

Mixing numeral systems in Timor-Leste

Catharina WILLIAMS-VAN KLINKEN^a and John HAJEK^b

Dili Institute of Technology^a, University of Melbourne^b

Speakers of Tetun Dili, in Timor-Leste, commonly use three sets of numerals, namely native Tetun as well as numerals from the languages of the two former colonial powers, Portuguese and Indonesian. Tetun numerals can modify nouns from any language, within native Tetun constructions. In contrast, borrowed numerals modify only nouns from the same source language, using source language order. Functional and sociolinguistic differences were sought based on an extensive corpus of texts as well as two surveys of numeral use. The choice of numeral language is complex. Contributing factors include the language of the head noun for attributive adjectives, the size of the number, the topic, and sociolinguistic factors such as formality, lectal type, and language of education of the speaker and of the preceding generation. Tetun numerals are primarily used for small numbers within traditional contexts, while Indonesian and Portuguese vie for position for larger numbers and modern contexts.

1. Introduction¹

In the extensive and expanding literature on language contact and borrowing, the borrowing of numerals and numeral systems has often been noted (e.g. Matras 2007, 2011, Klamer & Kratochvíl 2014b). However, relatively little detail has been provided about it or about the grammatical consequences of such borrowing and the social factors linked to it. Where borrowing of numerals is reported, it is usually restricted to borrowing from *one* other language, e.g. Chinese numerals adopted in Thai (Matras 2011). In this context Tetun Dili, one of the two official languages of Timor-Leste (also known as East Timor), is typologically unusual in that it has three numeral systems, namely native Tetun in addition to borrowed Portuguese and Indonesian systems. Many people use all three, readily using numerals from two languages within the one sentence, as illustrated by (1) and (2).²

- (1) *Kalsa tahan ida deit, lima puluh dolar.*
pants.P sheet.T 1.T just 5.I tens.I dollar
'Just one pair of pants costs \$50.' (Young man in conversation, 2011)
- (2) *Ami tama tuku walu, sai dozi i meia.*
1PL.EXCL enter o'clock 8.T exit 12.P and.P half.P
'We start (school) at 8 o'clock, and finish at 12.30.' (Student in survey, 2012)

In this study, we explore the nature and behaviour of the three numeral systems in Tetun Dili, in order to understand how each interacts with the grammar of the language, and what factors influence the choice of language for numerals. Given the unusual sociolinguistic and socio-political situation of Timor-Leste, which has undergone significant change at least twice since the 1970s, we are also interested in exploring the

¹ Much earlier versions of parts of this paper were presented as Williams-van Klinken (2012) and published in much less detailed form in Tetun as Ximenes & Williams-van Klinken (2014).

² In the examples, T indicates Tetun, P Portuguese, and I Indonesian. Spelling of Portuguese numerals follows Tetun conventions.

relationship between societal change and the use of different numeral systems over time and across the generations.

The following sections present the historical background of Tetun Dili, a review of patterns of borrowing numerals cross-linguistically, and an overview of the structure of native and borrowed cardinal numerals within Tetun Dili, as well as ordinals and numeral classifiers.

We then turn in subsequent sections to three overarching questions about the uses of the three sets of numerals found in Tetun Dili.

Firstly, how do Tetun, Indonesian and Portuguese numerals fit into the structure of Tetun Dili? What are the grammatical differences between them? We will note that while Tetun numerals can be used with head nouns from any language, and are used in the full range of native Tetun morphological and grammatical structures, Indonesian and Portuguese numerals, when used attributively, can only modify nouns from their own language, following most of the phrase-level grammar of their own language. These borrowed numerals are also much more likely to stand alone or fulfil peripheral functions within a clause than to take part in subjects or objects.

A second question is this: What are the functional differences between numerals from these three languages? The results, based primarily on an extensive database of texts and two surveys, show that the factors affecting choice of language are complex, including not only the language of the head noun, but also the size of the number, the topic, and sociolinguistic factors such as age, formality, lectal type and the language of education of the speaker and of the previous generation.

The third question is: What changes have there been over time in terms of the language used for numerals? The results over the last twenty years indicate a stable role for Tetun numerals in everyday use but significant increase in its function in formal situations, while changes in the relative positions of Portuguese and Indonesian are not consistent.

The overall result is the most detailed study of language choice for numerals that we are aware of to date.

2. Historical background to Tetun Dili

The presence of three parallel numeral systems in Tetun Dili is a consequence of the specific colonial and linguistic history of Timor-Leste. Tetun Dili is the major lingua franca of the country and a co-official language alongside Portuguese. It has its roots in the Austronesian vernacular Tetun Terik, which was spoken in two ancient kingdoms, Wehali in the west and Luca in the east. Tetun Terik is just one of at least fifteen vernacular languages spoken in what is now Timor-Leste (Hull 2002, Lewis, Simons & Fennig 2018). However, as the major language of intercultural communication throughout much of the eastern half of the island of Timor for at least the last five hundred years, Tetun Dili has diverged from Tetun Terik to such an extent that mutual intelligibility is low.

Portuguese involvement with Timor started in the early 1500s, initially with trade visits, and later through a permanent colonial presence which lasted until 1975. Portugal's initial base was in Lifao, to the west of the region in which Tetun was used. However, in 1769, Portugal moved its capital to the current site of Dili, within the Tetun-speaking domain. Although the language of education and government was strictly Portuguese, the tiny new capital was soon bilingual in Tetun and Portuguese (Taylor-Leech 2007). This bilingualism presumably marked the beginning of the intensive influence of

Portuguese on Tetun Dili (see for instance Hajek (2006) and Williams-van Klinken & Hajek (2018) for more detail on language contact and its effects).

Once Tetun became one of the languages of the new capital Dili, it came under the influence not only of Portuguese, but also of Mambae, the original language of the area in which Dili is situated. Mambae is related to Tetun Terik, but even more isolating. Its influence is seen in a number of grammatical simplifications (Hajek 2006) as well as the near loss of the glottal stop and /w/ (Hull 2000). It did not result, however, in Tetun Dili adopting Mambae's quinary numerical system (Hull 2001: 14). This is not surprising given that Tetun Terik's decimal system coincides with the decimal systems of Portuguese and Malay, another lingua franca on the north coast during the early years of the Dili settlement.

In December 1975, Indonesia invaded Timor-Leste, and immediately imposed Indonesian as the language of education and government, while trying to eradicate Portuguese. By the end of their 24-year rule, many people, especially in towns, had learned Indonesian to some level. According to the 2004 census (National Statistics Directorate 2006), which was the first census taken after the Indonesian government and most of its population withdrew from the territory, 43% of East Timorese claimed a knowledge of Indonesian, rising to 71% in Dili district. Although not everyone could speak the language, it seems almost all had learned to use Indonesian numerals.

In 1999, Timor-Leste voted overwhelmingly for independence, which was achieved in 2002. Tetun and Portuguese were selected as co-official languages, and were phased into the education system replacing Indonesian, starting immediately with the first grade of primary school. Nowadays, after a period of transition, Indonesian is still used by many tertiary education institutions, and Indonesian shows (including children's cartoons) are popular on satellite TV, but outside of these domains Indonesian is little used, and the general level of proficiency in the language has noticeably declined, especially amongst younger East Timorese. Most areas of public life, including parliament, church, and the media, use Tetun, while the government uses both official languages.

Despite the constitutional status of Portuguese in Timor-Leste, it is not at present widely spoken. According to the 2004 census (National Statistics Directorate 2006), nationwide only 14% claimed to speak it, rising to 19% in Dili. According to the 2015 census, only 0.1% of the population spoke Portuguese as their mother tongue (Direcção Geral de Estatística Timor-Leste 2016), while 31% claimed to be literate in it (Eelens 2017), the latter figure being very generous based on our experience.

One result of this chequered history is that there are large differences between the various lectal types in Tetun Dili. The Portuguese-educated elite, especially in formal situations, speaks Tetun with a large proportion of Portuguese terms, readily reaching 30%, with virtually no Indonesian loans. In the mesolect, the proportion of Portuguese is lower, and some Indonesian loans are used. In the basilect, the proportion of Portuguese is lower still, below 10%, again with some Indonesian loans. Formal Tetun is pulled in two directions. That of the press normally favours a high proportion of Portuguese, reaching to over 40% of word tokens, while the liturgical Tetun used in the Catholic church favours loans from Tetun Terik, with only about 10% of Portuguese (Williams-van Klinken & Hajek 2018: 13).

3. Borrowing of numerals as a cross-linguistic phenomenon

Borrowing of numerals is quite common cross-linguistically in language contact situations. For instance, of the 27 languages mentioned by Matras (2007: 50), over two-thirds show some borrowing of numerals, these often being used alongside native numerals. Patterns of borrowing vary widely. Some languages borrow only certain numbers but not others, such as the borrowing of the numerals six to nine from Arabic into Swahili or the numbers 70, 80 and 90 from Turkish into the Pharas dialect of Greek (Matras 2011: 213). Some languages borrow entire patterns, for instance Mosestén is reported to have taken on a decimal system as a result of language contact (Matras 2007: 50). Sometimes entire native numeral systems are replaced with borrowed systems, for instance Chamorro has replaced its entire inherited numeral system (which was still attested in the early 20th century) with Spanish numerals (Matras 2011: 213).

Higher numerals are more easily borrowed than lower ones, in part because higher numerals are used in formal contexts more often than in everyday life, with everyday life favouring lower numerals (Matras 2011: 213). Another reason is that many traditional languages do not have numbers as high as a million. ‘Zero’ too is easily borrowed, as it is a formal mathematical concept for which many languages do not have a term (Matras 2011).

Within the East Nusantara region of eastern Indonesia and Timor-Leste, Klamer & Kratochvíl (2014a: 11) note “[t]he indigenous expressions for number as discussed in this volume are currently under pressure from the major languages of interethnic trade and national education, namely Indonesian in Indonesia, and Tetun and Portuguese in East Timor.” Within this region too, patterns vary. Speakers of some languages are known to be resistant to using Indonesian numbers, e.g. Helong on Semau (Balle & Cameron 2014: 44). Some languages borrow *juta* ‘million’ but otherwise use their own system, e.g. Tetun Fehan (van Klinken 1999: 100) and Tolaki (Donohue & Edwards 2014: 29). In yet others, such as Biak (van den Heuvel 2007: 338), borrowing of numerals is common especially for numbers greater than ten.

Borrowed numerals enter many languages through the dominance of the official language in education as well as in other high-level settings. Such borrowing is more common in formal contexts such as mathematics and finance than in informal contexts such as telling one’s age (Matras 2007). When a language has two sets of numerals in use, these are often associated with different domains. In particular, numerals from the language associated with economic or institutional domains tend to be used for economic transactions. Matras (2011: 212) gives the examples of Hokkien Chinese numerals being used in Betawi only for money, Chinese numerals being used in formal contexts in Vietnamese, and Romani using numerals from the politically dominant language for dates.

