Patterns of variation in subject-indexing prefixes in Vatlongos, Southeast Ambrym

Eleanor Ridge
Massey University

This paper describes four patterns of variation in the subject-indexing paradigm of Vatlongos (Oceanic, Vanuatu). It explores their quantitative distribution in a corpus of monologic texts from speakers in three different communities: Mele Maat, a relocated peri-urban community; Endu, which has a distinct dialect; and Ase-Taveak, the other villages of Southeast Ambrym. Speakers in Mele Maat are more likely to use a zero variant of the third person singular Non-future prefix and shorter syllabic variants, and less likely to use the paucal number category. All three patterns suggest formal simplification in the Mele Maat community. The latter two patterns are also associated with higher levels of education, and consequent greater exposure to national and international languages. Overall these patterns suggest that changes which involve reduction of distinctions in morphological paradigms are more likely to be accelerated in contexts of language endangerment than phonological change or phonological reduction.

Keywords: morphological variation, overabundance, simplification, urban language use, relocation, dialectal variation, Vanuatu, Oceanic, documentation corpora

1. Introduction

Vatlongos (also known as Southeast Ambrym) is an Oceanic language of Vanuatu. Most of the ~3000 speakers live in the Southeast region of Ambrym island, and in Mele Maat, a peri-urban community on the outskirts of the capital city Port Vila on Efate island. Like closely related Paamese, it has relatively complex verbal morphology for an Oceanic language (Crowley, 1991). There are multiple paradigms of subject-indexing prefixes marking relative tense and modal categories, and in parts of these paradigms there are variants expressing the same set of features, which are not always, or not only, phonologically conditioned.

https://doi.org/10.1075/aplv.19006.rid
Asia-Pacific Language Variation 8:1 (2022), pp. 72–105. ISSN 2215-1354 | E-ISSN 2215-1362

This article uses a variationist sociolinguistic approach to explore these patterns of variation in a language documentation corpus, in the spirit of other research seeking to draw together data, methods and insights from these two subdisciplines of linguistics. Bringing data from lesser-studied languages into the broader project of variationist sociolinguistics allows its central findings to be tested and refined against a broader range of language structures and social contexts (Labov, 2015; Stanford, 2016). Paying careful attention to variation in language documentation allows a more holistic view of language in its social context, and leads to more rigorous descriptive decisions (Meyerhoff, 2017). Studying variation in contexts of language endangerment also helps us to understand how threats to linguistic vitality can ~~effect~~ patterns of language use, and address puristic language attitudes which equate all signs of variation and change with degradation (Meyerhoff, 2017; Nagy, 2009, 2017).

Variationist approaches generally try to identify areas in the linguistic system where there is more than one way of "saying the same thing", and then examine the distribution of these different variants expressing the same semantic and grammatical information, looking for correlation with factors in the social context of language use. A challenge in applying sociolinguistic techniques in a documentation context is a paucity of evidence for the trajectory of changes, especially when there is no written record. Evidence from historical reconstruction is often necessary to fill these gaps. Historical evidence is considered in relation to the variation discussed in ~~Sections 0 and 0~~. A useful sociolinguistic approach is looking at how variation is distributed across age groups or generations, which allows the identification of linguistic changes through a snapshot of synchronic language use, based on the principle that adult speakers are slow to change their pattern of language use (Sankoff, 2006). Community shift to an innovative variant tends to proceed slowly at first, accelerates as the innovation is accepted by the majority of speakers, and decelerates as the new variant approaches saturation, a distribution known as an S-curve (Labov, 2001, pp. 447–454). However, typical documentation corpora are not as extensive or balanced as corpora for major languages. Evidence for change in apparent time, as discussed in Section ~~0~~, should therefore be treated with caution.

