WORKING PAPERS

IN

LINGUISTICS

The notes and articles in this series are progress reports on work being carried on by students and faculty in the Department. Because these papers are not finished products, readers are asked not to cite from them without noting their preliminary nature. The authors welcome any comments and suggestions that readers might offer.

DEPARTMENT OF LINGUISTICS

UNIVERSITY OF HAWAI'I AT MĀNOA

HONOLULU 96822

An Equal Opportunity/Affirmative Action Institution

DEPARTMENT OF LINGUISTICS FACULTY

2015

TETUN DILI AND CREOLES: ANOTHER LOOK

YEN-LING CHEN

This paper revisits the linguistic classification of Tetun Dili, an Austronesian language spoken in and around the city of Dili, East Timor. Since at least the late 1990s several writers have claimed that this variant of the more widely-spoken Tetun language is a creole. However, close examination of the relevant data for the Dili dialect in relation both to the more conservative Tetun Fehan, and to Portuguese, shows little or no supporting evidence for the creole hypothesis. It is concluded that Tetun Dili is simply an Austronesian language with fairly numerous loanwords from Portuguese, particularly in semantic domains that reflect the nature of the contact situation during the colonial period.

1. INTRODUCTION. This paper briefly discusses the status of Tetun Dili, a.k.a. Tetun Prasa, which is spoken in Dili, the capital of the Democratic Republic of Timor-Leste, or East Timor. According to the Constitution of the Democratic Republic of Timor-Leste, the official languages of East Timor include Tetun and Portuguese. In addition to the official languages, Indonesian and English are also used in many domains.

There are two major versions of Tetun in East Timor, i.e., Tetun Terik and Tetun Dili. Tetun Terik, used primarily in rural areas of southern East Timor, is an Austronesian (AN) language that is little influenced by other foreign languages (Lewis et al. 2015). In contrast, Tetun Dili has been classified as a creole heavily influenced by Portuguese and Mambae/Mambai.[1] It is not clear whether Tetun Dili and Tetun Terik are mutually intelligible. Grimes et al. (1997) reported that Tetun Dili is not mutually intelligible with Tetun Fehan and Tetun Foho, two varieties of Tetun spoken in Indonesian Timor. The mutual intelligibility between Tetun Terik and Tetun Fehan, or Tetun Terik and Tetun Foho requires further investigation.

Generally speaking, the evidence given to support the classification of Tetun Dili as a creole consists of little more than the statement that Tetun Dili is "a language of inter-ethnic communication" which has abundant Portuguese loans and lacks subject agreement, and when compared to other Tetun varieties, Tetun Dili favors periphrastic constructions more. Williams-van Klinken (2002:1–2) also follows Grimes et al. (1997) in claiming that Tetun Dili is a creole, but with the reservation that "Tetun Dili does not, however, exhibit all the features of classic creoles as strictly defined …". Lewis et al. 2015 lists Tetun Dili as a Tetun-based creole as well. The motivation to revisit the classification of Tetun Dili came from the lack of supporting evidence in the published materials that called it a creole.

In this paper I argue that Tetun Dili is not a creole, but rather a Tetun variety with abundant Portuguese loans. While Tetun Dili is geographically closest to Tetun Terik, which is also spoken in East Timor, there is no available reference grammar of Tetun Terik. Hence, Tetun Fehan, a well-described variety of Tetun (van Klinken 1999), will be used as the basis of comparison with Tetun Dili. The term "Tetun" will be used (1) as a cover term for all Tetun varieties of AN origin when discussing the sociohistorical background of the formation of Tetun Dili, and (2) to refer to native AN words in Tetun Dili when discussing the domains where borrowing takes place.

This section gives a brief summary of the traditional views on how creoles developed, whereas non-traditional views will be presented in section 5. Unlike pidgins, which are by definition spoken as second languages used in limited domains, creoles are considered full-fledged languages, and function as first languages (Thomason and Kaufman 1988; Winford 2003). Creoles are formed only in contact situations in which pidgins first result. However, Thomason and Kaufman (1988) also mention that it is possible for a creole to emerge without first undergoing pidginization, as in the case of Île de France Creole and some of the Caribbean creoles. They refer to such situations as "abrupt creolization" that developed among displaced slaves when a common language was urgently needed (Thomason and Kaufman 1988:148).

---

[1] The original text is as follows – "*Language Classification*: Creole, Tetun based, with heavy Portuguese and Mambae influence."

Since the slaves came from various language backgrounds, they needed a lingua franca in order to communicate. Contact situations that encouraged pidgins and creoles to form included trading contacts, and colonization or homestead settings such as farms or plantations, where slaves or laborers of different ethnicities were relocated from their home countries to a foreign land.

According to Thomason and Kaufman (1988:37–39), Winford (2003:321), and Matras (2009:277–278), the lexicon of a creole is predominantly based on the superstrate language(s), i.e. the dominant, usually European, language(s) in the contact settings, with supplement from substrate languages and innovations. The phonological and morphosyntactic systems of a creole are often simpler than those of the superstrate languages. In other words, compared to its source language, a creole tends to have a smaller phoneme inventory and little morphology. Matras (2009:281–87) further specifies that pidgins and creoles tend to avoid consonant clusters, to have SVO word order and little inflectional or derivational morphology such as case or agreement marking, and to employ juxtaposition possession, though it is true that creoles, compared to pidgins, have a more expanded grammar via internal change because of the need for more sophisticated communication among speakers. In sections 3 and 4, various aspects of Tetun Dili grammar will be reviewed based on these criteria.

The following section introduces the historical background of Timor, and sections 3 and 4 describe the phonology, lexicon, morphology, and syntax of Tetun Fehan, Tetun Dili, and Portuguese. Section 5 then evaluates both the linguistic features and the sociohistorical background of Tetun Dili within the frameworks of three non-traditional views of creole formation (the universalist approach, the substratist approach, and the superstratist approach), and section 6 concludes the paper.

2. A BIRD'S-EYE VIEW OF THE HISTORY OF TIMOR. According to Fox (1977:61–65), Timor is known for its high-quality sandalwood, which for centuries attracted Indonesian and Chinese merchants. Traditionally the island was divided into several feudal states under local hereditary chiefs. The rulers of each chiefdom were very influential and they were the ones who controlled the sandalwood trade long before Portuguese colonization. The Portuguese began to explore Timor and nearby islands after their conquest of Malacca in 1511. They set up their base on the Island of Solor, located to the north of Timor. These Portuguese soldiers and officers began to intermarry with local interpreters, and converted many of the local people to Catholicism. Their mixed offspring were the so-called "Topasses," a.k.a. Black Portuguese, or Gente de Chapeo. The Portuguese in Solor were not influential and posed no threats to the local chiefs. However, the Black Portuguese, who were a separate force from the Portuguese, spoke Portuguese and local languages, and gradually gained power. They were based in Larantuka, east Flores, were allied with local rulers, and became dominant in sandalwood trading, with the main trading port located in Lifau. Because of the rise of the Black Portuguese in Lifau, the Portuguese Viceroy in Goa declared that anyone who was able to control the Black Portuguese would be appointed as the Captain Major of Timor. However, no Portuguese official succeeded. The Black Portuguese were finally defeated by the Dutch, who soon signed treaties with local rulers after the victory.

A Portuguese appointee in Lifau moved the garrison to Dili (see map 1) in 1769 in order to avoid the zones that were controlled by the Topasses. The Portuguese consequently controlled the eastern part of Timor. However, the market for sandalwood started to decline during this period, and only the Chinese merchants stayed. In the following 200 years, Dili and the eastern part of Timor were colonized by the Portuguese, but no record shows that any other group of mixed-race Portuguese formed in eastern Timor. It is noteworthy that even after 200 years of Portuguese colonization, Timorese people who are bilingual in Portuguese are mostly literate and educated (Leonardo Germano Tourinho de Souza Rosa, 2014 pers. comm.).

MAP 1. Map of Timor[2]
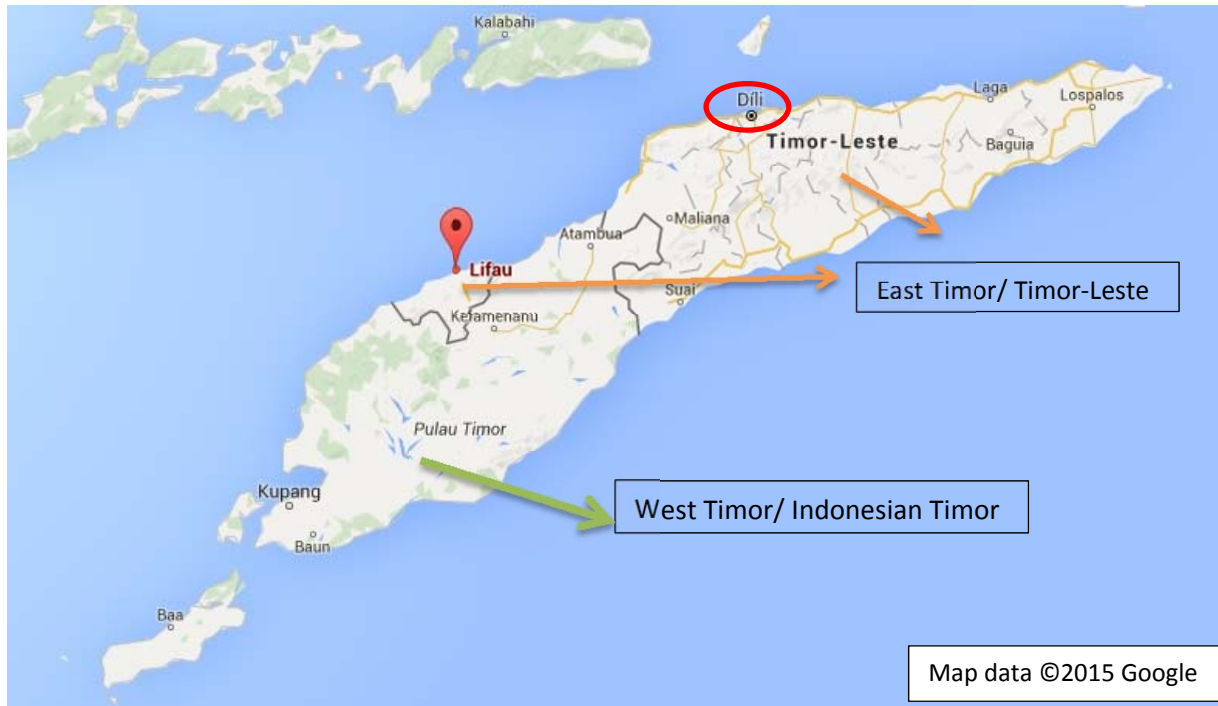
Hicks (1976) observed that, in addition to the chiefdom system, local Timorese people had their own kin and marriage system, which is very different from Western systems. Clans are divided into wife-taking groups and wife-giving groups, and one's lineage is very important with regard to one's social status. Traditionally the Timorese also practiced bridewealth/bride price, which required that the groom's clan offer the bride's clan goods "in exchange of a bundle of rights" in a marriage. We do not know if intermarriage between the Portuguese and the local Timorese was common, since it was not reported in the literature.