However, while many linguists have commented on borrowed numerals being used for modern functions, in formal contexts, or for higher numerals and ‘zero’, few have given detailed accounts of which functions the numerals of each language are used for. One notable exception is Schachter and Otnes’ (1972: 200–14) description of numerals in Tagalog which, like Tetun Dili, has a three-way numeral system. In Tagalog this consists of native Tagalog alongside borrowed Spanish and English numerals. As is the case for Timor-Leste, the Philippines’ complex colonial history has been the driver for the tripartite system. The Philippines were for centuries under Spanish control, before passing for a time into English-speaking American control (1898–1946) and then moved to eventual independence. However, judging by Schachter and Otnes’

description, the choice of language for numerals in Tagalog is simpler and diachronically more stable than is the case in Tetun Dili. The authors mention only two drivers for the choice of language, namely topic and formality. The details are spelled out in Table 1. As can be seen, where topics are handled in more than one language, formal situations favour Tagalog, while colloquial ones favour English.

Table 1. Drivers of language choice for numerals in Tagalog

Topic	Language	Impact of formality
quantity, measurement	Tagalog	
clock time	Spanish, Tagalog	Formal Tagalog
money, percentage, age	Tagalog, Spanish	
school grade, page number	Tagalog, English	Colloquial English
numbers within addresses	Spanish, English	Colloquial English
dates	Spanish, English, Tagalog	Formal Tagalog

Many authors have commented on the reasons for borrowing numerals (e.g. Matras 2007). Firstly, some borrowing is to fill lexical gaps, usually for ‘zero’ or large numbers. Secondly, it is common throughout the world to borrow numerals from the language of education, which brings a preoccupation with numbers, including price, time and date, and measurements. Since many of the tasks for which numbers are used in the modern world were introduced from outside of traditional cultures, they are often handled in the language with which these concepts were brought in. So, for instance, Arabic numbers are used for telling the time in Berber languages (Souag 2007), and English numbers for dealing with money in the Oceanic language Mele-Fila (Clark 2004). Such borrowing is then reinforced by the education system, which teaches mathematics in official languages.

A related factor in many places is the prestige of the source language (e.g. Matras 2007).

Another reason sometimes given for borrowing is that the borrowed numerals are shorter than the native ones. For instance, Arabic ordinal numbers are shorter than Berber ordinal expressions (Souag 2007) and English numbers are shorter and so more “convenient” than Mele ones (Clark 2004).

4. Methodology and corpus

Data for this paper were collected from a number of sources, including a large database of texts, two surveys on numeral use, conversations on the topic, and direct observation.

The database was a searchable corpus of over 3.6 million words of Tetun text, from a diverse range of genres and sources. This included over 500,000 words of oral text, ranging across political debate and campaigns, news, sermons, personal interviews by the authors and assistants on a wide range of topics, through to children playing cards. There were over 2.1 million words of written texts, including emails, published reports and news articles, academic articles, how-to books, short novels, poetry and stories for children. Additionally, there were over a million words of translated texts, which included laws and regulations, a wide range of reports, Bible portions and liturgical texts. These texts were stored within Toolbox (SIL International 2011). This database was supplemented by searches of much more written and translated material in the internet.

Two surveys were conducted in the capital Dili, one in 2012 and a second in January 2018, to determine how primary school children, university students and high-school educated adults over the age of forty responded to questions in Tetun about issues such as their age, date and time, prices, and phone numbers. The questioners were all tertiary-educated East Timorese in their 20s and 30s.³ Respondents were each asked only a few questions, in Tetun Dili. At a minimum, the language in which they gave the numerals in their replies was noted down, though in many cases the actual answer was also recorded. In 2018 data was noted separately for male and female respondents, but no overall difference was observed, so gender will not be commented on further below.

The list of survey questions is included as Appendix 1. There were three questions which were asked of all three age groups in both years, namely their age, low prices, and time to start and finish school or work. An additional six questions were repeated for both university students and adults in both surveys, about their phone number, high prices, current time, and dates (today's date, birthday, and date of the restoration of independence). Other questions varied by age group, and some new ones were added to the 2018 survey.

Table 2 shows the overall number of responses collected for each age group and survey year. In practice, the number of responses per question varied significantly as questioners continued only until their page was full, or the tendencies clear. Despite this issue, the survey results allow for a general quantification of which languages were used, where the main variables were the speaker's age and the topic.

Table 2. Number of responses collected for each age group and survey year

Respondent age group	year	year	Total
	2012	2018	
Primary school children	180	355	535
University students	518	701	1219
High-school educated adults over 40	97	607	704
Total	795	1663	2458

At the time of both surveys, children had (at least in theory) done all their education in Tetun and Portuguese.⁴ In 2012, university students had done all their schooling in Indonesian, while by 2018 the younger ones had (again, in theory at least) been

³ In 2012, data was collected by Dili Institute of Technology language lecturers Abreu Ximenes, Alexandre Fernandes Cham, Anabela Maia Santos, the late Jacinta Canossa Soares, Natalino de Jesus Dias and Alcina Soares. Margarida Godinho, a lecturer from Portugal, offered to help, but it soon became apparent that respondents avoided using Indonesian in responses to her, just as they did to the first author, an Australian, when she trialled the questions. Hence data from these two sources was not included. In 2018, data collection was done by Dili Institute of Technology graduates Nona Paixeco and Margarida Maria Soares.

⁴ According to the Basic Law of Education (*Lei de Base de Educação*, 2008: Law 14/2008: article 8), "The teaching languages of the Timorese education system are Tetun and Portuguese." However in practice very few teachers master Portuguese (6% according to self-reporting in 2004 (Ross 2012:2)), and teachers are receiving mixed and changing messages from the Ministry of Education as to the role of Tetun in schools (Ross 2012:2; Taylor-Leech 2007:172–175). At lower levels of primary school, some teachers turn to local vernaculars to aid communication, while at higher levels, especially at university, much teaching is still in Indonesian, or a mixture of Tetun and Indonesian. Mathematics tends to be taught in Portuguese, with much use of rote learning, regardless of how well the teachers master this language.

educated in Tetun and Portuguese. In 2012, some of the adults over the age of forty would have done some schooling in Portuguese prior to the Indonesian invasion in 1975, but the most part would have been in Indonesian.

Where there were differences across age groups or survey years in terms of which language was used, a Chi squared test was done using the statistical program GenStat to determine the statistical significance (p value) of these differences.

Two final sources of information on language choice were conversations with many people over the years about the use of numerals in Tetun Dili,⁵ as well as our own direct observations over the last sixteen years.

5. Numerals in Tetun Dili

5.1 Cardinal numerals

Tetun Dili uses a native system taken directly from Tetun Terik, as well as loans from Indonesian and Portuguese. All three are decimal systems. The basic numerals are shown in Table 3. The question word *hira* ‘how many, how much’ is included because it is used in most of the same positions as Tetun numerals. Indonesian and Portuguese question words are not normally used in Tetun.

Table 3. Basic numerals in Tetun Dili

Tetun	Indonesian loans	Portuguese loans	English
<i>ida</i>	<i>satu</i>	<i>ún/uma</i>	‘one’
<i>rua</i>	<i>dua</i>	<i>dois/duas</i>	‘two’
<i>tolu</i>	<i>tiga</i>	<i>tréz</i>	‘three’
<i>haat</i>	<i>empat</i>	<i>kuatru</i>	‘four’
<i>lima</i>	<i>lima</i>	<i>sinku</i>	‘five’
<i>neen</i>	<i>enam</i>	<i>seis</i>	‘six’
<i>hitu</i>	<i>tujuh</i>	<i>seti</i>	‘seven’
<i>walu</i>	<i>delapan</i>	<i>oitu</i>	‘eight’
<i>sia</i>	<i>sembilan</i>	<i>novi</i>	‘nine’
<i>sanulu</i> ⁶	<i>sepuluh</i>	<i>déz</i>	‘ten’
<i>-nulu</i>	<i>puluh</i>	-	‘tens’
<i>atus</i>	<i>ratus</i>	<i>sein, sentu</i>	‘hundred’
<i>rihun</i>	<i>ribu</i>	<i>míl</i>	‘thousand’
<i>hira</i>			‘how many, how much’

As is quite common in East Nusantara (Klamer & Kratochvíl 2014a: 5) and indeed cross-linguistically (Givón 1981), *ida* ‘one’ can be used for indefinite reference, e.g., *ema ida* (lit. person one) ‘someone, a person’.

For numerals over ten, all three languages use the same basic order, mentioning billions, millions, thousands, hundreds, tens and digits in that order. A major difference between the three languages is that in Tetun, modifiers follow the head, while in Indonesian and

⁵ Abreu Ximenes co-authored an earlier, much more limited paper on the use of numerals in Tetun (Ximenes & Williams-van Klinken 2014). Others who gave particularly useful input included Dili Institute of Technology language lecturers Justino da Silva, Dulce Marques Freitas and Norberto Gonçalves.

⁶ The apparent prefix *sa-* is not synchronically used in any other derivation that we know of.

Portuguese modifiers precede the head. So for instance, ‘four thousand’ is expressed by *rihun haat* (lit. thousand four) in Tetun, but by *empat ribu* and *kuatru mil* (lit. four thousand) in Indonesian and Portuguese.

The following example (3) illustrates the similarities and differences in the structures of numerals in the three languages, using the number ‘3,562’.

(3)	Language	thousands	hundreds	-tens	digits
	Tetun	<i>rihun tolu</i> ‘thousand 3	<i>atus lima</i> hundred 5	<i>neen-nulu</i> 6-tens	<i>resin rua</i> extra 2’
	Indonesian	<i>tiga ribu</i> ‘3 thousand	<i>lima ratus</i> 5 hundred	<i>enam puluh</i> 6 tens	<i>dua</i> 2’
	Portuguese	<i>trés mil</i> ‘3 thousand	<i>kinhentas</i> 500.PL’	<i>sesenta</i> 60	<i>i dois</i> and 2’

While Tetun and Indonesian higher numerals are all regular, many Portuguese numerals are phonologically irregular, as shown by the series *sinku* ‘5’, *kinzi* ‘15’, *sinkuenta* ‘50’ and *kinhentas/as* ‘500’. Another difference is that some Portuguese numerals have masculine and feminine forms, which agree in gender with the head noun, namely *ún/uma* ‘one (M/F)’, *dois/duas* ‘two (M/F)’, and numerals based on *-sentus/as* ‘hundreds’.

5.2 Numerals at the extremes: ‘million’, ‘billion’ and ‘zero’

The number for ‘million’ is as yet unstable in Tetun Dili, with three options in use:

- (i) *tokon* is presumably a back-derivation from Soibada Tetun Terik *tokon ba tokon* ‘a massive number, zillions’. Definitions of *tokon* given in dictionaries vary widely. Hull’s (1999) dictionary of Tetun Dili (mixed with Tetun Terik) lists it as a million, as does Tilman’s (1996: 47) guide for Tetun Terik, based on the Soibada dialect.⁷ Costa’s (2000) dictionary of Soibada Tetun lists it as a hundred thousand, and Morris’ (1984) dictionary of (mainly) Soibada Tetun opts for “the next number above *reben* (possibly one hundred thousand or one million)”, while defining *reben* (a number we have never heard in Tetun Dili) firstly (for the area of Alas) as “very numerous” and secondly as “ten thousand, the number above *rihun*”. *Tokon* is not mentioned in van Klinken’s (1999) grammar of Wehali Tetun nor in her unpublished dictionary. To date, *tokon* is mainly used in writing and formal situations.

