PROJECT MUSE®

Language Contact and Gender in Tetun Dili: What Happens When
Austronesian Meets Romance?

John Hajek, Catharina Williams-van Klinken

Oceanic Linguistics, Volume 58, Number 1, June 2019, pp. 59-91 (Article)

Published by University of Hawai'i Press

➔ For additional information about this article

<https://muse.jhu.edu/article/730688>

Language Contact and Gender in Tetun Dili: What Happens When Austronesian Meets Romance?

John Hajek and Catharina Williams-van Klinken

MELBOURNE UNIVERSITY AND DILI INSTITUTE OF TECHNOLOGY

Tetun Dili is an Austronesian language and one of the two official languages in Timor-Leste, alongside Portuguese with which it has been in contact for centuries. In this detailed study, we describe various aspects of gender in Tetun Dili, in both native and borrowed vocabulary. We identify marked differences between gender marking in native and borrowed Portuguese words, with Tetun prioritizing the feminine, and Portuguese the masculine. We show that through contact with Portuguese, Tetun Dili has also developed “marginal” gender, with grammatical gender agreement being optional but increasingly common for Portuguese loans, mainly in the acrolect. Grammatical agreement is also spreading to a handful of non-Portuguese words, to which there is evidence of incipient grammatical gender assignment. Some comparisons are made with developments in Chamorro and Tagalog, two Austronesian languages heavily influenced by Spanish (closely related to Portuguese), for which gender contact effects have previously been described in some detail. We conclude with a set of more general typological observations.

1. INTRODUCTION. The arrival of the Spanish and Portuguese in the Asia-Pacific region in the late 15th and 16th centuries led to hundreds of years of contact between their respective Romance languages and a wide range of Austronesian languages in Southeast Asia and the Pacific, including in the Philippines, Timor-Leste (also known as East Timor), and parts of eastern Indonesia and Micronesia, where Spanish and Portuguese colonial influence was most evident. To date, linguists have often focused on describing the significant amount of lexical borrowing from Spanish and Portuguese into various Austronesian languages, e.g., Abduramachman (1972:70) on Portuguese loanwords in Ambonese Malay, Quilis (1976) on Hispanic influence in Cebuano, Rodríguez-Ponga Salamanca (2006/2007) on Hispanic loans in Saipan Carolinian, and Hull (2002b:ix–xi) on Portuguese influence in Tetun Dili.¹

1. Where relevant in examples, .I indicates an Indonesian loan, .P a Portuguese loan, .T native Tetun.

There are also an increasing number of studies which focus more specifically on Romance (typically Spanish) grammatical influence on Austronesian, such as T. Stolz (2012) on gender borrowing in Chamorro and Philippine languages, Quilis and Casado-Fresnillo (2008) on the influence of Spanish in Philippine languages, Makihara (2001) on Spanish contact with Rapanui, and Pagel (2010, 2019) who focuses on Spanish influence on Chamorro. Nevertheless, much remains to be explored about the effects of such contact on the grammar of these Austronesian languages.

One interesting area of typological influence of Romance languages on some Austronesian languages is in the treatment of Romance-style gender marking, triggered in the first instance by large-scale lexical borrowing. In Portuguese and Spanish, differences in gender (masculine versus feminine) are fundamental in many parts of speech. For instance, gender is consistently marked in nouns, regardless of semantic content, and adjectives must agree in gender with their head nouns. However, grammatical gender on nouns or adjectives is not native to Austronesian languages (Blust 2009, Himmelmann 2005).² To the best of our knowledge, and of Stolz and Levkovich (in press), the only in-depth study of Romance influence on the gender system of any Austronesian language is that of T. Stolz (2012), focusing mainly on Chamorro.

Tetun Dili (often referred to simply as Tetun), an isolating Austronesian language that is the primary lingua franca in Timor-Leste, is a particularly interesting case with respect to language contact effects on gender. It has a long history of contact with Portuguese, a morphologically rich Romance language, with which it has shared co-official status since Timor-Leste's independence in 2002. This particular set of circumstances allows us to explore gender in Tetun with respect to both native and borrowed lexicon and grammar, in a long-term language contact situation that is still unfolding between two typologically very distinct languages.

In this study we first make general observations about gender borrowing in contact situations, primarily in Austronesia, before turning our attention specifically to Tetun in its historical context and to Portuguese influence on that language. We then give an overview of Tetun and Portuguese grammar, focusing especially on gender, and look in detail at how natural and grammatical gender are expressed in native and borrowed Portuguese vocabulary, noting that the two systems are entirely separate. We also outline more subtle gender differences, not previously discussed in the literature, such as the interpretation of plural gender-marked phrases and the ordering of gendered pairs, before providing a detailed account of natural gender and gender agreement. We then identify some minor local innovations in Portuguese word-forms, gender agreement in loans, as well as signs of incipient gender agreement with

2. Schapper (2010) discusses evidence of a semantic neuter v. non-neuter gender distinction in pronominals and demonstratives in some Austronesian languages of Eastern Indonesia as a result of contact with Papuan languages in the region. Given the very different nature of this gender distinction, we do not discuss this particular phenomenon further.

non-Portuguese nouns. Gender-related developments in Tetun are then compared with those reported in Chamorro and Tagalog, also with histories of intense contact with Romance (specifically Spanish), albeit the contact has in those cases not continued into the present. The evidence suggests that Tetun Dili is unusual, if not unique, in showing Romance-style gender agreement, at least for some speakers, in contexts not previously reported for other Austronesian languages. We are also able to make a number of typological observations of more general value, based on the data presented.

2. GENDER IN CONTACT. Gender has been characterized in a number of ways. “Grammatical gender” is defined by Corbett (2013a:1) as follows: “The defining characteristic of gender is agreement: a language has a gender system only if we find different agreements ultimately dependent on nouns of different types.” The basis on which these various types of nouns are distinguished varies cross-linguistically. Portuguese and Spanish, like other Romance languages, are examples of languages with a “sex-based gender system” (Corbett 2013b), in which the core of the masculine gender is nouns referring to males, and the core of feminine is nouns referring to females, even though both masculine and feminine categories contain many other nouns as well.

T. Stolz (2012:96), in his study of gender borrowing primarily in Chamorro but also in Tagalog, modifies Corbett’s grammatical definition of gender to allow for borderline cases of gender, which he calls “proto-gender”. These are “systems in which formal (mostly sex-based) distinctions are established on nominals that do not trigger agreement phenomena.” He uses the word “proto-gender” to suggest that these gender systems could potentially develop into grammatical genders that trigger agreement. A term with the same denotation, but not implying the potential for development, is “natural gender” (T. Stolz 2012). In a sex-based system, natural gender is the gender one would expect a word to have based on the sex of its referent.

When languages with different gender systems come into contact, T. Stolz (2012:95) concludes based on the 27 case studies of grammatical borrowing in Matras and Sakel (2007) that gender is quite often affected in some way. He presents a useful overview of the literature, with examples drawn from a range of languages. He observes that as a result of contact, it is common for existing gender systems to recede, disappear or alternatively to be reshaped (see also C. Stolz [2008, 2009]). T. Stolz (2012:95) also notes that in this same corpus of case studies, there are a few cases where incipient gender marking of some kind, as a result of borrowing from gendered languages, may now be evident in languages that previously did not show signs of grammatical gender. The main example he gives of such a process involves Indonesian.

With respect to contact-induced changes within (formerly) grammatically genderless languages specifically in the Austronesian language family, Indonesian is at the lower end of gender incorporation. It has adopted masculine and feminine forms of some nouns from Sanskrit, Arabic and Dutch, where

these forms reflect human sex distinctions. In some common kinship terms, such as the lexically distinct pair *opa/oma* ‘grandfather/grandmother’ from Dutch, the sex distinction is strictly maintained by all speakers, while for most other gendered pairs the feminine is mainly used in the acrolect, with the masculine functioning as the unmarked member of the pair in colloquial speech (Tadmor 2007:312).

Tadmor mentions one native Indonesian word that has taken on a masculine/feminine distinction on the basis of a parallel with Sanskrit loanwords ending in *-a* for males and *-i* for females, e.g., *putra/putri* ‘son/daughter’. This is *pemuda*, which has been reinterpreted from ‘young person’ to ‘young man’, with *pemudi* being created to mean ‘young woman’ by analogy with the Sanskrit loans. Indonesian has also borrowed the Sanskrit agentive suffixes *-wan/man* (m) and *-wati* (f), which can be attached productively to non-Sanskrit roots (Matras 2011:311). Elsewhere, Rapanui, the Austronesian language of Easter Island in the eastern Pacific, has introduced natural gender distinctions through adopting the gender-sensitive Spanish diminutive *-ito/ita* in borrowed terms as well as on some native roots (Chamorean 2012:73, Fischer 2007:391). Neither Indonesian nor Rapanui, however, have any claimed gender agreement, such as between head noun and modifying adjective.

Some Austronesian languages go beyond borrowing words or morphemes to acquiring a partial gender system. Although this is not common, in each case it involves intensive long-term contact either with Spanish, or in the case of Tetun, with Portuguese. In T. Stolz’s (2012) account of Spanish gender-marking in Tagalog in particular, as well as more briefly in his discussion of closely related Cebuano and Ilocano, natural gender is maintained in Spanish nouns referring to humans, as well as in a few adjectives describing humans (see also Reid and Schachter [2009:852]). Based on the limited description by Schachter and Otnes (1972:96–7, 196–8), T. Stolz (2012:100) concludes that Tagalog has “marginal” contact-induced gender, with a few nouns and adjectives that take gender marking. Stolz notes three restrictions: (a) gender only applies to Spanish loans; (b) agreement only applies for nouns with human controllers (thus reflecting natural gender); and (c) masculine/feminine distinctions are neutralized in the plural (see 10.1 for further discussion).

Baklanova (2016), after retesting these restrictions for Tagalog, agrees in general with T. Stolz’s (2012) observations, except that she concludes that gender applies not only to Spanish loans but also to blended words with non-Spanish roots (native and more recent English loans) derived using Spanish derivational affixes, e.g., *lasenggo/a* ‘male/female drunkard’ from Tagalog *lasing* ‘drunk’. She notes too (2016:31) that “the marginal character of the gender phenomenon in Tagalog is manifested in the *inconsistency of the gender agreement*”, with the masculine often being used for feminine human controllers even when a feminine form exists. Her corpus also includes some cases of gender-sensitive adverbial modifiers and gender-marked adjectives describing an object that is related to or possessed by the controller (see section 8.2.1 below). This all suggests

that “gender agreement” in Tagalog may not really be a grammatical phenomenon, but rather a somewhat optional semantic one.