Another frequent finding in traditional sociolinguistic research is that changes are often led by women (Labov, 1990, 2001, pp. 366–384). This tendency could be linked to the interaction between gender and class in the Western industrialised contexts where the majority of variationist sociolinguistic research has taken place, so it is useful to continue to test this observation in other contexts, while paying attention to the distinctive role of gender within different social structures (Meyerhoff, 2014). Social life in Vanuatu is often organised around gender divisions, especially in institutional contexts with a colonial legacy, such as

schools and churches (Meyerhoff, 2003). Vatlongos speaking communities are traditionally patrilocal, so women who move to another village when they marry could have an important role in disseminating innovative variants between communities.

On the other hand, social structure in Vanuatu is relatively egalitarian, without the established, highly-stratified class systems familiar in sociolinguistic work in Western contexts. However, there are direct linguistic consequences to participating in education and formal employment in Vanuatu's complex language situation, entailing removal from local-language-speaking communities and knowledge of the colonial languages English and French, as well as the English-lexifier creole Bislama which functions as a national lingua franca. There is also a distinction between rural and urban lifestyles, often designated with the Bislama terms *aelan* 'island' and *taon* 'town'. People in urban areas have easier access to education and employment opportunities, and more exposure to the national languages and other local languages, but less access to land and subsistence livelihoods. Vatlongos is spoken in an urban community who relocated from Ambrym Island in 1951, so this article aims to examine the sociolinguistic differences between rural and urban communities.

The relocation of the Mele Maat community to a peri-urban environment in the 1950s allows for a comparison of variation across rural and urban speaker communities, potentially shedding light on processes of linguistic divergence in separated communities, and potential consequences of sociolinguistic pressures associated with urban environments, especially language shift and simplification. This work therefore builds on other studies looking at variation between heritage varieties used in urban or metropolitan migrant communities and language use in homelands (Lim, 2016; Nagy, 2015, 2017; Rassool, 2013), and more generally accounts of language change in endangerment contexts (Aikhenvald, 2012; Chamoreau & Léglise, 2012). The sociolinguistic context of each speaker-community is outlined in Section 0. The corpus and the methods used to explore variation within it are described in Section 0.

The coexistence of two or more forms expressing the same set of grammatical features without phonetic, morphological or dialectal conditioning has been described as overabundance, and poses challenges for many models of morphology (Bermel et al., 2018; Bonami & Stump, 2016, p. 469; Thornton, 2012a, 2012b, 2018, 2019). The subject-indexing prefixes of Vatlongos exhibit unusually systematic overabundance, affecting all verb lexemes and multiple cells of the paradigm, rather than being restricted to certain lexemes. The subject-indexing paradigm is therefore good place to look for conditioning of variation by social factors. The subject-indexing prefixes and their role in Vatlongos verbal morphology is

described in Section 0. Section 0 reports on four patterns of variation within the subject-indexing paradigms, and their implications are discussed in Section 0.

2. Vatlongos speaker-communities

This article will distinguish three communities of Vatlongos speakers: Mele Maat, Endu and Ase-Taveak. Mele Maat is a peri-urban community whose original members relocated from Southeast Ambrym to Efate island, outside the capital city Port Vila, following a volcanic eruption in 1950 (Figure 1). The entire population of Southeast Ambrym was relocated to Epi Island following months of heavy ashfall, but when the Epi settlement was struck by a cyclone, tsunami and landfalls, most returned to Ambrym in 1951. Elder Solomon from Maat village arranged for members of his village to move to Efate island to work on a plantation. They founded the village of Mele Maat, a combination of the name of their village on Ambrym, and their closest neighbouring village on Efate (Figure 2). The circumstances of this migration and the subsequent progress of the Mele Maat community have been the subject of detailed anthropological documentation (Tonkinson, 1968, 1979, 1981, 1985). In this satellite community to the capital city, there is a greater pressure to shift towards Bislama, the national lingua franca, and clearer economic incentives to learn English and, to a lesser extent, French, which are required for education and employment opportunities in Port Vila and internationally. Several generations in this new sociolinguistic environment have allowed different patterns of language use to emerge in Mele Maat (Ridge, 2019).