3.   THE PHONOLOGY AND THE LEXICON OF TETUN DILI. This section examines the phonology and the lexicon of Tetun Dili to see whether it exhibits properties of creoles discussed in the literature. By comparing Tetun Dili with Tetun Fehan, a closely related Austronesian language spoken in Indonesian Timor, and/or with Portuguese, we will have a better understanding of the extent of Portuguese influence on Tetun Dili. Note that Tetun Dili has not undergone any socially recognized standardization. According to Williams-van Klinken et al. (2002:5), "A significant issue in any description of Tetun Dili is the unusually high levels of interspeaker variation—seemingly across all areas of the grammar, including the lexicon and phonology." For consistent speaker judgment, all the Tetun Dili data in this paper are from my own fieldwork (Chen 2014a and 2014b, unless otherwise noted), carried out in Hawai'i with Leonardo Germano Tourinho de Souza Rosa, a native speaker of Tetun Dili, who was born and grew up in Dili before coming to Hawai'i for higher education.[3]

Generally speaking, Tetun Dili has five vowels, which show a three-way contrast in terms of height (high, mid, low), and backness (front, central, back) (see Table 1). The vowel inventory of Tetun Fehan is identical to that of Tetun Dili, except that the mid vowels can be either close-mid or open-mid, depending on the phonetic environments. By contrast, Portuguese has seven distinct vowels, five of which also have

---

[2] Adapted from Google Map.

[3] For the grammar, based on a language corpus, including data from non-native speakers of Tetun Dili, please refer to Williams-van Klinken et al. 2002.

a nasal counterpart (see Table 3). Unlike Tetun Dili or Tetun Fehan, Portuguese shows a four-way contrast in height, where mid can be further divided in to high-mid and low-mid. Considering that Proto-Malayo-Polynesian has only four contrastive vowels: /i, ə, a, u/ (Blust 2013) and that Tetun Fehan has only five vowels, it is not plausible to regard the vowel inventory of Tetun Dili as a simplified version of Portuguese.

TABLE 1. The vowel inventory of Tetun Dili

	Front	Central	Back
high	i		u
mid	ɛ		ɔ
low		ɑ	

TABLE 2. The vowel inventory of Tetun Fehan[4]

	Front	Central	Back
high	i		u
mid	e [e, ɛ]		o [o, ɔ]
low		ɑ	

TABLE 3. The vowel inventory of Portuguese[5]

	Front	Central	Back
high	i, ĩ		u, ũ
	e, ẽ		o, õ
mid		ɐ, ɐ̃	
	ɛ		ɔ
low		a	

As for consonants, it is questionable if the Tetun Dili system should be regarded as a simplified version of either Portuguese or Tetun Fehan, given that Tetun Dili and Tetun Fehan share 12 native consonants. As presented in Table 4, Tetun Dili has 19 consonant phonemes, 12 of which are indigenous and 7 of which occur only in loanwords with either Portuguese or Indonesian sources (/p/ and /g/ occur both in Portuguese and Indonesian loans, and only /z/, /v/, /ʃ/, /ʒ/ and /ɲ/ are exclusively from Portuguese). As shown in Table 5, Tetun Fehan is influenced by Indonesian and has imported 5 phonemes, i.e., /p/, /c/, /ʤ/, /ŋ/ and /g/, along with the Indonesian loanwords. The indigenous phonemes in Tetun Fehan resemble those found in Tetun Dili with the possible exception of the glottal stop. More specifically, the glottal stop is not considered a phoneme in Tetun Dili because it always occurs intervocalically, and omitting it does not change the meaning of a word. The glottal stop always occurs between vowels in Tetun Fehan as well, but it is not clear whether the glottal stop can be omitted, or if the omission changes the meaning of a word.

---

[4] All the Tetun Fehan data in this paper are adapted from van Klinken 1999 unless otherwise noted.
[5] Portuguese data in table 3 and table 6 are adapted from Cruz-Ferreira 1995.

TABLE 4. The consonant inventory of Tetun Dili (Parentheses indicate that these phonemes are borrowed)

	bilabial	labiodental	dental	alveolar	palatal	velar	glottal
nasal	m		n		(ɲ)		
stop	(p) b		t	d		k (g)	
fricative		f (v)		s (z)	(ʃ) (ʒ)		h
affricate							
trill				r			
lateral			l				
approximant	w						

TABLE 5. The consonant inventory of Tetun Fehan (Parentheses indicate that these phonemes are borrowed)

	bilabial	labiodental	dental	alveolar	palatal	velar	glottal
nasal	m		n			(ŋ)	ʔ
stop	(p) b		t	d	(c)	k (g)	
fricative		f	s				h
affricate					(dʒ)		
trill				r			
lateral			l				
approximant	w						

TABLE 6. The consonant inventory of Portuguese

	bilabial	labiodental	dental	alveolar	palato-alveolar	palatal	velar	uvular
nasal	m		n			ɲ		
stop	p b		t d				k g	
fricative		f v		s z	ʃ ʒ			ʁ
affricate								
trill								
tap				ɾ				
lateral			l			ʎ		

Table 7 presents a list of personal and demonstrative pronouns and *wh-* question words in Tetun Dili, Tetun Fehan, and Portuguese. Given that the forms in Tetun Dili and Tetun Fehan share the same etyma, there is little doubt that Tetun Dili employs Austronesian words for personal and demonstrative pronouns and *wh-* words, and no Portuguese borrowing is observed. In the pronominal systems of Baba Malay (Lee 2014), Tok Pisin (Dutton 2012), Bislama (Crowley 2004), and Chabacano (Forman 1993), where at least two languages were involved in the formation of creoles, mixed systems are observed. By contrast, the personal pronouns in Tetun Dili are purely Austronesian, which casts doubt on the claim that Tetun Dili is a creolized language.

TABLE 7. Pronouns and *wh-* words in Tetun Dili, Tetun Fehan, and Portuguese

Gloss	Tetun Dili[6,7]	Tetun Fehan	Portuguese
1SG	hau	ha'u	eu
2SG	o	ó	tu, você
3SG	nia, ni[8]	nia	ele (M), ela (F)
1PL.INCL	ita	ita	nos
1PL.EXCL	ami	ami	
2PL	imi	emi	vocês
3PL	sira	sia, sira	eles (M), elas (F)
this	ne'e	ne'e	este
that	neba'a	nebá	esse
who	se'e	sé	quem
what	saida	sá	que
where	nebe'e	nabé	onde
when	wainhira/bainhira (generic); hor wainhira (in the past)	hori.hirak (past; in recent time); wain.hira (future)	quando

As for the kinship system in Tetun Dili, it resembles the Austronesian kinship system found in Tetun Fehan (see table 8). The inconsistency of the final nasal in the words for 'father', 'mother', and 'sibling' between Tetun Dili and Tetun Fehan is due to the inalienable possessive suffix *-n*. In this semantic category, Portuguese left its trace in the words for 'aunt', 'uncle', 'cousin', 'niece', and 'nephew'. One may wonder if the use of Portuguese terms implies intermarriage between the Timorese and the Portuguese in the past. However, although Tetun Dili and Tetun Fehan employ different terms for 'husband' and 'wife', neither of these words in Tetun Dili is from Portuguese or is translated literally from Portuguese. *Katuas oan* 'husband' in Tetun Dili literally means 'old man + child' while *ferik oan* 'wife' literally means 'old woman + child'. Moreover, while Tetun Dili has *kaben* for 'spouse', it does not have a real semantic equivalent for Portuguese *noiva* 'bride' and *noivo* 'groom'. Instead, the Tetun Dili words associated with weddings are *barlaki* 'bride price' or *folin* 'price', which refers to the money that the groom needs to pay in order to marry the bride, and *be'e manas ai tukan* as the gifts or pre-payment for *barlaki* from a man's family to the bride's family at the engagement. In either case, the word choice indicates that it is less likely that frequent intermarriage existed between Tetun Dili speakers and the Portuguese in the past, given that the wedding tradition is predominantly Tetun.

---

[6] In this paper, the working orthography of indigenous Tetun Dili words together with equivalent IPA symbols is as follows: i /i/, e /ɛ/, a /ɑ/, o /ɔ/, u /u/, m /m/, n /n/, b /b/, t /t/, d /d/, k /k/, g /g/, f /f/, s /s/, h /h/, r /r/, w /w/, l /l/, and the *'okina* (') for [ʔ]. As in daily life, in written Tetun Dili, loanwords are predominantly spelled following the standard spellings of the source languages, and the pronunciations of these loans vary from speaker to speaker, depending on their fluency of the source languages (Leonardo Germano Tourinho de Souza Rosa, 2014 pers. comm.).

[7] According to Melody A. Ross (2015 pers. comm.), the medium of instruction in East Timor is currently Portuguese, though many teachers may not be able to speak it (well). The national curriculum in Tetun is expected to be reformed soon. Regarding the standard orthographies, there are two sets for Tetun Dili, with Tetun Oficial used by the government, and the other one used by commoners and mass media. Book publishers might use either spelling system.

[8] *ni* '3SG' is employed only in the possessive construction.

TABLE 8. Kinship terms in Tetun Dili (The shading indicates Portuguese loans.)

Gloss	Tetun Dili	Tetun Fehan[9]	Portuguese
child	oan; labarik	oan; lawarik	criança
mother	ama, inan[10]	ina(n)	mãe
father	apa, aman	ama(n)	pai
daughter	oan feto	--	filha
son	oan mane	--	filho
younger sister	alin (feto)	ali	irmãzinha
younger brother	alin (mane)	ali	maninho
husband	katuas oan	la'en	marido
wife	ferik oan	feen	mulher
aunt	tia	--	tia
uncle	tiu	--	tio
female cousin	prima	--	prima
male cousin	primu	--	primo
niece	subrinha	--	sobrinha
nephew	subrinhu	--	sobrinho
bride	N/A	--	noiva
groom	N/A	--	noivo

Starting from the establishment of Dili in the late 1700s until 1975, Tetun Dili speakers and the Portuguese were in intensive contact (Williams-van Klinken et al. 2002). A majority of East Timorese nowadays are Catholic, with Latinized names, after being converted by the Portuguese. Consequently, it is not surprising that the words for months and days in Tetun Dili, as a result of the past colonization, are all Portuguese loans (see table 9). Even now, Tetun Dili speakers use Portuguese for dates as in *dez de Outubro de dois mil e dois* for '10/Oct/2002'. Only in delivering formal speeches or in literature will people use native Tetun numbers to name the date, such as *loron sanulu fulan outubro tinan rihun rua sanulu resin ha'at* for '10/Oct/2014' (Chen 2014b).

However, though the notion for telling time on a clock is relatively recent, the words related to time are native Tetun Dili expressions, where *tuku* 'to strike' is used as an index. So 'one o'clock' is *tuku ida* 'strike one' and two o'clock is *tuku rua* 'strike two' (cf. in the expression of *agora tuku hira (*lit. 'now strike when')? 'What time is it?', the word *agora* 'now' is Portuguese). The word for 'season' in Tetun Dili is borrowed from Portuguese *tempo* 'season'. But the ideas of different seasons (spring, summer, fall, and winter) do not really exist, due to the geographical location of Timor. Though Tetun Dili speakers distinguish *bai loron* 'dry season' from *tempo udan* 'wet season', only *bai loron,* lit. 'time sun'*,* is predominantly employed by every Tetun Dili speaker (Chen 2014b). This indicates that borrowings usually take place only when new notions are introduced.

