Catharina Williams-van Klinken* and John Hajek

Language contact and functional expansion in Tetun Dili: The evolution of a new press register

<https://doi.org/10.1515/multi-2017-0109>

Abstract: Tetun Dili, an Austronesian language spoken in East Timor, was until 1999 primarily an oral language of intercultural communication. Since the 1999 vote on independence from Indonesia, Tetun Dili has become the dominant language of public life, including the government, education and the media, as well as becoming an official language alongside Portuguese, a former colonial language. The rapidly evolving press register of Tetun shows significant impact from language contact. Portuguese influence is seen primarily in extensive lexical borrowing, brought in by the Portuguese-educated elite as well as by translators and writers. Indonesian influence is seen in several calques for expressing anaphora, brought in by Indonesian-educated writers, and an adversative passive. Other new constructions, including a more general passive and final quote margins, have come about through the combined influence of Portuguese, Indonesian and English, the last as a source language for much literal translation into Tetun. Some discourse features of press Tetun, such as high information density and events being told out of chronological order, are the results of international journalistic style. While a rapid evolution of new written registers is not uncommon, the East Timor situation is unique in its combination of contact languages, and the significant number of grammatical innovations.

Keywords: Tetun Dili, media, language contact, language change, Timor-Leste

1 Introduction

After hundreds of years of colonisation, first by Portugal and then by Indonesia, East Timor (officially Timor-Leste) achieved independence in

*Corresponding author: Catharina Williams-van Klinken, Centre for Language Studies, Dili Institute of Technology, Dili, Timor-Leste, E-mail: cwvk@livedream.tl

John Hajek, Research Unit for Multilingualism and Cross-cultural Communication, University of Melbourne, Melbourne, Australia, E-mail: j.hajek@language.unimelb.edu.au

2002. Its constitution names two official languages, the Austronesian language Tetun, and the Romance language Portuguese. Before independence, Tetun (also known as Tetun Dili or Tetun Prasa) was largely an oral language of intercultural communication, used in informal settings, the chief exception being its use by the Catholic Church in the mass. However, beginning with the United Nations run ballot on independence from Indonesia in 1999, Tetun was propelled quickly to the fore. Since that time it has developed rapidly as a language of public life, including the press and other media, government and education.

Within the press, there has been a rapid development of a specific register, with characteristic vocabulary, grammar and discourse. This new press register shows strong influence from the former colonial languages Portuguese and Indonesian, as well as from journalistic writing in general. Portuguese influence, most immediately evident at the level of new vocabulary, enters to a large extent through the Portuguese-educated elite, who constitute a major source of stories and quotes in newspapers. Indonesian influence where it occurs is most likely to enter through the Tetun used by Indonesian-educated reporters who dominate local journalism. A number of recent grammatical innovations in Tetun writing we have identified below occur similarly in Portuguese, Indonesian and English writing, and are also found in the ubiquitous literal translations into Tetun of English and Portuguese (and to a much lesser extent Indonesian) documents.

As such, Tetun provides an interesting case study of the development of a new press register within a very short period of time, as well as of the role of the press in both reflecting and influencing linguistic variation and innovation (Percy 2012). Press writing is known to show its own linguistic characteristics driven by discourse needs and journalistic conventions (e.g. Biber and Hared 1994; Wallace 1977), some of which can then filter into other registers, both spoken and written (e.g. Reder 1981). The extensive literature on language contact (e.g. Matras and Sakel 2007) has made very little mention of the role of press writing. An exception is Shlesinger (cited in Rosenhouse and Fisherman (2008: 141–142)), who sees Hebrew journalese with its high rate of English loans as the mediator between foreign languages and Hebrew. Writing in general has been more widely recognized as important in mediating change, including structural change. For instance, Thomason and Kaufman (1988: 79–80) mention the impact of Sanskrit on the structure of literary Dravidian languages, and of literary Chinese on Japanese. Kalmár (1985: 161) describes the innovation of subordination in the Inuit language Inuktitut through contact with English, which he attributes to the influence of writing. Often, however, changes brought about through writing are found in the frequency of constructions rather than

the development of new constructions (e.g. Romaine (1994) on Tok Pisin and Kilham (1987) on languages in the South Pacific).

In the case of Tetun, this is a language that has only recently been formally codified and introduced into the education system, where written texts are increasingly available to schoolchildren and students. For adults, newspapers are likely to be the only published material they interact with on any regular basis. For these citizens, therefore, press writing is the primary written context which could influence their language use more generally. However, it should be noted that newspaper consumption remains relatively low, and many people, especially outside Dili, cannot afford to buy a newspaper.

The focus of this study is to investigate the linguistic evolution of Tetun's press register. We seek to identify and describe the lexical and structural characteristics of Tetun press writing, and to identify the source of these characteristics. In particular, we examine the impact of contact with the two colonial languages Portuguese and Indonesian, in addition to English, which is a latecomer to East Timor and has a much more limited role. We also look at the influence of the journalistic genre *per se*.

In the sections that follow we present our methodology and corpus, outline the history of the language situation in East Timor, examine which languages have been used in the press in recent decades, and then present some of the linguistic characteristics of Tetun press writing.

1.1 Methodology and corpus

The data for this paper was collected in various ways over a period of almost twenty years. For studying the characteristics of press language, we primarily used a large computerised corpus of newspaper articles. Almost all of these date from 1999 onwards, since this was the first time newspapers used much Tetun, but there are a few earlier sources as well. The corpus also includes translations, other written materials and oral texts, enabling comparison with other genres. This corpus was supplemented by searches of the increasing volume of online newspapers and texts.

To determine the languages used in newspapers, one edition of each newspaper was analysed in 2009, and one edition in 2016. Dili Institute of Technology language lecturers handled data collection, counting the number of pages published in each language in each edition, noting what functions each language was used for, and checking which articles had been obtained from the internet.¹

¹ Data handling in 2009 was by Hendriana da Costa Marçal, Alexandre Fernandes Cham, Anabela Maia Santos, Helio Brites da Silva, Guido Diamantino and the late Jacinta Canossa Soares. In 2016 it was undertaken by Alexandre Fernandes Cham and Marta Cardoso.

For attitudes to innovations, the first author talked with her language teacher colleagues and Tetun language students at Dili Institute of Technology, with fellow linguists, with journalists at Suara Timor Lorosae, and with translators who attended various translation training courses that she led.

Data at the word level, such as estimates of the percentages of loans for each part of speech, were derived from the first author's extensively annotated lexical database. A selection of the more commonly known lexemes in this database is published as Williams-van Klinken (2011).

1.2 The language situation in East Timor prior to 1999

The linguistic situation in East Timor is complicated, and reflects its unusual colonial and post-colonial history. Estimates of the number of indigenous languages in East Timor vary, depending partly on how one distinguishes between languages and dialects. Hull (2002a: 381) places the number at 16, while Lewis et al. (2016: 191–192) put it at 19.

One of the Austronesian languages, Tetun Terik, formed the basis for a lingua franca version of Tetun which was already in use throughout much of the territory when Portugal first established a colonial presence in the 1500s (Thomaz 1981). The Tetun now used as an official language is ultimately derived from that lingua franca. In common parlance this variety of modern Tetun is also known as Tetun Dili (named after the capital Dili) or Tetun Prasa (after the old term for the capital).

After centuries of trade and indirect rule, Portugal ruled East Timor fairly intensively for over a century until 1975. Throughout this period, Portuguese was the primary language of government, church, education and the media, and virtually the sole language of writing. Although hardly used in writing at the time, Tetun was the language most widely used for interethnic spoken communication. This lengthy and extensive contact with Portuguese left an indelible mark on Tetun Dili, mainly through large-scale borrowing.

The first known written materials in Tetun appeared at the end of the nineteenth century, with the Catholic priest da Silva producing a catechism (1885) and a 431-page Portuguese-Tetun dictionary (da Silva, 1889). Over the following century, the vast majority of Tetun writing was religious, including prayer books, catechisms and Scripture portions, although some folk-tales and songs in Tetun were also published (Hull 1996: 46). The result was the development of a liturgical variety of Tetun, which shows a strong influence of Tetun Terik (Hull 1996: 46; Williams-van Klinken 2002).

Throughout the Indonesian occupation which started with the 1975 invasion, the official and public functions of Portuguese were taken over by Indonesian. Indonesian policy was focused on rapid Indonesianisation of the population, through mass education, a huge influx of Indonesian speakers and exclusive use of Indonesian in all official domains. Tetun Dili, however, remained the preferred spoken lingua franca amongst the East Timorese themselves. Portuguese continued to be used by the Catholic Church in the mass until the early 1980s, after which it adopted Tetun as its only liturgical language, under Indonesian pressure to eliminate Portuguese entirely. During the initial years of the occupation, Tetun writing received a boost through the Fretilin literacy program, with educators in areas as yet outside Indonesian control using Tetun for literacy instruction (Hull 1996: 50). Within the resistance movement, Tetun was also used, alongside Portuguese, in letters between Fretilin members, in lists (such as records of equipment), in song lyrics and anthems, in announcements and declarations, and in reports of human rights violations, appeals and petitions (Cabral and Martin-Jones 2008: 166).

2 The language situation from 1999

In 1999, East Timorese voted for independence from Indonesia in a ballot run by the United Nations (UN). It was then placed under United Nations supervision, which lasted from late 1999 until independence in May 2002.