The Southeast Ambrym villages are further divided to distinguish Endu, the northernmost village, from other villages in Southeast Ambrym (Figure 3). Vatlongos speakers recognise Endu-Vatlongos as a separate dialect, mainly distinguished by several high-frequency lexical variants (Ridge, 2018a, 2019). Many members of the Endu community believe Endu-Vatlongos to be a conservative dialect, framing it as the original language of Southeast Ambrym and claiming that South Vatlongos has been tainted by mixing with Paamese. Compared to the other villages, Endu is more closely integrated with Northern Ambrym communities through intermarriage and bilingualism in North Ambrym language. It also faces different sociolinguistic pressures, due to international tourists staying in the village to visit the adjacent volcano.

I treat the other villages of Southeast Ambrym as a single community, Ase Taveak, named after the first and last villages on the road travelling from North to South (Figure 3). Although this masks some inter-village dialectal variation, these villages are all closely linked by ties of intermarriage, and members of

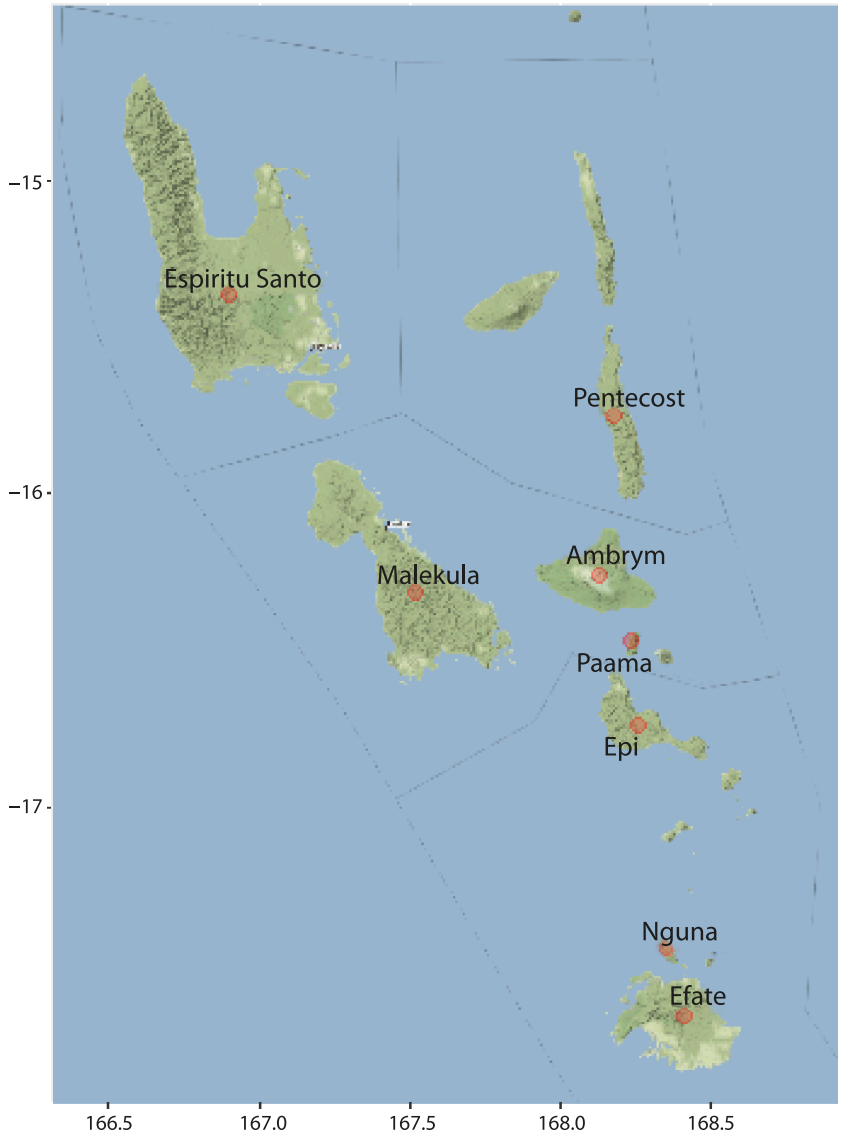

Figure 1. Map of Central Vanuatu, showing Ambrym and Efate islands[1]

different villages regularly come together for events and rituals such as weddings, funerals, circumcision ceremonies, New Year celebrations, fundraisers and