---

[9] In this paper, the Tetun Fehan equivalents are provided only when the data are available.

[10] It is fine to use *inan* 'mother' or *aman* 'father' to refer to one's own mother or father. However, when used to denote other people's mother or father, *inan* and *aman* are only used in cursing contexts.

TABLE 9. A comparison of time expressions in Tetun Dili and in Portuguese

Gloss	Tetun Dili	Portuguese
Monday	segunda-feira	segunda-feira
Tuesday	terça-feira	terça-feira
Wednesday	quarta-feira	quarta-feira
Thursday	quinta-feira	quinta-feira
Friday	sexta-feira	sexta-feira
Saturday	sabado	sabado
Sunday	domingo	domingo
January	janeiro	janeiro
February	fevereiro	fevereiro
March	março	março
April	abril	abril
May	maio	maio
June	junho	junho
July	julho	julho
August	agosto	agosto
September	setembro	setembro
October	outubro	outubro
November	novembro	novembro
December	dezembro	dezembro
1 o'clock	tuku ida	uma hora
2 o'clock	tuku rua	duas horas

Table 10 demonstrates that the numeral system in Tetun Dili is, without doubt, Austronesian. However, different numeral systems are used in different domains (Chen 2014b). Though Tetun Dili speakers use Tetun Dili numerals in serial counting, they use Indonesian numerals for naming prices and bargaining in markets, but Portuguese ones in mathematics class (except in Indonesian schools). They also use Indonesian numerals when giving out phone numbers.

TABLE 10. Basic numerals in Tetun Dili, Portuguese, and Indonesian

Gloss	PAN[11]	Tetun Dili	Portuguese	Indonesian
1	*esa/isa	ida	um, uma	satu
2	*duSa	rua	dois, duas	dua
3	*telu	tolu	três	tiga
4	*Sepat	ha'at	quatro	empat
5	*lima	lima	cinco	lima
6	*enem	ne'en	seis	enam
7	*pitu	hitu	sete	tujuh
8	*walu	walu	oito	delapan
9	*Siwa	sia	nove	sembilan
10	*sa-puluq	sanulu	dez	sepuluh
11	--	sanulu resin ida	onze	sebelas
12	--	sanulu resin rua	doze	dua belas
100	*RaCus	atus ida	cem	seratus
1000	(PMP: *Ribu)	rihun ida[12]	mil	seribu

TABLE 11 presents a list of words belonging to different semantic categories, such as household items and accessories, buildings, school, modern technology, and sports. However, the existence of massive borrowings in certain semantic domains does not mean that the recipient language is a creolized version of the donor language (cf. Bidau Portuguese Creole with Portuguese as its lexifier; for more details, see Baxter 1990). For example, abundant Latin loans can be found in the following semantic fields in English:[13] (1) the church life: altar, angel, candle, hymn, martyr, psalm, relic; (2) education: school, master, grammar, verse, gloss; (3) domestic life and clothing: cap, sock, silk, purple, mat, sack), but English is not considered to be a creolized version of Latin. Although Grimes et al. (1997:52) note that "Tetun Prasa has a significantly higher portion of Portuguese loanwords …," they fail to capture the nature of the domains where the borrowing has taken place. If we use their standard to define a creole, English, Japanese, and Tarifiyt Berber should be regarded as creoles as well, considering 42% of the English lexicon, 36% of the Japanese lexicon, and 53% of the Tarifiyt Berber lexicon are borrowed (Grant 2009; Schmidt 2009; Kossmann 2009).

In Tetun Dili it is not surprising that massive borrowings occur in words for material culture and education. The majority of the words associated with school are from Portuguese, though the words for 'homework assignment' have two different forms, depending on which school that one goes to. It is worth mentioning that items such as 'clothes', 'bag', 'bowl', 'plate, 'spoon', 'house', 'roof' and 'ship' already existed in the traditional society before the arrival of the Portuguese. Thus, the original Tetun Dili words are retained, given that there is little need for borrowing. Also, according to my language consultant Leonardo Germano Tourinho de Souza Rosa (2014, pers. comm.), although most Tetun Dili speakers have converted to Catholicism and have become westernized nowadays, the importance of *knua* 'the house-altar' has never declined, and the keeper of the family house has to be able to tell the family history during the family gathering.

---

[11] All the PAN and PMP forms cited in this paper are from *the Austronesian Comparative Dictionary* (Blust and Trussel, ongoing; accessible online at http://www.trussel2.com/ACD).

[12] The word *rihun* 'thousand' is very likely an Indonesian loan, given that PAN *R in *RaCus 'hundred' is dropped, and that *r* is the regular reflex of PMP *R in Malay/Indonesian.

[13] Excerpts from Baugh and Cable 2001.

TABLE 11. Words for modern culture in Tetun Dili

Category	Gloss	Tetun Dili
household items and accessories	table	meza (< Portuguese)
	chair	kadeira (< Portuguese); turfatin (< Tetun)[14]
	bed	kama (< Portuguese)
	clothes	hena (< Tetun)
	shoes	sapatu (< Portuguese)
	socks	meias (< Portuguese)
	cup	kopu (< Portuguese)
	pants	kalsa naruk (pants long; < Portuguese + Tetun)
	shorts	kalsa badak (pants short; < Portuguese + Tetun)
	plate	bikan (< Indonesian)[15]
	bowl	maŋkok (< Tetun)
	coffee	café (< Portuguese)
	tea	xa (< Portuguese)
	spoon	kanuru (< Tetun)
	kettle	buli (< Tetun)
	fork	garfu (< Portuguese)
	bag; rice sack	karon (< Tetun)
building	fan	vintunha (< Portuguese)
	roof	uma kakuluk (< Tetun)
	house	uma (< Tetun)
	church	igreja (< Portuguese)
	house-altar	knua (< Tetun)
school	glasses	oklu (< Portuguese)
	backpack	pasta (< Portuguese)
	pencil	lapis (< Portuguese)
	pen	lapizeira (< Portuguese)
	book	libru (< Portuguese)
	paper	surat tahan (< Indonesian + Tetun)
	glue	lem (< Indonesian)
	staple	hekter (< Indonesian)
	homework assignment	T.P.C. (trabalho para casa) (< Portuguese); P.R. (pekerjaan rumah) (< Indonesian)
	school subject	materia (< Portuguese)
	mathematics	matematika (< Portuguese)
	science	ciencia (< Portuguese)
	school	eskola (< Portuguese)
	business card	cartão (< Portuguese), kartu (< Portuguese)
modern	printer	printer (< English)

---

[14] *Turfatin* in Tetun Dili refers to things that people can sit on, not just chairs.
[15] The Tetun Dili *bikan* 'plate' is borrowed from Indonesian *pinggan* 'plate', which was borrowed from Tamil.

technology	computer	komputador (< Portuguese)
	car	kareta (< Portuguese)
	bike	bicicleta (< Portuguese)
	motorcycle	motor (< Portuguese)
	bus	bis (< Indonesian)[16]
	taxi	taxi (< Portuguese)
	airplane	avião /aviaun (< Portuguese)
	ship	ro (< Tetun)
sports	basketball	basket (< English)
	soccer	tebe bola[17] (< Tetun + Portuguese) / joga bola (< Portuguese)
	tennis	tennis (< English)

In addition to the nouns, basic verbs in Tetun Dili are also predominantly of Austronesian origin (see table 12). However, a few verbs seem to have only loaned forms. Verbs related to politics and education, such as, like *vota* 'to vote' and *estuda* 'to study', are usually Portuguese loans, given that Portugal colonized East Timor until 1975.

As for *nonton* 'to watch (TV, movie, sports)', it is borrowed from Indonesian. The occupation of East Timor by Indonesia took place from 1975 to 1999, which was the period during which televisions and movies were introduced. Hence, the Indonesian verb could be introduced along with the new fashion. Though Tetun Dili has its own word *hadomi* for 'to love', it still borrowed *gosta* 'to like' and *sente* 'to feel' from Portuguese. A similar type of borrowing is attested in English, in which the basic words *enjoy* and *carry*, among others, were borrowed from French. Therefore, it is not convincing to claim that Tetun Dili employs Portuguese as its lexifier simply because Portuguese loans are observed.

TABLE 12. Verbs in Tetun Dili and Portuguese

Gloss	Tetun Dili	Portuguese
to drink	hemu	beber
to eat	han	comer
to bite	tata	morder
to see	haré	ver
to hear	rona	ouvir
to know	hatene	conhecer
to die	mate	morrer
to kill	oho	matar
to swim	nani	nadar
to fly	semo	voar
to go	ba	ir
to walk	la'o	andar
to come	mai	vir
to sit	tu'ur	sentar
to stand	hamrik	pé

---

[16] Indonesian borrowed this word from Dutch *bus*.
[17] *Tebe* means 'to kick' in Tetun Dili.

to give	fo'o	dar
to love	hadomi	amar
to watch/view	nonton (< Indonesian)	observar
to like	gosta (< Portuguese)	gostar
to feel	sente (< Portuguese)	sentir
to vote	vota (< Portuguese)	votar
to study	estuda (< Portuguese)	estudar

Common property-denoting words of Portuguese origin are attested even though they are not all associated with any socio-cultural innovations (see table 13). For example, for 'happy' and 'sad', both Tetun and Portuguese forms are used frequently, whereas Tetun Dili speakers only employ Portuguese loans for 'delicious, tasty' and 'excited'. Nevertheless, the majority of frequently used property-denoting words are of Tetun origin.

TABLE 13. Property-denoting words in Tetun Dili

Gloss	Tetun Dili
ugly	oin a'at (lit. 'face bad') (< Tetun)
pretty	furak (< Tetun), bonita (< Portuguese)
fat	bokur (< Tetun)
skinny	krekas (< Tetun)
chubby	bebe bokur (lit. 'baby fat') (< Tetun)
happy	ksolok (< Tetun), kontente (< Portuguese)
sad	hakraik (< Tetun), nakraik (< Tetun), tristi (< Portuguese)
angry	hirus (< Tetun)
hungry	hamlaha (< Tetun)
excited	kontente (< Portuguese)
jealous	laran moras (< Tetun)
greedy	karak (< Tetun)
talkative	koalia barak (< Tetun)
smart	matenek (< Tetun)
stupid	beik ten (< Tetun)
delicious, tasty	gostu (< Portuguese)
sweet	midar (< Tetun)
easy	gampang (< Indonesian); fasil (< Portuguese); la susar 'not difficult' (< Tetun)
difficult	susar (< Tetun)
sick, painful	moras (< Tetun)
spicy, hot	manas (< Tetun)
bitter	moruk (< Tetun)
sour	sin (< Tetun)

Basic color terms in Tetun Dili are given in table 14. According to Blust (2013:301), only three colors (i.e., 'white', 'black', and 'red') are reconstructible to Proto-Austronesian (PAN) and Proto-Malayo-Polynesian (PMP), but five colors (i.e., 'white', 'black', 'red', 'grue (green and blue)' and 'yellow') are reconstructible to Proto-Oceanic (POC). It is quite clear that Tetun Dili words for 'white', 'black', 'red', 'yellow', and 'green' are reflexes of the reconstructed PMP forms, while the word for 'blue' is not.