Further along the scale of gender influence is Chamorro, spoken in Guam and the Northern Marianas. T. Stolz (2012) presents a detailed analysis of its gender system, which, like Tagalog, he classifies as marginal. Chamorro has borrowed hundreds of pairs of masculine and feminine forms of nouns and adjectives from Spanish, all reflecting real-life male/female distinctions. Unlike in Indonesian, masculine/feminine forms of loaned nouns are generally used to refer to male/female humans, and unlike Tagalog, masculine/feminine adjectival loans too are generally distinguished with human referents. That is, there is gender agreement, but only for descriptions of humans. In addition, there are some set phrases which have been borrowed from Spanish complete with gender agreement morphology. A further development, similar to what Baklanova (2016) reports for Tagalog, is that some blended forms in Chamorro with native or English roots and Spanish derivational affixes have retained gender marking on the borrowed affixes to distinguish male and female, e.g., *gamboleru/a* ‘gambler (M/F)’, from English ‘gambler’ and the Spanish agentive suffix *-ero/a*. To date we are not aware of any Austronesian language referred to in the existing literature that has adopted gender to such an extent that gender agreement of the Romance type applies even when reference is not to humans. As we shall see, Tetun Dili has a noticeable amount of grammatical gender agreement of a kind not found in Chamorro and Tagalog, at least within the acrolect, as well as signs of an incipient process of gender assignment to a few non-Portuguese words. Overall, developments in Tetun Dili appear to be highly unusual, if not unique, within the Austronesian world.

Another difference between Tetun Dili and the languages discussed above (Indonesian, Tagalog and other Philippines languages, and Chamorro) is in the currency of the language contact. In these other Austronesian languages, contact with the gendered donor language has long ceased. Such contact-induced innovations in gender marking as exist have, therefore, become part of the language. In Tetun Dili, in contrast, contact with Portuguese is current, and a range of contact-induced innovations have been observed entering the language and spreading fast over the last two decades since independence (Hajek 2006, Williams-van Klinken and Hajek 2018a,b).

3. THE HISTORICAL CONTEXT OF TETUN DILI. In order to better understand language contact effects with respect to gender in Tetun Dili, it is important to have an overview of the relevant historical context. Tetun Dili is an Austronesian language spoken in Timor-Leste, in the eastern part of the Indonesian archipelago. It ultimately derives from Tetun Terik, a vernacular language spoken along the border between East and West Timor, and further east along the south coast. Probably as a result of the influence of two Tetun-speaking kingdoms located in these two regions, Tetun developed into a lingua franca which has been spoken throughout much of the eastern half of the island of

Timor for at least 500 years (Thomaz 1981). This lingua franca co-existed with over a dozen indigenous languages, both Austronesian and non-Austronesian.³

The Portuguese arrived in Timor in the early 1500s, initially making visits, and later adopting Timor-Leste as an ‘overseas province’. Five hundred years of Portuguese contact was followed by twenty-four years of Indonesian occupation from late 1975 to 1999, followed by United Nations control until full independence was achieved in 2002. At this point Tetun and Portuguese became the two co-official languages of the new country.

During the Portuguese colonial period, Portuguese was the exclusive language of schools and officialdom. Nevertheless, by the end of Portuguese rule in 1975, only people of Portuguese descent and so-called *mestiços* (of mixed Portuguese-Timorese descent) spoke Portuguese as a mother tongue, and only about 15–20% of the population spoke any Portuguese at all (Thomaz 2002:140, 2). According to Marcos (1995), 5–10% were literate in Portuguese. By the end of the 1960s the Portuguese administration had begun serious efforts at early mass literacy in Portuguese, so that some contact with Portuguese was becoming increasingly common (Hajek 2000). A Portuguese-based creole previously spoken in parts of Dili was by then almost extinct (Baxter 1990).

With the start of Indonesian control in late 1975, Portuguese was quickly removed from public settings and vigorously replaced in education and government contexts by Indonesian, such that East Timorese attending school during this period were almost all educated in Indonesian. Although Indonesian was used in all government functions, knowledge of Tetun Dili, as a symbol of East Timorese solidarity, spread throughout this time. One factor was that the Catholic Church, under pressure to stop using Portuguese in the liturgy in favor of Indonesian, switched to using Tetun instead. While few spoke Tetun Dili as a first language, it was undoubtedly the most widely spoken language during all of this time (Hajek 2000) and has continued to spread rapidly since the end of Indonesian occupation in 1999 (Williams-van Klinken and Williams 2015).

Since independence, Indonesian has been replaced in schooling and officialdom by Portuguese and Tetun, and its role in Timor-Leste society has greatly reduced. By the 2015 census, the population of Timor-Leste was 1.1 million, and the proportion claiming to be able to speak Tetun Dili was 87%, including 31% speaking it as a home language and 56% speaking it as a second or third language (Direcção Geral de Estatística Timor-Leste 2016). The figures on knowledge of Indonesian and Portuguese have not yet been published for 2015, but in 2010 they were reported to be 55% for Indonesian and 50% for Portuguese. In practice, however, knowledge of the latter is still mostly passive with little active use of the language in daily life, except amongst the elite. Portuguese influence on Tetun thus largely comes about through the influence of this elite, with innovations to some extent spreading to others who do not master Portuguese, in a pattern described for other non-Romance languages

3. Hull (2002a) estimates the number of such languages to be 16, while Lewis et al. (2018) put the number at 19.

by T. Stolz (2008:2). The type of Portuguese taught and used in official contexts is based on the standard European variety used in Portugal.

Although the majority of East Timorese can speak Indonesian, it has had little obvious influence on formal Tetun Dili, as noted earlier by Hull (1999:xi) and Hajek (2006), although it has some contact-influenced structures used in press writing (Williams-van Klinken and Hajek 2018a). Indonesian loans are still common in informal speech of some people, but are reducing, and Indonesian loans are avoided in formal contexts in favor of Portuguese and Tetun. Knowledge and use of Indonesian have declined significantly since independence, especially among younger East Timorese.

4. THE INFLUENCE OF PORTUGUESE ON TETUN DILI. Tetun Dili shows much influence from Portuguese, most obviously seen in extensive borrowing that occurred during colonial times and which is occurring with renewed vigor this century. This is especially marked in some types of formal speech and writing, where new loans are continually being added to handle the many new functions for which Tetun Dili has been used since Timor-Leste's vote for independence from Indonesia in 1999, including most political debate, the media, conferences, as well as much teaching at all levels of the education system. (For instance, see Williams-van Klinken and Hajek [2018a] on Tetun in the press.) While Tetun and Portuguese are co-official languages, the latter has higher formal status, given its role as the primary language for law-making and legal interpretation, and its high status within the education system.

The degree of Portuguese influence in Tetun Dili varies not only by context, but also varies widely amongst speakers. Tetun Dili speakers are aware of the proportion of Portuguese loans used by others, and tend to associate sociolinguistic judgments with it, for instance linking high Portuguese content with someone who is "elite, educated, and probably well-employed, but distant and disconnected from the general Timorese population" (Ross 2017:56).

Three socially marked varieties of Tetun Dili can be identified, linked also to relative formality and education, as noted also by Hull and Eccles (2001). The acrolectal variety of Tetun Dili is spoken by the traditional Portuguese-educated elite, who speak Portuguese fluently, and for whom this is (or at least was) their preferred language of writing.⁴ As the senior leadership of the young country, they are the source of much of the news and public commentary, so exercising great influence on the Tetun used in the media, including television, radio and the press, and thereby also modeling their variety of Tetun Dili to a broader public. There is an almost universal agreement within Timor-Leste that the appropriate language for new loans is Portuguese, rather than Indonesian (as often happened during the Indonesian occupation) or English (which was also introduced through the arrival

4. It is important to note that proficiency in Portuguese within the acrolect is not uniform: it varies from native bilingual to near-native as well as some imperfect command. Some acrolectal speakers learnt Portuguese as adult resistance fighters in the mountains during the Indonesian occupation (1976-1999). Amongst mesolectal speakers, proficiency is much more variable.

of the United Nations and aid organizations in 1999). Indonesian- and English-educated⁵ Timorese therefore also use a high percentage of Portuguese loans in their Tetun, especially in formal contexts, where, based on evaluation of our large-scale corpus (see below) the proportion of Portuguese tokens can easily reach 35% or more. In newspaper articles and technical writing it can rise to over 40%, with headlines often almost entirely in Portuguese, as seen in the recent example (1), in which only the final word *rua* ‘two’ is a native Tetun form.

- (1) Timor-Leste/Eleisaun: Tribunál Rekursu rejeta
 Timor-east/election court appeal reject
 kandidatura partidáriu rua
 candidature party.ADJ two
 ‘Timor-Leste/Election: The court of appeal rejects two party candidates.’
 (Newspaper Sapo, 29/3/2018)

At the other extreme is the basilectal variety of Tetun Dili, spoken as a second language by rural people. This shows very little Portuguese influence, and has a much lower proportion of Portuguese loans, at around 5%. This reflects a general tendency of limited literacy, lack of knowledge of Portuguese, and little use of technical terms and formal registers.

In between these two extremes is the mesolectal variety used by the non-elite residents of the national capital Dili. In their informal urban speech, the percentage of Portuguese tokens is in the range of 10-25%, although this proportion increases significantly when talking of technical matters.

5. DATA SOURCES. Data for this paper have been collected during intensive involvement with Tetun Dili over most of the last eighteen years, including direct observation by both authors.

Much of the lexical data was from the second author’s extensively annotated database of over 16,000 lexemes, managed in Toolbox (SIL International 2011). A selection of the more commonly known lexemes in this database is published as Williams-van Klinken (2019).

Textual data was mainly obtained through the second author’s searchable Toolbox 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, 2.1 million words of written texts, and 1 million words of translations. This was supplemented by searches of the large and growing body of written and translated texts available on the internet

Finally, the second author discussed many of the issues with language lecturers at Dili Institute of Technology.⁶

5. English-educated Timorese include former refugees educated in Australia, often also fluent in Portuguese, who have returned in small numbers to work in Timor-Leste. Apart from the diaspora, there are a few Timorese who have done higher education in English-speaking institutions overseas.
6. Those whose observations were particularly useful included Leoneto da Silva Ribeiro, Abreu Ximenes and Dulce Marques Freitas.

6. LANGUAGE OVERVIEW OF TETUN DILI AND PORTUGUESE.

In order to properly understand gender-related effects in Tetun Dili as a result of contact with Portuguese, we first briefly sketch the basic grammatical features of both languages (following T. Stolz [2012] for Chamorro and Spanish).