1. Maps were produced using terrain map tiles from Stamen (2018), and the following R packages: ggmap (Kahle & Wickham, 2013); ggplot2 (Wickham, 2016); ggrepel (Slowikowski, 2018).

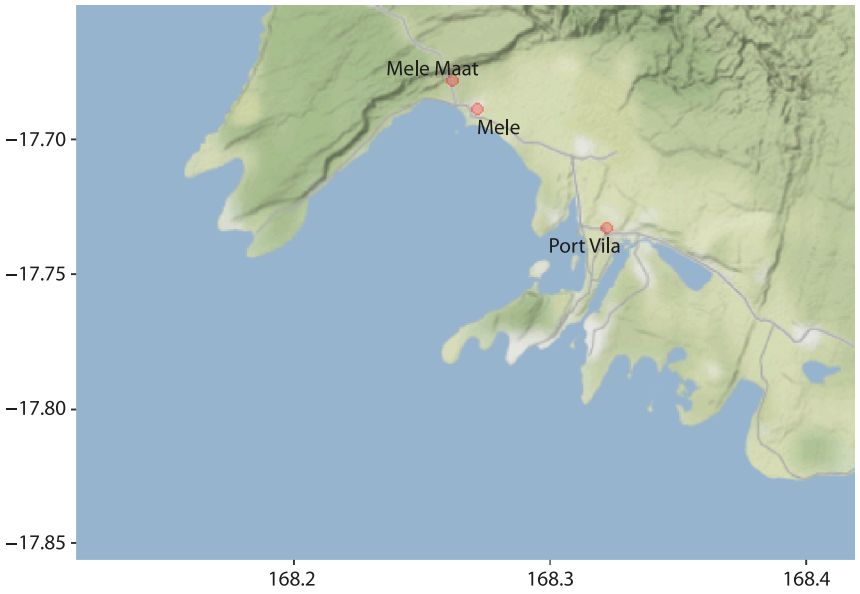

Figure 2. Map of Efate, showing Mele Maat, Mele and Port Vila

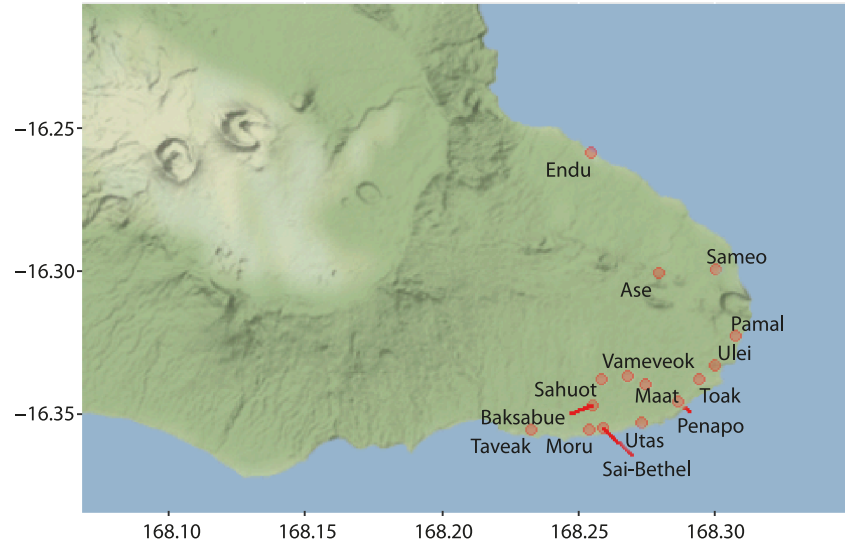

Figure 3. Map of villages in Southeast Ambrym

church activities. These villages share a similar sociolinguistic context, and the vitality of Vatlongos is more secure than in Mele Maat or Endu. Despite near total bilingualism with Bislama, there is less pressure to use Bislama across all

domains than in Mele Maat, and less competition from other local languages in key family domains than in Endu (Ridge, 2018a, 2019).