⁷ Tilman’s (1996:47) “Guide to Tetun” lists many otherwise unknown words for Tetun Terik. For ‘zero’, he lists *mamuk* ‘empty’, for ‘million’ he lists both *tokon* and *nanun*, for ‘1,000 million’ he gives *rutokon* and *runanun* (with the *ru-* presumably coming from *rua* ‘two’), and for ‘million million’ *toltokon* and *tolnanun* (with *tol-* presumably deriving from *tolu* ‘three’). The latter four words go against the phonemic rules of Tetun Terik, in each case having a vowel other than /a/ in the antepenultimate syllable, and in two cases having otherwise unknown consonant clusters starting with /l/. Many of the other words in Tilman’s book similarly break the phonemic rules of the language, pointing to neologisms on the part of the author.

We have no evidence for Tetun Dili using the numeral *be’in*, which is listed as a Tetun Terik numeral for ‘10,000’ in Mathijsen (1906:9)’s dictionary for Tetun in West Timor with the spelling *beein* and by Schapper (2010:99) who spells it *beqin*. Van Klinken (1999:100, footnote 2) notes that several knowledgeable Tetun Terik speakers from Betun identified it to her as being a large number, but were uncertain of its value. We have not noted it in Tetun Dili, except for one speaker who identified it as meaning ‘billion’, and it is not listed in the dictionaries of Morris (1984) or Hull (1999).

- (ii) *milhaun* is from Portuguese *milhão* ‘million’. When first used, some people considered it easily confused with Indonesian *milyar* ‘1,000,000,000’ and Portuguese *mil* ‘thousand’. In a numeral expressed solely with Portuguese elements, the plural *milhoens* is used for more than one unit of millions, e.g. *déz milhoens* ‘ten million’. However with a following Tetun numeral, the singular is used, e.g. *milhaun sanulu* (lit. million ten). This loan is mainly used in writing and formal situations.
- (iii) *juta* is from Indonesian, but pronounced with a soft Portuguese/Tetun initial fricative, rather than affricate as in the source, e.g. *juta rua* ‘two million’.⁸ Like other Indonesian numbers, it is mainly used in informal oral speech.

All three of these options show signs of nativization in that they can be used with postposed Tetun numeral modifiers, e.g. *tokon tolu, milhaun tolu* or *juta tolu*, all of which mean ‘three million’, thus indicating a degree of integration into the Tetun grammatical system. At the same time, Portuguese *milhaun* and Indonesian *juta* can also be modified by numerals from their source language, e.g., *trés milhoens* or *tiga juta* ‘three million’.

For ‘billion’ (thousand million), people use either Portuguese *bilhaun* (plural *bilhoens*) or Indonesian *milyar*. As with the terms for ‘million’, either one can be modified by preceding numerals from the source language (e.g., Portuguese *dois bilhoens* or Indonesian *dua milyar* ‘two billion’) or by a following Tetun numeral (e.g., *bilhaun rua, milyar rua*).

Like many languages (Matras 2007: 50), Tetun has no native word for the mathematical concept ‘zero’. This too is often borrowed from Portuguese (*zero*). In less formal situations it can also be borrowed from Indonesian, either as *nol*, or, in sequences of Indonesian numbers such as in telephone numbers or bus numbers, as *kosong* ‘empty’, e.g. *mikrolét kosong enam* ‘minibus 06’. Occasionally one hears the Tetun word *mamuk* ‘empty’ in telephone numbers, for instance in notices read out on television, presumably as a calque on Indonesian *kosong*.

5.3 Ordinal numerals

While our focus in this study is on cardinal numbers, it is worth noting that Tetun Dili, like Tetun Terik, has no native ordinal numbers. The only ordinals in common use are the Portuguese loans *primeiru* ‘first’, *segundu* ‘second’ and, to a lesser extent, *terseiru* ‘third’. Higher Portuguese ordinals are borrowed mainly in formal language, in which case they follow Portuguese rules of word order and gender agreement, e.g., *kintu kongresu* ‘fifth.M congress.M’. Indonesian ordinals are occasionally used in informal speaking. Other options introduced in recent years are to use *ba dala X* (lit. for time X), mainly for things that recur, such as conferences, e.g., *konferénsia ba dala lima* (lit. conference for time five) ‘fifth conference’. This is sometimes abbreviated to a prefix *da-*, e.g., *dalima* ‘fifth’. Since each of these options has limited currency, and the latter two have significant issues of acceptance, we will not consider them further in this article. A further option, when modifying borrowed nouns, is to use a postponed numeral from the same language, in a pattern described in section 6.2.

⁸ Very occasionally Indonesian *juta* is even mixed with a preceding Portuguese numeral and a plural suffix *-s*, particularly *déz juta-s* ‘ten million-s’, as in the following question asked by someone who wasn’t sure how much the Portuguese word *milhoens* represented: *Déz milhoens nee déz juta-s ka?* (lit. ten.P million.PL.P this ten.P million.I-PL or) ‘Does this ten million.P mean ten.P million.I?’

5.4 Numeral classifiers

Tetun Dili has a numeral classifier system, including both mensural classifiers (e.g., *bee kopu ida* (lit. water glass one) ‘one glass of water’) and sortal classifiers. The only commonly used sortal classifier is *nain* (lit. owner, lord) used when enumerating humans, as in example (4). This classifier is very commonly used for numerals from two to about ten, above which it becomes less common, to the extent that it hardly occurs with numbers over a hundred. The word *ida* ‘one’ seldom takes a classifier. If it does have one, as in example (5), then the classifier forces a numeric interpretation rather than an indefinite singular interpretation.

- (4) *Família ida nee iha oan nain neen.*
 family.P one.T this have child.T CLF:human 6.T
 ‘This family had six children.’ (Woman explaining child-rearing, 2000)
- (5) *Iha dadeer-saan nee ema nain ida mate.*
 LOC early.morning this person CLF:human one die
 ‘Early in the morning, one person died.’ (Young man recounting an attack, 2006)

There are a few other classifiers in addition to *nain*, which depend on the shape and size of the item being counted, but they are optional and used infrequently. The options are *musan* (lit. seed) for small roundish objects such as tablets, *fuan* (lit. fruit) for larger roundish objects such as eggs or bread rolls, *lolon* (lit. trunk) for long objects such as lengths of wood or guns, and *tahan* (lit. leaf) for thin flat objects such as shirts and roofing iron, illustrated in (1).

This is the same set of classifiers noted by van Klinken (1999: 105) for the Tetun Terik of Wehali, except that *matan* (lit. eye) is not used in Tetun Dili. This loss could be because *matan* is used in Tetun Terik for counting pigs and buffalo, two animals very important in cultural life. In Dili, these animals retain their importance to culture. However, where possible, cultural negotiations (e.g., about bride price) tend to be conducted in vernacular languages rather than in Tetun Dili. Reduction of a classifier system is in any case not uncommon as a result of language contact. Portuguese has no sortal classifiers, and Mambae (the original language of the area where Dili is located) has an extremely limited set used even less than in Tetun Dili (Hajek 2006).

6. Grammatical differences between Tetun Dili and borrowed numerals

6.1 Phrase-level differences

There is a sharp distinction between native Tetun numerals and numerals from Indonesian and Portuguese. One major distinction from which several differences follow is that while Tetun numerals can modify nouns from any language, Indonesian and Portuguese numerals can only modify nouns from their own languages. Another underlying distinction is that Tetun numerals are used with native Tetun constructions, while borrowed numerals are used with much of the phrase-level grammar of their own language. The resulting differences are summarised in Table 4, and discussed in turn below.

Table 4. Properties of indigenous and borrowed numerals within Tetun Dili

	Tetun	Indonesian	Portuguese
Co-occurrence:			
Can modify nouns from language	any	Indon. only	Port. only
Can take sortal classifier	yes	Indon. only	no
In compounds ‘X days from now’	yes	no	no
NP grammar follows source language:			
Position within NP	follows head	precedes head	precedes head
Some agree in gender with head	no	no	yes
Noun takes plural marking	no	no	optional
Number ‘plus’	<i>resin</i>	<i>lebih</i>	<i>i tal</i>
Native Tetun patterns for Tetun only:			
Can reduplicate to mean ‘groups’	yes	no	no
Specify range for approximation	yes	no	no
<i>X ba X</i> for indefinite large number	yes	no	no
Borrowed patterns for borrowings:			
Used as ordinal	v. limited	yes: with Indon. noun	yes: with Port. noun

Tetun numerals readily modify nouns from any source language. Examples are Tetun *uma rua* (lit. house two) ‘two houses’, Portuguese *kareta rua* (lit. car two) ‘two cars’ and Indonesian *hansip nain rua* (lit. civil.guard CLF:human two) ‘two civil guards’. In sharp contrast, borrowed numerals can only modify nouns from the same language, with the numeral preceding the noun as per the source language phrase structure, e.g., Portuguese *trinta dolar-es* ‘thirty dollar-PL’, Indonesian *seribu dolar* ‘thousand dollar’. A few such borrowed nouns are only used with borrowed numerals, with the Tetun synonym being used in all other contexts. For instance, Portuguese *anu* ‘year’ and *pesoas* ‘people’ are commonly used with Portuguese numerals in official reporting and parliamentary debate, while in all other contexts speakers use native Tetun *tinan* ‘year’ and *ema* ‘person’. Using borrowed numerals only with head nouns from the same source language restricts what can be enumerated using Indonesian and Portuguese attributive numerals to things for which the noun from these languages is both known and used in Tetun. So, for instance, one can use these languages to enumerate years but not buffalo, or to count total population but not the number of brothers in a family, since buffalo and brothers are always discussed using Tetun nouns. In this case Indonesian and Portuguese numerals can be used only if they are non-attributive, for instance in the predicate, as in example (14) below.

One consequence of borrowed numerals only modifying nouns from their own language is that when a native sortal classifier is used, one must use a Tetun numeral, since all classifiers are Tetun. Examples include *tahan* in example (1), and *nain* in (5). Portuguese has no classifiers. Indonesian does have them, and they are occasionally used with Indonesian numerals, e.g., *Ami sembilan orang* (lit. 1PL.EXCL 9.I person.I) ‘There were nine of us.’

Another result is that only Tetun numerals can be used in compounds meaning ‘X days from now’, using the native bound root *bain* ‘day’, e.g., *bain-rua* (lit. day-two) ‘in two days time’, *hori bain-haat* (lit. at.past day-four) ‘four days ago’.