Though Tetun Dili has its own word for 'blue', the Portuguese loan *azul* is always favored, probably due to the similarity between *matak* 'green' and *modok* 'blue'.

TABLE 14. Color terms in Tetun Dili, PMP, and Portuguese

Gloss	PMP	Tetun Dili	Portuguese
white	*ma-putiq	mutin	branco
black	*ma-qitem	metan	preto
red	*ma-iRaq	me'an	vermelho
yellow	*kunij 'turmeric'	kinur	amarelo
green	*ma-qataq[18] (PAN)	matak	verde
blue	--	modok (< Tetun); azul (< Portuguese)	azul

It is not uncommon for Tetun Dili speakers to use both Portuguese and Tetun greetings, implying that long-term contact of Tetun speakers with Portuguese speakers created the need for them to learn another set of greetings. Since Tetun Dili has its own forms for greeting in addition to the Portuguese borrowings, it is again questionable to claim that Tetun Dili is a creolized language.

TABLE 15. Greetings

Gloss	Tetun Dili	Portuguese loans
Hello; How are you?	diak ka lae '(lit.) good or not'	ola
Good morning	dade'er diak '(lit.) morning good'	bom dia
Good afternoon	(use Portuguese)	boa tarde
Good evening/night	kalan diak '(lit.) night good'	boa noite
Good day	loron diak '(lit.) day good'	--
Thank you	(use Portuguese)	obrigado/a
Goodbye	(use Portuguese)	adeus
Sorry	(use Portuguese)	deskulpe

4. THE MORPHOSYNTAX OF TETUN FEHAN AND TETUN DILI. In general, Austronesian languages spoken in mainland Southeast Asia and the Lesser Sundas employ a relatively low degree of morphological markings (Blust 2013:360). Tetun Dili and Tetun Fehan are no exception to this typology.

This section provides a preliminary description and comparison of morphology and syntax in Tetun Dili, Tetun Fehan, and Portuguese. Issues to be covered include word order, affixation, possessive construction, relative clauses, and conditional clauses. The Tetun Fehan data are excerpted from van Klinken 1999[19] while the Tetun Dili data are from Chen 2014a, 2014b. Portuguese data are from Perini 2002 and Carolina Aragon (2014, pers. comm.).

4.1 BASIC WORD ORDER. Both Tetun Fehan and Tetun Dili are head-initial languages with the head followed by its modifier, such as N-A and N[head]-N for noun phrases (as in (1), (2), (5), and (6)),[20] Preposition-N for prepositional phrases (as in (3) and (7)), and V-O for verb phrases (as in (4) and (8)), although exceptions have also been observed. Since both Tetun varieties have little morphology, the word order is predominantly fixed except for the sentences with a relativizer and a focus marker.

---

[18] It means 'raw, unripe', instead of 'green' in PAN.

[19] I have, however, modified the interlinear glosses used in van Klinken 1999 in accordance with the Leipzig glossing rules.

[20] A, standing for adjectives, actually refers to the property-denoting words in Tetun Dili. It is noteworthy that most Austronesian languages do not have true adjectives.

- Tetun Fehan

(1) feto   ki'ik   á
     woman   small/young   DEF.ART
     'this young woman' (adapted from van Klinken 1999:133)

(2) na'an   fahi
     meat   pig
     'pig meat' (adapted from van Klinken 1999:134)

(3) iha   uma
     LOC   house
     'at home' (adapted from van Klinken 1999:160)

(4) hasoru nia
     meet   3SG
     'meet him/her' (adapted from van Klinken 1999:172)

- Tetun Dili

(5) tinan   lima
     year   five
     'five years'

(6) du'ut   matak
     grass   green
     'green grass'

(7) husi   Timor
     from   Timor
     'from East Timor'

(8) tuku   nia
     punch 3SG
     'punch him/her'

Portuguese is also predominantly head-initial, in that the head precedes its modifier, such as N[head]-de-N (as in (9)), N-A (as in (10)) and Preposition-N (as in (11)). However, example (12) shows that when a noun is modified by a numeral, the numeral has to precede the head noun it modifies.

- Portuguese

(9) carne   de porco
     meat   of pig
     'meat of pig (pork)'

(10) uma   janela   aberta
      INDF.SG.F window   open.F
      'an open window' (adapted from Perini 2002:86)

(11) em   casa
      LOC   home
      'at home'

(12) duas   pessoas
      two   people
      'two people'

As for constituent order in sentences, Portuguese is canonically SVO (as in (13)), although it shows SOV when sentences contain clitic pronouns (as in (14)). Moreover, Portuguese is a pro-drop language, so the subject can be dropped without posing problems for deciphering a sentence in Portuguese, as shown in (15).

(13) Eu   gosto   de maça
     1SG.NOM   1SG.like   of apple
     'I like apples.'

(14) Alex   me   assustou
     Alex   1SG.ACC   3SG.scare.PST
     'Alex scared me.'

(15) (Eu)   estou   com   medo   do Alex
     1SG.NOM   1SG.COP   with   fear   of Alex
     'I am afraid of Alex.'

4.2 SUBJECT AGREEMENT. Table 16 shows that verbs in Tetun Fehan agree with subjects. The agreement marker for 1SG, 2SG, and 3SG are *k-*, *m-*, and *n-*, respectively. As for the plural subjects, only 3PL is specified on the verbs using *r-*, which can alternate with *n-* (from the agreement marker for 3SG). According to van Klinken (1999:172–173), subject agreement in general can occur only on verbs that begin with *h-*, which will then be replaced by the agreement marker. Take *há* 'to eat' for instance. After prefixation, *há* becomes *ká* '1SG.eat', *má* '2SG.eat', *ná* '3SG.eat' and *rá* '3PL.eat'. Example (16) shows that the initial consonant of *hasoru* 'to meet' is also replaced by the first singular subject agreement marker *k-*, resulting in *kasoru* '1SG.meet'.

TABLE 16. Subject agreement in Tetun Fehan[21]

Gloss	Full pronoun	Subject agreement	*há* 'to eat'
'1SG'	ha'u	k-	*k*á
'2SG'	ó	m-	*m*á
'3SG'	nia	n-	*n*á
'1PL.INCL'	ita	--	*h*á
'1PL.EXCL'	ami	--	*h*á
'2PL'	emi	--	*h*á
'3PL'	sia	r- (n-)	*r*á (*n*á)

(16) ha'u   **k**.sai   mai   **k**.asoru   nia
     1SG   1SG.exit   come   1SG.meet   3SG
     'I came out and met him.' (adapted from van Klinken 1999:172)

(17) ha'u   **k**-bá   nono   wé   á
     1SG   1SG-go   heat.liquid   water   DEF.ART
     'I went and boiled water …' (adapted from van Klinken 1999:175)

(18) *ha'u   **k**-uru   surat
     1SG   1SG-recollect   paper
     'I recollect the playing cards.' (adapted from van Klinken 1999:172)

By contrast, Tetun Dili does not have subject agreement, as shown in (19) and (20), where *ba* 'go' is not prefixed with *k-*, cf. Tetun Fehan ***k*-bá** '1SG-go' in (17), and the subject agreement marker *r-* for '3PL'

---

[21] Adapted from van Klinken 1999:173.

in Tetun Fehan is not observed on *han* 'eat' in Tetun Dili, which contains an *h*- initial; cf. Tetun Fehan *kasoru* '1SG.meet' in (16).

(19) hau   ba igreja ho   nia
     1SG   go church with   3SG
     'I go to the church with him.'

(20) sia   han   etu   ho   nan
     3PL   eat   rice   with   meat
     'They eat rice with meat.'

Subject agreement, as in Tetun Fehan, is also observed in Portuguese, although a different strategy is employed. By comparing (21)–(23), we see that in Portuguese the past tense forms of the verb 'to scare' change according to the person and number of the subject.

(21) Nós   **assustamos**   Alex
     1PL.NOM   1PL.scare.PST   Alex
     'We scared Alex.'

(22) Eu   **assustei**   Alex
     1SG.NOM   1SG.scare.PST   Alex
     'I scared Alex.'

(23) Alex   me   **assustou**
     Alex   1SG.ACC   3SG.scare.PST
     'Alex scared me.'

In sum, Portuguese uses inflectional morphology for subject agreement in verbs, while Tetun Fehan employs subject agreement prefixes on verbs beginning with *h*-. As for Tetun Dili, it shows no subject agreement at all. Van Klinken (1999:172) points out that subject agreement in Tetun shows dialectal variation. While all agreement markers occur only on *h*-initial verbs in Tetun Fehan and Tetun Foho, similar agreement markers are said to occur on *o*-initial verbs as well in Tetun Terik, spoken in southern East Timor. That is to say, subject-agreement in Tetun Fehan and Tetun Foho is simpler compared to the Tetun variety spoken in southern East Timor. Tetun Dili can probably be regarded as a further simplified version of Tetun Fehan and Tetun Foho on this subject agreement variation continuum. No Portuguese influence is observed.

4.3 THE PREFIXES *HA*- AND *HAK*-. Van Klinken (1999:59–61) states that there are two common prefixes, *ha*- and *hak*-, that are employed to derive verbs from verbs, adjectives, or nouns in Tetun Fehan. Generally speaking, the prefix *ha*- is used to derive transitive causative verbs from adjectives or intransitives (see Table 17). It is likely that the prefix *ha*- originated from PAN *pa*- 'causative of dynamic verbs' (for more details, see Blust 2003; Zeitoun and Huang 2000).

If the stems begin with a consonant cluster, the first member of the cluster will be deleted after *ha*-prefixation, such as **ha**badak 'to shorten' (< *ha*- + **k**badak 'short') and **ha**mukit 'to impoverish' (< *ha*- + **k**mukit 'poor').

TABLE 17. Words containing the prefix *ha*- in Tetun Fehan[22]

Stem	Gloss	*ha*-	Gloss
kbadak	'short'	**ha**badak	'to shorten'
kmukit	'poor'	**ha**mukit	'to impoverish'
lós	'straight'	**ha**lós	'to straighten'
butan	'worthless'	**ha**butan	'to make worthless'
sa'e	'to ascend'	**ha**sa'e	'to raise'

---

[22] Adapted from van Klinken 1999:59–60.

toba	'to lie down'	**ha**toba	'to lay down'
sai	'to exit'	**ha**sai	'to remove'

Example (24) contains an intransitive verb *tún* 'to descend' which, after prefixation, becomes a causative verb **hatún** 'to lower' (lit., 'make (something) descend'), as shown in (25).

(24) nia   tún   **n**ika   bá rai   klaran
     3SG   descend   3SG.back   go earth   middle
     'She descended back to earth (from heaven)' (adapted from van Klinken 1999:60).

(25) ita   **ha**-tún   kair   bá wé
     1PL.INCL   make-descend   fishing.line   go water
     'We (INCL) lower the fishing line into the water' (adapted from van Klinken 1999:60).

As for the prefix *hak-*, which might originate from PAN *pa-ka- 'causative of stative verbs' (see Blust 2003; Zeitoun and Huang 2000), it is used to change transitive verbs into intransitives in Tetun Fehan. Some of the stems (e.g. *laik*) may not be free forms in contemporary Tetun Fehan (see Table 18).