There is also greater institutional support for Vatlongos in Ase-Taveak than in Mele Maat or Endu. In religious domains the Presbyterian Church is dominant in these communities, and in recent years the Presbyterian Church has been the most influential religious group involved in the production of the Vatlongos Bible translation (Wycliffe Bible Translators, 2015), encouraging the use of Vatlongos in church services and activities. Endu has resisted this change and continues to use Bislama in church services, both in the Presbyterian church and other religious groups active in Endu, to avoid excluding speakers of Northern Ambrym languages in the congregation. Similarly, Ase-Taveak communities have increased the use of Vatlongos in pre-school education, and increasingly in the first three years of primary education, in line with the implementation of Vanuatu's language policy (Vanuatu Ministry of Education, 2012). The primary school in Endu has instead chosen to use Bislama for the first three years, again because many children speak Northern Ambrym languages rather than Vatlongos at home.

3. Corpus and methodology

Section 0 explores patterns of variation within a corpus of spontaneous and semi-spontaneous monologic texts collected as part of a wider language documentation project, recorded between 2014 and 2017. Most of these texts are available via the Pangloss collection, according to speaker preferences (Ridge, 2018b). The corpus consists of 48,509 words in 108 texts from 73 speakers. Each speaker who contributed to the corpus also took part in a survey which covered basic demographic details and language attitudes (Ridge, 2018a, 2019, Chapter 2). Each text in the corpus was coded for speaker community membership, years of education, age group, and gender, as well as text genre.[2]

The content of the corpus was determined by community goals and the priorities of individual contributors. As a result, the corpus is necessarily not balanced, especially by word count. The well-represented genres by word count are employment histories (12,685 words), custom stories (10,642), disaster narratives (7,941), children's stories with animal protagonists (5,301), and procedural texts (4,245). The only prompted texts in the corpus are two frog stories (2005 words), narra-

2. See Ridge, 2019, Chapter 9 for motivation of these categories and cut off points. All respondents categorised themselves as either male or female for gender.

tives prompted with pictures which quite closely resembled the genre children's stories with animal protagonists.[3]

Table 1 shows the word count for each subsection of the corpus by community membership, years of education, gender and age group. More than half of the corpus consists of texts by speakers in Ase-Taveak, and speakers with less than seven years of formal education, which is a reasonable reflection of the demographics of the speaker population. Although women and men contributed a similar number of texts (F: 52, M: 56), men tended to contribute much longer texts, resulting in the imbalance in word count by gender. Speakers under the age of 31 contributed fewer, shorter texts, so this is the smallest subsection of the corpus, limiting the representativeness of trends in apparent time discussed in this article. There are also fewer younger speakers from Endu and Mele Maat represented in the corpus, and the Mele Maat subcorpus has more texts from speakers in the oldest age bracket.

Table 1. Word count by community, years of education, gender and age group

Community	Word count	Years of education	Word count	Gender	Word count	Age group	Word count
Endu	10,501	<7	25,044	F	19,044	<31	5,527
Ase-Taveak	23,271	7–10	13,195	M	29,465	31–45	14,938
Mele Maat	14,737	>10	10,270			46–65	15,340
						>65	12,704

The analyses of variation here are based solely on the corpus of transcribed texts in Vatlongos orthography, and lexical annotations. This means that variation in finer phonetic details cannot be explored here. It would be useful to conduct an acoustic analysis of tokens of the some of the patterns of variation discussed here, especially the variation between /o/ and /a/ in the second person singular prefixes (Section 0), and the (C)VC- and (C)VCV- prefixes (Section 0). The patterns of variation described here should guide that future work.