At this time, East Timorese started speaking Tetun in formal public contexts, although most still wrote in either Portuguese or Indonesian, depending on the language of education of the writer. Schools also immediately started a now completed phase-in of Tetun and Portuguese to replace Indonesian as the languages of instruction. As a result, younger Timorese are significantly less proficient in Indonesian than those who were educated in that language during the Indonesian occupation.

With full independence in 2002, East Timor's constitution (Democratic Republic of Timor-Leste 2002) specified Tetun and Portuguese as the country's official languages (article 13), with Indonesian and English accepted as working languages to be used in the civil service alongside the official languages "for as long as deemed necessary" (article 159).

In 2002, too, the National Institute of Linguistics (*Instituto Nacional de Linguística*, or INL) was set up. Government decree 1/2004 (April 2004) states that "The INL is the scientific custodian of Official Tetum" (article 4), defining "Official Tetum" as follows: "The variety of Tetum recognized as the official

and [first] national language is Official Tetum, a modern literary form of the vernacular most widespread in the country and based on Tetum-Praça” (article 2). The INL has continued the tradition of Hull (1999) in focusing on acrolectal Tetun Dili, with numerous loans from Portuguese and Tetun Terik, as evidenced by its large monolingual Tetun dictionary (Correia and Guterres 2005). Hull’s 1999 dictionary included an appendix of Indonesian loans in common use, but this was removed in the 2002 edition (Hull 2002b), in keeping with the almost universally held view that Indonesian loans are not appropriate in formal contexts. In addition to adopting words from Portuguese and Tetun Terik, the INL has created some new expressions (e.g. for ordinals) based on native Tetun morphemes.

Since 1999, the knowledge and use of Tetun Dili has spread rapidly. In 1999 approximately 7 % of East Timorese spoke Tetun Dili at home. By the 2004 census the percentage claiming to speak it at home was 24 %, by the 2010 census it had risen to 36 % (Williams-van Klinken et al. 2016).

Estimates as to what proportion of the population speak or write each of the four official and working languages vary widely. According to the 2010 census (National Statistics Directorate and United Nations Population Fund 2011: xvii), 56 % of the population over the age of 15 years claim to be literate in Tetun, 45 % in Indonesian, 25 % in Portuguese, and 15 % in English. Based on the authors’ own experience within a university context as well as in rural areas, the percentages for Portuguese, and particularly for English, seem very generous. Literacy rates are higher in urban areas than rural ones, and higher amongst men than women, with the result that 81 % of men in Dili claim literacy in Tetun, while only 39 % of rural women do. Literacy in Tetun has been rising markedly, being much higher amongst the young than the old.

As a result of the changing official languages over the last decades, we are left with an unusual dichotomy in language capacity. The vast majority of the Timorese political elite were educated in Portuguese before 1975, and the Tetun they use shows significantly greater influence of Portuguese than that of other East Timorese. Many who were overseas or who were in the guerrilla force during the Indonesian occupation speak good Portuguese but have limited, if any, capacity in Indonesian. On the other hand, the vast majority of journalists are relatively young and were educated during the Indonesian occupation. Very few speak Portuguese well. This linguistic distinction underlies press reporting in Tetun, the only common language of both parties, and, as we aim to show below, affects linguistic developments in that context. The influence of the elite is huge, as they are given substantial space in press reporting, with widespread interviewing and direct quoting. Alongside this is narrative and commentary by Indonesian-educated journalists.

Another influence on Tetun Dili writing was the arrival of the UN and other English-speaking institutions after 1999. This resulted, especially in the early years, in copious and often rushed translations of written texts into Tetun (and often also into Indonesian and Portuguese) from English, as the working language of the UN. This was an entirely new phenomenon, and encouraged the development of a recognizable translationese.

The fact that the written registers of Portuguese, Indonesian and English share many features, such as the marked presence of passives, means that all are potentially influencing Tetun in the same ways, making it harder to identify the specific source of an innovation in that language – a point we return to in the final discussion.

3 Language choice in the newspapers

Prior to 1999, all newspapers were in the colonial languages. Towards the end of Portuguese rule, there were a few locally-produced Timorese periodicals in Portuguese, including the local Catholic diocesan periodical *Seara*, an army journal *A Província de Timor* ‘The Province of Timor’, and the official journal of the provincial government *A Voz de Timor* ‘The Voice of Timor’ (Pires 2001).

From 1975 to 1999 all newspapers were in Indonesian. The only locally-produced newspaper was *Suara Timor Timur* ‘The Voice of East Timor’, established in 1993 (Hopper 2008: 490). Although in Indonesian, it included an occasional human-interest story in Tetun. The student resistance group RENETIL published a Tetun-language underground bulletin called *Loriku Lian* ‘Voice of the Lorikeet’ (Hopper 2008: 500), although in the circumstances it is likely to have appeared irregularly and with very limited distribution. The major producer of written Tetun language materials at this time remained the Catholic Church, which translated many materials into Tetun, including the New Testament (Diocese de Dili 2000). In doing so they further developed the liturgical variety of Tetun. This variety has remained largely restricted to the religious sphere, with some flow-on into formal speech, but without having much impact on the later development of a press register or on informal spoken Tetun.

This situation changed dramatically when East Timor voted for independence from Indonesia in 1999. Tetun immediately began to be used in a much expanded role in newspapers, as well as on radio and television. It was one of the four languages used by the 65 journalists employed by the UN Transitional Authority in East Timor to work in print media, radio and television (Hopper

2008: 488). The first newspaper to be printed in Tetun, *Timor Lorosae Foun*, appeared in fourteen editions from 22 October 1999 until 21 January 2000. It was published by the United Nations in Dili alongside its equivalent Indonesian edition *Timor Timur Baru*, as translations of the English language *The New East Timor*. This was quickly followed by another periodical *Tais Timor* also established by the United Nations. It was published until 2002 in the original English with Tetun, Indonesian and Portuguese editions prepared by translators rather than journalists.

In May 2000 the first locally managed and edited newspaper appeared with a bilingual masthead, namely *A Voz de/Suara Timor Lorosae* (STL). This was the reborn edition of *Suara Timor Timur* which had folded in early September 1999 as a result of the violent post-ballot chaos. Although this newspaper still published mainly in Indonesian, it also did substantial local reporting in Tetun – the first time that articles were written directly in Tetun.

Since then, despite having a population of only about a million, East Timor has seen the rise of a number of daily and weekly newspapers, some of which (such as the English-language weekly *Timor Sun*) have since folded. Almost all have included a substantial proportion of Tetun. Many have included Indonesian, the language of education for all journalists and most readers. Some have included Portuguese, as an official language and the primary language of the ruling elite. And one or two have focused on English, the primary language of the United Nations and other international organizations working in East Timor.

By May 2009, there were three daily newspapers (*A Voz de/Suara Timor Lorosae*, *Timor Post*, and *Jornal Nacional Diário*) and two weekly newspapers (*Tempo Semanal* and *The Dili Weekly*). By November 2016, there were two additional daily newspapers (*Diariu Timoroman* and *Diariu Independente*) and one weekly (*Business Timor*). Most also have an online presence. In addition there are now news sources that are only available online, such as *Tatoli* and *Forum Haksasuk*.

The language practices of the newspapers initially differed markedly, but have converged significantly, with a much increased role for Tetun. For the purposes of comparison, we analyzed one edition of each of the daily and weekly newspapers in 2009, and one edition of each newspaper in 2016.

In 2009, the newspapers differed greatly in their language practices. As detailed in Table 1, the three daily newspapers had most articles in Tetun or Indonesian, while *Dili Weekly* printed all articles in both Tetun and English, and *Tempo Semanal* (which published only local news) was

Table 1: Languages used in newspapers in 2009. (Note that sport and entertainment were all international.).

Newspaper	Date	Tetun	Indonesian	Portuguese	English
Suara Timor Lorosae	2/6	4.5 pages: national, editorial	6.5 pages: national, international	–	–
Tempo Semanal	20/5	all 15 pages: national, opinion	–	–	–
Timor Post	2/6	7 pages: national, editorial	6.5 pages: international, opinion on Timor	1 page: international	0.5 pages: international
Dili Weekly	14–20/5	7.5 pages: national, editorial	–	0.5 page: opinion	8 pages: (Same articles as in Tetun)
Jornal Nacional Diário	2/6	6 pages: national, editorial	4 pages: international	4 pages: international	–

entirely in Tetun. In all cases Tetun was used only for national news, editorials and opinion pieces. Any international news, including sport and entertainment, was in Indonesian, in English (in the case of *Dili Weekly*), or (mainly in *Jornal Nacional Diário*) in Portuguese. In those editions of the newspapers that we analyzed, all international news articles were taken directly from the internet. This is merely a low-budget version of what most news outlets do worldwide, with nearly all international news being generated by only four international news agencies (Bell 1991). At that time, only *A Voz de/Suara Timor Lorosae* still had journalists regularly writing some national news in Indonesian. No newspapers regularly published original material of any kind in Portuguese or English except for the bilingual *Dili Weekly*.

By 2016 the language practices were much more uniform, and the percentage of Tetun had further increased. As can be seen in Table 2, all newspapers had a large majority of writing in Tetun, with the exception of *Business Timor*, which had almost as much Indonesian as Tetun. As in 2009, all used Tetun for national news, editorials, and opinion pieces. The major development was that two also used Tetun for international news, and, on the date of our survey, there were two locally written articles in Portuguese (in *Timor Post*) and one

Table 2: Languages used in newspapers in 2016.