TABLE 18. The prefix hak- in Tetun Fehan[23]

Stem	Gloss	**hak-**	Gloss
doko	'to rock (e.g. baby)'	**hak**doko	'(to be) rocked, tremble'
tesi	'to break'	**hak**tesi	'(to be) broken'
tutan	'to join (pieces)'	**hak**tutan	'(to be) connected'
*laik	--	**hak**laik	'(to be) partly dry'

Example (26) provides a contrast between the *ha-* and the *hak-* prefixes. The causative verb *ha-laik* 'to partly dry (cause something to be partly dry)' is formed by the stem *laik* and the causative prefix *ha-*, while the intransitive verb *hak-laik* 'to be partly dry' is formed by the same stem base prefixed by *hak-*, which turns verbs into intransitives. The word-initial *h-* onset is then replaced by the subject agreement marker *n-* for '3sg'.

(26) ita   hawai   **ha**laik   na'an   lai   na'an   **na**klaik
     1PL.INCL   dry   partly.dry   meat   now   meat   3SG.partly.dry
     'We (INCL) first dry partly-dry the meat. (So) the meat is partly dry' (adapted from van Klinken 1999:61).

As for Tetun Dili, it also employs *ha-* to form morphological causatives. Interestingly, *ha-*, being reanalyzed as deriving from the lexical verb *halo* 'to make; to do' by the present-day Tetun Dili speakers, seems to have a variant *hak-* in which the origin of *k-* remains unknown synchronically. Table 19 presents a list of stems that are prefixed with *ha-* and/or *hak-*.

TABLE 19. The prefix *ha(k)-* in Tetun Dili

Stem	Gloss	**ha(k)-**	Gloss
bokon	'wet'	**ha**bokon	'to wet'
bosu	'full'	**ha**bosu	'to make oneself full'
fo'er	'dirty'	**ha**fo'er	'to make something dirty'
moris	'alive'	**ha**horis	'to make alive; to give birth to'
mamuk	'empty'	**ha**mamuk	'to empty'

---

[23] Adapted from van Klinken 1999:59–61.

mo'os	'clean'	**ha**mo'os	'to clean'
tauk	'afraid'	**ha**tauk	'to make someone afraid'
toba	'to lie down; sleep'	**ha**toba	'to lay down'
monu	'to fall'	**ha**monu	'to cause something to fall'
*filak	--	**hak**filak	'to cause something to transform'

Example (27) is an example of a simple intransitive sentence with the intransitive verb *monu* 'to fall' and its subject *didin* 'wall'. When being causativized, the causative prefix *ha-* is attached to the intransitive verb *monu* 'to fall' to form **ha**monu 'to cause something to fall', as shown in (28).

(27) didin   monu
     wall   fall
     'The wall collapsed.'

(28) hau   **ha**monu   nia
     1SG   CAUS.fall   3SG
     'I made it fall.'

Similarly, the pair of examples in (29) and (30) are also used to illustrate that the intransitive *tauk* 'to be afraid' becomes a transitive causative verb 'to scare (to make someone afraid)' in (30) after adding the prefix *ha-*.

(29) hau   tauk   Alex
     1SG   afraid   Alex
     'I am afraid of Alex.'

(30) hau   **ha**tauk   Alex
     1SG   CAUS.afraid   Alex
     'I scared Alex.'

The use of *hak-* in Tetun Dili seems to carry the same function as in Tetun Fehan, converting verbs into intransitives, as exemplified in (32) where *fodak* becomes **hak**fodak 'to be surprised.' The major difference between examples (31) and (32) is that *ha-* turns the stem *fodak* into a transitive causative verb, whereas *hak-* turns *fodak* in to an intransitive, which is expected.

(31) hau   **ha**fodak   nia
     1SG   CAUS.surprised   3SG
     'I surprised him.'

(32) nia   halo   Leo   **hak**fodak
     3SG   make   Leo   surprised
     'He made Leo surprised'

However, unlike the prefix *hak-* in Tetun Fehan, which seems to function only as a detransitivier (since it converts verbs into intransitives), the prefix *hak-* in Tetun Dili, like *ha-*, also functions as a causativizer (see (33)),[24] which is equivalent to the syntactic causative construction containing the causative verb *halo* followed by a lexical verb (see (34)).

(33) labarik **hak**filak   lafaek
     child   CAUS.transform crocodile
     'The child transformed the crocodile.'

(34) labarik **halo**   filak   lafaek
     child   make   transform   crocodile
     'The child transformed the crocodile.'

---

[24] Few bare nouns begin with *kf-* in Tetun Dili.

4.4 THE POSSESSIVE CONSTRUCTION. Unlike other noun phrases which are head-initial (Head-Modifier), the possessive construction in Tetun Fehan predominantly follows the Possessor-Possessum order (Modifier-Head) with the possessive markers attached to either possessors or possessums, depending on whether the possessums end in a consonant or a vowel. If the possessum ends in a vowel, the possessum marker *-n* occurs on the possessum,[25] exemplified in (35) where *-n* is attached to *fé* 'wife'.

(35) ó   fé-**n**
      2SG   wife-GEN
      'your (SG) wife' (adapted from van Klinken 1999:143)

     Another possessum marker is *–r*, which occurs on the possessum only when the possessor is plural. In this case, the possessum marker *-r* can alternate with *-n*, as in (36). Van Klinken (1999:150) points out that *-n* is favored by educated people and younger generations, whereas *-r* is favored by illiterate native speakers. Moreover, the marker *-r*, contrary to *-n*, can never co-occur with an affixed possessor, as in (37), where the possessor *ami* '1PL.EXCL' is marked by the possessor marker *-kan*.

(36) (a) sa   ata-**r**
           3PL   slave-GEN
           'slave of theirs'

      (b) sa   ata-**n**
           3PL   slave-GEN
           'slave of theirs'

(37) ami-**kan**   bei-**n**
      1PL.EXCL-POSS ancestor-GEN
      'our (EXCL) ancestor'

     Unlike examples (35)–(37), if the possessums end with a consonant, they cannot be affixed, as demonstrated in (38) and (39). In this case, Tetun Fehan employs a possessor marker *-kan* to specify the possessive relation between the two nouns. Example (38) shows that since *ha'u* '1SG' is marked by *-kan*, so *ha'u* must be the possessor. (39) and (40) further illustrate that if the possessor is a common noun, a corresponding pronoun has to co-occur with the common noun for the possessor marker *-kan* to land on. Because *ina-ama* 'parents' and *tamukun* 'village.head' are common nouns, the corresponding pronouns *sia* '3PL' and *nia* '3SG' are employed here.

(38) ha'u-**kan**   tais
      1SG-POSS   cloth
      'my cloth'

(39) ina-ama   sia-**kan**   sasoin
      mother-father   3PL-POSS   wealth
      'parents' wealth' (adapted from van Klinken 1999:143)

(40) tamukun   nia-**kan**   fé-**n**
      village.head   3SG-POSS   wife-GEN
      'village head's wife' (adapted from van Klinken 1999:143)

     Although the possessive construction is by default head-final, van Klinken (1999:142) states that around 20% of the texts in her corpus have the Possessum-Possessor (Head-Modifier) order, with the help

---

[25] Van Klinken (1999) tags *-k* and *-kan* as possessive markers, and *-n* and *-r* as genitive clitics. I refer to her possessive markers as possessor markers because those markers occur on possessors, and her genitive clitics as possessum markers, for a similar reason. And I gloss possessor markers using POSS and possessum markers using GEN.

of the possessor marker -*k*. In example (41), the head *asu* 'dog' precedes the possessor *ó* '2SG', which is affixed by the possessor marker -*k*. If the possessor is a common noun, parallel to the possessor marker -*kan*, the possessor marker, -*k*, which might be derived from -*kan*, needs to land on a pronoun, resulting in the use of *nia* '3SG', which co-references with the possessor *feto* 'woman', as in (42).

(41) asu   ó-**k**
     dog   2SG-POSS
     'your (SG) dog' (adapted from van Klinken 1999:143)

(42) inan-aman   feto   nia-**k**
     mother-father   woman   3SG-POSS
     'the girl's parents' (adapted from van Klinken 1999:143)

The use of possessor and possessum markers in Tetun Fehan is summarized in Table 20.

TABLE 20. Possessive markers in Tetun Fehan

	-n	*-r*	*-k*	*-kan*
possessor marker vs. possessum marker?	possessum marker (by default)	possessum marker (when the possessum is plural)	possessor marker	possessor marker
needs a co-referential pronoun to land on?	N/A	N/A	Yes	Yes
used in what ordering?	possessor-possessum	possessor-possessum	possessum-possessor	possessor-possessum
co-occurrence with a possessor marker	Yes	No	N/A	N/A

The possessive construction in Tetun Dili is morphosyntactically more straightforward, compared to that in Tetun Fehan. Tetun Dili employs a possessive marker *nia*, which is identical with the lexical noun *nia* '3SG'. Though Tetun Dili is canonically head-initial, its possessive construction is head-final (as shown in (43)–(46)).[26] In (43), the noun phrase shows the order "Possessor-*nia*-Possessum," in which the possessed item *asu* 'dog' follows the possessive marker, while the possessor *o* '2SG' precedes *nia*.

(43) o   **nia**   asu
     2SG   GEN   dog
     'your (SG) dog'

(44) ni   **nia**   asu
     3SG   GEN   dog
     'his/her dog'

(45) ita   **nia**   hahan
     1PL.INCL   GEN   food
     'our (INCL) food'

(46) hau   **nia**   mau   **nia**   asu
     1SG   GEN   brother GEN   dog
     'my brother's dog'

---

[26] The reversed order is not attested in the corpus I have.

Tetun Dili also uses the absolute possessive pronoun *nian*, like 'mine' and 'hers' in English (as in (47) and (48)).

(47) asu ne'e nia **nian**
     dog this 3SG absolute.possessive
     'This dog is his.'

(48) hahan ne'e ita **nian**
     food this 1PL.INCL absolute.possessive
     'This food is ours (INCL).'

Moreover, the possessum marker *-n* in Tetun Fehan also occurs in cases where no possessive relation is involved. This is the so-called inalienable possession, which includes items like body parts (as in *ulun* 'head' and *kabun* 'stomach'), spatial relations (as in *laran* 'interior' and *leten* 'top'), kinship terms (as in *oan* 'child' and *aman* 'father'), and close and significant belongings (as in *rain* 'land' and *loron* 'sun; day'). This inalienable possession is also observed in Tetun Dili, but not in Portuguese. A list of inalienable possession in Tetun Dili is given in Table 21. This again shows that Tetun Dili is morphologically closer to Tetun Fehan, which is classified as an Austronesian language.