Time-aligned transcriptions were produced using SayMore (Moeller, 2014) and ELAN (MPI Nijmegen, 2018). Initial transcriptions and translations into Bislama were produced by a team of three transcribers: Bell Mansen, Simeon Ben and Madleen Ben. These transcriptions were checked against the audio recordings and translated into English by the researcher, in discussion with the transcription team. Texts were interlinearised and cross-referenced with a lexical

3. See discussion in Ridge, 2019, Chapter 9. The points about the composition of the corpus largely apply here as well, although four additional texts from Endu speakers are included here.

database in Fieldworks Language Explorer (FLEx Development Team, 2018). The interlinearised corpus was exported from FLEx into R (R Core Team, 2018) using the interlineaR package (Loiseau, 2018/2018), and the quantitative analysis was conducted in RStudio (RStudio Team, 2018). Chi-squared tests were used to test the significance of relationships between choice of variants and speaker community membership, level of education, gender and age group.

The qualitative description in Section 0 is also supported by evidence from a larger corpus of texts which includes elicited and translated data, in addition to the spontaneous texts in the main corpus, totalling 69,171 words.

4. Vatlongos verbal morphology

4.1 Overview and descriptive decisions

I describe the Vatlongos verbal affix template as having two slots for prefixes and one for suffixes (Table 2). The initial prefix slot is obligatory if zero morphemes 'fill' the slot. Zero marks the third-person singular Non-future in certain phonological environments discussed below, the third-person singular before any of the prefixes in the second slot, and the second-person singular Imperative.

Table 2. Vatlongos verbal affix template

Subject indexing	**Negative polarity and apprehensive**	**Verb root**	**Object/transitivity**
Multiple TAM paradigms (Table)	**Negative:** Non-future: *taa-* Elsewhere: *naa-* **Apprehensive:** *na-*		Object pro-index (with transitive verb class) Transitiviser: *-ni* (with intransitive verb class)

I describe the subject-indexing prefixes as separate paradigms for each TAM category, as cumulative formatives expressing both subject person-number and TAM. The paradigms for different TAM contexts could be analysed as further segmentable into two or three affix slots, with portmanteau forms in many cells due to phonological processes resulting in morphological fusion (see Parker, 1968 for such an analysis; Crowley, 2002 for arguments in favour of 3 prefix slots; and Ridge, 2019, Chapter 5 for arguments for 2).

The subject-indexing prefixes can co-occur with subject noun phrases (1), including independent pronouns (2), but can also occur by themselves to index a

subject (3). They are therefore examples of 'cross-indexes' in Haspelmath's (2013) typology of argument-indexing.

(1) *horamue tei xal mama nan lu-di*[4]
boy one with mum 3SG.POSS 3DU.NFUT-NFUT.stay[5]
'a boy and his Mum stayed' [https://doi.org/10.24397/pangloss-0000145#S4][6]

(2) *xalu lu-ba*
3DU 3DU.NFUT-NFUT.go
'they went' [https://doi.org/10.24397/pangloss-0000089#S71]

(3) *lu-ba*
3DU.NFUT-NFUT.go
'they went' [https://doi.org/10.24397/pangloss-0000065#S3]

4.2 Subject-indexing prefixes

There are complete paradigms of subject-indexing prefixes for the four major TAM categories: Prior, Non-future, Immediate future and Distant future relative tenses. This paradigm most likely developed from a more typically Oceanic realis-irrealis split: the Non-future closely resembles the realis category in Paamese and is used for a similar range of contexts; Ridge (2019, Chapter 4) provides a more detailed analysis of TAM categories in context.