6.1 OVERVIEW OF TETUN DILI GRAMMAR. Tetun Dili has a strict subject-verb order, with the exception of presentational clauses, and a few other intransitive verbs (e.g., *mosu* ‘appear’) which, as a result of Portuguese influence, can precede the sole argument. The default order for transitive clauses is subject-verb-object, although the object can be fronted. Tetun traditionally has no passive, but in recent years one has developed within the acrolect as a result of language contact (Williams-van Klinken and Hajek 2018a). Time and aspect are optionally indicated by separate words.

The majority of modifiers follow the head noun. Exceptions which precede the nouns are most possessors, Portuguese numerals, and the Portuguese loan *kada* ‘each’, e.g., *kada tinan* ‘every year’. There is no inflectional morphology, and traditionally no marking of gender or number on nouns, adjectives, or verbs. Tetun Terik, the source language of Tetun Dili, has subject marking on the verb (van Klinken 1999), but this was lost at some stage in the development of Tetun Dili, as a result of local language contact on its way to becoming a lingua franca even before Portuguese influence began (Hajek 2006).

Tetun Dili has only very limited derivational morphology, having lost most of the derivational options found in its source language Tetun Terik. The main morphology retained is compounding, causative *ha-*, a limited use of detransitivizing *nak-*, and minimal reduplication (Williams-van Klinken, Hajek, and Nordlinger 2002:17–24).

As in other Austronesian languages, pronouns distinguish singular and plural, but not gender. Since gender is not shown on pronouns, nor on most nouns, it is not uncommon to tell an entire story without ever showing whether the story is about a male or a female.

6.2 OVERVIEW OF PORTUGUESE GRAMMAR. Portuguese clauses default to subject-verb-object, but have significant word order freedom. In part this is because verbs are inflected for the person and number of the subject, and predicative adjectives and nouns agree in gender and number with the subject. Another contributor is that third person pronouns and determiners are inflected for gender and number, such that participants can often be distinguished without the need for strict word order. Tense-aspect is obligatorily indicated within the verbal paradigm.

Within noun phrases, determiners and numerals precede the head noun, while relative clauses follow it. Ordinals and possessive adjectives normally precede but can also follow the head noun, while adjectives normally follow but can precede in some cases. Adjectives, ordinals, possessive pronouns, determiners and numerals (to the extent that they distinguish gender) agree

obligatorily in gender and number with the referent noun (see [Hutchinson 2003](#) for more information).

The contrast in grammatical structure and morphological complexity between Tetun and Portuguese can be seen by comparing the Tetun sentence in (2) with its Portuguese equivalent in (3).

- (2) Horiseik hau nia organiza^{saun} sosa motór kiik rua.
 yesterday 1SG POSS organization buy motorbike small two
 ‘Yesterday my organization bought two small motorbikes.’
- (3) Ontem a minha organiza^{ção} comprou
 yesterday DEF.F 1SG.POSS.F organization.F buy.3SG.PST
 duas pequena-s motorizada-s.
 two.F small.F-PL motorbike.F-PL
 ‘Yesterday my organization bought two small motorbikes.’

As is typical for Romance languages, Portuguese has a rich derivational morphology. East Timorese, even those who do not speak Portuguese, are familiar with a number of Portuguese affixes as a result of borrowing sets of related words, particularly in the technical and abstract lexicon, e.g., *administradór* ‘administrator’, *administrasaun* ‘administration’, and *administrativu/a* ‘administrative’, all derived from the Portuguese verb *administrar* ‘administer’ which appears in Tetun in its third person singular form *administra*.

6.3 INTRODUCTION TO GENDER IN TETUN AND PORTUGUESE.

Nouns and adjectives, the two classes of words discussed in this article, are categories in which Portuguese influence is particularly great in Tetun Dili. In the previously referred to database of Tetun Dili lexemes, 52% of the 9400 nouns are from Portuguese, as are 67% of the 1180 adjectives.

In these classes of nouns and adjectives, there is a major clash between Tetun and Portuguese grammar. As already noted, unlike Portuguese where natural and grammatical gender marking is pervasive, Tetun does not traditionally mark grammatical gender, although, as we shall see, there is limited lexical expression of natural gender (as noted also by [Choupina 2011](#)).

The gender assignment system ([Corbett 2013c](#)) in Portuguese is in the first instance sex-based, with nouns denoting males being masculine, and those denoting females being feminine. For other nouns gender is a purely grammatical category. In the vast majority of cases, the gender for inanimate nouns can be predicted from phonological form. Many masculine and feminine forms are distinguished by final *-o* versus *-a*, e.g., *brasileiro/a* ‘male/female Brazilian’ or *integrado/a* ‘integrated (M/F)’.⁷ Another common pattern is for the masculine to end in *-or*, and the feminine to have a final *-ora*, e.g., *moderador/moderadora* ‘chairman/chairwoman’, *pastor/pastora* ‘male/female pastor’.⁸ Some gender assignment is morphological, for instance with abstract nouns derived using

7. To facilitate glossing, we do not normally indicate singular number, unless particularly relevant.

8. We use Portuguese spelling conventions in this paragraph to exemplify Portuguese. However, for Portuguese loans used within Tetun we follow Tetun spelling conventions, even where the original source of the example might have used mixed spelling. For instance, Portuguese final *-o* is spelled *-u* in Tetun.

-*ação* being feminine. For most other words, when referring to non-animates, gender is not self-evident, e.g., *mar* ‘sea.m’ versus *paz* ‘peace.f’.

With these general rules of thumb, and the marked presence of Portuguese loans in most varieties of Tetun Dili, even many East Timorese who do not master Portuguese grammar have some idea about gender marking in that language.

7. GENDER MARKING IN TETUN DILI

7.1 NATURAL GENDER IN TETUN NOUNS. A few native Tetun kinship terms distinguish between male and female, e.g., *inan* / *aman* ‘mother / father’, *biin* / *maun* ‘older sister / older brother’. There are very few other Tetun nouns that do so. For a more complete list see table 1. The exceptions we are aware of are *ferik* / *katuas* ‘mature woman / man’ (not uncommonly used to refer to one’s parents or spouse), *bui-huu* / *mau-huu* (lit. older.sister / older.brother-blow), informal terms for female / male informers to the Indonesian occupiers, and *bui-bere* / *mau-bere* (lit. older.sister / older.brother Bere (a traditional name)) ‘female / male native Timorese’. The latter two are terms adopted by the FRETILIN resistance party to honor the ‘common people’ during the early days of resistance to Indonesian occupation.

For other terms, instead of specifying gender within the noun, it is normally specified, if at all, by a following noun *mane* ‘male’ or *feto* ‘female’ for humans, e.g., *oan mane* (lit. offspring male) ‘son’, and *labarik feto* (lit. child female) ‘girl’. For animals, male and female terms can be derived using the nouns *aman* (lit. father) and *inan* (lit. mother), e.g., *manu aman* (lit. bird/chicken father) ‘rooster’, and *bibi inan* (lit. goat mother) ‘nanny-goat’.

7.2 NATURAL GENDER IN BORROWED NOUNS. All kinship terms borrowed from Portuguese are gender-specific for all speakers of Tetun Dili, regardless of sociolect, e.g., *apaa* / *amaa* ‘dad / mum’ (native Tetun *aman* / *inan*), *kunhadu/a* ‘brother- / sister-in-law’, *padrastu* / *madrasta* ‘stepfather / stepmother’ (native Tetun *aman kiik* / *inan kiik*), *primu/a* ‘male / female cousin’, *subrinhu/a* ‘nephew / niece’, and *tiu/a* ‘uncle / aunt’. Note that most of these pairs have a shared root, differing only in final *-u/-a*. Some do not have native Tetun equivalents because traditional Timorese family structures are so different to Portuguese ones.

The one exception in which a kin term borrowed from Portuguese has lost its gender distinction in Tetun is *avoo* ‘grandparent’. This merger has occurred as a result of neutralization of the distinction between close-mid /o/ in Portuguese *avô* ‘grandfather’ and open-mid /ɔ/ in *avó* ‘grandmother’, due to Tetun having only one mid back vowel.

Apart from kin terms, the vast majority of nouns referring to people are traditionally borrowed from Portuguese in the masculine form as gender-neutral terms, e.g., *funsionáriu* ‘office worker’, used in Tetun to refer to either men or women. This tendency is sufficiently strong that *bapa* (from Indonesian *bapak* ‘father, respectful title for a man’) has been borrowed as a gender-neutral term

for Indonesians in general. Since independence, the English word *mister* has been adopted as an informal term of address for white English-speaking foreigners, with some speakers using it gender-neutrally, while others distinguish between male *mister* and female *misís*.

There are, however, a few non-kin relationship terms from Portuguese for which all Tetun speakers distinguish between men and women, in particular *senhór/senhora* 'sir / madam', *padre/madre* 'priest / nun', *irmaun/irman* '(religious/social) brother / sister', *mestri / mestra* 'male / female teacher', *profesór / profesora* 'male / female teacher' and *enfermeiru/a* 'male / female nurse'. These are all high-frequency nouns which even young people would have had to use correctly to address senior figures during the Portuguese era. In the case of teachers and nurses, a significant proportion is – and was – in any case women.

The number of feminine Portuguese nouns in use within Tetun Dili to refer to females is however rising significantly, as the media follows Portuguese and elite Timorese practice in using feminine nouns to refer to women. So, for instance, all ambassadors, regardless of whether they were male or female, used to be commonly referred to as *embaixadór*, e.g., *Senhora Joana embaixadór Australia* 'Mrs. Joana is Australian ambassador.M'. Increasingly, however, acrolectal and mesolectal speakers are distinguishing between male *embaixadór* and female *embaixadora*, thus now preferring *Senhora Joana embaixadora Australia* 'Mrs. Joana is Australian ambassador.F'.

To the best of our knowledge, there is only one feminine noun which in Portuguese is restricted to females but which in Tetun is gender-neutral. This is *baka/vaka*, from Portuguese *vaca* 'cow'. In Tetun, *karau baka/vaka* (lit. buffalo cow) indicates cattle (regardless of sex) as opposed to *karau (Timor)* 'water buffalo'. There appear to be no word pairs, whether native or borrowed, lexicalizing gender distinctions for animals in Tetun, in marked contrast to Portuguese which has, for example, *vaca* 'cow' versus *touro* 'bull'.

7.3 NATURAL GENDER IN ADJECTIVES. Native Tetun adjectives do not distinguish between male and female. When adjectives are borrowed from Portuguese, it is (as with nouns) nearly always the masculine singular form that is borrowed, and used neutrally for both men and women, whether individually or in groups, e.g., *fraku* 'weak' or *riku* 'rich'.