Although borrowed numerals precede a head noun as per the source language, they can – though rarely do – co-occur with native Tetun modifiers, such as the determiner *nee*

‘this’ or relative clauses. These retain their usual Tetun position, as illustrated by (6), where the Portuguese ordinal *primeiru* precedes the head as per Portuguese rules, and the determiner *nenee* follows it as per Tetun rules.

- (6) *Primeiru sinku anu-s ne-nee, ema la hatene buat ida.*
 first.P five.P year-PL.P this-this person not know thing one.T
 ‘During the first five years, people don’t know anything.’ (Priest in discussion on TV, 2006)

As noted above, numerals are used with much of the phrase-level grammar of their source language. One result is that when Portuguese nouns are modified by Portuguese numerals, there can be number and gender agreement. The use of Portuguese numerals complete with agreement phenomena is quite common in high-level formal speech of the elite, and in writing. Thus Portuguese numbers greater than one are usually used with plural nouns as per Portuguese number agreement rules, though for non-Portuguese speakers, singular nouns also occur, e.g. *sinku dolar-es / sinku dolar* ‘five dollar(s)’. Some numerals take gender agreement with the head noun, following Portuguese rules, e.g., feminine singular *uma ora* ‘one o’clock’ versus masculine *ún anu* ‘one year’, and feminine plural *duas oras* ‘two o’clock’ versus masculine *dois anos* ‘two years’. For acrolectal speakers, at least, such agreement is a productive process, not dependent on borrowing whole phrases, and with no indication of code-switching. After all, in oral acrolectal Tetun Dili, the proportion of Portuguese tokens readily reaches 30%, with the percentage reaching over 40% in some articles in the press (Williams-van Klinken & Hajek 2018: 13).

When Portuguese nouns are modified by Tetun numerals, in contrast, the noun does not take plural marking. So, for instance, one can pluralise *dois livru-s* ‘two book-s’ with a Portuguese numeral, but one cannot normally pluralise the noun in the equivalent *livru rua* (lit. book two), which has a Tetun numeral.

There is one noun for which the interpretation depends on the language of the numeral. This is *oras*, from the Portuguese plural noun *horas* ‘hours’, but used in this form even for a single hour, e.g., *oras ida* (lit. hour one). When followed by a Tetun numeral, it means ‘hour’, e.g., *oras haat* (lit. hour four) ‘four hours’. When preceded by a Portuguese numeral following Portuguese grammar, however, it means ‘o’clock’, e.g., *kuatru oras* ‘four o’clock’. To state the time using a Tetun numeral, the noun *tuku* (from the verb *tuku* meaning ‘punch’) must be used, e.g., *tuku haat* ‘four o’clock’.

Another way in which numerals are tied to the phrase-level grammar of the source language is in the expression ‘plus, or more’, which has to be of the same language as the numeral. Tetun numerals take *resin* ‘extra’, e.g., *rua-nulu resin* ‘twenty plus’. Portuguese numerals are used with Portuguese *i tal* ‘and so on’, e.g., *vinti i tal*, while Indonesian numerals are followed by Indonesian *lebih* ‘more’, e.g., *dua puluh lebih*. In each case, the construction implies ‘more than 20, but not more than 30’.

There are several native Tetun constructions which are used only with native numerals. One is reduplication. Tetun numerals can be reduplicated to mean ‘in groups of X’, e.g., *lao rua-rua* (lit. walk two-two) ‘walk two by two’. This is not done with borrowed numerals. Despite the fact that Indonesian does have such reduplication (e.g., *dua-dua*), such reduplicated Indonesian numerals are not used in Tetun.

Another is the use of two Tetun numerals within a range to express approximation, e.g., *ida-rua* ‘one (or) two’, *rua ka tolu* ‘two or three’. This range can in informal speech be followed by *ida* ‘one’, as in (7). Both constructions are limited to Tetun numerals.

- (7) *La too lima-nulu, atus ida, kala netik rua-nulu, tolu-nulu ida.*
 not reach 5-tens.T hundred.T one.T perhaps a.bit 2-tens.T 3-tens.T one
 ‘There weren’t as many as 50 or 100 (students), maybe twenty or thirty.’ (Middle-aged scholarship manager in informal meeting, 2018)

For indefinite large numbers, one can use Tetun expressions *atus ba atus* (lit. hundred to hundred) ‘hundreds’, *rihun ba rihun* ‘thousands’, the slang term *rihun-rahun* (based on *rihun* ‘thousand’) ‘thousands, a great many’, and *tokon ba tokon* meaning something like ‘a massive number, zillions’. Again, this construction is not possible in Tetun with Portuguese or Indonesian numbers.

6.2 Postponed borrowed nouns as ordinals

There is one construction which exists for Indonesian and Portuguese numerals but not for Tetun ones. This is the use of a borrowed cardinal numeral following a head noun of the same language, to function as an ordinal, e.g., Portuguese *tópiku ún* ‘topic one’, Indonesian *semester tujuh* ‘semester seven’ in example (17). For most nouns there is no effective strategy for doing this using Tetun numerals, as a following numeral would most readily be interpreted as quantity, as shown in the following two comparisons.

- (8) *semester rua / semester dua*
 semester 2.T semester 2.I
 ‘two semesters’ ‘semester two’
- (9) *rejiaun ida / rejiaun ún*
 region 1.T region 1.P
 ‘a region’ ‘region one’

There are however two exceptions in which Tetun cardinal numerals are used as ordinals without ambiguity. The first involves dates. For instance *loron lima fulan rua* (lit. day 5.T month 2.T) refers to the fifth day of February as the second month (not five days two months), presumably as the date context clarifies the interpretation. The other exception is the Portuguese loan *andár* ‘storey’, for which *andár tolu* (lit. storey 3.T) can mean either ‘three storeys’ (e.g., if a building has three storeys), or ‘third storey’ (e.g., if one is ascending to the third storey). Again context clarifies the intended meaning.

6.3 Relative integration into clauses

In addition to the above differences between indigenous and borrowed numerals, there is a tendency for phrases with Tetun numerals to be more integrated into clauses than borrowed numerals are. This pattern is however a tendency rather than a rule. Indonesian and Portuguese numerals are very common for years, dates, prices, sports scores, telephone numbers and population counts. Most phrases containing them are therefore not tightly integrated into clauses, being standalone phrases as in (10) or peripheral prepositional phrases as in (11). They can however readily be objects, as in (12) and the second clause of (13). Much less commonly they are found in subjects, as in the start of (13) or predicates, as in (14).

- (10) Q: *Maun agora tinan hira?* A: *Dua puluh lima tahun.*
 older.brother now year.T how.many 2.I tens 5.I year.I
 ‘How old are you (older brother)?’ ‘Twenty-five years.’
 (Survey, 2018)

- (11) *Portantu hau koalia kona ba korenta milhoens de dolar-es*
 so 1SG talk about 40.P million.PL.P of.P dollar-PL.P
 ‘So I’m talking about forty million dollars
ba ema kiak sira, déz milhoens ba estudante i joven sira.
 for person poor DEF.PL ten.P million.PL.P for student and youth DEF.PL
 for the poor, (and) ten million for students and young people.’ (National president
 in speech on TV, 2007)
- (12) *Ha-tuun mesin ninian, maizumenus hau husu sepuluh dolar*
 make-descend engine POS more-or-less 1SG ask ten.I dolar.I
 ‘For removing the engine, I ask about ten dollars,
hansa nee, balun sira hanoin mak foo lima belas, lima belas dolar.
 like this some 3PL think FOC give 5.I teen.I 5.I teen.I dollar.I
 some take pity (on me) and give fifteen – fifteen dollars. (Mechanical engineering
 student chatting with friend, 2006)
- (13) *Artigu vinti nee dehan saida? Lee deit número ún!*
 article.P 20.P this say what read just number.P one.P
 ‘What does this article twenty say? Just read (line) number one!’ (Parliamentarian
 during budget debate, 2011)
- (14) *Nee.duni ami hamutuk dozi pesoa-s, ...*
 so 1PL.EXCL together 12.P person-PL.P
 ‘So altogether there are twelve of us (brothers and sisters), [but one has died].’
 (Young man talking about his family, 2006)

7. Functional differences between numerals from the three languages

The language used for numerals in Tetun Dili depends on a range of factors, including not only the head noun for attributive adjectives (as noted in section 6.1), but also the size of the number, the topic, the context (especially formality), the lectal type, and the language of education of the speaker and in some cases of the previous generation. In the following sections, we first overview the large distinction between Tetun and borrowed numerals, using data from the first author’s corpus of texts. We then use the results of the surveys, supplemented by corpus data and direct observation, to analyse in more detail the functions for which numerals from each of the languages are used.

7.1 Functional differences between Tetun and borrowed numerals

Within Tetun Dili, Indonesian and Portuguese numerals fill much the same functional slot, which is clearly differentiated from the function of native Tetun numerals. The distinction in everyday life is largely that between traditional functions (for Tetun) and modern functions (for Indonesian and Portuguese) – a pattern that has been noticed in many other languages which have borrowed numerals (Matras 2007). In particular, Tetun is used for quantities (e.g., number of people or items), as well as for some times and dates. Indonesian and Portuguese are used for prices, time and date, phone numbers, measurements, sequence, mathematics and sports scores, amongst others.

Orally, Tetun is mainly used for small numbers, up to ten, with larger numbers usually being expressed in Indonesian or Portuguese. The media and written texts use Tetun numerals much more, however. If numerals are spelled out in written texts, they are almost always in Tetun. If they are written as digits with a following percentage sign or abbreviation, e.g., 70% or 75mm, then they can be read in any of the three languages,

since the abbreviations for most units of measurement are the same in all three. However if numbers quantify a preceding head noun, they must be read in Tetun, since only Tetun has numbers following the head noun, e.g. *kaixa 7* must be read as *kaixa hitu* (lit. box 7.T) ‘seven boxes’. Alternatively, if the number quantifies a following head noun, it must be read in the language of that noun. Both options are illustrated in example (15). Note that in this example, the Portuguese loan *tonelada* ‘ton’ is used in the plural even with Tetun numerals, despite this being very rare. Presumably the plural form is a copy of the plural in the first clause.

- (15) ... *Maliana hetan foos 17,7 tonelada-s, Atabae tonelada-s 4 ou 5*
 ... (place) get rice 17.7 ton-PL (place) ton-PL 4 or 5
 ‘... Maliana received 17.7.P tons of rice, Atabae 4 or 5.T tons,
 no Balibo 17 tonelada-s.
 and (place) 17 ton-PL
 and Balibo 17.P tons.’ (Newspaper Suara Timor Lorosae 3/5/2007)

The limited use of Tetun for high numbers is reflected in the difficulty that both children and adults have when reading numbers over a thousand in Tetun, a fact we have often noted in personal experience. If the number is presented as digits (e.g., ‘144.000’), most people can read it in Indonesian for instance, but not in Tetun. On the other hand, if it is spelled out (e.g., *rihun atus ida haat-nulu resin haat*), readers can sound it out, but often do not know what quantity this represents.