Vatlongos subject-indexing prefixes distinguish four persons and four numbers. Like most Oceanic languages (Lynch et al., 2002, p. 35), Vatlongos makes a distinction between inclusive (including the addressee) and exclusive (excluding the addressee) first person in non-singular numbers. It has four number categories: singular, dual, paucal and plural. Though most Oceanic languages dis-

4. Orthography aligns with IPA with the following exceptions: voiced stops are prenasalised ⟨ b ⟩ /ᵐb/, ⟨ d ⟩ /ⁿd/, ⟨ g ⟩ /ᵑg/; ⟨ v ⟩ can be realised as [v] or [β], the digraph ⟨ ng ⟩ represents /ŋ/, ⟨ j ⟩ represents the affricate /dʒ/ which only occurs in loan words from Bislama and English. Capitalisation and punctuation follow practices in English, the language of education for most Vatlongos speakers.

5. Non-Leipzig abbreviations in glossing: IFUT, immediate future relative tense; NFUT, non-future relative tense; PC, paucal; PRI, prior relative tense.

6. Example codes consist of the date of the recording event in YYYYMMDD format (20170222), an identifying letter for each session recorded that day (b), an underscore, a letter indexing broad genre (n = narrative, p = procedural, x = elicitation), a number identifying the recording within the session (01), a letter indicating the speaker community (e = Endu, m = Mele Maat, s = Ase Taveak), a speaker code (151), an underscore, and a number identifying the pause unit within the recording (04). Doi links are provided for examples that are available in Ridge, 2018b.

tinguish dual number, the paucal category is less common (Lynch et al., 2002, p. 35; Pearce, 2012). Vatlongos has a generous paucal category that can be used for fairly large groups of people, with a rhetorical effect of familiarity or minimisation, similar to that described for closely-related Paamese (Crowley, 1982, p. 81) and Daakie (South Ambrym) (Krifka, 2018, 2011, p. 6). There is therefore a choice whether to index a group of people as paucal or plural, and this is an area that shows variation between communities, discussed in Section 0.

Another place where number marking does not necessarily align with the semantic number of subject referents relates to animacy. In the third person, inanimate subjects are often coded with singular subject agreement on the verb, even if they are semantically plural, dual or paucal. Example (4) describes events during an earthquake. The trees are explicitly marked plural with quantifiers in the subject noun phrases, but are indexed with a singular subject prefix on the verb, while the animate boys are marked with a plural subject prefix.

(4) *Vatit vetei xil ak mu-lul. Mago xil ak mu-lul.*
tree breadfruit PL PROX 3SG.NFUT-shake mango PL PROX 3SG.NFUT-shake
Tutut horamue xil la-muis.
little boy PL 3PL.NFUT-cry
'These breadfruit trees were shaking; these mangos were shaking; the little boys were crying.' [https://doi.org/10.24397/pangloss-0000091#S15]

This kind of variation is widely attested cross-linguistically (Corbett, 2000, pp. 55–61), in Vanuatu (Schnell, 2019) and in other Ambrym languages (Hopperdietzel, 2018; von Prince, 2015, p. 293).

Table shows the subject-indexing prefixes in Prior, Non-future, Immediate Future and Distant Future relative tenses, including the patterns of variation that will be discussed below. Section 0 discusses the variation between /o/ and /a/ forms in the second person singular, and Section 0 the variation between *mi-* and zero in the third person singular Non-future. The zero variant can only occur before verb roots with initial /p/ or a prenasalised consonant, but *mi-* can also occur in these environments. These variants are displayed one below the other within each cell. The variation between (C)VCV and (C)VC prefixes discussed in Section 0 is displayed by splits in the cells to allow the resulting patterns of syncretism to be seen more clearly. The second /t/ in the paucal Prior forms is in brackets to indicate that this is optionally realised as a geminate or simple consonant. This is most likely a phonological process, but this pattern is not discussed in detail here as it would require more accurate transcription based on more detailed phonetic analysis than has been conducted at this stage in the research.