Newspaper	Date	Tetun	Indonesian	Portuguese	English
Suara Timor Lorosae	8/11	10 pages: national, editorial, profile	3.5 pages: opinion, international	1 page: international	–
Diariu Timoroman	8/11	6 pages: national, editorial	4 pages: international, opinion	0.25 page: national	–
Jornal Nacional Diário	8/11	10.5 pages: national, editorial	4.5 pages: international	1 page: international	0.5 pages: opinion
Timor Post	8/11	9.5 pages: national, international, editorial, opinion	4.25 pages: international	0.5 pages: national	0.5 pages: international
Tempo Semanal	8/11	6 pages: national	4.5 pages: international,	1 page: international	1 page: international
Diariu Independente	8/11	6 pages: national, editorial, opinion	2.5 pages: international	1 page: international	1.5 page: national, international, opinion
Business Timor Weekly	7–13/11	10.5 pages: national, editorial, international	10 pages: opinion, international	1 page: national, international	1 page: national

locally written article in English (in *Diariu Independente*). Most international news remained in Indonesian, taken directly from the internet. All seven newspapers gave a page or less to Portuguese, while five allocated approximately a page to English.

4 Development of a press register

Even when Tetun Dili still functioned primarily as a spoken lingua franca amongst the East Timorese, it already showed a fairly high degree of language contact with Portuguese, falling into the middle category of a five-point scale of borrowing proposed by Thomason and Kaufman (1988: 74), in that it has many words borrowed from Portuguese, including prepositions and one derivational suffix *-dor* (Hajek and Williams-van Klinken 2003).

Until 1999, however, there was very little structural borrowing, a characteristic of Thomason and Kaufman's fourth category of "strong cultural pressure". The only grammatical influence we are aware of prior to that time is the possibility of putting some intransitive verbs before the subject, on the model of Portuguese, as seen in example (1).

- (1) *Ikus ne'e, iha Cidade Dili mosu Warung barak, special*
 end this LOC City Dili appear local.restaurant many especially
faan naan Asu.
 sell meat dog
 'These days, in the city of Dili, many Indonesian-style restaurants have appeared, in particular those selling dog meat.' (Suara Timor Timur 04/09/1998)²

Within a very short time of Tetun becoming a regular press language, a distinct press register had developed (Williams-van Klinken 2002), illustrating Ure's (1982: 6) comment that "[t]he register change of a language is one of the most immediate ways in which it responds to social change." This new register was characterized by an elevated proportion of Portuguese loans, avoidance of Indonesian loans and the presence of final quote margins.

Since that time there have been further developments in the press register of Tetun, a number of which are outlined below. The following sections discuss in turn vocabulary, including the appearance of new sub-classes of adjectives and adverbs, an expanded choice of verbs of speaking, grammatical innovations, and discourse characteristics.

5 Vocabulary

5.1 Loans and innovations

As early as 2002, Williams-van Klinken (2002) noted that "[t]he primary characteristic of press Tetun is extensive Portuguese influence," primarily in vocabulary.

Tetun Dili has a large proportion of words borrowed from other languages, mainly Portuguese. Most of the basic vocabulary is native Austronesian, with

² Spelling in Tetun Dili is variable. In the examples, the spelling follows that of the source.

less than 5 % of the 207 words in the Swadesh list of basic vocabulary coming from Portuguese. However, in technical and high level registers, the percentage of Portuguese increases markedly. Many of the Portuguese loans in everyday use were introduced into Tetun during the lengthy period of Portuguese rule. During the Indonesian period (1975–1999), borrowing from Indonesian became common, and borrowing from Portuguese was largely interrupted, although some people concentrated on increasing their Portuguese vocabulary as a mark of resistance (Edio da Costa, pers. comm.). After 1999, borrowing from Portuguese recommenced with renewed vigour, especially in more formal registers, where borrowing from Indonesian is almost universally considered to be inappropriate.

There are a number of reasons why new high-level Portuguese loans are constantly being introduced into Tetun Dili. The first is the need to efficiently express new and complex concepts which were previously handled in the colonial languages, in areas such as politics, economics, and technology. This is a common issue in languages with newly expanded roles. Romaine (1994), for example, reports the massive borrowing of English loanwords in Tok Pisin newspaper sports writing, even though in this case the language has the capacity to coin new words. In the case of Tetun Dili, it has very little productive morphology, so that the options for expressing new abstract and technical concepts are usually either to borrow lexical items or to use lengthy paraphrases, an issue discussed by Williams-van Klinken (2004) with regard to electoral terminology. Since the press gives priority to saving print space, even at the cost of possibly reducing reader comprehension, massive borrowing is essential, and by general consensus it must be from the co-official language Portuguese.

Another factor supporting heavy use of Portuguese loans in the press is that Portuguese was the language of education of the national elite, which means they are both equipped and prone to using Portuguese words when speaking or writing Tetun. Similarly, translators and writers consciously looking for equivalents for terms in other languages, mainly English and Indonesian, turn to Portuguese loanwords. This often allows them to find words which exactly parallel the source terms semantically, and in many cases even phonologically, thanks to shared Latin roots, e.g. English *education* ~ Portuguese *educação*, the latter appearing in Tetun orthography as *edukasaun*. New loans, especially when used by the elite and cited in press reports, are then effectively mediated to the public by journalists. As with Dutch loans in Indonesian (Tadmor 2007: 303), these Portuguese loans are not triggered by wide-spread bilingualism in Portuguese, with only a small minority of East Timorese being fluent in the language.

Portuguese loans are traditionally adapted to Tetun grammar, with singular masculine forms of nouns and adjectives normally used regardless of the number or gender of the referent. Increasingly, though, journalists and other writers, and indeed speakers, are imitating the tendency of heavily Portuguese-influenced speakers to use Portuguese rules of number (singular v. plural), gender (masculine v. feminine) and word order on Portuguese noun phrases. One example is using a Portuguese phrase with preposed ordinal adjective agreeing with the noun, as in *primeira faze* (lit. ‘initial.FEM phase’), instead of *faze primeira* (lit. ‘phase first.MASC’) ‘initial phase’ with standard Tetun word order and lack of agreement. However, Portuguese agreement rules are not always followed correctly, such that, for instance, plural adjectives may be used with singular nouns, something impermissible in Portuguese, e.g. *área rurais* ‘rural area(s)’ (lit. ‘area rural.PL’), as opposed to Portuguese singular *área rural* or plural *áreas rurais*.

A particular challenge for Indonesian-educated journalists is that when they come across unknown Portuguese words as they interview the Portuguese-educated elite, they have no comprehensive Portuguese-Indonesian dictionary in which to seek the meaning. As a result, they necessarily use some new words without having an in-depth understanding of their meaning and usage themselves.

In the first author’s database of Tetun Dili lexemes, 52% of the 9,400 nouns are from Portuguese, as are 42% of 2,500 verbs. The lower percentage for verbs than nouns is consistent with Matras’ (2007: 48) observation that verbs are harder to borrow than nouns, even if a language has no verbal morphology. The only categories of words that are not borrowed into Tetun are pronouns, determiners, quantifiers, and grammatical markers such as possessive markers, the focus marker, negatives and tense-aspect markers.

Approximately one third of all word tokens in press Tetun are Portuguese loans, although the percentage varies according to topic and style. In a selection of seven articles from varied newspapers in 2009, an average of 32% of word tokens were Portuguese loans. Most articles were in the range 29–41%, while one written about Christmas in simple readily comprehensible Tetun had only 18% Portuguese. In 2016, a sample of 8 articles from a range of newspapers gave an average of 35% Portuguese. One article, about the impact of the delayed wet season on farmers, had only 24% Portuguese, while the others were in the range 29–47%. Such variability is to be expected given that what we are calling the ‘press register’ has multiple sub-registers (Biber and Hared 1994).

This compares to some 10–25% Portuguese loans for informal urban speech, up to 30% for formal speech (Williams-van Klinken 2010a), about 10% for liturgical Tetun, and less than 10% for rural spoken Tetun.

For open class words (nouns, verbs and adjectives), the percentage of Portuguese loans in the sampled newspapers from 2009 was as high as 75% in one article. In the 2016 sample it averaged 71%, with a range from 55–80%. The result is that the written text looks like Portuguese words linked together by a mixture of Tetun and Portuguese function words. This is illustrated by example (2), in which Portuguese loans are underlined.

- (2) *Komunidade sira preokupa estrada diresaun Suku Lauhata*
 community DEF.PL concerned road direction village (name)
liga
 connect
 ‘The community is concerned that the road linking Lauhata village
ba Postu Bazartete Munisipiu Likisá too agora
 to subdistrict (name) district (name) until now
kontinua *aat*
 continue bad
 with the subdistrict town of Bazartete continues to be bad,
hodi afeta movimentu transporte ba mai.
 to/and affect movement transport go come
 affecting the movement of transport to and fro.’ (Independente 22/5/2017)

Based on informal comments made to us over many years, many East Timorese are aware that the press uses a high proportion of Portuguese, although not quite the extent of it until that point is brought to their attention. This was illustrated by the staff member doing the mark-up of word sources in the 2009 newspaper corpus, who exclaimed in amazement, “There’s no Tetun here!”