Natural gender, with masculine used for males and feminine for females, is used consistently by all speakers only for *bonitu/a* 'handsome.M / pretty.F', which in Tetun can be applied only to humans, in contrast to the more general Portuguese source *bonito/a* 'nice, handsome, pretty'. All speakers distinguish male and female *bonitu/a* even though native terms for the concept are gender-neutral, e.g., *oin diak* (lit. face good) 'good-looking, handsome, pretty'. Note that the *bonitu/a* distinction involves natural or semantic gender marking rather than grammatical agreement, it being very possible for the feminine form to modify gender-neutral Tetun nouns when describing females, e.g., *ema bonita* (lit. person pretty.F) 'pretty woman', or *oin bonita* 'pretty face' in example (4). The latter example illustrates that the adjective does not necessarily agree with

the noun it modifies (here gender-neutral *oin* ‘face’), but rather with the sex of the person being described. Baklanova (2016:32f) in her description of gender in Tagalog also notes some instances of gender-sensitive adverbial modifiers or adjectives modifying “an *object* related/possessed by the [feminine human] controller”, e.g., *Graciosang umupo ang babae* ‘The woman sat gracefully.F.’

- (4) Nusaa “Noi Oan” sira ho tinan sei
 how.come dear.F offspring DEF.PL with year still
 kiik, isin sei nurak teb~tebes,
 small body still young ADV~true
 ‘Why is it that these “young dears” who are still young in years, their
 bodies still very young,

oin bonita tanan~tanan deit, buka fali vida ho prostituisaun?
 face pretty.F ADV~bare just seek instead life with prostitution
 their faces all pretty, go looking for a life of prostitution?’
 (quote in the newspaper Suara Timor Lorosae, 2007)

It is presumably because *bonitu/a* are restricted to human referents, and because beauty is perceived differently for men than for women, that gender distinction can be obligatory for this word pair. This could explain why *bonito/a* (from Spanish) is also the most common one to retain this distinction in Chamorro (T. Stolz 2012).

Another adjective for which many Tetun Dili speakers regularly distinguish gender is *kuriozu/a* ‘interested, curious’. This is a relatively new high-level loan, which presumably retains its gender distinction because it is normally used to refer to humans and is mostly used by people with Portuguese influence (Dulce Marques, pers. comm.). For other adjectives describing humans, the feminine form is very rare, but even that is possible if the language is sufficiently Portuguese-influenced, as illustrated by example (5), describing a historically important nun. It seems that gender marking is more likely when speaking of highly respected people.

- (5) Nia hetan isin-diak maibee sai fraka liu.
 3SG get body-good but become weak.F very
 ‘She regained her health, but became very weak.’ (An article on the
 co-founder of FMA, a religious order, 2012)

All the above cases of gender are of ‘natural gender’, in which masculine applies to males, and feminine to females. ‘Grammatical gender’, for which the defining characteristic is the presence of agreement phenomena (Corbett 2013a) is discussed in section 9 below.

In a specifically local minor innovation, there are also quite a few adjectives for which Tetun uses a Portuguese form ending in *-a* without this being specifically feminine gender in Tetun, or acceptable as a feminine adjective in Portuguese. Some are adjectives derived from borrowed feminine

Portuguese nouns, e.g., *kalma* ‘calm’ from the noun *calma* ‘calm’, and *forsa* ‘strong’ from the noun *força* ‘strength’. Some have final *-a* because the cognate in Indonesian (at this stage the better-known non-local language) does, e.g., Adj./N. *karnívora* ‘carnivorous’ (Portuguese adjective *carnívoro/a* and noun *carnívoro*, Indonesian *karnivora*), and *ibrida* Adj./N. ‘hybrid’ (Portuguese adjective *híbrido/a* and noun *híbrido* with silent ‘h’, Indonesian *hibrida* with pronounced ‘h’). Yet others have final *-a* for most speakers because they are most often heard in conjunction with feminine Portuguese nouns, as noted also by Greksakova (2018:215), e.g., *morna* ‘tepid’, typically collocated with *água* ‘water.F’, and *klimátika* ‘climatic’, usually heard in conjunction with feminine *mudansa* ‘change’ or *alterasaun* ‘change’. Finally, some adjectives have final *-a* because they derive from the 3SG present tense form of Portuguese verbs, this being the form always used when verbs are borrowed. Examples include *aperta* ‘tight’ from Portuguese Vt *apertar* ‘tighten’ (alongside more acrolectal *apertadu* from the Portuguese adjective *apertado* ‘tight’), *keima* ‘blown (e.g., of fuse)’ from Portuguese Vt *queimar* ‘burn’ (alongside more acrolectal *keimadu* from the Portuguese adjective *queimado* ‘burnt’).

One adjective, *morena* ‘light skinned’, is by most speakers used only in the feminine, simply because lightness of skin is spoken of for women rather than for men (Leoneto Ribeiro, Dulce Marque, Abreu Ximenes pers. comm.).

8. SUBTLE DISTINCTIONS: TETUN PRIORITIZES FEMININE, PORTUGUESE MASCULINE

8.1 GENDER INTERPRETATION OF MASCULINE PLURAL NOUN PHRASES. Portuguese gender patterns have impacted Tetun Dili not only in the adoption of gendered pairs of loanwords, but also in more subtle ways. We have noted three ways in which Portuguese loans prioritize the masculine, whereas Tetun prioritizes feminine. One is in the interpretation of plural masculine nouns, for which there is a sharp distinction between the interpretation of borrowed and indigenous nouns.

For native Tetun nouns which refer specifically to males, the plural does not include females. For instance, *maun sira* (lit. older.brother DEF.PLUR) ‘the older brothers’ includes only older brothers, not older sisters, and *katuas sira* (lit. mature.man DEF.PLUR) ‘the older men’ includes only men, not women.⁹ The exception is *maun-alin sira* (lit. older.brother-younger.sibling DEF.PLUR) ‘brothers and sisters’, which for most speakers includes older sisters even though the term does not specifically mention them. The partially borrowed expression *apaa boot sira* (lit. father big DEF.PLUR) ‘father’s older brothers’ too includes only men. Presumably the Tetun adjective *boot* makes this a Tetun expression from a gender inclusion point of view, despite the Portuguese head *apaa*.

9. The interpretation is different when *sira* is used as an associative plural to mean ‘and associated people’ as opposed to being a definite plural marker. In this case *maun sira* would be interpreted as *maun* (a particular older brother) and his wife or other associated people – something which is not possible in Portuguese.

(6) Ita boot sira hotu hakilar mai ami
1PL.INCL big DEF.PL all shout come 1PL.EXCL
deputadu sira.
parliamentarian.M DEF.PL
‘You¹⁰ are all shouting at us parliamentarians.’ (Said by a female
parliamentarian addressing parliament in 2011.)

8.2 ORDERING OF GENDERED WORD PAIRS. Another area in which the effect of Portuguese gender is seen is in the order of gendered pairs of words. Here too there is a sharp discrepancy between native pairs of terms distinguishing male and female, and borrowed Portuguese pairs. And here too Tetun prioritizes the feminine, while Portuguese prioritizes the masculine. In Tetun pairs, the female is named first, as shown by the examples in table 1 below.

TABLE 1. DEFAULT ORDER OF MALE/FEMALE PAIRS

Tetun pairs		Portuguese pairs	
<i>inan - aman</i>	'mother - father'	<i>apaa - amaa</i>	'dad - mum'
<i>feen - laen</i>	'wife - husband'	<i>pai - main</i>	'father - mother'
<i>feton - naan</i>	'sister (of man) – brother'	<i>irmaun - irmán</i>	'brother - sister' (religious/social)
<i>ferik - katuas</i>	'mature woman - man'	<i>tiiu - tia</i>	'uncle - aunt'
<i>feto - mane</i>	'female - male'	<i>kompadre - komadre</i>	'co-father – co-mother' ¹¹
<i>oan feto – oan mane</i>	'daughter - son'	<i>kunhadu – kunhada</i>	'brother-in-law / sister-in-law'
		<i>padrastu - madrastra</i>	'step-father / step-mother'
		<i>padrinhu – madrinha</i>	'godfather / godmother'
		<i>noivu - noiva</i>	'groom - bride'
		<i>primu - prima</i>	'male / female cousin'
		<i>subrinhu - subrinha</i>	'nephew - niece'
		<i>senhór - senhora</i>	'sir - madam'
		<i>mestri - mestra</i>	'male / female teacher'
		<i>profesor – profesora</i>	'male / female teacher'
		<i>deputadu – deputada</i>	'male / female parliamentarian'
		<i>enfermeiru - enfermeira</i>	'male / female nurse'
		<i>santu - santa</i>	'male / female saint'
	'girl – boy [slang] (lit. small / large gecko)		

11. One's *komadre* and *kompadre* are the godparents of one's child, or the parents of one's godchild.

Note that some pairs are normally compounded, e.g., *inan-aman* (lit. mother-father) to mean 'parents', while others are normally separated by a coordinator, e.g., *feton ho naan* (lit. sister.of.man and brother.of.woman') to mean 'brother and sister'.

In Portuguese loan pairs, in contrast, the male is mentioned first, as per Portuguese, even for very common loans such as *apaa-amaa* 'dad-mum'. The Portuguese list is much longer than the Tetun one because lexicalized gender distinctions are standard in Portuguese and uncommon in Tetun.

There are two exceptions in which Tetun normally mentions the male first. One is the relatively recent pair *mauhuu-buihuu* 'male / female informer to the Indonesians', presumably as male informers far outnumbered female ones. The other is *maun-alin inan-feton sira* (lit. older.brother-younger.sibling mother-sister.of.man DEF.PL) 'ladies and gentlemen', a formula sometimes used in speeches to address the audience.

There are many masculine/feminine pairs which are not normally spelled out specifically. For instance, one does not hear *embaixadôr-embaixadora* to include male and female ambassadors, or *doutôr-doutora* to include all doctors. This may be in part because these roles were traditionally held by men (Dulce Marques, pers. comm.). However, it is also because, for these terms, the plural is normally indicated using just the masculine form, as per Portuguese.

Indonesian pairs, when used, follow the Indonesian order, *ibu-bapak* 'madam-sir' and *mama-papa* 'mummy-daddy', which in both these cases puts the female first. The only English pair we are aware of in Tetun follows the English and Portuguese order of putting the male first, namely *mister-misis* 'Mister (and) Mrs.'