As an example of the differences between the three languages, we take the number ‘three’ as it is found in the first author’s database of oral, written and translated texts. In this corpus, there are 1966 examples of Tetun *tolu* from all genres, most within numbers modifying a preceding head noun. The vast majority give the number of people, usually using the human classifier *nain*, e.g., *labarik nain tolu* ‘three children’. There are also many quantities (e.g., *litru tolu* ‘three litres.P’, *lata tolu* ‘three tins.P’), periods of time (e.g., *tinan tolu* ‘three years’) and many instances of *dala tolu* ‘three times’. It also quantifies other nouns, e.g., *golú tolu* ‘three goals.P’. This native Tetun numeral is common in oral texts (accounting for 68% of 784 examples of the numeral ‘three’). When the number is spelled out in written and translated texts, it is virtually always in Tetun (99% of 1091 written examples of ‘three’, 100% of 365 translated examples).

For Indonesian *tiga*, there are 144 examples in the corpus. One is from an Indonesian phrase within a written text. All the rest are from informal oral texts, except for a few instances of lexicalised expressions, such as *Tiga Roda* (lit. three wheel.I) to refer to vendors with pushcarts. Speakers include a wide range of ages, but not many acrolectal speakers. In addition to fixed expressions, *tiga* is used in dates, phone numbers, soccer scores, prices, and in sequencing, as described in section 6.2 with the function of ordinals (e.g., *juara tiga* (lit. champion.I 3.I) ‘third place getter’, *gigi tiga* (lit. gear.I 3.I) ‘third gear’). A few examples enumerate Indonesian nouns which are not normally used in Tetun, suggesting code-switching, e.g., *tiga bahasa* ‘three language(s)’.

The corpus contains 136 examples of Portuguese *três* (including spellings *tres*, *trez* and *três*). Only five are in written texts, all of these within fixed expressions, such as a company name *Tres Amigos* ‘Three Friends’. The rest are oral, from a range of registers and speakers, but including a significant proportion of acrolectal speakers. Indeed, Portuguese numerals are uncommon in basilectal Tetun. Most occur in dates, with others being found in phone numbers, soccer scores, prices, and a few in the sequence *númeru três* ‘number three’, e.g., *kandidatu númeru três* ‘candidate number three’.

In addition to spelled-out numerals, there are many written and translated texts that use the digit ‘3’ in contexts which force or encourage a Portuguese reading, such as in prices, percentages, or population figures.

7.2 Functions of the three languages: Overview of survey results

We turn now to data from the two surveys of numeral use. Data for the two years have been combined in Table 5, where they are presented according to language preference (Tetun, Indonesian, Portuguese or mixed) and then context. As can be seen in this table, the language preference for numerals is strongly influenced by the topic, with Tetun, Indonesian and Portuguese each having their own small niches, while for other topics the language preference is not clear. There were only a few questions for which language preference reflected the language of education. Each of these topics will be discussed in turn, with examples and additional information from other sources also provided where relevant. As noted earlier, the number of respondents (“Total No.”) for each question varied substantially.

Table 5. Languages in which numerals were given in response to survey questions (with 2012 and 2018 results combined)

Question	Topic	Total No.	% Tet.	% Ind.	% Port	% Mixed
Tetun strongly preferred						
How many brothers and sisters do you have?	home	113	92	6	1	1
Indonesian strongly preferred						
What semester are you in?	school	63	3	97	0	0
What class are you in?	school	48	2	90	8	0
How many credit points are you taking?	school	65	8	75	14	3
How many CC is your motorbike?	tech	35	3	97	0	0
What is your telephone number?	tech	117	4	78	15	3
What does ... cost? (low prices)	price	99	6	87	3	4
What does ... cost? (high prices, e.g. cost of tuition, house, motorbike, car)	price	104	4	84	13	0
Fractions (adults). Read: 1/2, 3/4	maths	71	0	79	19	2
Portuguese strongly preferred						
Political date: What date is the restoration of independence?	date	91	0	18	82	0
Fractions (children): Read: 1/2, 3/4	maths	62	0	0	100	0
Mixed preferences (none over 75%)						
How many kilograms is a sack of rice?	weight	59	2	71	27	0
How heavy are you?	weight	32	3	53	44	0
Count my fingers / these papers.	count	135	53	30	17	1
How old are you?	home	173	10	69	19	2
Date: What is the date today?	date	126	1	58	37	5
Date: What date were you born?	date	107	1	59	32	8
Time: Tomorrow, what time do you start and finish school/work? / What time do you eat?	time	147	12	68	5	15
Time: What is the time now?	time	106	13	67	16	4
Read: 56	maths	150	5	53	40	1
Read: 256 (in 2012), 238 (in 2018)	maths	133	9	67	23	2
Read: 3685	maths	138	10	44	45	1
Read: 2,6	maths	99	0	54	44	2
Read: 86%	maths	111	0	62	32	5
Fractions (students). Read: 1/2, 3/4	maths	62	0	40	58	2

7.3 Topics for which Tetun is strongly preferred

The only survey question for which the vast majority of answers used Tetun numerals was the question about the respondents’ number of brothers and sisters, for which 92% used Tetun, as in (16). This fits with Tetun being used for traditional contexts, and low numbers, with most people having no more than ten siblings. An additional factor

favouring Tetun for family size is that for small numbers of people in complete sentences, the classifier *nain* is usually used, as in (4), and this co-occurs only with Tetun numerals. Native Tetun kin terms too, force Tetun numerals, when these are used attributively. Nevertheless, Indonesian and Portuguese numerals do occur in giving family sizes, as in (14).

(16) Q: *Ita iha maun-alin nain hira?*
 2SG have older.brother-younger.sibling CLF:human how.many
 ‘How many brothers and sisters do you have?’

A: *Sia. Feto lima, mane haat.*
 9.T female 5.T male 4.T
 ‘Nine. Five sisters and four brothers.’ (Survey, 2018)

7.4 Topics for which Indonesian is strongly preferred

Topics for which Indonesian is strongly preferred include grade and semester in school, the number of credit points taken at university, engine sizes, telephone numbers, and prices. All (except telephone numbers) are commonly discussed using Indonesian or ambiguously Indonesian/Portuguese nouns, such as *semester* ‘semester.I’ and *dolar* ‘dollar.I/P’, which both reflects the Indonesian-language influence in these fields and further encourages the use of Indonesian numerals.

The use of Indonesian numerals reflects the ongoing impact of the Indonesian occupation that ended in 1999. In addition to obligatory Indonesian-medium education, large numbers of Indonesians settled in Timor-Leste during the 24-year occupation, playing a major role in commerce and daily life as traders and customers. Even children who have grown up since independence use Indonesian numbers, as a result of socialization at home and in the community as well as by interactions at shops and markets.

7.4.1 Grade and semester in school, credit points

When asked for their semester at university, 97% of 63 students replied in Indonesian, just as in example (17) from a chat amongst students, even though for most the semester is a low number not more than ten. The remainder used Tetun. Perhaps Indonesian was so strongly favoured over Tetun because (as noted above) Indonesian *semester dua* ‘semester two’ unambiguously indicates ‘second semester’, whereas Tetun *semestre rua* is more likely to be interpreted as ‘two semesters’.

(17) *Hansa hitungan kona.ba nia CC sira ne-nee,*
 like count.I about POS CC.I/P DEF.PL this-this
 ‘Such as calculations on its CC and so on,
uluk ita hetan iha sekitar semester tujuh.
 formerly 1PL.INCL get LOC about.I semester.I 7.I
 we did that in about seventh semester.’ (Mechanical engineering student at a repair shop, 2006)

In contrast, when asked the number of credit points they were taking (mentioning the commonly-used Indonesian term *SKS* ‘credit point’, only 75% of the 65 respondents answered in Indonesian. The rest replied in Portuguese or Tetun, with Portuguese being possible so long as the Indonesian head noun isn’t repeated.

When asked what year they were in at primary school, 90% of 48 children replied in Indonesian (84% in 2012, rising slightly to 93% in 2018). Most of the remainder replied in Portuguese, using a numeral such as *kuatru* ‘four’, as in (18). None mentioned the ordinal numbers typically used to identify grades within the education ministry. Outside of the ministry, however, ordinals are only used when people talk about school years under the Portuguese system, prior to the 1975 Indonesian invasion, e.g., *primeira klase* ‘first.F grade.F’, *terseiru anu* ‘third.M year.M’.

- (18) Q: *Alin agora klas hira?* A: *Klas empat. / Kuatru.*
 younger.sibling now.P class.I how.many class.I 4.I 4.P
 ‘What grade are you in?’ ‘Grade four.’ ‘Four.’
 (Survey, 2018)

7.4.2 Engine size

In the survey, when asked the CC rating of motorbikes, an overwhelming 95% (of 35) replied in Indonesian. This reflects Indonesian being the language of motorcycle shops and many repair workshops, most of which are owned by Indonesians, or at least source their supplies from Indonesia.

7.4.3 Telephone numbers

In the survey, the majority gave their telephone number in Indonesian (83% of 70 students, 67% of 43 adults), illustrated by (19). Nearly all the rest gave their phone number in Portuguese. Before about 2008 only the elite used Portuguese for phone numbers. After that, Portuguese started to come in, even for young people. However, the percentage of Portuguese does not seem to be rising, if anything, it fell during this period for both students and adults, although the sample sizes were not large enough to make these statistically significant. The continuance of Indonesian is despite the fact that Portuguese is typically used to read telephone numbers on television, for instance in the frequently screened advertisements encouraging current secondary school leavers to enrol in tertiary institutions.

- (19) A: *Hau bele hatene ita nia númeru telefone?* B: *Tujuh tujuh tiga ...*
 1SG can know 2SG POS number telephone 7.I 7.I 3.I
 ‘May I know your telephone number?’ ‘773...’ (Lecturer, 2018)

The vast majority of people cannot give their phone number in Tetun unless they are reading it aloud or translating it in their heads.

7.4.4 Prices

Low prices, such as for items sold in markets and kiosks, are usually in Indonesian, as in (20). This is so regardless of the age of the speaker. Note that the word for ‘price’ is normally Tetun *folin*, never Indonesian *harga*. Since 2008 low prices are occasionally given in Portuguese, though this is still uncommon. In the survey, 87% of respondents gave the price of cheap items in Indonesian.

- (20) *Mina nee, litru ida satu dolar.*
 oil/petrol.T this litre.P one.T one.I dollar.I/P
 ‘Petrol costs a dollar a litre.’ (Student telling a story, 2006)

High prices, for items such as houses and cars, are in everyday speech usually given in Indonesian, regardless of the language used to name the item being priced, as illustrated by (21). It can however also be in Portuguese, especially for older well-educated adults.

In the survey (which did not include the elite), 84% responded in Indonesian, with little influence of age.