TABLE 21. The *-n* suffix in Tetun Dili

Tetun Dili Form	PMP	Gloss	Semantic categories
rain, rai	*daReq	'land'	
ai fuan	*buaq	'fruit'	personal belongings
kutun	*kutu	'head louse'	
ikun	*ikuR	'tail'	
ulun	*qulu	'head'	
fulun	*bulu	'body hair'	
matan	*mata	'eye'	
teen	*taqi	'feces'	
liman	*lima	'hand'	
susun	*susu	'female breast'	body parts
ran	*daRaq	'blood'	
aten	*qatay	'liver'	
sikun	*siku	'elbow; joint'	
kukun	*kuku	'nail'	
ruin	*duRi	'bone'	
ain	*qaqay	'leg'	
susu ben	*wahiR	'breast milk'	
alin	*huaji	'younger sibling'	
fen	*bahi	'wife'	family members
laen	*laki	'husband'	
rihun	*Ribu	'thousand'	
mean	*ma-iRaq	'red'	
mutin	*ma-putiq	'white'	others
maran	*maja	'dry, dry up'	
tuan	*tuqah	'old (things)'	
foun	*baqeRu	'new'	

The possessive construction in Portuguese can be either head-initial or head-final, depending on whether the possessors are nouns or pronouns. In (49) and (50), both the possessors and the possessums are common nouns, so the possessums *receita* 'recipe' and *marido* 'husband' are followed by the

possessors *vovó* 'grandma' and *Carol* 'Carol', respectively, rendering in the order "Possessum-*da*-Possessor," which is head-initial.

(49) receita **da** vovó
     recipe from.ART.F grandma
     'Grandma's recipe'

(50) marido **da** Carol
     husband GEN Carol
     'Carol's husband'

When the possessors are pronouns, the possessor *meu* 'my' is followed by the possessum *livro* 'book' as in (51). The word order is "Possessor-Possessum," which is head-final.

(51) Aquele é **meu** livro
     that 3SG.COP 1SG.POSS book
     'That is my book.'

Portuguese, like English and Tetun Dili, also employs absolute possessive pronouns, as illustrated by *deles* 'theirs' in (52).

(52) Aquela casa é **deles**
     that house 3SG.COP from.3PL
     'That house is theirs.'

Because the word order and the markers in the possessive construction in Tetun Dili and Portuguese are so divergent, it is not possible to claim that the difference between Tetun Dili and Tetun Fehan with regard to the possessive construction is a result of influence from Portuguese grammar. If Portuguese had influenced Tetun Dili, it would be expected that Tetun Dili align with Portuguese in either word order or possessive construction, neither of which is attested. Moreover, it is noteworthy that according to Hull (2003:12), the possessive construction in Southern Mambai strikingly resembles that found in Tetun Dili (as shown in (53)–(55)), which implies that the less complicated possessive construction in Tetun Dili might be attributed to the influence of Mambai instead.[27]

(53) hin **ni** um
     woman GEN house
     'the woman's house'

(54) o ama **ni** kud
     2SG.POSS father GEN horse
     'your (SG) father's horse'

(55) arbau **ni** oe
     buffalo GEN foot
     'the buffalo's foot'

4.5 RELATIVE CLAUSES. Languages of the world differ in strategies of relativization and the position that can be relativized. Keenan and Comrie (1977) proposed an Accessibility Hierarchy for relativization, in which they claim that (1) if there is only one position that can be relativized in a language, that position must be the subject, and (2) if a lower position (towards the right end of this continuum) permits relativization using the gap strategy, all the positions to the left must employ the gap strategy for relativization as well (see (56)). The relativization in Tetun Fehan, Tetun Dili, and Portuguese will be discussed in this section based on the given hierarchy.

---

[27] Traube (1980:292) states that Dili was once a Mambai-speaking region.

(56) The Accessibility Hierarchy for Relativization
subject > direct object > indirect object > oblique > genitive > object of comparison

Tetun Fehan employs the relativizer *ma'ak/mak* to form relative clauses. This relativizer is used to introduce the clause that modifies the relativized noun. In general, the gap strategy is used in the relativization of subjects (as in (57)) and direct objects (as in (58)). Example (57) shows that the relativizer *ma'ak/mak* follows the antecedent *kabau rua* 'two buffalos', which serves as the subject of the verb *lakon* 'disappear'. By contrast, the relativizer *ma'ak/mak* follows the antecedent *sá-sá* 'what', which is the direct object of the verb *fola-folan* 'to swallow something completely' in the relative clause in (58).

- **Relativization of a subject**

(57) kabau rua [ma'ak ____ lakon] ne'e
buffalo two REL disappear this
'the two buffalo which disappeared' (adapted from van Klinken 1999:137)

- **Relativization of a direct object**

(58) sá-sá [ma'ak nia fola-folan _____] ne'e sia
RDP-what REL 3SG RDP-swallow.whole this PL
'the various things that she had swallowed' (adapted from van Klinken 1999:318)

Van Klinken (1999:319) also points out that it is uncommon for relativization to take place in grammatical relations other than subjects and direct objects. In order to relativize a locative, a resumptive pronoun *ne'e* 'this' must be employed, as in (59).[28] The gap strategy is not applicable in this situation.

- **Relativization of a locative**

(59) … sia boi n-ola tasi ida [mak sia atu n-akdiuk iha **ne'e**]
3PL choose 3SG-take sea one REL 3PL about.to 3SG-play LOC this
'[Then] they chose a (part of the) sea in which they would play' (adapted from van Klinken 1999:319).

On the other hand, Tetun Dili employs *nebe'e* 'who, which' as its relativizer to introduce a clause that is used to modify the noun. Relative clauses in Tetun Dili follow the noun they modify, i.e., these relative clauses are post-nominal (N-REL), as expected from the V-NP order found in Tetun Dili. Take (60) for instance. *Feto* 'woman' is the noun that the relative clause 'whom I met yesterday' modifies; hence the former is followed by the latter.

(60) feto [nebe'e hau hasoru ___ horseik] mak Eka
woman REL 1SG meet yesterday FOC Eka
'The girl whom I met ___ yesterday is Eka.'

The relativizer *nebe'e* can be used to modify nouns that are [+animate] or [-animate]. The relativized position can be a subject, as in (61) and (62), a direct object as in (63) and (64), an indirect object, as in (65) or an oblique, as in (66).

- **Relativization of a subject**

(61) mane [nebe'e ____ haksolok] haré hau
man REL happy see 1SG
'The man who is happy sees me.'

(62) hau haré mane ida [nebe'e _____ haksolok]
1SG see man one REL happy

---

[28] The resumptive pronouns used here are referred to as pronoun retention by typologists.

'I see the man who is happy.'

- **Relativization of a direct object**

(63) hau   haré   feto   ida [nebe'e o   gosta __]   ohin   dade'er
     1SG   see   woman one REL   2SG like   today   morning
     'I saw the girl whom you (SG) like this morning.'

(64) libru   [nebe'e Maria fo'o   ___   ba Leo]   todan
     book   REL   Maria give   to Leo   heavy
     'The book that Maria gave to Leo is heavy.'

- **Relativization of an indirect object**

(65) mane   [nebe'e hau   fo'o   libru ba ____]   a'as
     man   REL   1SG   give   book   to   tall
     'The man whom I give a book to is tall.'

- **Relativization of an oblique**

(66) feto   ida [nebe'e o   mai   ho ____]   furak   lo'os
     woman one REL   2SG   come   with   beautiful very
     'The girl whom you came with is very pretty.'

In addition, relativization also occurs with a genitive, as in (67), and with the object of comparison, as in (68). We thus conclude that Tetun Dili allows relativization of all positions using the gap strategy, where an empty position (gap) is co-indexed with the head with the help of the relativizer *nebe'e*.

- **Relativization of a possessor**

(67) mane   [nebe'e ___   nia   naran John]   gosta hau   nia   alin   feto
     man   REL   GEN   name John   like   1SG   GEN   younger sister
     'The man whose name is John likes my younger sister.'

- **Relativization of an object of comparison**

(68) labarik feto   [nebe'e hau   bo'ot liu ____]   iha   busa
     child   female REL   1SG   big   than   have   cat
     'The girl who I am older than has a cat.'

As for Portuguese, it canonically employs *que*, though with some variants, as its relativizer to form relative clauses in which the head noun is followed by the clause that modifies it. Examples (69)–(73) show that Portuguese uses the gap strategy to relativize subjects, direct objects, indirect objects, and obliques. For instance, in (69) the subject *garota* 'girl' is followed by the relative clause, which is introduced by the relativizer *que*. In (70), the antecedent *garota* 'girl' is an indirect object of *gosta* 'like'. The gap strategy is also used in this relative clause.

- **Relativization of a subject**

(69) A   garota [que ___   gosta   do   Alex]   é   pequena
     DEF.F   girl   REL   3SG.like from.ART.M   Alex   3SG.COP   short.F
     'The girl who likes Alex is short.'

- **Relativization of a direct object**

(70) Ontem,   eu   vi   a   garota [que   Alex   gosta   ___]
     yesterday 1SG.NOM   1SG.see.PST DEF.F   girl   REL   Alex   3SG.like
     'Yesterday I saw the girl whom Alex likes.'

(71) O   livro   [que Maria deu   ___   ao   Leo] é   pesado
      DEF.M book   REL Maria 3SG.give.PST   to.ART.M Leo 3SG.COP heavy.M
      'The book that Maria gave to Leo is heavy.'

- **Relativization of an indirect object**

(72) O   homem [a   ___   quem dou   meu   livro] é   alto
      DEF.M man   to   REL   1SG.give 1SG.POSS book 3SG.COP tall.M
      'The man whom I give my book to is tall.'

- **Relativization of a locative**

(73) A   cidade [que   estou   morando   ___ por três anos] é
      DEF.F city   REL   1SG.COP live.PROG   for 3   years 3SG.COP
      o   Honolulu
      DEF.M Honolulu
      'The city in which I have lived for 3 years is Honolulu.'

However, in cases where relativization takes place in an oblique, a resumptive pronoun must be employed in the relative clause, as illustrated in (74), where *ela* '3SG.ACC.F' agrees with the sentential subject *garota* 'girl' in person, number, and gender.

- **Relativization of an oblique**

(74) A   garota [que   você   chegou   com   **ela**]
      DEF.M girl   REL   2SG.NOM 2SG.arrive.PST with   3SG.ACC.F
      é   muito   bonita
      3SG.COP   very   pretty.F
      'The girl whom you came with is very pretty.' (literally, 'the girl that you arrived with her is very pretty')

In sum, Tetun Fehan, Tetun Dili, and Portuguese employ different relativizers in forming relative clauses. While Tetun Dili employs *nebe'e* to be the relativizer, Tetun Fehan uses *mak* instead. Portuguese, on the other hand, employs *que* as its relativizer. The gap strategy is used to relativize all the positions in Tetun Dili. However, Tetun Fehan seeks the help of a resumptive pronoun in order to relativize a locative, whereas the gap strategy is used to relativize a subject and an object. By contrast, Portuguese uses the gap strategy to relativize grammatical relations higher than an oblique in the Accessibility Hierarchy proposed by Keenan and Comrie (1977), but the pronoun strategy is used when an oblique is relativized. In this case, Tetun Fehan, compared to Tetun Dili, is more similar to Portuguese in employing two strategies for relativization.

4.6 CONDITIONAL CLAUSES. Conditional clauses in Tetun Fehan make use of the conditional marker *kalo* 'if', which is borrowed from Indonesian *kalau* 'if'. Although the Tetun Fehan speakers are aware of the origin of *kalo*, the use of *kalo* has become nativized in Tetun Fehan (van Klinken 1999:312). As demonstrated in (75), the subordinate clause containing *kalo* 'if' canonically precedes the matrix clause 'we get sleepy' to introduce the hypothetical situation 'If we don't chew betel and don't chew tobacco'.