8.2.1 Gender suffixes on English loans and bigendered human nouns in Portuguese. The priority given to masculine forms in Portuguese loans is evident even in the few words adopted from English which have been made to look like Portuguese or 'Lusified' (without wide acceptance) by the addition of final *-u*, e.g., *environmentu* 'environment', *asesimentu* 'assessment', and *komitmentu* 'commitment' (adopted via Indonesian *komitmen*). These examples are patterned on the many cognates in which Portuguese has final *-mentu* while English has *-ment*, e.g., *sakramentu* 'sacrament'. So too the Indonesian adjective *rapi* 'well-dressed', is sometimes 'Lusified' as *râpidu*.¹² We are not aware of any such nouns or adjectives being adapted by the addition of final feminine *-a*, nor of any native Tetun nouns being made to look Portuguese in this way using final *-a*.

It is worth noting here that there are some human nouns in Portuguese which end in *-a*, but which are bigendered, that is, they are masculine for male referents and feminine for female referents, e.g., *motorista* 'motorcyclist, driver (M/F)' and *diplomata* 'diplomat (M/F)'. We have never noted Tetun speakers misinterpreting these forms as underlyingly feminine and then attempting to form 'masculine' nouns by replacing the final *-a* with *-u*.

12. The unrelated but well-known Portuguese loan *râpidu* 'rapid' provides a phonological model for this innovation.

9. GENDER AGREEMENT

9.1 AGREEMENT WITH PORTUGUESE NOUNS IN TETUN DILI.

Having looked at the expression of natural gender, we turn now to agreement phenomena involving Portuguese loans in Tetun Dili. Fixed phrases borrowed from Portuguese come complete with Portuguese gender agreement, and are used by all Tetun speakers in this form, e.g., *komisaun organizadora* (lit. committee.F organizing.F) ‘organizing committee’. These fixed phrases also follow Portuguese word order, which in certain circumstances puts the adjective before the noun, e.g., *primeira dama* (lit. first.F lady.F) ‘first lady’, whereas in Tetun all adjectives traditionally follow the noun.

When a Portuguese adjective modifies a Portuguese noun outside of fixed phrases, patterns vary, although in speaking most people use the standard masculine singular form of borrowed adjectives regardless of the Portuguese or natural gender of the head noun. Twenty years ago this pattern was so dominant in the mesolectal variety of Tetun Dili that the possibility of applying Portuguese agreement rules outside of fixed phrases was not even mentioned in our 2002 grammar (Williams-van Klinken et al. 2002).

Grammatical gender agreement is, however, always potentially possible, even if not necessarily used, in the acrolect, and indeed it is becoming increasingly common in it, with some flow-on into the mesolect. Hull (1999:xxiv) noted such gender agreement as early as 1999, as did Hajek (2006), who nevertheless observed that it was avoided by most people. While most people still say *nasaun demokrátiku* (lit. nation.F democratic.M) ‘democratic country’, acrolectal speakers tend to favor *nasaun demokrátika* with feminine agreement. Gender agreement applies mainly to high-level vocabulary, and is conditioned by such factors as relative formality, register, context and speaker status, leading to significant variation in the acrolect – whether spoken or written. The second author’s lexical database currently lists masculine/feminine forms for some 230 adjectives, less than 10% of which could describe humans, e.g., *Australianu/a* ‘Australian.M/F’, *kapturadu/a* ‘captured M/F’. In reality, though, almost any Portuguese adjective can be used in the feminine form if the Tetun is Portuguese-influenced enough. Normative dictionaries such as Hull (2002b), and Correia et al. (2005) list contrasting masculine and feminine forms, where they exist, for Portuguese nominal and adjectival loans, a clear indication that gender agreement is acceptable to these normative dictionary writers, even if it is not necessarily activated in practice.

Outside of fixed phrases, the order of noun and borrowed adjective almost always follows the Tetun pattern, with the noun occurring initially, as in *faze primeiru* (lit. phase.F first.M) ‘initial phase’. In formal written texts and highly Portuguese-influenced speech, however, the reverse order is possible when Portuguese allows it, e.g., *primeira faze* (lit. first.F phase.F). In this case there is always gender agreement, consistent with this particular structure being so heavily Portuguese-influenced.

nouns took the default masculine adjectives, regardless of whether these were attributive or predicative. The remainder, which took feminine adjectives, were all in written texts, for which she posits an explanation of bad editing. For feminine nouns, 75% of attributive adjectives were feminine, mostly in fixed expressions, which as noted above always have agreement. In other constructions, the proportion of feminine agreement was lower, being 18% for predicates, and 8% for adjectives within relative clauses. It is unclear why her figures for predicates were higher than for relative clauses, whereas our observations are that the reverse are true; the most likely explanation is the small number of examples.

Greksakova found a clear distinction in oral texts between speakers who claimed to speak Portuguese, and those who claimed not to. Those who claimed to speak Portuguese had 95% feminine agreement within NPs, and 50% in predicates and relative clauses, while those who claimed not to speak Portuguese had 67% feminine agreement within NPs (mostly with preceding ordinals and in fixed expressions), and no agreement within predicates or relative clauses. Another distinction was that feminine agreement was higher for written than oral texts. Both observations are in line with our own.

While the basilect has no gender agreement, and the acrolect has a tendency to use it, journalistic practice which combines acrolectal and mesolectal Tetun, is inconsistent or mixed. Journalists sometimes use a standard masculine singular adjective, sometimes apply gender (and number) agreement, and sometimes use erroneous agreement, often within the one article. Example (10), for instance, has a feminine singular head noun, albeit with clear plural reference, with gender-unmarked plural, singular feminine and singular masculine adjectives modifying it.

- (10) ... eskola sira mak hanesan teknika profesionais
 school.F DEF.PL FOC like technical.F-SG professional.PL
 '[The reality clearly shows that] the schools such as vocational
 technical (schools)

tantu privada no publiku sira, ...
 both private.F-SG and public.M-SG DEF.PL
 both private and public [are mainly interested in agriculture ...]' (Tubun
 Foin Sae program, Radio Timor, 2009)

9.2 INCIPIENT GENDER ASSIGNMENT TO TETUN NOUNS AND TO AN INDONESIAN NUMERAL. The discussion so far has been of gender agreement with borrowed Portuguese nouns. Here we turn to an incipient process of gender assignment to a few nouns not of Portuguese origin, a process which has been evident in recent years.

Sometimes a feminine adjective is used when modifying a Tetun noun which is a synonym for a well-known Portuguese feminine noun, especially

lian portugeza (lit. language Portuguese.F) ‘Portuguese language’, imitating the frequently used Portuguese expression *língua portugeza* (lit. language.F Portuguese.F), and *uma kreda katólíka* (lit. house faith Catholic.F) ‘Catholic church’¹³ paralleling the common Portuguese expression *igreja katólíka* (lit. church.F Catholic.F).

The feminine form of *katólíka* is so predominant in texts that many people use it even for *ema katólíka* (lit. person Catholic.F) ‘Catholic person’, regardless of the sex of the referent. In fact, while an informal survey of seven university lecturers showed that they all accepted only *ema katólíku* without agreement as correct, a Google search resulted in 3040 examples of *ema katólíka*, versus only 175 for *ema katólíku*, indicating that feminine marking in this case is now completely normal, even if not yet accepted as normative in this almost exclusively Catholic nation.

There is evidence to suggest that the noun *ema* ‘person’ may be undergoing a more general process of feminine gender assignment in Tetun, beyond this one expression. While forms such as *ema Portugés* ‘Portuguese person’ and *ema Australianu* ‘Australian person’ without agreement are usual, one can now find examples of *ema Portugeza/Australiana* with the feminine form of the adjective, regardless of the sex of the referent, as illustrated in (11). There appear to be several drivers for this new pattern. Firstly, *pessoa*, the Portuguese equivalent of *ema*, is feminine and is frequently used in formal registers when identifying numbers of participants (e.g., *trinta pessoas* ‘thirty people’), though rarely with modifying adjectives. Secondly, the high frequency of *ema katólíka* serves as a model. Finally, it is more than likely that the final /a/ in *ema* encourages a feminine interpretation given the clear phonological association between that particular vowel and feminine gender in the Portuguese gender system.

- (11) Peter Coleman, ema Australiana, uluk hanesan
 Peter Coleman person Australian.F formerly like
 vise-prezidente ...
 vice-president
 ‘Peter Coleman, an Australian, used to be vice-president [of a large
 company ...]’ (Sapo, 14/5/2011)

Amongst Portuguese loans used in Tetun, there are a few nouns which are masculine despite ending in *-a*. There is evidence that these are being analyzed as feminine by some speakers of Tetun. For instance, of the 4,896 examples of *problema(s)* ‘problem.M) in our textual database, 60 have gendered Portuguese attributive adjectives. Of these, only 63% have the default masculine form which would also be correct according to Portuguese grammar, e.g., *problema politíku* ‘political problem’. A full 25% have a feminine adjective, such as *problema polítíka*, and a further 13% have the final *-u/a* omitted to restore

13. The use of feminine *katólíka* with Tetun *kreda* was noted over twenty years ago by Hull (1994:357). Greksakova (2018:217) notes that feminine adjectives are occasionally used with Tetun words, speculating that this could be because these nouns are feminine in Portuguese.

Tetun-style penultimate stress, e.g., *problema politik*. Other examples from written texts in Google searches support this tendency. For example, of 170 instances of *sistema demokrátiku/a* ‘democratic system’, 57% have masculine agreement, while 43% have feminine non-agreement. Feminine gender assignment to *problema* and *sistema* does not only indicate lack of mastery of Portuguese gender exceptions, a common enough error to make for non-fluent speakers. More importantly in this context, it shows that some people who have not fully mastered Portuguese gender are nevertheless attempting Portuguese-style gender agreement in Tetun Dili based on what they assume to be completely predictable phonological form. It illustrates the cross-linguistic pattern of a word having ‘double gender’ for a period when there is a clash between two gender assignment systems (Corbett 1991:183), in this case between following Portuguese gender and assigning feminine gender to words with final *-a*.

Finally, we have a handful of examples (one oral, four from the press) of the Indonesian numeral *juta* ‘million’ being analyzed as feminine with a preceding Portuguese feminine *uma* ‘one’, e.g., *uma juta i tal* (lit. one.F million.I and more.P) ‘a million plus’. The fact that the form is modified by a Portuguese numeral is an indication that some speakers perceive the head noun as a Portuguese loan (with inherent gender), since only these should appear with Portuguese numerals in Tetun (Williams-van Klinken and Hajek 2018b). We have found no examples of *juta* occurring with masculine *um*, despite the fact that the commonly-known Portuguese synonym *milhão* is masculine, suggesting that for this word, at least, phonology (final *-a*), takes precedence over synonyms in gender assignment.

The above examples show that Tetun Dili speakers use two strategies for assigning gender to non-Portuguese nouns, namely:

1. Phonology, with words ending in *-a* defaulting to feminine, and *-u* to masculine, on the model of Portuguese.
2. Semantics, with words taking on the gender of a well-known synonym in Portuguese.