- (21) *Ita baa buka took bibi dikur ida hanesan nee-nee, seratus, ...*
 IPL.INCL go search try goat.T horn one like this-this hundred.I
 ‘If we go and look for a goat with a horn like this, (it costs) a hundred, [the price of a buffalo!] (Zodiac radio program, 2006)

Discussions in parliament and government on the budget mainly use Portuguese numbers, as illustrated in (11), with some Tetun. So too do media discussions of government revenue.

In the early years after independence, prices were often still given in Indonesian rupiahs, especially in rural areas. Even today it is very occasionally heard, with an assumed conversion rate of 10,000 rupiahs to the US dollar. These prices are always given in Indonesian.

7.5 Topics for which Portuguese is strongly preferred

The only question in the survey to which a large majority of responses used Portuguese numerals was a request for the date of the restoration of independence, a date that has been mentioned frequently in the media and official speeches, normally in Portuguese. Responses were overwhelmingly in Portuguese, namely *vinti de Maiu* (lit. twenty of May), with the remainder in Indonesian. The percentage of Portuguese rose by age and year. In 2012 69% of 32 students used Portuguese and 82% of 11 adults, rising to 90% of 29 students and 100% of 16 adults in 2018.

7.6 Topics for which language preferences reflect language of education

There were a few categories in which there were clear differences between the age groups in the survey. Of these, fractions reflect the language of education of the respondent. So too does counting to high numbers, something not tested in this survey. Both are mathematical concepts.

Two other topics to some extent reflect the language of education of the previous generation, especially for children. For age, children favour Indonesian, while adults favour Portuguese, with only university students favouring their own language of education, namely Indonesian. For percentages, about half of children use Indonesian, the language of education of the previous generation. However most students and adults (who were educated in Indonesian) favour Indonesian too.

7.6.1 Counting

When asked to count the questioner’s fingers or a small number of papers, about two thirds of children (69% of 52) and students (67% of 35) used Tetun, with the remainder using Indonesian or Portuguese. In contrast, two-thirds of adults (65% of 43) counted in Indonesian, with the remainder using Portuguese, not Tetun. The relative prevalence of Tetun is consistent with this being an everyday task, with numbers under ten.

In contrast, our own direct observation, supported by comments from Timorese colleagues, indicates that when counting larger quantities (such as large amounts of money, or large numbers of people), most people use the language of their education.

7.6.2 Fractions

Fractions play a very small role in East Timorese life, even for urban Tetun Dili speakers. The only fractions that are used in everyday Tetun Dili are a few terms meaning approximately ‘half’. These are:

- (i) *balu(n)* ‘half/some’, e.g., *tinan haat ho balu* (lit. year four and half/some) ‘four and a half years’. This is an approximation, not an exact mathematical ‘half’.
- (ii) *metade* ‘half’, from Portuguese, is used mainly in writing and formal contexts.
- (iii) *meiu/a* ‘half’ from Portuguese, is used almost exclusively in telling the time in Portuguese (e.g., *sinku i meia* (five and half) ‘5.30’, *meia ora* ‘half hour’) or in other fixed Portuguese expressions, e.g. *meudia* ‘midday’.

In the 2018 survey, respondents were asked to read “1/2” and “3/4”. The language used in responses strongly reflected the language of education of the respondents. All 62 children’s readings used Portuguese numerals. Only one read using the term *metade* ‘half’, all the rest reading the two digits separated by *por* ‘out of’ (e.g., *ún por dois* ‘one out of two’) or *bara* ‘slash’ (e.g., *trés bara kuartru* ‘three slash four’). The former appears to be a calque on Indonesian *tiga per-empat* (lit. three out.of four), while the latter is the standard way of reading ‘3/4’ when it is not a fraction, for instance in reference numbers. Surprisingly, none of the children replied using the fractions as they would be said in Portuguese, namely *trés kuartrus* ‘three quarters’. Even many secondary school graduates do not recognise fractions in Portuguese, despite notionally having completed their mathematics education in secondary school in Portuguese.

For university students (of whom the younger ones were educated in Portuguese and the older in Indonesian), 58% of the 62 respondents replied in Portuguese, using the same system as the children, and the rest in Indonesian. Of the 71 adult readings, 79% were in Indonesian, with the rest using Portuguese numerals.

No respondents of any age attempted Tetun. This reflects the fact that fractions are a concept learned in school, and are in practical terms for most speakers restricted to that domain.

There are however a few fractions that have been learned in specific contexts with specific meanings, having been introduced by Portuguese-educated leaders. One is *dizimu* (from Portuguese *dízimo* ‘tenth part’), which Protestants use to refer to tithes. Another specific fraction came into use in Tetun in late 2017, when the annual state budget was not approved. It is *duodésimu* (from Portuguese *duodécimo* ‘twelfth’), used to refer to an amount equivalent to a twelfth of the budget, which can be used for to cover a month’s expenses when the annual budget has not yet been passed.

7.6.3 Age

When asked their age in the survey, there was a difference between age groups, but not a significant change from 2012 to 2018. For specification of age, there seems to be a time lag, with many respondents using the language of education of the previous generation. In particular, two-thirds of the children (68% of 68) gave their age in Indonesian (the language of education of their parents), with the rest mainly in Portuguese. For students, whose parents were educated in Indonesian if at all, the percentage of Indonesian was higher (though not significantly so), at 84% of 61 respondents, as in example (10). In contrast, only 48% of 40 adult respondents used Indonesian (so differing from children with a significance of $p < 0.001$). Nearly all the rest used Portuguese (in which only older adults were educated). Tetun was mainly used

by children up to the age of ten, reflecting the generalisation that Tetun is mostly used for numbers in this range.

7.6.4 Percentages

In the 2018 survey, when asked to read “86%” aloud, two-thirds (62% of 111) opted for Indonesian, with all the rest replying in Portuguese or a mixture of the two (e.g., a Portuguese number followed by Indonesian *persen*). There were marked differences in age, the least Indonesian being used by students (50% of 36) and children (58% of 40), and adults having a much higher percentage (80% of 35, with the difference between students and adults being significant at $p=0.03$). None read the percentages in Tetun.

In the first author’s database of oral texts, however, the three languages are almost evenly split. There are 21 Portuguese numerals plus *porsentu*, as in (22), all of which are formal except one *sein porsentu* (lit. 100%) ‘totally’. There are 23 *porsentu* followed by a Tetun numeral (all formal, including presentation of results at a seminar, and radio programs on science), and 27 Indonesian numerals with *persen* (mostly but not all informal). The relatively lower proportion of Indonesian, and the higher proportion of Tetun in these oral texts than in the survey reflects the fact that a significant number of these instances were in formal situations, where Indonesian is avoided, and Tetun and Portuguese numerals are favoured.

- (22) *Hahuu hosi doiz mil i seti,*
 start from 2.P thousand.P and.P 7.P
 ‘As of 2007,
ema nebee tau naran ba eskola, sae ba noventa porsentu.
 person REL put name for school ascent to 90.P percent.P
 school registrations rose to 90%.’ (Speech by national president, 2007)

7.7 Topics for which there were mixed responses

Two topics for which there was not one language with an overwhelming majority were date and time. Both can be in any language, but preferences vary.

7.7.1 Date

Dates can be in any language. In the survey, about two-thirds of the students used Indonesian for stating today’s date or their birthday, as in (23) (65% of 69 in 2012, 64% of 74 in 2018), with most of the rest using Portuguese, illustrated by example (24) from the corpus. 9% answered in mixed languages, with the usual mixture being month names in Portuguese, years in Indonesian, and days in Portuguese or Indonesian.

Educated adults favoured Portuguese in dates (73% of 22 in 2012, 63% of 62 in 2018).

- (23) Q: *Maun moris iha saa data?*
 older.brother born LOC what.T date.P
 ‘What date were you born?’
 ,
 A: *Dua puluh tiga bulan lima sembilan belas sembilan puluh.*
 2.I tens.I 3.I month.I 5.I 9.I teen 9.I tens.I
 ‘23rd of the fifth month, 1990.’ (Survey, 2018)

- (24) *Hau moris iha Zumalai iha dia déz de Abril,*
 1SG born LOC (place) LOC day.P ten.P of.P April.P
 ‘I was born in Zumalai on tenth April

míl *novi-sentus* *setenta i* *trés.*
 thousand.P 9-hundred-PL.P 70.P and.P 3.P
 1973.’ (Young woman introducing herself for a recording, 2000)

There was only one Tetun date given in our survey, although they do occur in texts and in real life, especially in formal contexts. This can be illustrated by counts of dates in the database of oral texts. Of the 290 examples of the years ‘two thousand ...’, 22% use Tetun *rihun rua* (mostly formal), 66% use Portuguese *dois míl* (mostly formal), and 12% use Indonesian *dua ribu* (all informal).

7.7.2 Time

Time can be in any language, with Indonesian being the most common. In keeping with Tetun being used mostly for everyday functions and for small numbers, Tetun is more common for hours up to ten o’clock than for eleven or twelve o’clock, and is rarely used where minutes are specified. In this case Indonesian (or occasionally Portuguese) is used instead, as in example (25).

In our survey, some 68% gave the start and finish time of their classes or work in Indonesian, with a big increase in Indonesian responses over the two survey dates, but almost no impact of age. In 2012, 46% of 67 respondents gave the times in Indonesian. The rest were split between Tetun (19%), Portuguese (12%) and a further 22% who used two different languages for two different times. The most common mix was to use Tetun for a number up to ten, and Indonesian (or Portuguese) for a time after ten, as in example (2), or for one in which minutes are specified. In 2018, the percentage of Indonesian rose markedly ($p < 0.001$), to 87% of 79 responses. The reasons for this leap are unknown, especially in light of the otherwise general decline in use of Indonesian and Indonesian loans in Timor-Leste.

For the current time, 67% answered in Indonesian, as per example (25). In 2012 the average was 59% Indonesian. In 2018 it diverged by age, 55% (of 31) university students replying in Indonesian, and the percentage for adults over 40 years being much higher at 94% (of 32), most of the remainder replying in Portuguese.

(25) Q: *Agora tuku hira?* A: *Satu lewat kosong dua.*
 now o’clock how.many? One.I pass.I empty.I 2.I
 ‘What time is it?’ ‘It’s 1.02.’ (Survey, 2018)

In practice, time is usually to the previous hour or half hour. There is no term for ‘quarter to/past’ in common use.

Times can be further specified by a following (or occasionally preceding) *dadeer-saan* ‘early morning’ or *dadeer* ‘morning’, *meudia* ‘noon’ (from Portuguese *meio-dia*), *lokraik* ‘afternoon’, *kalan* ‘evening, night’ and *kalan boot* (lit. night big) ‘midnight or thereabouts’, as in example (26). There is no abbreviation comparable to English ‘am’ or ‘pm’.

(26) *Ami hahuu votasaun hosi tuku neen dadeer,*
 IPL.EXCL start voting from o’clock.T 6.T morning
 ‘We started the voting at 6am,
remata iha tuku tolu lokraik.
 finish LOC o’clock.T 3.T afternoon
 and finished at 3pm.’ (Woman telling of referendum, 2000)

7.7.3 Reading numbers

In the survey, when asked to read numbers of two, three and four digits (56, 238 and 3685), the results were quite consistent across age groups and length of number. Of the 421 readings, 55% were in Indonesian, 36% Portuguese, 8% Tetun, and the remainder used a combination of languages.