(75) [**kalo** la   mama,   la   món],   matan dukur
      if   NEG   chew (betel) NEG   chew.tobacco eye   sleep
      'If (we) don't chew betel (and) don't chew tobacco, (we) get sleepy' (adapted from van Klinken 1999:312).

(76) [**kalo** ha'u feto] ha'u la bele k-akur tasi wé-n
     if 1SG woman 1SG NEG can 1SG-cross sea water-GEN
     mai buka ó
     come seek 2SG
     'If I were a woman, I wouldn't be able to cross the sea to come and look for you (SG)'
     (adapted from van Klinken 1999:312).

Tetun Dili, instead, uses *se* 'if' and *karik* 'if; maybe' to construct conditional clauses. The word *se* 'if' indicates that the occurrence of the event in the matrix clause is dependent on the hypothetical situation depicted in the subordinate clause. In (77) 'I won't go to school' is entirely dependent on 'I am sick', presupposed in the subordinate clause.

(77) [**se** aban hau moras] hau sei la ba eskola
     if tomorrow 1SG sick 1SG will NEG go school
     'If I get sick tomorrow, I won't go to school.'

The word *karik* 'maybe; if' must co-occur with *se*, which also means 'if', within the same clauses as in (78)–(79). *Se*, by itself, can occur either at the beginning or at the end of a subordinate clause, as in (77) and (80). However, when *se* occurs at the end of a subordinate clause, *karik* cannot occur with it, as illustrated in (80).

(78) [**se karik** hau halo hau nia T.P.C.], hau sei hetan valor diak
     if if 1SG do 1SG GEN homework 1SG will get grades good
     'If I had done my homework, I would have gotten good grades.'

(79) [**se** hau halo hau nia depese (**karik**)], hau sei hetan valor diak
     if 1SG do 1SG GEN homework if 1SG will get grades good
     'If I had done my homework, I would have gotten good grades.'

(80) [hau sei la-moras **se**], hau han tiha ona
     1SG will NEG-sick if 1SG eat already
     'If I would have eaten, I would not have been sick.'

Interestingly, the conditional word in Portuguese is also *se* 'if' (Perini 2002:193). It is very likely that the form *se* used in conditional clauses in Tetun Dili is borrowed from Portuguese. Examples (81) and (82) show that *se* 'if' is used in Portuguese subordinate clauses to introduce a hypothetical situation; this is identical to the function of *kalo* 'if' in Tetun Fehan and *se* 'if' in Tetun Dili. However, Portuguese *se* 'if,' unlike *se* 'if' in Tetun Dili, can occur only at the beginning of a subordinate clause, cf. (80). In this regard, the relative position of Portuguese conditional marker resembles its counterpart in Tetun Fehan. By contrast, it is grammatical to place *se* 'if' at the end of a subordinate clause in Tetun Dili. Although *se* 'if' in Tetun Dili might be a loan from Portuguese, it follows the Tetun convention. It is worth mentioning that the borrowing of the function words for 'if' is also observed in Swahili, Hausa, Thai, Bezhta, Ceq Wong, and Q'eqchi' (Haspelmath and Tadmor 2009), none of which is a creole.

(81) [**se** você pedir], ele certamente vai te ajudar
     if 2SG ask 3SG.M certainly will 2SG help
     'If you ask (him), he will certainly help you' (adapted from Perini 2002:194).

(82) [**se** Marcos fez isso], errou
     if Marcos did that erred
     'If Marcos did that, he was wrong' (adapted from Perini 2002:194).

5. TETUN DILI UNDER OTHER VIEWS OF CREOLE FORMATION. I have demonstrated that Tetun Dili does not fit into the category of creoles when examined using the traditional view of creole formation. This section further discusses the linguistic status of Tetun Dili using three creole formation theories, i.e., the universalist approach, the substratist approach, and the superstratist approach.

Derek Bickerton's Language Bioprogram Hypothesis (LBH), which is based on the case of Hawaiʻi Creole English, is one of the universalist approaches. LBH suggests that the structural similarities of creoles can be explained using first-language acquisition (Bickerton 1981, 1984). In general, pidgins arose abruptly as a way to communicate on plantations, where adult laborers and slaves speaking different languages had no need or desire to learn each other's language. Given this circumstance, children acquiring this pidgin as their first language had to depend on the influence of universal grammar to develop the pidgin into a full-fledged language, i.e., a creole. Bickerton (1981) presents a list of characteristics shared by creoles, including the Subject-Verb-Object (SVO) word order, relativization with no overt relativizer, interrogative sentences without question particles, etc. Tetun Dili will be evaluated using the criteria Bickerton proposed to see if the structure of Tetun Dili fits into the universalist's view.

The word order of Tetun Dili is SVO, which seems to fit Bickerton's claim that creoles have SVO word order. However, both Tetun Fehan and Portuguese are SVO languages. If Tetun Dili is a creole, the word order of Tetun Dili can be directly attributed to its source languages instead of universal grammar. Moreover, 488 out of 1377 languages surveyed by Dryer (2013) show the SVO word order. It is hence questionable whether SVO word order can be a criterion to classify languages as creoles. Furthermore, contrary to Bickerton's proposal on universal traits of creoles, Tetun Dili does employ an overt relativizer *nebe'e* to form relative clauses, as in (83) (also see section 4.5) and a question particle *ka* to form yes-no questions, as in (84) and (85). In this case, Tetun Dili does not conform to Bickerton's creole universals. That is to say, Tetun Dili cannot be viewed as a creole under the framework of the universalist approach.

(83) mane [**nebe'e** haksolok] haré hau
      man REL happy see 1SG
      'The man who is happy sees me.'

(84) o moras **ka**
      2SG sick Q
      'Are you sick?'

(85) o bele nani **ka**
      2SG be.able.to swim Q
      'Can you swim?'

      lo'os, hau bele nani
      yes 1SG be.able.to swim
      'Yes, I can swim.'

Proponents of the substratist approach include Stewart (1962), Muysken (1981), Lefebvre (1998), and Siegel (2000), among others, with a focus on relexification and substrate transfer. In short, they define relexification as a process where the lexicon of the substrate language is replaced by the "phonetic strings" from the lexifier language, although the grammatical structure of the substrate language remains unchanged. Substrate transfer, by contrast, emphasizes the effect of L1 onto L2. As Lee (2014:366) points out, the substratist approach differs fundamentally from Bickerton's universalist approach because the substratist approach focuses on "the influence of the creator's first language (L1) or languages" whereas in the latter "the role of creole formation is left to children, and influence from the first language is overtly not taken into account."

Where Tetun Dili is concerned, based on Grimes et al. 1997, the lexifier language is presumably Portuguese and the substrate language is supposed to be Tetun. Structural features of Tetun Dili shared exclusively between Tetun Dili and Tetun Fehan include using causative prefixes *ha-/hak-*, as in (86) (also see section 4.3), preserving the inalienable possessive suffix *-n*, as in *ulu**n*** 'head', and employing of the gap strategy in certain relative constructions, as in (87) (also see section 4.5). However, since no Portuguese lexical item is observed on the abovementioned constructions (causatives, possessives, and

relative clauses), we cannot find any L1 effect on L2 to claim that certain structures can be attributed to substrate transfer.

(86) hau   **ha**monu   nia
      1SG   CAUS.fall   3SG
      'I made it fall.'

(87) hau   haré   mane   ida   [**nebe'e**   haksolok]
      1SG   see   man   one   REL   happy
      'I see the man who is happy.'

With respect to conditional clauses, when *se* is examined, the phonetic string *se* is Portuguese but the usage is still Tetun, considering that *se* in Portuguese always occurs in the beginning of the sentence (see examples (88) and (89)). Using *se* shall be regarded as a borrowing, rather than relexification, where the majority of lexical items in Tetun Dili are expected to be replaced by Portuguese loans.

(88) [hau   sei   la-moras   **se**], hau   han   tiha ona
      1SG   will   NEG-sick   if 1SG   eat   already
      'If I would have eaten, I would not have been sick.'

(89) [**se** Marcos fez isso],   errou
      if   Marcos did that   erred
      'If Marcos did that, he was wrong.' (Perini 2002:194)

Given that substrate transfer and relexification are not observed in Tetun Dili, it is not convincing to conclude that Tetun Dili is a creole language. Moreover, the substratist's view cannot account for resultant structures in Tetun Dili, which lack subject agreement and do not employ the pronoun strategy in certain relative constructions, unlike Tetun Fehan.

The third theory of creole formation presented here is the superstratist approach, in which Chaudenson (1974), Mufwene (1996, 2001, 2008), and DeGraff (2001, 2003), among others, hold the view that creoles should be regarded as restructured varieties of their lexifier languages which could often be regional dialects of the standard languages. They think that creoles can be classified as having a single ancestor, whereas the traditional view considers creole languages to have multiple parents. They also propose that pidgins and creoles arise from different circumstances. In particular, Mufwene's founder-effect principle (1996, 2001, 2008) focuses on "how structural features of creoles have been predetermined to a large extent (though not exclusively!) by characteristics of the vernaculars spoken by the population that founded the colonies in which they developed" and how some features are "favored by new ecological conditions in the colony" (2008:29). That is to say, the superstratists value sociohistorical background more than the other two approaches.

Under the framework of the founder-effect principle, the founder population of Tetun Dili should be the Tetun-speaking people, given that Tetun was already the local lingua franca before the Portuguese colonizers arrived. Thus, the structural features of Tetun Dili were predetermined by the Austronesian speakers of Tetun, rather than the Indo-European speakers of Portuguese. However, because written records are lacking, it is not clear which Tetun variety was the dominant one when Tetun Dili started to develop. Even though the superstratists emphasize the importance of the founder-effect principle and the ecological conditions in the colony, questions, such as why certain linguistic features are favored by its speakers, remain to be solved. For instance, why is subject agreement absent in Tetun Dili whereas the causative prefixes *ha-* and *hak-* are preserved? Consequently, the superstratist approach is not helpful in deciding whether Tetun Dili is a creole. Besides, it is still debatable whether a creole can be regarded as genetically related to its lexifier language (Campbell 2013:315).

6. THE STATUS OF TETUN DILI: CONCLUSION. Based on the linguistic facts and socio-historical factors, I argue that it is not possible to consider Tetun Dili to be a creolized Tetun variety. First of all, Tetun Dili has never been reported to have undergone the pidgin stage. It would also not have undergone "abrupt

creolization," considering that Tetun had functioned as a local lingua franca before the arrival of the Portuguese. Second, that the basic vocabulary of Tetun Dili is primarily of Tetun origin could be construed as evidence for the Founder Effect. Yet, this also contradicts another notion in creolistics, which is the observation that the basic vocabulary of Tetun Dili is primarily of Tetun origin, which is contrary to the traditional view that the socio-economically dominant languages supply the vocabulary in forming pidgins and creoles. Third, there is no evidence of substrate influence. The words of Portuguese origin in Tetun Dili appear to have entered the language through borrowing, since most of them refer to items or ideas for which there were no indigenous words. Substrate transfer or relexification is not attested, given that borrowing in Tetun Dili usually took place when new items or ideas were introduced. This is akin borrowing in most languages, when there exist no indigenous words for these newly-introduced concepts.