These appear to be common processes cross-linguistically, being amongst the five strategies listed by C. Stolz (2008:400) for (re)assigning gender to loans in already gendered European languages, and amongst the strategies discussed by Corbett (1991:71–82) for gender assignment in loans cross-linguistically, though he warns that semantic analogies should be handled carefully lest they obscure more general patterns.

10. DISCUSSION. In this study, we have seen a number of ways in which gender marking in Tetun Dili has been influenced by contact with Portuguese, as summarized in table 2, which compares native and borrowed expressions of gender.

**TABLE 2. SUMMARY OF GENERAL DIFFERENCES IN GENDER
BETWEEN NATIVE AND BORROWED PORTUGUESE
VOCABULARY IN TETUN**

Characteristic	Native Tetun vocabulary	Portuguese loans (all speakers and sociolects)	Portuguese loans (additional features of acrolect)	Portuguese rule
Human nouns indicate sex of referent	Rarely. To show sex, usually add <i>feto/mane</i> 'female/male'.	Obligatory for kin and common professions.	Also optionally for all human nouns.	Obligatory
Interpretation of masculine plural	Includes males only.	Includes males and females.		Includes males and females.
Order of gendered pairs	Female – male.	Male – female.		Male – female
Mark natural gender in adjectives	No.	Obligatory for <i>bonitu/a</i> 'handsome/pretty'.	Common but optional for <i>kuriozu/a</i> 'curious', rarely marked on other adjectives, but potentially possible.	Obligatory on all adjectives.
Grammatical gender agreement of adjectives with controller nouns	No gender.	Obligatory in fixed phrases.	Obligatory when adjective precedes noun (typically acrolectal construction). Optional elsewhere.	Obligatory.

When considering the table, a cline with two clearly separated end points (native Tetun and Portuguese) is clearly evident. As one moves along the cline, the application of Portuguese gender-marking in Portuguese loans increases with sociolectal status.

10.1 INNOVATIONS SHARED WITH SOME OTHER AUSTRONESIAN LANGUAGES. In terms of innovations in other Austronesian languages prompted by language contact with a Romance language, Tetun Dili shares some features with Tagalog (and other Philippine languages), as well as with Chamorro, spoken in Guam and the Northern Marianas. As already noted in sections 1 and 2 these particular languages are heavily influenced by the former colonial language Spanish. As in Timor-Leste, there has been massive lexical borrowing from the dominant Romance language, which is particularly high in nouns and adjectives, and much higher in the acrolect than the basilect. In all three regions, it was acrolectal speakers who traditionally were educated and fluent in the Romance language, and hence most likely to borrow lexicon as well as to maintain the gender distinctions that come with it (T. Stolz 2012). For useful historical overviews of the sociolinguistic context of these countries, see, for instance, Quilis and Casado-Fresnillo (2008) on the Philippines, Pagel (2010) on Guam, and Thomaz (1981) on Timor-Leste. Comments below on gender in languages

of these regions are, as partly reported in Section 2, based on T. Stolz (2012), who discusses both Chamorro and Philippine languages, as well as T. Stolz (2003) on Chamorro, and Baklanova (2016) and Reid and Schachter (2009:852) on Tagalog. More recently, as part of a broader study on Romancization in Chamorro and Tetun Dili, Stolz and Levkovych (in press) have postulated very different grammatical gender rules for each, in a manner consistent with our findings.

One significant difference to be borne in mind in any comparison between Tetun on the one hand, and Chamorro, Tagalog and other Philippine languages on the other is the very different historical sociolinguistic contexts they find themselves in today, as already alluded to above. In the case of Tetun, direct contact with Portuguese was maintained until 1975 and has been significantly renewed since 1999, since which time it has also been reinforced by use in the media and in schooling. In the case of the other languages, direct Spanish control was lost after the American-Spanish War of 1898, at which point Guam and the Philippines passed into American hands, at a time well before mass literacy and media. This means that any borrowed gender patterns have for some time now not received reinforcement by extensive knowledge and use of Spanish.

One innovation shared between Tetun Dili, Tagalog and Chamorro is the borrowing of a significant number of gendered pairs of words from Portuguese or Spanish respectively (section 7.2), even though the native vocabulary has very few such pairs (section 7.1). Based on data in T. Stolz (2012), it appears that Chamorro, like Tetun, has adopted many more pairs of gendered words than appears to be the case for Tagalog.¹⁴ The core in all three languages appears to be borrowed kin terms plus a limited number of basic professions. In both Tetun Dili and Chamorro, the only adjective to obligatorily show sex distinctions is the Spanish/Portuguese adjective *bonito/a* 'handsome/pretty'.

One question arises as to why Tetun Dili should adopt so many gendered pairs when the native vocabulary has very few. A critical fact is that the Portuguese category of gender is at its core sex-based. This means that for human nouns it exactly matches the semantic distinction that can be made in Tetun through the addition of *feto* 'female' or *mane* 'male'. A second factor is that the Portuguese gender system is largely phonologically transparent. Thirdly, there has been the previously mentioned large-scale adoption of kinship terminology from Portuguese, with its clearly gender-marked terminology, e.g., *tia/tiu* 'aunt/uncle'. These high-frequency loans provide an important entry point into Tetun for the concept of lexical gender-marking on the Portuguese model. The spread of gender distinctions from kin terms and common professions into other professions is being led by the elite and spread, as previously noted, through such things as the media, a

14. T. Stolz (2012:100) notes discrepancies in previous reports of Spanish-style gender marking in Tagalog, in that Schachter and Otones (1972:96, 197) insist that the class of items with a gender distinction is very small, while Wolff (2001:241) indicates some productivity.

reminder of the huge influence of sociolect and register in Tetun Dili gender marking.

A second innovation also shared at least with Chamorro (T. Stolz 2012:120) is the retention of gender (and number) agreement within borrowed fixed phrases, such as the Portuguese loan in Tetun *Nasoens Unidas* (lit. nation-F.PL united-F.PL) ‘United Nations’ (section 9.1). This is the case for all Tetun speakers in all contexts, showing that these expressions are fully lexicalized.

As previously mention in section 2, T. Stolz (2012) formulates three restrictions on grammatical gender in Tagalog, against which we will compare the Tetun Dili situation, showing significant differences:

1. Gender only applies to Spanish loans. Baklanova (2016) modifies this to include blended words with non-Spanish roots and gendered Spanish derivational affixes.
2. Agreement only applies for nouns with human controllers (thus reflecting natural gender).
3. Masculine/feminine distinctions are neutralised in the plural.

The first restriction is largely true in Tetun Dili as well, in that gender only applies to Portuguese loans, but is showing signs of breaking down, with the recent appearance of a few native words, e.g., *uma kreda*, *ema* which show signs of grammatical gender categorization (see section 9.2 above).

The second restriction noted by Stolz for Tagalog is largely true of basilectal speech in Tetun Dili, in which there is only one adjective that distinguishes gender (*bonitu/a* ‘handsome/pretty’), this adjective agreeing semantically rather than grammatically with a human controller noun. The two major departures from this restriction in Tetun Dili are the retention of agreement markers in fixed phrases (not mentioned for Tagalog but also found in Chamorro), and the optional use of productive Portuguese agreement with non-human controllers in acrolectal speech, and to a much lesser extent in the mesolect.

Stolz’s third restriction does not apply in Tetun Dili. That is, gender restrictions are not neutralized in the plural. Groups of women would be described by all speakers using feminine *bonita* ‘pretty’ rather than masculine *bonitu*. For highly Portuguese-influenced speakers, grammatical gender agreement can readily be indicated on pluralized adjectives, e.g., *desizoens drástikas* (lit. decision.F.PL drastic.F.PL) ‘drastic decisions’.

10.2 INNOVATIONS NOT NOTED IN OTHER AUSTRONESIAN LANGUAGES. There are two innovations which we have not noted in descriptions of other Romance-influenced Austronesian languages, namely purely grammatical gender agreement outside of fixed phrases, and incipient grammatical gender assignment to non-Romance lexemes. There is a third characteristic which we have not seen described, at least in the Austronesian

context, namely the generally sharp distinction maintained between gender in native and borrowed terms.

10.2.1 Gender agreement. One major innovation, found primarily in the acrolect, is optional grammatical gender agreement between Portuguese adjectives and Portuguese nouns, even outside of fixed phrases, e.g., *kestaun ideolójika* (lit. issue.F ideological.F) ‘ideological issue’ (section 9.1). This is purely grammatical gender agreement, with no basis in real-world gender. We have not noted genuine grammatical agreement (without any correlation to sex distinctions) of the type found in Tetun in the descriptions of any other Austronesian languages. Indeed, worldwide, adoption of grammatical gender agreement is extremely rare (Nichols 2003:303, Aikhenvald 2000:388), so much so that borrowing is not mentioned as a source for grammatical agreement systems in Corbett’s (1991:137–43, 309–14) discussion of sources of gender systems.

Such grammatical agreement is quite common in the acrolect, but totally absent in the basilect. It is common for high-level vocabulary, but extremely rare for common everyday vocabulary. It is obligatory when the adjective precedes the noun in the Portuguese pattern, but optional when the adjective follows the noun in the native Tetun pattern. It is also more likely when the adjective is in close proximity to the controller noun than if there are other words separating the two.

Why should such agreement be adopted? And why should there be such strong sociolinguistic and grammatical influences on it? A major factor is the high degree of Portuguese influence on acrolectal Tetun Dili speakers, much higher than for mesolectal and especially basilectal speakers. According to the five point borrowing scale proposed by Thomason and Kaufman (1988), acrolectal Tetun Dili is at level four on the scale. This level is characterized by strong cultural pressure, with moderate structural borrowing, such as is found in Tetun Dili, for instance in the adoption of verb-subject order for certain verbs, and in the adoption of passives (Williams-van Klinken and Hajek 2018a). In this study structural borrowing is noted in the adoption of grammatical agreement, which is absent in the much less Portuguese-influenced basilect.

Cross-linguistically, level 4 on the borrowing scale is characterized by “very extensive bilingualism among borrowing language speakers, social factors strongly favoring borrowing” (Thomason 2001:70). In Timor-Leste, it is primarily the older elite who are fluent Tetun-Portuguese bilinguals, with the vast majority of the middle-aged and younger elite and virtually all journalists having been educated in Indonesian. However, the Portuguese-educated elite dominate the media and news, so modeling Portuguese-influenced speech to the wider community. In addition, many Indonesian-educated people have consciously adopted Portuguese loans in preference to the better-known Indonesian ones, as a sign of East Timorese nationalism. This probably carries over into adoption of Portuguese-style grammatical features. Higher

Portuguese influence also suggests knowledge of Portuguese, and higher social status, with Portuguese being the prestige language in Timor-Leste.