Many of the open class words of Tetun origin that do appear in the press have low semantic contribution to the sentence. Some are semantically redundant, such as *loron* ‘day’ in *loron Sabadu* ‘day (Tetun) Saturday (Portuguese)’. Others are generic verbs like *halo* ‘do’, *halao* ‘carry out’ or *iha* ‘have’, for which most of the meaning is contributed by the following object, such as *halo avaliasaun* ‘do evaluation’ ‘evaluate’.

In contrast to Portuguese, Indonesian loans are strongly avoided by the media, except where they come up in interviews. Only about 0.3% of tokens in the seven newspaper articles analyzed from 2009 were from Indonesian. In the 2016 sample of eight articles, given the general consensus on avoiding Indonesian loanwords in writing, one article on agriculture had 1.7% Indonesian, while all the rest had none. This compares to some 7% Indonesian in casual urban youth speech (Williams-van Klinken 2010a).

English appears even more rarely in press writing. Although rare in the press, English and English-based words are quite commonly used by some NGOs (non-government organizations) and, as a result, may appear in press articles for which these NGOs were interviewed. The words are sometimes adopted via Indonesian, and may be adapted to sound more like Portuguese, resulting in words such as *komitmentu* (Indonesian *komitmen*) to mean ‘commitment’ and *asesmentu* for ‘assessment’ (Godinho 2014).

A fourth source of loans is the vernacular rural variety of Tetun, known as Tetun Terik. A small number of these have entered formal Tetun and the press register through the conservative liturgical Tetun, including *no* ‘and’ (from *n-o* ‘3SG-and/with’), *katak* ‘say’ (in Tetun Dili it is normally the conjunction ‘that’), *tenik* ‘say’, and *hafoin* ‘then’.

There are also a few Tetun innovations, used either instead of or in addition to the more usual Portuguese loans. These include *uma fukun* ‘clan house’ used in parallel with *parlamentu nasionál* ‘national parliament’ to mean ‘parliament house, parliament’, and *sorumutu* (from *hasoru hamutuk* ‘meet together’) instead of the common Portuguese loan *enkontru* ‘meeting’.

Another source of vocabulary comes from the new terms proposed by the National Institute of Linguistics (Correia and Guterres 2005; Hull and Eccles 2001). A few of these have gained traction in the press and other media, particularly *ba dala* ‘for time’ and the invented prefix *da-* to create an equivalent of ordinal numerals, such as *ba dala lima* ‘for time five’ and *da-lima* ‘ORDINAL-five’ to mean ‘fifth (time)’.

5.2 New sub-classes of adjectives and adverbs

As noted above, some 52% of nouns are borrowed from Portuguese. The percentage of loans is even higher for adjectives, with 67% of 1,180 adjectives in the first author’s lexical database being from Portuguese. Many of these are high-level vocabulary. Amongst these loans, there are two new sub-categories of adjective. Both are used almost exclusively in press and technical writing and in formal speech by educated people.

The first sub-category involves adjectival variants for what would normally be expressed by nouns in Tetun. Examples include *agríkola* ‘agriculture-’ as opposed to *agrikultura nian* ‘agriculture POS’, *Aziátiku* ‘Asian’ instead of *hosi Ázia* ‘from Asia’, and *infantíl* ‘infant-’ instead of *labarik nian* ‘child POS’.

The second sub-category consists of past participle forms of Portuguese verbs which are used adjectivally in Tetun (as they can be in Portuguese), such as *aprovadu* ‘approved’, where everyday Tetun would use *nebee*

aprova tiha ona (lit. REL approve already PRF) ‘which (someone) has approved’.

Of the 250 adverbs in the first author’s lexical database, 36 % are borrowed from Portuguese. Most of these loans (67 %) end with the Portuguese adverbialising suffix *-mente*, such as *globalmente* ‘globally’ and *ofisialmente* ‘officially’. In our corpus of oral texts, only *diretamente* was found before 2006. Since that time, however, *-mente* adverbs have become very common in press writing, other written texts and formal discourse, though not yet in informal speech or amongst less educated people.

These three new sub-categories of adjectives and adverbs allow for more direct translation from Portuguese, Indonesian and English, and in some cases allow one to say in a single word what would traditionally have required a separate phrase or clause. For instance, some *-mente* adverbs enable one to give a comment on a statement while one is making it (such as *infelizmente* ‘unfortunately’ in (3)), or to specify the domain about which one is talking, using a term such as *ekolojikamente* ‘ecologically’. Their appearance in press writing in the early 2000s was sudden and striking to the outside observer. They have obvious functional value in reducing the word count in newspapers where space is at a premium.

- (3) *Infelizmente, projeitu lei ida ikus nee mos nia* *kualidade*
 unfortunately project law one end this also 3SG.POSS quality
laduun diak.
 not.very good
 ‘Unfortunately, this latest legal bill too is of poor quality.’ (Suara Timor
 Lorosae 31/3/2007)

5.3 Increased range of verbs of speaking

One of the characteristics of journalism in general is a frequent use of direct quotes. In quote margins, the Tetun press uses verbs that are in regular use in the community. In addition, as in journalism in other languages (Wallace 1977), journalists use many verbs that are seldom used elsewhere, to avoid repetition. As can be seen in the following list, most of these are from Portuguese, but some are from Tetun Terik, brought in through the conservative liturgical register of Tetun. Many potentially allow fine distinctions of meaning that are not available in spoken Tetun, although in practice some, such as *salienta* and *sublinha* are

often used to mean just ‘say’. Some of these verbs also allow writers to indicate speaker attitude, e.g. *lamenta* ‘lament’.

In common use:

dehan ‘say’
deklara ‘declare’ (Portuguese)
esplika ‘explain’ (Portuguese)
foo sai ‘reveal’ (lit. give exit)
hatete ‘say’
hatoo ‘pass on’
hatutan (tan) ‘add (more)’

husu ‘ask’
informa ‘inform’ (Portuguese)
tuir X katak ‘according to X’ (lit. follow X that)

Tetun Terik-based terms mainly used in media:

haktuir ‘tell, recount’
katak ‘say’ (this is usually the conjunction ‘that’)
tenik ‘say’ = *dehan*
kasu lia(fuan) hirak nee ‘said this’ (lit. take.off
 word these.specific this)³

Portuguese terms mainly used in media:

afirma ‘affirm, state’
akresenta ‘add’ = *hatutan tan*
alega ‘allege’
anunsia ‘announce’ = *foo sai*
apela ‘appeal’
esklaresu ‘explain’ = *esplika*
informa kestaun nee ‘speak on
 this issue’
kestiona ‘question’
komenta ‘make a comment’
lamenta ‘lament’
mensiona ‘mention’ = *temi*
reforsa ‘stress’
relata ‘give an account of’
salienta ‘stress, point out’
sublinha ‘underline, stress’
sujere ‘suggest’

5.4 The impact of little-known words on text comprehension

When journalists first started writing in Tetun, they used a high proportion of words which were not known in the general community. The vast majority of these were from Portuguese. In 2001, the first author tested the intelligibility of newspaper articles with tertiary-educated native speakers of Tetun Dili, and found that “it was not uncommon to find several words in a single paragraph for which they could not even hazard a guess at the meaning” (Williams-van Klinken 2002: 6). A comparable situation is noted in the French creole Kwéyòl by Garrett (1999: 89), who goes so far as to suggest that “mastery of high Kwéyòl presupposes mastery of standard English.”

This issue of comprehensibility of new terms was recognized by newspaper editors. In the early 2000s, a frequent press strategy was to insert a gloss in

³ Traditionally *kasu lia* is used for oral invitations to ceremonies.

Indonesian immediately after a Portuguese loan to assist the reader. This practice has declined over time.

In 2009 the first author again tested newspaper articles with tertiary-educated East Timorese under the age of 30, and found that the proportion of unrecognized words had greatly reduced. On average, in every five paragraphs, there were two words which the readers did not understand. This reflects the increased Portuguese vocabulary of educated readers in Dili, rather than a simplification in the vocabulary used by journalists.

On further testing intelligibility in 2016, the results were very similar. Three tertiary-educated East Timorese residents of Dili found 27 words they did not know in 47 paragraphs, that is, just over half the paragraphs had words they did not recognize. These were all high-level Portuguese words, with the exception of one Tetun Terik word (*uma fukun* ‘clan house’, used to refer to parliament) and one word invented by the National Institute of Linguistics (*da-hira-k* ‘ORDINAL-how.many-ADJ’, used to mean ‘several’).

6 Grammatical innovations

There is no doubt that journalists have been instrumental in introducing or spreading a number of grammatical innovations. As will be seen, some appear to be percolating into use outside of the media, while others have been used daily in the media for years without any adoption by other writers or speakers of the language.

In the following sections, we first provide a brief orientation to Tetun grammar, then discuss a range of innovations that appear to have occurred since 1999. These are two new passive constructions, final and medial quote margins, new ways of expressing anaphora, and a new way of presenting events out of chronological order. There is also an avoidance of constructions that are not found in the other official and working languages, and significant changes in discourse structure.

With such a strong presence of Portuguese loans in press Tetun (as indeed in other varieties of written Tetun), it is easy to jump to the conclusion that such structural changes are also exclusively of Portuguese origin. As we shall see, this is not necessarily the case.