7.7.4 Weights and measures

In the survey, when asked the weight of a bag of rice, 71% of 59 replied in Indonesian, the rest in Portuguese. The percentage of Indonesian was lower for adults (61% of 31) than for students (82% of 28). When asked their own weight, a similar percentage of adults replied in Indonesian (53% of 32), almost all the rest in Portuguese (we did not ask students or children this question).

(27) Q: *Foos saka ida nia todan hira?* A: *Dua puluh lima kilo.*
 rice bag.P one.T POS weight how.many 2.I tens.I 5.I kilo.I
 ‘How heavy is a sack of rice?’ ‘25 kg.’ (Survey, 2018)

In oral texts, there are not many references to weights and measures. Of 14 references to *kilo.P/I* ‘kilogram’, *kilograma.P / kilogram.I* ‘kilogram’ or *kilometru.P / kilometer.I*, 50% are in Tetun, and the rest spread between Portuguese and Indonesian, with the language of the numeral matching the language of the head noun.

8. Functional differences between age groups and over time

To enable comparison across age groups and across survey years, we used only the questions that were the same for both, and totalled them. This was nine identical questions for adults and university students, with three of these also being asked of children. Table 6 shows the results.

Table 6. Languages in which numerals were given in response to common survey questions

Year	Age group	Total	% Tetun	% Indonesian	% Portuguese	% Mixed
2012	primary	103	28	57	6	9
2012	university	298	8	67	20	5
2012	adult over 40	97	6	45	48	0
2018	primary	120	26	66	6	3
2018	university	266	5	67	22	6
2018	adult over 40	252	1	69	27	2

The biggest change from the 2012 to the 2018 survey was a significant increase ($p < 0.001$) of 24% in Indonesian numerals for adults over 40, mostly at the expense of Portuguese, which dropped 21%. The reasons for this change are not clear. It goes against our expectation that the proportion of Portuguese numbers would increase over time replacing Indonesian, but matches our informal observation that there appears to have been an increase in the use of Indonesian numerals amongst adults, even though the use of other Indonesian loans does not appear to be increasing. We suspect that this reflects a partial relaxation of earlier pressure among adults to avoid Indonesianisms, coupled with the persistence of Indonesian numerals in everyday commerce, amongst other things.

For children too there was an increase ($p=0.001$) in Indonesian of 9%, however this was mostly at the expense of mixed answers, which dropped 6%. For students, however, the percentages of each language stayed almost constant.

As can be seen, children used a much higher percentage of Tetun and lower percentage of Portuguese than students and adults did. This was partly because they were asked less technical questions, and no questions requiring large numbers.

In 2012, university students used more Indonesian and less Portuguese ($p<0.001$) than adults did. In 2018, however, the results for university students and adults over the age of 40 were very similar, as the Portuguese percentage for adults had reduced so much.

9. Reasons for borrowing numerals in Tetun Dili

Only in the case of ‘zero’ and numbers from a million up do borrowed numerals in Tetun Dili fill lexical gaps. All remaining borrowing of numbers is “unnecessary” borrowing (Clark 2004). This is clearly shown by the fact that in rural Tetun Terik in the old kingdom of Wehali located in Indonesian West Timor, just twenty years ago, Indonesian numbers were not used except for *juta* ‘million’ (Williams-van Klinken 1993–1995 field data).

Most of the reasons for “unnecessary” borrowing of numerals in Tetun Dili are the same as those posited for other languages (e.g. Matras 2007), as discussed in section 3. Firstly, Portuguese and Indonesian numerals predominate for concepts introduced from the outside world through the colonial languages, and are reinforced in those areas within the education system, which uses the official languages. In Timor-Leste, this was Portuguese up to 1975, Indonesian from 1976–1999, and Portuguese and Tetun from 2000.

Secondly, the predominance of Portuguese numerals in formal situations is supported by the prestige of Portuguese as an official language and the favoured language of the nation’s elite.

Thirdly, some borrowed numerals are shorter than Tetun ones. This is particularly so for Tetun numerals from 11–19, in which the *resin* contributes to making the Tetun numerals significantly longer than Indonesian and Portuguese numerals, as shown by the comparison in Table 7 below. Many Tetun speakers have commented to us over the years that the longer Tetun numerals are less convenient for this reason. In addition, Indonesian allows short forms for years, in which the *puluh* ‘tens’ is omitted, e.g., *sembilan sembilan* (lit. 9 9) for ‘1999’.

Table 7. Comparison of numeral length in Tetun, Indonesian and Portuguese, with number of syllables in brackets

Number	15	320	325
Tetun	<i>sanulu resin lima</i> (7)	<i>atus tolu rua nulu</i> (8)	<i>atus tolu rua nulu resin lima</i> (12)
Indon.	<i>lima belas</i> (4)	<i>tiga ratus dua puluh</i> (8)	<i>tiga ratus dua puluh lima</i> (10)
Port.	<i>kinzi</i> (2)	<i>trezentus i vinti</i> (6)	<i>trezentus i vinti sinku</i> (8)

A final reason for using borrowed numerals, which we have not previously noticed in the literature, but which we have discussed in section 6.2, is that borrowing allows a distinction between specifying quantity and specifying ordinal relations, so distinguishing, for instance *kandidatu rua* ‘2.T candidates’ and *kandidatu dois* ‘candidate 2.P’, i.e., ‘the second candidate’.

10. Conclusion

10.1 Factors affecting language choice for numerals

Based on the detailed analysis of numeral systems for Tetun Dili presented above, it is evident that the choice and use of numerals from several source languages is complex. This complexity is seen in issues relating to the relative integration of these numerals into the native grammatical system, as well as in the choice of numerals to use in specific contexts, and changes over time. There are many factors influencing the choice of which numerals to use. These include the following list, which is ordered according to their relative importance.

- (i) The head noun for attributive numerals: The only rule which is almost exceptionless is that Tetun numerals can modify nouns from any language, while Indonesian and Portuguese numerals can only modify nouns sourced from their own language. This means that to use borrowed numerals, the borrowed head noun must be one which is used within Tetun, or the numeral must be used outside of a noun phrase, for instance in the predicate.
- (ii) The size of the number: In everyday speech, Tetun is mainly used for numbers up to ten, with borrowed numerals being used above that. In the media and some formal speech, however, Tetun numerals are used alongside Portuguese ones even for very large numbers, in recognition of the co-official and therefore formal status of Tetun.
- (iii) The topic: In general, Tetun is used for traditional concepts such as quantities, while Indonesian and Portuguese are used for modern ones. Some topics strongly favour Indonesian, including prices, and motorbike sizes (as almost all motorcycle shops source their products from Indonesia). Indonesian is also favoured for school data (such as semester or year, and credit points) even though the schools now officially operate in Portuguese and Tetun, presumably showing the influence of the home, of teachers used to the Indonesian system and of previous generations of students (which can include older siblings) on younger students moving through the school system today. One context strongly favouring Portuguese involves dates of official events, which are frequently heard on the news.
- (iv) Formality: In informal speech, numerals from all three languages can be used. In formal contexts, Indonesian numerals are normally avoided, just as other Indonesian loans are avoided in formal contexts and in writing (Williams-van Klinken & Hajek 2018). In television reporting, Tetun numerals are favoured over Portuguese, even for large numbers, for which most people normally avoid Tetun. On the other hand, the voiceover for written announcements on television make overwhelming use of Portuguese for items such as telephone numbers, dates and prices.
- (v) The lectal type: At one extreme, the acrolect used by Portuguese-educated elite and officialdom strongly favours Portuguese and Tetun numerals, avoiding Indonesian. At the other extreme, the basilectal variety of uneducated rural speakers is much less influenced by Portuguese in general, and favours Indonesian and Tetun numerals. In between, speakers of the mesolectal variety use all three languages. Note that lectal type and formality are closely related, with formal speech mostly tending towards the acrolect, and thus either favouring

Portuguese numerals (in keeping with the high rate of Portuguese loans in the speech of the elite and in the press) or Tetun (as per liturgical Tetun).

- (vi) The language of schooling of the speaker or of the previous generation: In addition to the effect on school data mentioned in the previous paragraph, the impact of the language of education of the speaker is seen most obviously in fractions, which in Timor are used virtually only within the school context. Counting to high numbers also follows the school language. For other topics, such as prices or telephone numbers, older Portuguese-educated speakers are more likely to use Portuguese than Indonesian, but children (who are studying mathematics in Portuguese and Tetun) do not. In yet other topics, there seems to be a time lag, with people using the language of education of the previous generation. This influence is strongest in telling age.

10.2 Changes over time

When looked at over time, the position of Tetun numerals seems to have been relatively stable, at least for the last twenty years. The one obvious exception is the use of higher Tetun numerals on television, but then again Tetun was not used in the media prior to the 1999 referendum on independence.

For borrowed numerals, however, there has been some shift over the last twenty years from Indonesian to Portuguese. When Indonesian control of Timor ended in 1999, almost all numerals in everyday use by the non-elite were in Indonesian or Tetun. Portuguese numerals were used enough to make their way into dictionaries (e.g. Hull 1999) and grammars (e.g. Williams-van Klinken, Hajek & Nordlinger 2002), but seldom enough in everyday life that the Tetun language course developed for Peace Corps volunteers (Williams-van Klinken 2003) waited until chapter 48 before teaching Portuguese numerals, long after Indonesian ones were taught in chapter 19.

Based on the authors' observation, Portuguese numerals started to filter into popular urban use some eight years later, in about 2007–8, when primary school education in Portuguese had become more established and Portuguese language media more prominent. Given this start, and the fact that Portuguese numerals are favoured in the education system and in the media, we had expected to see a significant rise in the use of Portuguese numerals between the 2012 and 2018 surveys. However, this turned out to not be the case. Rather, there was a significant increase in Indonesian for adults (at the expense of Portuguese) and for primary school children (mainly at the expense of mixed answers), with university students showing no change in language preference during this period. In light of the general decline in Indonesian use and proficiency since 1999, the reasons for this unexpected pattern in our 2018 data remain unclear, although we have suggested a partial relaxation of pressure to avoid Indonesianisms amongst adults. They do however highlight how ingrained the use of a previously influential language can be, even after it has exited bureaucracy and the education system and has lost official status. Our results point instead to the importance of socialization of numbers in a specific language amongst children by, for instance, Indonesian-educated parents and older siblings in the home, as well as in important everyday settings such as markets and shops.

Another factor is likely to be lack of mastery of higher Portuguese numerals by many Tetun Dili speakers. As noted in section 5.1, Portuguese higher numerals, in contrast to Tetun and Indonesian ones, are not regular. Added to this is the fact that (as noted in section 2, the majority of East Timorese do not have significant personal proficiency in

Portuguese. For this majority, most of their exposure to Portuguese numerals comes through Tetun, not through direct knowledge of Portuguese.