With regard to the phonological system, Tetun Dili and Tetun Fehan share almost identical consonant inventories (exclusive of the glottal stop) if the loaned phonemes are excluded. The phonological system of Tetun Dili thus should not be regarded as a corrupted version of Portuguese, or Tetun from imperfect learning. In terms of word order, Tetun Dili resembles Tetun Fehan in being a head-initial, non-pro-drop language with a fixed word order. For the conditional clauses, the words for 'if' in Tetun Fehan and Tetun Dili are both loans. The borrowing of this function word cannot serve as evidence for classifying Tetun Dili as a creolized Tetun.

Regarding morphology, though Tetun Dili does not employ subject agreement, van Klinken (1999) has specified that subject agreement shows dialectal variations in other Tetun varieties. The absence of subject agreement in Tetun Dili could be regarded as being on one end of this continuum of subject marking, instead of a simplified/creolized Tetun version due to imperfect learning after language contact, as Grimes et al. (1997) have implied. Also, *ha-* and *hak-* morphemes are found only in Tetun Dili and Tetun Fehan, and morphological causatives in Tetun Dili usually have periphrastic counterparts (cf. Grimes et al. (1997); for a longer list, see Appendix 1).

To conclude, neither the social background of language development nor the language structure of Tetun Dili supports the claim that Tetun Dili is a Tetun-based creole/creolized Tetun. Instead, Tetun Dili should be regarded as an Austronesian language with many Portuguese loans.

APPENDIX 1. A LIST OF CAUSATIVIZED VERBS IN TETUN DILI[29]

ha-	halo V	gloss	Base
habokon	halo bokon	'to wet'	bokon
habosu	halo bosu	'to make oneself full'	bosu
hadomi	--	'to love'	domin
hafahe	(halo fahe)	'to divide'	fahe
hafo'er	halo fo'er	'to make something dirty'	fo'er
hahoris	--	'to make alive; to give birth to'	moris
hakarak	--	'to want'	karak
hafera	halo fera	'to break into pieces'	*fera
hakfilak	halo filak	'to cause something to transform'	(nakfilak)
hakfodak	halo fodak	'to surprise'	*fodak
hakonu	--	'to make something full'	(nakonu)
hakotu	halo kotu	'to break'	kotu
hamamuk	halo mamuk	'to empty'	mamuk
hamanas	halo manas	'to heat'	manas
hamaran	halo maran	'to dry'	maran
hamate	--	'to turn off'	mate
hamo'os	halo mo'os	'to clean'	mo'os
hamoris	--	'to turn on'	moris
hamonu	halo monu	'to cause something to fall'	monu
hatauk	halo tauk	'to make someone afraid'	tauk
hatoba	halo toba	'to make lay down'	toba
hakna'ar	--	'to make someone busy'	kna'ar

---

[29] Adapted from Chen 2014a.

APPENDIX 2. A LIST OF GLOSSING ABBREVIATIONS

1	first person	LOC	locative
2	second person	M	masculine
3	third person	NEG	negation
ACC	accusative	NOM	nominative
ART	article	PL	plural
CAUS	causative	POSS	possessive
COP	copula	PROG	progressive
DEF	definite	PST	past
EXCL	exclusive	Q	question
F	feminine	RDP	reduplication
GEN	genitive	REL	relative
INCL	inclusive	SG	singular
INDF	indefinite		

REFERENCES

BAUGH, ALBERT C., and THOMAS CABLE. 2001. *A history of the English language*, 5th ed. Englewood Cliffs, NJ: Prentice Hall International.

BAXTER, ALAN NORMAN. 1990. Notes on the Creole Portuguese of Bidau, East Timor. *Journal of Pidgin and Creole Languages* 5(1):1–38.

BLUST, ROBERT. 2003. Three notes on early Austronesian morphology. *Oceanic* Linguistics 42(2):438–78.

BLUST, ROBERT. 2013. *The Austronesian languages* (revised edition). Canberra: Pacific Linguistics.

BLUST, ROBERT, and STEPHEN TRUSSEL. Ongoing. *Austronesian comparative dictionary* (web version). Retrieved from: http://www.trussel2.com/ACD/.

BICKERTON, DEREK. 1981. *Roots of language*. Ann Arbor: Karoma.

BICKERTON, DEREK. 1984. The language bioprogram hypothesis. *Behavioral and Brain Sciences* 7:173–221.

CAMPBELL, LYLE. 2013. *Historical linguistics: An introduction*. 3rd ed. Cambridge, MA: The MIT Press.

CHAUDENSON, ROBERT. 1974. *Le lexique du parler créole de la Réunion*. Paris: Champion.

CHEN, YEN-LING. 2014a. A grammatical sketch of Tetun Dili. Unpublished MS, Honolulu.

CHEN, YEN-LING. 2014b. Field notes on Tetun Dili. Unpublished MS, Honolulu.

CROWLEY, TERRY. 2004. *Bislama reference grammar*. Oceanic Linguistics Special Publication No. 31. Honolulu: University of Hawai'i Press.

CRUZ-FERREIRA, MADALENA. 1995. European Portuguese. *Journal of the International Phonetic Association* 25 (2):90–4.

DEGRAFF, MICHAEL. 2001. On the origin of Creoles: A Cartesian critique of Neo-Darwinian linguistics. *Linguistic Typology* 5:213–310.

DEGRAFF, MICHAEL. 2003. Against Creole exceptionalism (discussion note). *Language* 79:391–410.

DUTTON, THOMAS. E. 2012. *A new course in Tok Pisin (New Guinea Pidgin)*. 2nd ed. Canberra: Pacific Linguistics.

KOSSMANN, MAARTEN. 2009. Tarifiyt Berber vocabulary. In *World loanword database*, ed. by Martin Haspelmath and Uri Tadmor. Leipzig: Max Planck Institute for Evolutionary Anthropology, 1688 entries. (Available online at http://wold.clld.org/vocabulary/96, Accessed on 2/10/2015.)

FORMAN, MICHAEL L. 1993. Verb serialization, word order typology, and Zamboangueño: A comparative approach. *Oceanic Linguistics* 32(1):163–82.

FOX, JAMES J. 1977. *Harvest of the palm: Ecological change in eastern Indonesia*. Cambridge, MA: Harvard University Press.

GOOGLE MAPS. 2015. [Lifau, Oecussi-Ambeno, Timor-Leste]. Retrieved from https://www.google.com/maps/place/Lifau,+Timor-Leste/@-9.0503947,124.4906162,8.79z/data=!4m2!3m1!1s0x2c556f67c8d7451d:0x8ad5ac04c313ea72

GRANT, ANTHONY. 2009. English vocabulary. In *World loanword database*, ed. by Martin Haspelmath and Uri Tadmor. Leipzig: Max Planck Institute for Evolutionary Anthropology, 1516 entries. (Available online at http://wold.clld.org/vocabulary/13, Accessed on 2/10/2015.)

GRIMES, CHARLES E.; TOM THERIK, BARBARA D. GRIMES; and JACOB MAX. 1997. *A guide to the people and languages of Nusa Tenggara*. Kupang: Artha Wacana Press.

HASPELMATH, MARTIN, and URI TADMOR (eds.) 2009. *World Loanword Database*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wold.clld.org, Accessed on 2/10/2015.)

HICKS, DAVID. 1976. *Tetum ghosts and kin: Fieldwork in an Indonesian community*. Palo Alto, CA: Mayfield Publishing Company.

HOLM, JOHN. 1988. *Pidgins and creoles*. Vol. 2. Cambridge: Cambridge University Press.

HULL, GEOFFREY. 2003. *Southern Mambai*. Dili, East Timor: Instituto Nacional de Linguistica, Universidade Nacional Timor Lorosa'e.

KEENAN, EDWARD L., and BERNARD COMRIE. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8:63–99.

LEE, NALA H. 2014. *A grammar of Baba Malay with sociophonetic considerations*. University of Hawai'i at Mānoa PhD dissertation.

LEFEBVRE, CLAIRE. 1998. *Creole genesis and the acquisition of grammar: The case of Haitian creole*. Cambridge: Cambridge University Press.

LEWIS, M. PAUL; GARY F. SIMONS; and CHARLES D. FENNING (eds.). 2015. *Ethnologue: Languages of the World (18th ed.)*. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

DRYER, MATTHEW S. 2013. Order of Subject, Object and Verb. In *The world atlas of language structures online*, ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/81, Accessed on 12/15/2014.)

MATRAS, YARON. 2009. *Language contact*. Cambridge: Cambridge University Press.

MUFWENE, SALIKOKO S. 1996. The founder principle in creole genesis. *Diachronica* 13:83–134.

MUFWENE, SALIKOKO S. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.

MUFWENE, SALIKOKO S. 2008. *Language evolution: Contact, competition and change.* London/ New York: Continuum International Publishing Group.

MUYSKEN, PIETER. 1981. Halfway through Quechua and Spanish: The case for relexification. In *Historicity and variation in creole studies*, ed. by Arnold R. Highfield and Albert Valdman, 52–78. Ann Arbor, MI: Karoma.

PERINI, MÁRIO A. 2002. *Modern Portuguese: A reference grammar*. New Haven and London: Yale University Press.

SCHMIDT, CHRISTOPHER K. 2009. Japanese vocabulary. In *World loanword database*, ed. by Martin Haspelmath and Uri Tadmor. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2131 entries. (Available online at http://wold.clld.org/vocabulary/21, Accessed on 2/10/2015.)

SIEGEL, JEFF. 2000. Substrate influence in Hawai'i Creole English. *Language in Society* 29:197–236.

STEWART, WILLIAM A. 1962. An outline of linguistic typology for describing multilingualism. In *Study of the role of second language in Asia Africa and Latin America*, ed. by Frank A. Rice, 15–26. Washington: Center for Applied Linguistics.

THOMASON, SARAH G., and TERRENCE KAUFMAN. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley, CA: University of California Press.

TRAUBE, ELIZABETH. 1980. Mambai rituals of black and white. In *The flow of life*, ed. by James J. Fox, 290–316. Cambridge, MA: Harvard University Press.

VAN KLINKEN, CATHARINA. 1999. *A grammar of the Fehan dialect of Tetun: An Austronesian language of West Timor*. Canberra: Pacific Linguistics.

WILLIAMS-VAN KLINKEN, CATHARINA; JOHN HAJEK; and RACHEL NORDLINGER. 2002. *Tetun Dili: A grammar of an East Timorese language*. Canberra: Pacific Linguistics.

WILLIAMS-VAN KLINKEN, CATHARINA. 2008. *Word-finder: English-Tetun, Tetun-Ingles*. Dili: Dili Institute of Technology.

WINFORD, DONALD. 2003. *An introduction to contact linguistics*. Oxford: Blackwell.

ZEITOUN, ELIZABETH, and LILLIAN M. HUANG. 2000. Concerning *ka-*, an overlooked marker of verbal derivation in Formosan languages. *Oceanic Linguistics* 39 (2):391–414.

yenling@hawaii.edu