6.1 Overview of Tetun grammar

Tetun has a basic word order of subject–verb–object, with fronting of the object possible to maintain topic continuity or for contrast. There is no passive voice,

and no marking of person, tense or aspect on verbs. Instead, time and aspect are indicated by separate words such as progressive *hela*. Portuguese verbs are always borrowed in the 3SG present tense form; for instance Portuguese *canta* ‘3SG.PRES sing’ becomes Tetun *kanta* ‘to sing, sings, sang’. There is no marking of gender or number on native Tetun nouns. Tetun has compounding, but very little productive derivational morphology. Ellipsis of arguments is frequent in spoken Tetun, with the result that context is critical in the comprehension of discourse. For more details of Tetun grammar see Williams-van Klinken et al. (2002).

6.2 A new passive construction

As noted above, Tetun traditionally has no passive construction. In this it follows the pattern of its source language, Tetun Terik (Van Klinken 1997: 13), and indeed of East Timorese and central-eastern Indonesian languages in general (Klamer 2002: 371).

Since independence, however, a new passive construction has arisen. In this, the undergoer is placed in initial subject position, followed by a verb, then *hosi* or *husi* ‘from’ plus the actor. Unlike passives in most languages including Portuguese, Indonesian and English (Keenan 1985: 255), there is neither marking on the verb nor an auxiliary to show that the construction is a passive.

This construction was completely unknown during Portuguese and Indonesian times. In 2002 an extensive search of press materials by the second author brought only a few examples, but was striking for the fact they had begun to appear. It was at that stage largely restricted to two contexts. The first was relative clauses within long complicated sentences, such as in the second line of example (4). In such contexts there is no native construction that would allow a new actor to be added without starting a new sentence.

- (4) ... *nebe maka mosu iha kompanha P ... ltd*
 REL FOC arise LOC company (name)
 ‘... which arose in the company P ... Ltd.
 nebe maka lidera husi ministra nia kaben.
 REL FOC lead from minister.FEM POSS spouse
 which is led by the minister’s husband.’ (Suara Timor Lorosae 16/12/08)

The second context was a standard phrase *Liafuan ne’e hatoo hosi* (lit. word this deliver from), used to introduce the speaker for a preceding quote. This is based

on Indonesian expressions such as *Hal tersebut disampaikan oleh ...* (lit. matter aforementioned PASS-deliver by ...) ‘This was stated by ...’

In 2009, Williams-van Klinken (2010b: 182–184) found about one instance of this passive in each newspaper edition, and also noted the first instances of it in speaking in formal contexts. It was, however, still largely restricted to the two linguistic contexts in which it was initially observed in the press.

By 2016 the passive construction had become much more frequent in the press, with 3 instances found in 100 paragraphs across a range of newspapers. It has now spread to other forms of writing. In fact, a search on Google returned 2,750 examples of *prepara hosi/husi* (lit. prepare from) ‘prepared by’. The construction is now also starting to occur in speaking. It is also being readily used in shorter, simpler sentences, and even in main clauses, as in (5).⁴

- (5) *Premiu sira ne’e sei entrega diretamentu hosi Xefi*
 prize DEF.PL this will hand.over directly from head
Governo
 government
 ‘These prizes will be handed out in person by the Head of Government
ba ema nain 6 ne’e.
 to person CLF:human 6 this
 to these six people.’ (Diariu 8/11/2016)

One of the great attractions of this construction is that it allows direct translation of passives from other languages, including Portuguese, Indonesian and English, all of which make frequent use of the passive in formal writing. Indeed, in our direct experience, it is typically used by students and inexperienced translators when translating passives, even by individuals who do not accept the construction as valid Tetun. Having a passive construction is useful not only for translators, but also for writers who are accustomed to writing in languages which have passives, as is the case with all adult writers, including journalists, in East Timor.

As this construction has grown more common, it has also gained increasing acceptance by Tetun speakers, at least in the capital Dili. In the first author’s large

⁴ Steele (2007) speculates that it is the lack of passives in Tetun that helped Timorese journalists at Suara Timor Timur during the Indonesian occupation to write with more “vigour” than she noted in other Indonesian-language newspapers from the same period. These others tended to use passives, such that “incidents ‘just happened’ without anyone being responsible,” whereas Timorese journalists wrote in a more active style.

informal sampling of students and writers, the construction was widely considered to be ‘not Tetun’ in 2004, even though it already occurred in the media at that time, and was common in translations of passives. However, by 2010 many considered it acceptable even though they themselves claimed not to use it, and a group of high-level translators were quick to defend it as a useful innovation. By 2015, even a high-status colleague who rejected the construction outright was starting to wonder if it were not inevitable that he would end up using it, when so many others were.

As mentioned earlier, there is in most instances neither passive marking on the verb nor any auxiliary to show that this is a passive. As a result, this construction is only used when the initial noun phrase cannot logically be interpreted as the actor. For instance, it can be used when referring to a company being led by a person (since companies cannot ‘lead’ people), but cannot be used of a person being led by another person, since the first person would automatically be interpreted as the one doing the leading. It is therefore not surprising that in almost all examples, the actor is animate, and the undergoer inanimate.

The verb is usually an unmarked form of a Portuguese loan verb, such as *lidera* ‘lead’ in (4) and *entrega* ‘hand out’ in (5). It can also be a native Tetun verb, such as *halo* ‘make’ in (6).

- (6) ... *relatorio ba Dezenvolvimentu Humano Nasional 2006 nian*
 report on development human national 2006 POSS
 ‘... the 2006 report on National Human Development
nebe mak halo hosi PNUD (Programa Nasoens Unidas
 REL FOC do from PNUD program nation-PL united-F.PL
ba Dezenvolvimentu).
 for development
 which was prepared by the PNUD (United Nations Development Program).’
 (Timor Post 11/3/2006)

Although still much less common, a more directly Portuguese-looking passive has now begun to appear in press writing. In this construction, unknown in the early 2000s, a Portuguese past participle is used, such as *aprovadu* ‘approved’ in (7). This passive form has been used, primarily by highly Portuguese-influenced speakers and writers, since at least 2006.

- (7) *Tuir Regulamentu kona-ba Prosedimentu Tekniku*
 follow regulation about procedure technical
 ‘According to the Regulations on Technical Procedures
ba Atualizasaun Resenseamentu Eleitoral nian, ne’ebe aprova-du
 to update registration electoral POS which approve-PASS

hosi CNE

from CNE

for the Update of Electoral Registrations, which was approved by CNE

no publika iha Jornal de República ...

and publish LOC Journal of Republic

and published in the Journal of the Republic, ...'

(Suara Timor Lorosae 30/4/2007)

Tetun Dili is, of course, not the only language to have developed a passive construction. Heine and Kuteva (2002) list eight paths by which new passives have been observed to develop in the languages of the world. They do not, however, list the construction used most often in Tetun Dili. This could be because, unlike prototypical passives, there is no morphology or auxiliary to mark the voice. Rather, this passive is derived from object-verb constructions, by the addition of an actor, in a pattern noted for some other languages by Shibatani (1999: 409).

6.3 A new adversative passive

A second passive construction, more restricted in meaning, has arisen in recent years, in which the verb is preceded by *hetan* 'get, find'. This is an adversative passive, used with verbs such as *tiru* 'shoot', *oho* 'kill' and *estraga* 'destroy', normally without an actor, as in (8). The undergoer can readily be human, presumably as *hetan* clearly shows that they are not the agent. It seems to be a calque on Malay *dapat* 'get, be able' (as described, for instance, in Litamahuputty (2012: 114)).

- (8) ... *nain* 2 *kona rama Ambon*, 1 *ema sona ho*
 ... CLF:human 2 touch bow (place) 1 person stab with

tudik, *nain* 4 *hetan baku*, ...

knife CLF:human 4 get bash

'... 2 people were shot with darts, 1 was stabbed with a knife, 4 were bashed, ...' (Suara Timor Lorosae 1/3/2007)

We first noticed this construction being used for agentless passives in 2004 by journalists and translators. It is now beginning to be used with an agent, as in (9), our first recorded examples of this being from 2006. It is also finding its way, though much less commonly, into speaking.

- (9) ... *populasaun familia 8 neebe hetan muda sai husi estadu ...*
 ... population family 8 REL get shift exit from state
 ‘... the 8 local families who were moved out by the state ...’ (Suara Timor
 Lorosae 8/11/2016)

6.4 Final quote margins

A second significant change in word order that is very evident in press writing has taken place in quotes. In Tetun, the default word order is subject–verb–object. Although the object can be fronted, the order of subject and verb is (with a few exceptions, such as *iha* ‘there is’) completely fixed. As such, the word order for presenting quotes is strictly speaker–verb–quote. In newspapers, however, only 4 % of quotes follow this order. In the remaining 96 %, the journalists place the quote first. Based on our own informal counts of quotes in on-line Indonesian and Portuguese newspapers, this proportion echoes that found in Indonesian newspaper writing, and is much higher than is found in Portuguese or English journalism.

When the quote is placed first, there are three options for the structure of the quote margin. 47 % of quotes use the order quote–speaker–verb, as in example (10).