Another factor favoring the adoption of Portuguese gender agreement in acrolectal Tetun Dili may involve processing requirements (Matras 2007, 1998). Merging the patterns of Tetun and Portuguese reduces the processing load on fluent Portuguese-Tetun bilinguals as they move between these two languages in daily life. Constant stripping of Portuguese gender marking and agreement when using Portuguese loans in Tetun would arguably place a higher processing load on these bilingual speakers.

10.2.2 Incipient grammatical gender assignment to native and other non-Portuguese lexemes. As noted in section 9.2, there is evidence of an incipient process of grammatical gender assignment to some Tetun nouns as well as the Indonesian numeral *juta* ‘million’. This gender assignment has no relation to any natural gender of the referent, being noted only in grammatical gender agreement on modifiers. Although the literature does include occasional examples of sex-based gender assignment to indigenous words (e.g., Tadmor 2007), and numerous examples of gender assignment to loans in languages which already have gender systems (e.g., Corbett 1991, C. Stolz 2008, 2009), we have not noted any examples of purely grammatical gender assignment in languages which have no native grammatical gender system.

10.2.3 Subtle distinctions between indigenous and borrowed gendered pairs. While discussions of gender in contact have traditionally focused on the borrowing of grammatical forms, detailed analysis of Tetun has also brought to the surface more subtle effects in the target language. Although Tetun Dili has adopted many Portuguese pairs of gendered words referring to humans, these are sharply distinguished from native Tetun pairs in two ways. Firstly, for masculine plural noun phrases, Tetun expressions include only males, while Portuguese terms are interpreted to include females, as per Portuguese rules (section 8.1). Secondly, when mentioning both the male and the female terms, for native Tetun pairs the female is mentioned first, while for Portuguese pairs the male is mentioned first, just as in Portuguese (section 8.2). We have not seen mention in the literature of such distinctions in other languages, though this could well reflect a paucity of detailed descriptions rather than uniqueness of Tetun Dili.

10.3 POTENTIAL INNOVATIONS NOT ADOPTED IN TETUN. Both Chamorro and Tagalog have productively adopted agentive and other human-related affixes from Spanish, complete with gender distinctions, e.g., Chamorro *danderu/a* ‘singer (M/F)’ from Chamorro *dandan* ‘sing’ and Spanish agentive *-ero/a*, and Tagalog *binatilyo/a* ‘teenager (M/F)’ from Tagalog *binata* ‘young person’ and Spanish diminutive *-illo/a*. Both are morphologically rich languages (Schachter and Otnes 1972, T. Stolz 2012).

Tetun Dili, in contrast, is a strongly isolating language. Despite centuries of contact, it has borrowed only one Portuguese affix, the agentive *-dor*. However it has adopted it not as an affix but as a separate lexeme, with restricted semantics, and without gender or number marking, e.g., *dukur door* ‘sleepyhead’ (M/F) from *dukur* ‘sleep’ (Williams-van Klinken and Hajek in press, [Hajek and Williams-van Klinken 2003](#)). This restricted behavior in Tetun Dili is in sharp contrast to the Portuguese affix *-dor*, and despite the fact that Tetun Dili has adopted well over a hundred Portuguese nouns ending in *-dor*, which can in principle be marked for gender and number in acrolectal Tetun, just as they are obligatorily in Portuguese, e.g., masculine versus feminine *administrador(es)* (M.SG/PL) / *administradora(s)* ‘administrator(s)’ (F.SG/PL). The fact that Tetun has not adopted separate singular and plural or masculine and feminine forms of this morpheme for derivations with native roots reflects the lack of these categories in native vocabulary and the marked resistance of native lexicon in a highly isolating language to borrowed affixation even under intense contact.

10.4 TYPOLOGICAL GENERALIZATIONS. Analysis of the Tetun Dili data allows us to make more general typological observations, in particular to formulate some implicational scales for contact-induced gender marking and agreement, which can be tested against other languages in Austronesia and elsewhere.

1. Contact-induced gender phenomena are very sensitive to lectal type, occurring much more in the acrolect than the mesolect, and more in the mesolect than in the basilect (see also [Tadmor 2007](#), [T. Stolz 2012](#)). Related to this, they occur more commonly in technical vocabulary, most widely available to and used by acrolectal speakers, who are relatively highly educated and comfortable with using such language, than for the everyday vocabulary used by all speakers, regardless of educational background.
2. Contact-induced gender phenomena occur almost exclusively in loans from the gendered language, spreading very little to native vocabulary. This includes grammatical gender agreement, which in Tetun has so far spread to native vocabulary in only a handful of cases.
3. Gender agreement is expected in borrowed fixed phrases. Indeed it is a sign that the phrase has been adopted as a lexeme.
4. Gendered pairs of human-related lexemes follow the following scale:

kin > common professions > other professions.

Kin terms, given their frequent use and importance in daily interactions, are not surprisingly a critical entry point for gender distinctions understood by all speakers, regardless of sociolect. Our data shows that in

Tetun gender distinction also applies to a small number of basic professions that all speakers are likely to have contact with. Evidence shows gendering is now spreading to other professional terms, typically in the acrolect but now also in the mesolect.

5. Adjectives which semantically distinguish essential male and female human traits such as attractiveness are more likely to show regular gender-marking than other human-related adjectives.

‘handsome/pretty/goodlooking’ > other adjectives describing humans

This tentative observation is based on Chamorro *bonito/a* ‘handsome/pretty’ showing exactly the same pattern of obligatory marking as *bonitu/a* in Tetun Dili. However so too does *kariñoso/a* ‘affectionate/loving/warm’ in the former and *kuriozu/a* ‘curious’ in the latter (used only the acrolect and mesolect), which while both being human traits, do not obviously show a high degree of gender distinction.

6. Agreement follows a predictable scale in terms of word order and syntactic structure:

attributive adjective in non-native position > attributive adjective in native position > attributive adjective in native position but separated from it by other words > predicative adjective

Our scale interacts with both social and grammatical factors. In Tetun Dili, the non-native prenominal position which is possible with some Portuguese adjectives is highly marked, and more likely to be used by those with greater command of Portuguese which always requires gender marking. More generally, eyeballing our corpus data shows that acrolectal speakers show a clear preference for agreement between directly juxtaposed noun and adjective in the noun phrase (NP) than in those contexts to the right in the scale. In addition to relative proximity to the nominal head (inside NP, i.e. ([N Adj_{PORT}] v [N X Adj_{PORT}])), the type of grammatical relationship (noun and adjective in the noun phrase ([N Adj_{PORT}]) v subject (S) + adjective in the predicate ([S=NP] [Predicate=Adj])) clearly has an effect. There are a number of plausible explanations for the latter effect. In the first instance, there is reduced syntactic cohesion between subject and predicate coupled with the frequent appearance of intervening elements, e.g., adverbs, negators or focus markers, between subject and predicative adjective. In the second instance, Matras (2015) argues that inflectional borrowing in the predicate is much less common because of identity effects involving bilingual speakers, who in his model show their affiliation to a language by the use or non-use of language structures in this critical context, e.g., affiliation to Tetun in this case, by avoiding non-native agreement here.

11. CONCLUSION AND FUTURE DIRECTIONS. Our detailed presentation and analysis of gender in Tetun Dili, in both native and borrowed lexicon and grammar, shows how different, complex and variable gender can be in the same language – as a result of language contact. The target (Tetun Dili) and source language (Portuguese) are typologically very different: the former is a largely isolating Austronesian language with no grammatical gender and limited lexical natural gender differentiation, while Portuguese is a morphologically complex Romance language with a well-developed sex-based grammatical gender system. Intense contact with Portuguese has led to large-scale borrowing into Tetun of gendered lexical items whose behavior, including gender agreement, is in practice conditioned in the first instance by sociolect, in addition to such things as context and relative formality.

Gender in Tetun also shows more subtle effects of contact, e.g., the presence of parallel systems of interpretation and ordering of gendered elements.

Detailed analysis shows the extent of Portuguese-style gender marking in Tetun, and also allows for comparisons with developments in other Austronesian languages such as Chamorro and Tagalog, which have also had centuries of contact with Romance gender via Spanish. While there are some similarities, there are also notable differences in outcomes – a reflection of such things as very different inherent grammatical type (isolating v. complex morphological type), and the related willingness to borrow gendered affixes and attach them to native and other non-Romance lexicon. At the same time, Tetun shows much greater ability to mark Portuguese grammatical gender agreement between Portuguese nouns and adjectives. It also shows some signs of incipient gender assignment and agreement with a very small number of high frequency non-Portuguese words for reasons of semantic but also apparent phonological alignment with gender assignment in Portuguese – without additional affixation.

We are also able based on the data we present to formulate a number of typological observations – that are likely to apply to other languages, or can at least be tested against them, to see how reliable and accurate they are.

What will happen to Portuguese gender in Tetun in the future? Tetun provides us with an unusual and exciting laboratory on language contact and change in real time (see also Williams-van Klinken & Hajek [2018a, b]). It is our prediction that gender agreement, especially between Portuguese elements and within and across constituents, will become more common, as speakers of Tetun become increasingly exposed to and comfortable with Portuguese grammar, both as a result of ongoing contact with acrolectal forms in the Tetun language media and in formal contexts, as well as through increased direct contact with acrolectal Tetun and Portuguese through schooling and contact with officialdom.

Despite the current irregular nature of gender agreement in journalistic writing, we note the importance of the media, in particular the press, in modeling language patterns and mediating change in Tetun to its readership

(see Williams-van Klinken and Hajek [2018a] for detailed discussion). We also predict that the practice of gender agreement in this setting will become more common, as journalists are being taught to use it by trainers from the National Institute of Linguistics whose dictionary of Tetun (Correia et al. 2005) includes Portuguese gender marking distinctions for all nouns and adjectives from Portuguese.

One area of special interest for ongoing monitoring by Austronesianists and typologists is the extent to which Portuguese-style gender assignment and agreement involving Tetun and other non-Portuguese nouns will become evident, beyond the small number of lexical items identified at present, within the tight constraints of Tetun's largely isolating grammar.

