise* (Vol.1). London: Macmillan.

Williams-van Klinken, C. (2002). High Registers of Tetun Dili: Portuguese Press and Purist Priests. In C. Allen (Ed.), *Proceedings of the 2001 Conference of the Australian Linguistic Society.* Retrieved October 31, 2017, from: http://www.als.asn.au/proceedings/als2001/williams-vanklinken.pdf.

Williams-van Klinken, C. (2007). *Spelling options in Tetun Dili: Awareness of 'rr' and glottal stop amongst tertiary students*. Retrieved October 31, 2017, from: http://www.tetundit.tl/Publications/Awareness%20of%20glottals%20and%20rr.pdf.

Williams-van Klinken, C. (2008). Boundaries of serialization: non-serialised verb sequences in Tetun Dili. In G. Senft (Ed.), *Serial verb constructions in Austronesian and Papuan languages* (pp. 99-111). Canberra: Pacific Linguistics.

Williams-van Klinken, C. (2010a). Voice and valency in Tetun Dili. In M. C. Ewing, & M. Klamer (Eds.), *Typological and areal analyses: contributions from East Nusantara* (pp. 167-185). Canberra: Pacific Linguistics.

Williams-van Klinken, C. (2010b, August). *Two Tetuns: Language change before your very eyes*. Paper presented at 6th International East Nusantara Conference on Language & Culture. Kupang, Indonesia.

Williams-van Klinken, C. (2011). *Tetun-English interactive dictionary*. Dili: Dili Institute of Technology.

Williams-van Klinken, C., & Hajek, J. (2009a, June). *Double agent, double cross? Or how a suffix changes sides in an isolating language: dór in Tetun Dili*. Paper presented at Eleventh International Conference on Austronesian Linguistics. Aussois, France.

Williams-van Klinken, C., & Hajek, J. (2009b, June). *Real-time language contact and change in the Austronesian world: Tetun as a new media language*. Paper presented at Eleventh International Conference on Austronesian Linguistics. Aussois, France.

Williams-van Klinken, C., & Hajek, J. (2016). Tetun-gés: Influénsia portugés ba estrutura Tetun. In S. Smith, N. C. Mendes, A. B. da Silva, A. da C. Ximenes, C. Fernandes, & M.

Leach (Eds.), *Timor-Leste: iha kontextu local, rejional no global / O local, regional e global / the local, the regional and the global / Lokal, regional dan global 2015* (Vol. 1, pp. 32-36). Hawthorn: Swinburne Press.

Williams van Klinken, C., Hajek, J., & Nordlinger, R. (2001). Serial verbs in Tetun-Dili: a preliminary account. *Studies in languages and cultures of East Timor, 4,* 48-60.

Williams van Klinken, C., Hajek, J., & Nordlinger, R. (2002a). *Tetun Dili: a grammar of an East Timorese language*. Canberra: Pacific Linguistics.

Williams van Klinken, C., Hajek, J., & Nordlinger, R. (2002b). *A short grammar of Tetun Dili*. Munich: Lincom Europa.

Winford, D. (2003). *An Introduction to Contact Linguistics*. Oxford: Blackwell.

Wouk, F., & Ross, M. (2002). *The history and typology of western Austronesian voice systems*. Canberra: Pacific Linguistics.

Wortelboer, W. (1955). Zur Sprache und Kultur der Belu (Timor). *Anthropos, 50*(1/3), 155-200.

Wurm, S. A., & Hattori, S. (Eds). (1981-3). *Language Atlas of the Pacific Area. Part 2: Japan Area, Philippines and Formosa, Mainland and Insular South-East Asia*. Canberra: Australian Academy of the Humanities in collaboration with the Japan Academy.

Zhang, X. (1937). *Dongxiyang kao*. [Investigations into the Eastern and Western Oceans] (originally published in 1617). Shanghai: Commercial Press.

Zhao, R. (1930). *Zhufan zhi jiaozhu* [Records of the Various Foreign Nations, with corrections and commentaries]. Changsha: Commercial Press.