- (10) “... *Problema tolu nee hanesan fator determinante ba paz no*
 problem three this like factor determining for peace and
estabilidade
 stability
 “[... petitioners, refugees, and armed rebels ...] These three problems are
 key factors for peace and stability
iha rai laran,” nia hatete.
 LOC country inside 3SG say
 within the nation,” he said.’ (Suara Timor Lorosae 5/12/2006)

In a further 37 %, the normally irreversible order of subject and verb is in fact reversed, giving the order quote–verb–speaker, as in the first three lines of example (11). This order follows the standard order of Indonesian and Portuguese news articles, even though in terms of Tetun, as one student exclaimed, “It’s back to front!” When asked whether this construction was acceptable, one teacher explained, “We have to accept it, as we come across it every day. But you can’t use this order anywhere else.” Indeed, this order is found in writing even outside of the media, such as in translations and

children's stories, considerably increasing the difficulty of interpretation for children. To date, however, it appears never to be used in speaking.

- (11) “*Eletrisidade no be ne’e sira uza selu tiha ona,*
 electricity and bee this 3PL use pay already PRF
 “‘This electricity and water, they have already used and paid for,
maibe tanba saida mak tenki selu fali husi Janeiru 2008
 but because what FOC must pay again from January 2008
ba leten.
 to above
 but why does it now have to be paid from January 2008 on?
Osan ne’e atu tama iha bolsu karik,”
 money this about.to enter LOC pocket perhaps
 Perhaps this money will go into (their own) pockets,”
dehan deputadu bankada Fretilin, Francisco Jeronimo
 say parliamentary bench (party) (name)
ba Jornalista sira
 to journalist DEF.PL
 said Fretilin bench parliamentary, Francisco Jeronimo to journalists
iha uma fukun Parlamentu Nasionál, Segunda (15/12). ...
 LOC house joint parliament national Monday (date)
 at the National Parliament, on Monday (15/12).’
 (Suara Timor Lorosae, 16/12/08)

A final quote-initial option is to present the quote followed by a full sentence explaining who said it, as in example (12).

- (12) “*Asesu ba justisa labele haree husi distansia hela fatin, kór,*
 access to justice not-can see from distance live place color
relijiaun rasa
 religion race
 “‘Access to justice must not depend on distance to one’s home, color,
 religion, race
no seluk tan,” Presidente Tribunal Rekursu Claudio Ximenes
 and other more president court appeal (name)
hateten ida ne’e,
 say one this
 or other things.” The president of the Court of Appeal, Claudio Ximenes,
 said this

semana kotuk iha Dili.
 week back LOC Dili
 last week in Dili.’ (The Dili Weekly, 14–20/5/2009)

None of these quote-initial orders have made inroads into oral Tetun.

For indirect quotes, the speaker is very occasionally specified within the quote itself, as in the first line of example (13). This too is a Portuguese, Indonesian and English construction. It is completely impossible in traditional Tetun, and does not seem to have found its way into the oral language, even on television or radio.

- (13) *Tempu udan oras nee, katak Ines, labarik barak kona ba*
 time rain hour this say (name) child many touch go
moras barak
 sick many
 ‘During this wet season, said Ines, many children are getting many illnesses
hanesan isin manas, inus metin, isin katar no moras
 like body hot nose tight body itch and sick
selu-seluk tan.
 REDUP-other more
 such as fever, blocked noses, itches and various other conditions.
Husi Departementu Saude seidauk ba iha nee
 from department health not.yet go LOC this
 (Staff) from the Department of Health have not yet gone there
hodi halo tratamentu ba moras sira nebe refere.
 to/and do treatment for sick DEF.PL REL refer
 and offered treatment for these illnesses.’ (Suara Timor Lorosae 13/3/2007)

6.5 New options for anaphora

In spoken Tetun, references to previously mentioned referents use *nee* ‘this’, e.g. *iha fatin nee* ‘at this place’. In press Tetun, however, a few other options have been added through literal translations from Indonesian.

One is to use *nebee hanesan* ‘REL same’ or simply *hanesan* ‘same’. This is used in two expressions. *Iha fatin nebee hanesan* (lit. LOC place REL same) ‘in the same place’, introduces a new story located in the same place as the previous one. In example (14), for instance, it comes after eight paragraphs of presenting what a student spokesperson said on a certain topic at a press conference, and introduces the ideas of a new speaker. *Iha tempu nebee hanesan* (lit. LOC time

REL same) ‘at the same time’ introduces a story which occurs at the same time as the previous one. This is a clear calque on Indonesian *pada tempat/waktu yang sama* (lit. LOC place/time REL same). Portuguese, like English, prefers simpler prepositional phrases, namely *no mesmo lugar* ‘in the same place’ and *ao mesmo tempo* ‘at the same time’.

- (14) *Iha fatin nebee hanesan, nudar estudante finalista husi*
 LOC place REL same as student finalist from
Universidade Dili,
 university Dili
 ‘In the same place, speaking as a finalist student of Dili University,
 L... B... *reforsa liu tan katak, orgaun soberanu iha nasaun*
 (name) stress further more that organ sovereign LOC nation
ida nee ...
 this one
 L... B... further stressed that the sovereign bodies in this nation ...’
 (Suara Timor Lorosae 6/3/2007)

With regard to attitudes towards *iha fatin/tempu nebee hanesan*, lecturers and students whom we asked about this in 2009 identified it as ‘translation Tetun’ or ‘newspaper Tetun’, which they considered unacceptable for use outside the media. Indeed, it has to date not caught on in the wider community.

Another option for anaphora, which we first noted in newspapers and on television in 2006, is to use *nebee refere* (REL refer) ‘aforementioned’, as in the last line of example (13) above. After a year or two it started to sometimes be abbreviated to *refere*. Even after three years of being used nightly on the television news, when we tested speakers on their comprehension of this expression in 2009, it still had them puzzled, and most people still didn’t know what *refere* on its own meant. Now it is recognized as media Tetun, and is also used in other kinds of writing, but still rarely occurs orally outside of the media. This expression is based on Indonesian (*yang tersebut* (lit. (REL) PASS-mention) ‘aforementioned’).

6.6 Non-chronological order: ‘after’

Traditionally in Tetun, events are nearly always recounted in the order in which they occur. There are, therefore, many ways to say “the police received a warrant, then searched the building.” To reverse the order and say “the police searched the

building after receiving a warrant,” one normally needs to use at least two separate markers of linkage or tense-aspect (Williams-van Klinken 2015: 192).

In press Tetun, however, events are often told out of chronological order. This is partly because, as in journalism elsewhere, the initial lead paragraph needs to grab the attention of the reader, and so does not normally start at the beginning of an event, but rather at its apex.

One means which some journalists have adopted to express the concept ‘after’ is to use liturgical *hafoin* ‘then’ to mean both ‘then’ and its antonym ‘after’, as shown by example (15). In this case, it is only logic that enables readers to determine that the police surrounded the suspect’s house after, and not before, receiving the court ruling.

- (15) *Polisia serku fatin nee hafoin simu despaixo tribunal ...*
 police surround place this then/after receive ruling court
 ‘The police surrounded the place after receiving the court ruling.’
 (Timor Post 2/6/2009)

This temporal reordering marked by a conjunction is parallel to what is found in Portuguese, Indonesian and English reporting.

Although *hafoin* has been used in this dual way in the press for at least a decade, it appears to not have spread outside it. In 2005 students and colleagues either claimed to not have heard it, or rejected it outright. By 2009 they recognized it as media Tetun, but found it confusing and not something they themselves would use. This judgment remained unchanged in 2016.

7 Avoidance of distinctively Tetun constructions

One characteristic of press Tetun as well as translated Tetun is that features which are common in spoken Tetun, but do not occur in Portuguese, Indonesian and English, are little used. Indeed, they are under-represented in Tetun writing in general. This includes serial verb constructions, final *ida* ‘one’ in negations, subject-object-verb constructions and tail-head linkage. (A similar avoidance of common Kwéyòl constructions when they aren’t found in English is noted by Garrett (1999: 82).)

In Tetun, serial verb constructions are common for talking about anything involving motion. Particularly common are *baa* ‘go’ or *mai* ‘come’ followed by an action verb, such as *baa haan* ‘go (and) eat’, and a motion

verb followed by a direction verb, for instance *monu tuun* (lit. fall descend) ‘fall down’. As noted by Hajek (2006) a decade ago, these are uncommon in newspapers, and most of the examples that do occur there are found in either direct or indirect quotes. For instance, the corpus contains 201 examples of *tuun mai* (lit. descend come) ‘descend towards us’. This is eleven times as common in oral texts as in written ones. The fact that oral texts are more likely to talk about motion accounts for only part of this difference, as one third of the newspaper occurrences were in direct or indirect quotes. The only serial verb constructions to be commonly used in the press are those which are lexicalized, such as *foo hatene* (lit. give know) ‘inform’ or *foo haan* (lit. give eat) ‘feed’.

One way of negating a verb emphatically in short oral clauses is to use a negator before the verb and *ida* ‘one’ after it, for example ‘*Hau la baa ida!*’ (lit. 1SG not go one) ‘I didn’t go!’ This is felt to be informal, and is not used in journalism or other writing.

Another informal construction not used in writing is placing single-word objects before the verb, in a form of object incorporation, as in (16).

- (16) *Nia Tetun la hatene.*
 3SG Tetun not know
 ‘S/he doesn’t know Tetun.’

Tail-head linkage (repeating part of the end of one sentence as the beginning of the next) is a common feature in Tetun Terik and many other languages of East Timor. Although it is common in the speech of some Tetun Dili speakers, it is avoided by others, and is not used at all in writing. Kilham (1987) noticed the avoidance of such repetitive structures in newly-written languages in the South Pacific too.