REFERENCES

- Abduramachman, Paramita R. 1972. *Some Portuguese loanwords in the vocabulary of speakers of Ambonese Malay in the Christian villages of Central Moluccas*. Jakarta: Lembaga Research Kebudayaan Nasional LIPI.
- Aikhenvald, Alexandra Y. 2000. *Classifiers. A typology of noun categorization devices*. Oxford: Oxford University Press.
- Baklanova, Ekaterina. 2016. On marginal gender in Tagalog: A case study. In *Proceedings on the XII International Conference "Languages of the Far East, South East Asia and Western Africa", LESEWA, Moscow, Nov. 16–17, 2016*, ed. by M. V. Lomonosova, 25–33. Moscow: Yaziky Narodov Mira.
- Baxter, Alan N. 1990. Notes on the Creole Portuguese of Bidau. *Journal of Pidgin and Creole Languages* 5 (1):1–38.
- Blust, Robert. 2009. *The Austronesian Languages*. Canberra: Pacific Linguistics.
- Chamoreau, Claudine. 2012. Spanish diminutive markers *-ito/-ita* in Mesoamerican languages: A challenge for acceptance of gender distinction. In *Morphologies in contact*, ed. by Martine Vanhove, Thomas Stolz, Aina Urdze & Hitomi Otsuka, 71–91. Berlin: Akademie-Verlag.
- Choupina, Celda Morgado. 2011. Reflexões sobre o género em português europeu e em tétum [Reflections on gender in European Portuguese and Tetun]. *ELINGUP: Revista electronica de linguística dos estudantes da Universidade do Porto* 3 (1).
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- 2013a. Number of genders. In *The World Atlas of Language Structures Online*, ed. by Matthew S. Dryer & Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- 2013b. Sex-based and non-sex-based gender systems. In *The World Atlas of Language Structures Online*, ed. by Matthew S. Dryer & Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- 2013c. Systems of gender assignment. In *The World Atlas of Language Structures Online*, ed. by Matthew S. Dryer & Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- 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. *Dicionário nacional ba Tetun ofisial [National dictionary for official Tetun]*. Dili: Instituto Nacional de Linguística, Universidade Nacional Timor Lorosa'e.

- 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.
- Fischer, Steven Roger. 2007. Grammatical borrowing in Rapanui. In *Grammatical borrowing in cross-linguistics perspective*, ed. by Yaron Matras & Jeanette Sakel, 387–401. Berlin: Mouton de Gruyter.
- Greksakova, Zuzana. 2018. Tetun in Timor-Leste: The role of language contact in its development. PhD, Faculdade de Letras, Universidade de Coimbra.
- Hajek, John. 2000. Language planning and the sociolinguistic environment in East Timor: Colonial practice and changing language ecologies. *Current Issues in Language Planning* 1:400–14.
- . 2006. Language contact and convergence in East Timor: the case of Tetun Dili. In *Grammars in contact: a cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald & Robert M. W. Dixon, 163–78. 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.
- Himmelman, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In *The Austronesian languages of Asia and Madagascar*, ed. by Alexander Adelaar & Nikolaus P. Himmelman, 110–201. London: Routledge.
- Hull, Geoffrey. 1999. *Standard Tetum-English dictionary*. 1st ed. St Leonards: Allen & Unwin in association with the University of Western Sydney Macarthur.
- . 2002a. *The languages of East Timor: some basic facts*. Dili: Instituto Nacional de Linguística, Universidade Nacional Timor Lorosa'e.
- . 2002b. *Standard Tetum-English dictionary*. 3rd ed. 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.
- Hutchinson, Amelia P. 2003. *Portuguese: an essential grammar*. London: Routledge.
- Lewis, M. Paul, Gary F. Simons & Charles D. Fennig, eds. 2018. *Ethnologue: languages of the world*. 21st ed. Dallas, Texas: SIL International.
- Makihara, Miki. 2001. Modern Rapanui adaptation of Spanish elements. *Oceanic Linguistics* 40 (2):191–224.
- Marcos, Artur. 1995. *Timor Timorese: com suas línguas, literaturas, lusofonia*. Lisbon: Edições Colibri.
- Matras, Yaron. 1998. Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36 (2):281–331.
- . 2007. The borrowability of structural categories. In *Grammatical borrowing in cross-linguistics perspective*, ed. by Yaron Matras & Jeanette Sakel, 32–73. Berlin: Mouton de Gruyter.
- . 2011. Universals of structural borrowing. In *Linguistic universals and language variation*, ed. by Peter Siemund, 200–29. Berlin: Mouton de Gruyter.
- . 2015. Why is the borrowing of inflectional morphology dispreferred? In *Borrowed morphology*, ed. by Francesco Gardani, Peter Arkadiev & Nino Amiridze, 47–80. Berlin: De Gruyter Mouton.
- Matras, Yaron & Jeanette Sakel. 2007. *Grammatical borrowing in cross-linguistic perspective*. New York: Mouton de Gruyter.
- Nichols, Johanna. 2003. Diversity and stability in language. In *The handbook of historical linguistics*, ed. by Brian D. Joseph & Richard D. Janda, 283–310. Malden, MA: Blackwell Publishing.
- Pagel, Steve. 2010. *Spanisch in Asien und Ozeanien*. Frankfurt am Main: Peter Lang.

- . 2019. The opposite of an anti-creole? Why Modern Chamorro is not a new language. In *Linguistic ecology and language contact* ed. by Ralph Ludwig, Peter Mühlhäusler & Steve Pagel, 264–294. Cambridge: Cambridge University Press.
- Quilis, Antonio. 1976. *Hispanismos en cebuano: Contribucion al estudio de la lengua espanola en Filipinas*. Madrid: Ediciones Alcala, D.L.
- Quilis, Antonio & Celia Casado-Fresnillo. 2008. *La lengua Española en Filipinas: Historia, situación actual. El Chabacano. Antología de textos*. Madrid: Consejo Superior de Investigaciones Científicas.
- Reid, Lawrence & Paul Schachter. 2009. Tagalog. In *The world's major languages*, ed. by Bernard Comrie, 833–65. London: Routledge.
- Rodríguez-Ponga Salamanca, Rafael. 2006/2007. Hispanismos en la lengua Carolina de Saipán. *Revista Española del Pacífico* 19–20:13–48.
- Ross, Melody Ann. 2017. Summary: Survey and interview findings from the doctoral dissertation titled Attitudes toward Tetun Dili: A language of East Timor. https://www.academia.edu/34023892/Attitudes_Toward_Tetun_Dili_Abridged_Non-Technical_Dissertation_Report_-_A4. Downloaded 1/8/2017.
- Schachter, Paul & Fe T. Otones. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.
- Schapper, Antoinette. 2010. Neuter gender in Eastern Indonesia. *Oceanic Linguistics* 49 (2):407–35.
- SIL International. 2011. *Field Linguist's Toolbox*. Dallas: SAG.
- Stolz, Christel. 2008. Loan word gender: A case of romancisation in Standard German and related enclave varieties. In *Aspects of language contact new theoretical, methodological and empirical findings with special focus on romancisation processes*, ed. by Thomas Stolz, Dik Bakker & Rosa Salas Palomo, 399–440. Berlin: Mouton de Gruyter.
- . 2009. A different kind of gender problem: Maltese loanword gender from a typological perspective. In *Introducing Maltese linguistics: Selected papers from the 1st International Conference on Maltese Linguistics, Bremen, 18-20 October, 2007*, ed. by Bernard Comrie, 321–55. Amsterdam: John Benjamins.
- Stolz, Thomas. 2002. General linguistic aspects of Spanish-indigenous language contacts with special focus on Austronesia. *Bulletin of Hispanic Studies* 79 (2): 133–58.
- . 2003. Not quite the right mixture: Chamorro and Malti as candidates for the status of mixed language. In *The mixed language debate: Theoretical and empirical advances*, ed. by Yaron Matras & Peter Bakker, 271–315. Berlin: Mouton de Gruyter.
- . 2008. Romancisation worldwide. In *Aspects of language contact. New theoretical, methodological and empirical findings with special focus on Romancisation processes*, ed. by Thomas Stolz, Dik Bakker & Rosa Salas Palomo, 1–42. Berlin, New York: Mouton de Gruyter.
- . 2012. Survival in a niche. On gender-copy in Chamorro (and sundry languages). In *Morphologies in contact*, ed. by Martine van Hove, Thomas Stolz, Aina Urdze & Hitomi Otsuka, 93–140. Berlin: de Gruyter.
- Stolz, Thomas & Nataliya Levkovych. in press. Parallel Romancisation: Chamorro and Tetun Dili – two heavy borrowers compared. In *New perspectives on mixed languages – From core to fringe*, ed. by Eeva Sippola & Maria Mazzoli. Berlin, Boston: De Gruyter.
- Tadmor, Uri. 2007. Grammatical borrowing in Indonesian. In *Grammatical borrowing in cross-linguistics perspective*, ed. by Y. Matras & J. Sakel, 301–28. Berlin: Mouton de Gruyter.
- Thomason, Sarah G. 2001. *Language contact: an introduction*. Edinburgh: Edinburgh University Press.

- Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Thomaz, Luis Filipe F.R. 1981. The formation of Tetun-Praça, vehicular language of East Timor. In *Papers on Indonesian languages and literatures*, ed. by Nigel Phillips & Anwar Khaidir, 54–83. Paris: Cahiers d'Archipel 13.
- Thomaz, Luís Filipe F.R. 2002. *Babel Loro Sa'e: O problema linguístico de Timor-Leste*. Lisboa: Instituto Camões.
- 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. 2019. *Interactive Tetun-English dictionary*. 3rd ed. Dili: Dili Institute of Technology.
- Williams-van Klinken, Catharina & John Hajek. 2018a. Language contact and functional expansion in Tetun Dili: The evolution of a new press register. *Multilingua* 37 (6):613–48.
- Williams-van Klinken, Catharina & John Hajek. 2018b. Mixing numeral systems in Timor-Leste. *NUSA: Linguistic Studies of Languages in and Around Indonesia* 64:65–94.
- . forthcoming. Double agent, double cross? Or how a suffix changes sides in an isolating language: *dór* in Tetun Dili. In *Austronesian undressed: How and why languages become isolating*, ed. by David Gil & Antoinette Schapper. Amsterdam: John Benjamins.
- Williams-van Klinken, Catharina, John Hajek & Rachel Nordlinger. 2002. *Tetun Dili: a grammar of an East Timorese language*. Canberra: Pacific Linguistics, 528.
- Williams-van Klinken, Catharina & Robert Williams. 2015. Mapping the mother tongue in Timor-Leste: Who spoke what where in 2010? Dili Institute of Technology. <http://www.tetundit.tl/Publications/Timor-Leste%20languages%202010.pdf>.
- Wolff, John. 2001. The influence of Spanish on Tagalog. In *Lo propio y lo ajeno en las lenguas austronésicas y amerindias. Procesos interculturales en el contacto de lenguas indígenas con el español en el Pacífico e Hispanoamérica*, ed. by Klaus Zimmermann & Thomas Stolz, 233–52. Frankfurt am Main: Vervuert.