Despite these factors, we do expect to see an increase in the use of Portuguese numerals in future, at the expense of Indonesian ones. After all, friends have reported that even well into the Indonesian occupation, children in Dili used Portuguese numerals in play despite doing all their schooling in Indonesian. Portuguese numerals persisted despite the fact that use of the Portuguese language was actively discouraged by the occupying forces, and despite the large number of Indonesian military, police and civil servants living in the territory. Nevertheless, by the end of the 24-year occupation, almost all transactions amongst the general population used Indonesian numerals, with very little use of Portuguese numerals. There was clearly a lag between government and educational policy and the use of numerals on the ground. Now the same slow process of top-down pressure to replace one language by a new official language is well under way in the reverse direction, especially in more formal contexts. Timor-Leste's recent history suggests that eventually the use of Indonesian numerals will contract, with its niche being filled mainly by Portuguese.

Abbreviations

.I	Indonesian	FOC	focus
.P	Portuguese	Indon.	Indonesian
.T	native Tetun	INCL	inclusive
1	first person	LOC	locative
2	second person	M	masculine
3	third person	PL	plural
CLF:human	classifier for human nouns	Port.	Portuguese
DEF	definite	POS	possessive
EXCL	exclusive	SG	singular
F	feminine		

Appendix 1: Survey questions

The following are the questions asked in the surveys, along with a fairly literal translation into English, showing the source language of each of the key terms. Note that *hira* 'how many, how much' is native Tetun. In each case, questioners varied the word for 'you' to be appropriate to context, using terms such as *tiu* 'uncle.P' or *alin* 'young sibling.T'. For the cost of items, they chose items which the respondent was likely to have or know about.

Table 8. Survey questions asked in 2012 and 2018, along with the category of respondents for each year

Purpose	Question	2012	2018
Age	<i>Ita agora tinan hira?</i> 'You are now.P how many years.T?'	all	all
Low price	<i>Lapizeira / hudi sasuit / kadernu ida... folin hira?</i> 'A pen.P / hand of bananas.T / notebook.P ... price.T is how much?'	all	all
Time	<i>Aban ita tama aula/eskola/servisu tuku hira? Sai tuku hira? / Haan tuku hira?</i> 'Tomorrow you enter.T class.P / school.P / work.P how many o'clock.T? Leave how many o'clock? / Eat.T how many o'clock?'	all	all
Phone	<i>Ita nia número/nomor telefone hira?</i> 'Your telephone.P number.P/I is how much?'	student, adult	student, adult

Purpose	Question	2012	2018
High price	<i>Semester nee, imi estudante selu hira?</i> 'This semester.I, you students.P pay.T how much?' <i>Uma / kareta / motór nee folin hira?</i> 'This house.T / car.P / motorbike.P price.T how much?'	student, adult	student, adult
Time	<i>Agora tuku hira?</i> 'Now.P is how many o'clock.T?'	student, adult	student, adult
Date	<i>Ohin saa data?</i> 'Today.T is what.T date.P?'	student, adult	student, adult
Date	<i>Ita moris iha saa data?</i> 'You were born.T on what.T date.P?'	student, adult	student, adult
Official date	<i>Loron restaurasaun independénsia monu iha saa data?</i> 'Day.T of restoration.P of independence.P falls.T on what.T date.P?'	student, adult	student, adult
Family size	<i>Ita iha maun-alin nain hira?</i> 'You have older.brother-younger.sibling.T CLS:human.T how many?'	child, student	child, student
Class	<i>Ita agora semester / klas hira?</i> 'You are now.P in semester.I / class.I how many?'	child, student	child, student
Read	<i>Favor lee: 56, 3685</i> 'Please.P read.P'	child, student	all
Read	<i>Favor lee: 1/2, 3/4</i> 'Please.P read.P'	-	all
Read	<i>Favor lee: 86%, 2.6</i> 'Please.P read.P'	-	all
Read	<i>Favor lee: 256</i> 'Please.P read.P'	all	student
Read	<i>Favor lee: 3.500.000</i> 'Please.P read.P'	student	-
Count	<i>Sura took hau nia liman fuan. (2012)</i> 'Count.T try.T my fingers.T.' <i>Sura took hau kaer surat tahan hira? (2018)</i> 'Count.T try.T, I am holding.T how many sheets.T of paper.T.'	child	all
Credit points	<i>Ita foti SKS hira?</i> 'You take.T how many credit points.I?'	student	student
CC	<i>Ita nia motór boot ka kiik? Nia se-se hira?</i> 'Is your motorbike.P big or small? It's CC.I/P is how much?'	-	student, adult
Weight	<i>Foos saka ida nia todan hira?</i> 'Rice.T bag.P one.T, it's weight.T is how much?'	-	student, adult
Weight	<i>Ita nia todan hira?</i> 'Your weight.T is how much?'	-	adult

References

- Balle, Misriani & Stuart Cameron. 2014. The interplay of quantifiers and number in Helong grammar. In Marian Klamer & František Kratochvíl (eds.), *Number and quantity in East Nusantara*, vol. 1. 43–59. Canberra: Asia-Pacific Linguistics.
- Clark, Ross. 2004. "Necessary" and "unnecessary" borrowing. In Jan Tent & Paul Geraghty (eds.), *Borrowing: a Pacific perspective*. 33–9. Canberra: Pacific Linguistics.
- Costa, Luís. 2000. *Dicionário de Tétum-Português*. Lisbon: Edições Colibri.
- Direcção Geral de Estatística Timor-Leste. 2016. *Timor-Leste population and housing census 2015: Population distribution by administrative area. Volume 2 (Language)*. Dili: República Democrática de Timor-Leste.
- Donohue, Mark & Owen Edwards. 2014. Number in Tolaki. In Marian Klamer & František Kratochvíl (eds.), *Number and quantity in East Nusantara*, vol. 1. 27–42. Canberra: Asia-Pacific Linguistics.
- Eelens, Frank 2017. *Education Monograph 2015* (Timor-Leste Population and Housing Census 2015 Thematic Report Vol 11). Dili: Direcção Geral de Estatística Timor-Leste.
- Givón, Talmy. 1981. On the development of the numeral 'one' as an indefinite marker. *Folia Linguistica Historica* 2.1. 35–53.
- Hajek, John. 2006. Language contact and convergence in East Timor: the case of Tetun Dili. In Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds.), *Grammars in contact: a cross-linguistic typology*. 163–78. Oxford: Oxford University Press.

- Hull, Geoffrey. 1999. *Standard Tetum-English dictionary*, 1st edn. St Leonards: Allen & Unwin in association with the University of Western Sydney Macarthur.
- Hull, Geoffrey. 2000. Historical phonology of Tetum. *Studies in Languages and Cultures of East Timor* 3. 158–212.
- Hull, Geoffrey. 2001. *Mambai language manual (Ainaro dialect)*. Sydney: Sebastião Aparícu da Silva Project.
- Hull, Geoffrey. 2002. The languages of East Timor: some basic facts. Ms. Dili. Reprint Edition.
- Klamer, Marian & František Kratochvíl. 2014a. The expression of number in languages of East Nusantara: An overview. In Marian Klamer & František Kratochvíl (eds.), *Number and quantity in East Nusantara*, vol. 1. 1–14. Canberra: Asia-Pacific Linguistics.
- Klamer, Marian & František Kratochvíl (eds.). 2014b. *Number and quantity in East Nusantara*, vol. 1 (Papers from 12-ICAL). Canberra: Asia-Pacific Linguistics.
- Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2018. *Ethnologue: languages of the world*, 21 edn. Dallas, Texas: SIL International.
- Mathijssen, A. 1906. *Tetum-Hollandsche woordenlijst met beknopte spraakkunst*. Batavia: Albrecht & Co.
- Matras, Yaron. 2007. The borrowability of structural categories. In Yaron Matras & Jeanette Sakel (eds.), *Grammatical borrowing in cross-linguistics perspective*. 32–73. Berlin: Mouton de Gruyter.
- Matras, Yaron. 2011. Universals of structural borrowing. In Peter Siemund (ed.), *Linguistic universals and language variation*. 200–29. Berlin: Mouton de Gruyter.
- Morris, Cliff. 1984. *Tetun - English dictionary*. Canberra: Pacific Linguistics, C-83.
- National Statistics Directorate. 2006. *Timor-Leste census of population of housing 2004: Atlas*. Dili: National Statistics Directorate.
- Ross, Melody Ann. 2012. Survey on Language Attitudes and Education in Timor-Leste. *Studies in the Linguistic Sciences* 674. 1–23.
- Schachter, Paul & Fe T. Otnes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.
- Schapper, Antoinette. 2009. *Bunaq: A Papuan language of central Timor*. dissertation, The Australian National University.
- SIL International. 2011. Field Linguist's Toolbox. Computer program. SAG. <http://www.sil.org/computing/toolbox>
- Souag, Lameen. 2007. The typology of number borrowing in Berber. In Naomi Hilton et al. (eds.), *Proceedings of the Fifth University of Cambridge Postgraduate Conference in Language Research*. 237–44. Cambridge: Cambridge Institute of Language Research.
- Taylor-Leech, Kerry. 2007. *The ecology of language planning in Timor-Leste. A study of language policy, planning and practices in identity construction*. dissertation, Griffiths University.
- Tilman, Armindo da Costa. 1996. *Matadalan nosi Tetun: Timór lian*. Lisbon: Armindo da Costa Tilman.
- van den Heuvel, Wilco. 2007. Grammatical borrowing in Biak. In Y. Matras & J. Sakel (eds.), *Grammatical borrowing in cross-linguistics perspective*. 329–42. Berlin: Mouton de Gruyter.
- van Klinken, Catharina. 1999. *A grammar of the Fehan dialect of Tetun, an Austronesian language of West Timor*. Canberra: Pacific Linguistics, C-155.
- Williams-van Klinken, Catharina. 2003. *Peace Corps East Timor Tetun language course*, 1st edn. Dili: Peace Corps East Timor.
- Williams-van Klinken, Catharina. 2012. *Akua rua, satu dolar: Mixing numeral systems in Timor Leste*. Paper presented at 12th International Conference on Austronesian Linguistics, Denpasar, Indonesia.
- Williams-van Klinken, Catharina & John Hajek. 2018. Language contact and functional expansion in Tetun Dili: The evolution of a new press register. *Multilingua*.
- Williams-van Klinken, Catharina, John Hajek & Rachel Nordlinger. 2002. *Tetun Dili: a grammar of an East Timorese language*. Canberra: Pacific Linguistics, 528.

Ximenes, Abreu & Catharina Williams-van Klinken. 2014. Futun ida, satu dolar: Kahur sistema numeru iha lingua Tetun (*One bunch, one dollar: Mixing numeral systems in Tetun*). In Hannah Loney et al. (eds.), *Understanding Timor-Leste 2013*, vol. 1. 66–71. Dili, Timor Leste: Timor Leste Studies Association.