8 Non-native discourse patterns

One feature that stands out in press Tetun is that, relative to spoken Tetun, it is what East Timorese call *ulun-ain* (lit. head-foot), that is, back to front. This has already been seen in the new passive constructions and in placing quote margins after the quote instead of before it. However, it goes well beyond the sentence level, with entire events reported out of chronological order and without traditional introductions or conclusions. There appear to be several reasons for the radically new discourse patterns found in the press.

One reason is that writers are following discourse patterns of their dominant language of writing, which in the case of East Timorese journalists is

Indonesian. By transferring these conventions into Tetun (just as English conventions are adopted wholesale in Tok Pisin sports reporting (Romaine 1994: 64)), writers reduce the processing load on themselves (Matras 2007). To some extent journalists may also be following the discourse patterns of the dominant formal language of the senior politicians whom they interview, which in most cases is Portuguese. There seems to be little accommodation for the dominant reading language of the general public, which for many has now become Tetun.

Another factor driving new discourse patterns is the expectation of increased information density in newspaper articles as opposed to speech. In Tetun Dili oral story-telling, most clauses present only one or two new pieces of information. Giving three pieces of new information is not uncommon, but in oral stories, four is largely restricted to the initial clause in a paragraph, which often starts with time. On average, clauses have about 1.5 chunks of new information per clause. In newspapers, in contrast, the average is about 2.5 chunks per clause. Quote margin clauses often contain six chunks. In the last two lines of example (11), for instance, the journalist specifies the speaker's role and name, the addressee, the place of speaking (using both the Tetun and Portuguese names for it), the day, and the date.

The pressure for high information density also leads to longer noun sequences than are found in everyday speech, where the number of consecutive nouns is rarely higher than three, as in *xefi departamentu Tetun* (lit. head department Tetun) 'head of the Tetun department'. In the press it easily rises to four, for instance *Enkontru Komisaun Dezenvolvimentu Distritu* (lit. meeting commission development district) 'district development commission meeting'. Such complex noun phrases are typical of technical and press writing elsewhere. This mirrors the situation of Tok Pisin in Papua New Guinea, where the appearance of complex noun phrases came about largely as the result of newspaper reporting (Romaine 1994: 75–76).

A third reason for non-native discourse patterns is that there is very little original material written in Tetun outside of newspapers and on-line news services to function as alternative models of writing. Various organizations have produced original children's books. Other than that, the main source of original writing is the burgeoning Facebook; it however has quite a different register.

As expected, written texts including in the press show a higher frequency of explicit clause linkage than do oral texts. Portuguese conjunctions are widely used, even for relationships for which there are equivalent Tetun conjunctions also in use (Hajek 2006). In newspapers, as in writing in general, approximately 55 % of clauses start with an explicit connector, with the range across newspaper

articles being about 42%–70%. This is significantly higher than the oral text average of 34% explicit linking, with a range of 21–47%. This increased explicitness partly reflects the journalistic need to clearly mark clausal relations when events and facts are no longer presented in chronological order (as described above), but is also typically associated with higher and more complex registers involving writing (Siegel 1983: 87–88).

9 Concluding remarks

As shown above, Tetun Dili has seen a rapid rise from a primarily oral language prior to 1999 (with the limited exception of the Catholic Church) to a language that is also used extensively in the written and oral media. In the context of press writing, this has been accompanied by a marked increase in the range and frequency of Portuguese loans, as well as a significant number of structural innovations, including final quote margins and two new passive constructions. A major influence on press Tetun is the Portuguese of the national elite, seen in the large number of loans from Portuguese, as well as in such things as limited number and gender agreement within Portuguese noun phrases. Strong Portuguese influence stresses one's membership amongst the elite, just as heavily anglicised Tok Pisin emphasises speakers' "membership in the class of educated people" (Wurm and Mühlhäusler 1982: 73).

A second major influence on the Tetun press register is the Indonesian education of the journalists, leading to such calques as *nebee refere* 'aforementioned', *Liafuan nee hatoo hosi* ... 'these words were delivered by ...', and *iha fatin/tempu nebee hanesan* 'at the same place/time'. These seem to have been specifically journalistic innovations, with clear functions in text cohesion. The first has spread to other forms of writing and to a lesser degree to formal speech, while the latter two have not. Another calque on informal Indonesian is *hetan* 'get' plus a transitive verb to derive adversative passives. This calque is not specifically a media innovation, having been used by the translators and writers as well as the media, and to a lesser extent by speakers.

A third influence on writing in Tetun, including press writing, is the effect of translations from English and Portuguese, which are often very literal. These translations make very high use of Portuguese loans, which are very often close to English both semantically and phonologically. Initially, an expanded use of Portuguese technical terms led to significant difficulties in comprehension for adult readers who, like East Timorese journalists, are mostly educated in Indonesian. (A similar loss of intelligibility was noted by Smith (1990: 206) for

heavily English-influenced Tok Pisin in Papua New Guinea.) A simple strategy used by journalists is to place Indonesian equivalents in brackets after some of these words. This was particularly common in the early years after 1999, and may have contributed to the increased knowledge of Portuguese loan words by many readers.

That Tetun has borrowed extensively and almost exclusively from Portuguese in terms of lexicon can easily lead to the impression that Portuguese on its own is likely also to impact on other aspects of Tetun grammar. However, specifically Portuguese constructions do not seem to be impacting directly on Tetun grammar beyond Portuguese noun phrases which are sometimes adopted complete as fixed phrases with number and gender agreement. Instead, structural changes which could in principle come from Portuguese only seem, based on the examples we have looked at, to actually take place in Tetun if they are also supported by Indonesian as the language of education for most journalists and writers, and English as a major source of translations. Where these three languages have similar constructions lacking in Tetun, the corresponding construction in Tetun seems to arise in the writing of Indonesian-educated journalists and writers, and in translations, without it being clear who started the innovation. An example of this is the passive, with the word order and function of Tetun's new passive construction approximately following that of all three international languages, i.e. Portuguese, English and Indonesian. Another is the use of final quote margins. Similarly, certain distinctively Tetun constructions, such as serial verb constructions, seem to be avoided in press writing because they are absent in all three languages.

A final influence on press writing in Tetun is international journalistic style, such as is also found in Portuguese, Indonesian, and English. One result is the strong preference for final quote margins to allow quotations to be foregrounded, and the development of passives to similarly allow foregrounding of events. Heavy use of borrowed technical terms that are precise in meaning allow writers to easily increase information density by avoiding potentially lengthy paraphrasing, thus saving on expensive print space. The frequent presentation of events in non-chronological order is designed to generate immediate interest among readers, but results in much more frequent and explicit clause linking than occurs in speech. Similar effects in press writing are of course found in other languages which have developed from low status oral languages to high status written languages, such as Tok Pisin in Papua New Guinea, under the influence of English (Siegel 1983; Siegel 1984). What distinguishes Tetun from Tok Pisin and other languages is the unusually complex sociolinguistic, linguistic and literacy environment

the former finds itself in, as well as the high number of documented grammatical innovations.

To date, only some of these changes have found their way into the wider community. In particular, many Portuguese loans have become recognized if not generally used, and the passive has started to be used more widely. Others are used daily in the press (and in many cases also radio and television and in translation), and have received grudging acceptance as ‘media Tetun’ or ‘translationese’, but have not gone further, thus showing that journalists (and translators) have initiated changes, rather than reflecting usage in the community.

It remains to be seen whether these phenomena become more common. Our prediction is that direct Portuguese influence will continue to increase, at least amongst the urban elite, as Portuguese works its way up the education system and as the press, other media and translators continue to promote Portuguese vocabulary within Tetun. Local official documents are increasingly being translated from Portuguese into Tetun, further promoting the adoption of Portuguese loanwords. In recent years wire services (available in Portuguese, Indonesian and English) have started to be translated into Tetun, particularly for television, though also to some extent in newspapers, whereas previously they were printed in the original language. In this case, the combined pressure of journalistic tradition and rather literal translation would continue to reinforce the changes above, and quite likely introduce new ones.

References

- Bell, Allan. 1991. *The language of news media*. New Jersey: Wiley-Blackwell.
- Biber, Douglas & Mohamed Hared. 1994. Linguistic correlates of the transition to literacy in Somali: Language adaptation in six press registers. In Douglas Biber & Edward Finegan (eds.), *Sociolinguistic Perspectives on Register*, 182–216. Oxford: Oxford University Press.
- Cabral, Estevao & Marilyn Martin-Jones. 2008. Writing the Resistance: Literacy in East Timor 1975–1999. *International Journal of Bilingual Education and Bilingualism* 11(2). 149–169.
- Correia, Adérito & José Guterres, Benjamim de Araújo e Corte-Real, Geoffrey Stephen Hull, George William Saunders, Rosa da Costa Tilman, Mário Adriano Soares & Domingos dos Santos. 2005. *Disionáriu nasiunál ba Tetun ofisiál* [National dictionary for official Tetun] Dili: Instituto Nacional de Linguística, Universidade Nacional Timor Lorosa'e.
- Da Silva, Sebastião Maria Apparicio. 1885. *Catecismo da doutrina Christa em Tétum*. Macau: Tipografia do Seminário.
- Da Silva, Sebastião Maria Aparicio. 1889. *Diccionario de Português-Tétum*. Macao: Typographia do Seminario.

- Democratic Republic of Timor-Leste. 2002. *Constitution of the Democratic Replublic of Timor-Leste*. Dili: Democratic Republic of Timor-Leste.
- Diocese de Dili. 2000. *Liafuan diak ba imi [Good news for you]*. Dili: Diocese de Dili.
- Garrett, Paul B. 1999. "High" Kwéyòl: The emergence of a formal creole register in St. Lucia. In John McWhorter (ed.), *Language change and language contact in pidgins and creoles*, 63–101. Amsterdam: John Benjamins Publishing Company.
- Godinho, Margarida & Maria dos Santos. 2014. "Environmentu kapaas!?: Surgimento de falsos amigos em tétum com ar a Português ["Environmentu kapaas!?: The rise of false friends in Tetun with a Portuguese flavour]. In Hannah Loney, Antero B. Da Silva, Nuno Canas Mendes, Alarico da Costa Ximenes & Clinton Fernandes (eds.), *Understanding Timor-Leste 2013*, vol. 1, 187–193. Dili, Timor Leste: Timor Leste Studies Association.
- Hajek, John. 2006. Language contact and convergence in East Timor: The case of Tetun Dili. In Alexandra Y. Aikhenvald & Dixon M. W. Robert (eds.), *Grammars in contact: A cross-linguistic typology*, 163–178. Oxford: Oxford University Press.
- Hajek, John & Catharina Williams-van Klinken. 2003. Um sufixo românico numa língua austronésia: *-dor* em Tetum [A Romance suffix in an Austronesian language: *-dor* in Tetun]. *Revue de linguistique romane* 67. 55–65.
- Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization* Cambridge: Cambridge University Press.
- Hopper, Ben. 2008. Timor Leste. In Indrajit Banerjee & Stephen Logan (eds.), *Asian Communication Handbook*, 482–525. Singapore: Asian Media Information and Communication Centre and Nanyang Technological University.
- Hull, Geoffrey. 1996. *A language policy for East Timor: Background and principles It's time to lead the way: Writings from a conference on East Timor and its people*, 38–59. Melbourne: East Timor Relief Association.
- Hull, Geoffrey. 1999. *Standard Tetum-English dictionary*. St Leonards: Allen & Unwin in association with the University of Western Sydney Macarthur.
- Hull, Geoffrey. 2002a. *The languages of East Timor: Some basic facts*. Dili: Instituto Nacional de Linguística, Universidade Nacional Timor Lorosa'e.
- Hull, Geoffrey. 2002b. *Standard Tetum-English dictionary*. Winston Hills: Sebastião Aparício da Silva Project in association with Instituto Nacional de Linguística.
- Hull, Geoffrey & Lance Eccles. 2001. *Tetum reference grammar*. Winston Hills: Sebastião Aparício da Silva Project, in conjunction with Dili: Instituto Nacional de Linguística, Universidade Nacional de Timor Lorosa'e.
- Kalmár, Ivan. 1985. Are there really no primitive languages? In D. R. Olson, N. Torrance & A. Hildyard (eds.), *Literacy, language, and learning: The nature and consequences of reading and writing*, 148–166. Cambridge: Cambridge University Press.
- Keenan, Edward L. 1985. Passive in the world's languages. In Timothy Shopen (ed.), *Language typology and syntactic description, volume 1: Clause structure*, 243–281. Cambridge: Cambridge University Press.
- Kilham, Christine A. 1987. A written style for oral communicators? *Notes on translation* 123(12). 36–52.
- Klamer, Marian. 2002. Typical features of Austronesian languages in central/eastern Indonesia. *Oceanic Linguistics* 41(2). 363–383.
- Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.) 2016. *Ethnologue: Languages of the world*. Dallas, Texas: SIL International.

- Litamahuputty, Bathseba & Helena Johanna. 2012. *Ternate Malay: Grammar and Texts*. Leiden: Universiteit Leiden.
- 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 & Jeanette Sakel. 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31(4). 829–865.
- National Statistics Directorate & United Nations Population Fund. 2011. *Population and Housing Census of Timor-Leste, 2010. Volume 3: Social and Economic Characteristics*. Dili: Democratic Republic of Timor-Leste.
- Percy, Carol. 2012. Early Advertising and Newspapers as Sources of Sociolinguistic Investigation. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The Handbook of Historical Sociolinguistics*, 191–210. West Sussex: John Wiley & Sons, Incorporated.
- Pires, Paulo. 2001. *A imprensa em Timor, antes do 25 de abril* [The Press in Timor before the 25th of April]. *Camões: Revista de Letras e Culturas Lusófonas* 14, 135–145.
- Reder, Stephen. 1981. The written and the spoken word: Influence of Vai literacy on Vai speech. In S. Scribner & M. Cole (eds.), *The psychology of literacy*, 187–199. Cambridge, MA: Harvard University Press.
- Romaine, Suzanne. 1994. On the Creation and Expansion of Registers: Sports Reporting in Tok Pisin. In Douglas Biber & Edward Finegan (eds.), *Sociolinguistic Perspectives on Register*, 59–81. Oxford: Oxford University Press.
- Rosenhouse, Judith & Haya Fisherman. 2008. Hebrew: Borrowing ideology and pragmatic aspects in a modern(ised) language. In Judith Rosenhouse & Rotem Kowner (eds.), *Globally speaking: Motives for adopting English vocabulary in other languages*, 121–144. Clevedon, UK: Multilingual Matters.
- Shibatani, Masayoshi. 1999. Voice. In Keith Brown & Jim Miller (eds.), *Concise encyclopedia of grammatical categories*, 406–412. Amsterdam: Elsevier.
- Siegel, Jeff. 1983. Media Tok Pisin *Papers in pidgin and creole linguistics* No. 3, 81–92. Canberra: Pacific Linguistics A-65.
- Siegel, Jeff. 1984. Current use and expansion of Tok Pisin in the mass media. In S.A. Wurm & P. Mühlhäusler (eds.), *Handbook of Tok Pisin (New Guinea Pidgin)*, 517–533. Canberra: Pacific Linguistics C-70.
- Smith, Geoff P. 1990. Creolized Tok Pisin: Uniformity and variation. In Richard B. Jr Baldauf & Allan Luke (eds.), *Language planning and education in Australasia and the South Pacific*, 197–209. Clevedon: Multilingual Matters Ltd.
- Steele, Janet. 2007. The voice of East Timor: Journalism, ideology, and the struggle for independence. *Asian Studies Review* 31. 262–282.
- Tadmor, Uri. 2007. Grammatical borrowing in Indonesian. In Y. Matras & J. Sakel (eds.), *Grammatical borrowing in cross-linguistics perspective*, 301–328. Berlin: Mouton de Gruyter.
- Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Thomaz, Luis & F.R. Filipe 1981. The formation of Tetun-Praça, vehicular language of East Timor. In Nigel Phillips & Anwar Khaidir (eds.), *Papers on Indonesian languages and literatures*, 54–83. Paris: Cahier d'Archipel 13.

- Ure, Jean. 1982. Introduction. In Jeffrey Ellis & Jean Ure (eds.), *Register range and change*, 5–23. Berlin: Mouton.
- Van Klinken, Catharina Lumien. 1997. *A grammar of the Fehan dialect of Tetun, an Austronesian language of West Timor*. Canberra: The Australian National University PhD.
- Wallace, William D. 1977. How registers register: A study in the language of news and sports. *Studies in the Linguistic Sciences* 7(1). 46–78.
- Williams-van Klinken, Catharina. 2002. High registers of Tetun Dili: Portuguese press and purist priests. Paper presented to the *Proceedings of the 2001 Conference of the Australian Linguistics Society*, Canberra.
- Williams-Van Klinken, Catharina. 2004. Developing electoral terminology for a new official language: Tetun in East Timor. *Current Issues in Language Planning* 5(2). 142–150.
- Williams-van Klinken, Catharina. 2010a. Two Tetuns: Language change before your very eyes. Paper presented at the The 6th ENUS Conference on Language and Culture, Kupang.
- Williams-van Klinken, Catharina. 2010b. Voice and valency in Tetun Dili. In Michael C. Ewing & Marian Klammer (eds.), *East Nusantara: Typological and areal analyses*, vol. PL 618, 165–183. Canberra: Pacific Linguistics.
- Williams-van Klinken, Catharina. 2011. *Interactive Tetun-English dictionary*. Dili: Dili Institute of Technology.
- Williams-van Klinken, Catharina. 2015. *Peace Corps East Timor Tetun language course* Dili: Peace Corps East Timor.
- Williams-van Klinken, Catharina, John Hajek & Rachel Nordlinger. 2002. *Tetun Dili: A grammar of an East Timorese language*, 528. Canberra: Pacific Linguistics.
- Williams-van Klinken, Catharina, Robert Williams & Helio Brites Da Silva. 2016. Tetun Dili laos Dili nian ona: Tetun nudar lian inan tuir dadus sensus nian [Tetun Dili no longer belongs to Dili: Tetun as mother tongue according to census data]. In Sarah Smith, Antero B. Da Silva, Nuno Canas Mendes, Alarico da Costa Ximenes, Clinton Fernandes & Michael Leach (eds.), *Timor-Leste: The local, the regional and the global*, vol. 1, 65–71. Dili, Timor Leste: Timor Leste Studies Association.
- Wurm, S.A. & P. Mühlhäusler. 1982. Registers in New Guinea Pidgin. In Jeffrey Ellis & Jean Ure (eds.), *Register range and change*, 69–86. Berlin: Mouton.