There are notable patterns of syncretism in this data. The Non-future and Immediate Future consistently have the same form in singular and plural num-

Table 3. Complete subject-indexing prefixes for major TAM categories

		1 Inclusive		1 Exclusive		2		3	
SG	PRI	n/a		*nate-*	*nat-*	*ote-* *ate-*	*ot-* *at-*	*te-*	
	NFUT			*na-*		*o-* *a-*		*mi-* *Ø-*	
	IFUT							*va-*	
	DFUT			*ni-*		*u-*		*i-*	
DU	PRI	*rute-*	*rut-*	*mate-*	*mat-*	*mute-*	*mut-*	*lute-*	*lut-*
	NFUT	*ru-*		*ma-*		*mu-*		*lu-*	
	IFUT	*ralo-*	*ral-*	*malo-*	*mal-*	*mulo-*	*mul-*	*lalo-*	*lal-*
	DFUT	*rali-*		*mali-*		*muli-*		*lali-*	
PC	PRI	*rat(t)e-*	*rat(t)-*	*mat(t)e-*	*mat(t)-*	*mut(t)e-*	*mut(t)-*	*lat(t)e-*	*lat(t)-*
	NFUT	*rata-*	*rat-*	*mata-*	*mat-*	*muta-*	*mut-*	*lata-*	*lat-*
	IFUT	*rato-*		*mato-*		*muto-*		*lato-*	
	DFUT	*rati-*		*mati-*		*muti-*		*lati-*	
PL	PRI	*rate-*	*rat-*	*mate-*	*mat-*	*mute-*	*mut-*	*late-*	*lat-*
	NFUT IFUT	*ra-*		*ma-*		*mu-*		*la-*	
	DFUT	*ri-*		*mu-*				*li-*	

bers, except in the third-person. However, the Non-future is always uniquely identified in the dual, where the Future tenses are optionally not distinguished from each other. In the paucal, all TAM categories can be distinguished with (C)VCV- forms, but the (C)VC- forms do not distinguish any TAM categories except the Prior, which is optionally distinguished with a geminate consonant.

Whereas the major TAM categories have full subject-indexing paradigms, Imperative (Table 4) and Prohibitive moods have partial paradigms. The Prohibitive subject-indexing prefixes co-occur with the negative clitic *ti*.

Table 4. Paradigm for subject-indexing prefixes in Imperative mood

	2	3
SG	*Ø-*	*Ø-*
DU	*lu-*	n/a
PC	*to-*	n/a
PL	*mu-*	n/a

Table 5. Paradigm for subject-indexing prefixes in Prohibitive mood

	2	3
SG	*ona-* *ana-* *on-* *an-*	*na-*
DU	*mulna-*	n/a
PC	*mutna-*	n/a
PL	*muna-*	n/a

The second-person singular Prohibitive has four variants in line with two of the patterns of variation discussed in this article: alternation between /o/ and /a/, and between (C)VCV and (C)VC syllable shapes. The second-person singular forms are the most frequent, with 29 examples in the full corpus (including elicited and written texts, 66,351 words). The VCV- forms are more frequent (18 compared to 11), and the forms based on /o/ are much more common than forms based on /a/ (24 compared to 4). The higher frequency of this cell might explain why only the second-person singular maintains this level of variation.

4.3 Interaction with other morphological processes

Some of the variation in the subject-indexing prefixes is straightforwardly conditioned by other processes in the verbal morphology of Vatlongos, especially co-occurrence with the negative prefixes and verb-initial consonant mutation.

The negative prefixes are *taa-* in the Non-future, and *naa-* for all other relative tenses (Prior, Immediate Future and Distant Future). In the third-person singular Non-future, *taa-* appears with the zero-morpheme form of the subject-indexing prefix *mi-* '3SG.NFUT'. *Naa-* occurs without the third-person singular Immediate Future prefix *va-*, while in the Distant Future *naa-* can optionally co-occur with *i-* '3SG.DFUT'. The negative prefixes cannot co-occur with the Imperative and Prohibitive subject prefixes, or with Apprehensive marking. In the corpus, the negative prefixes only occur with the (C)VC- variants of subject prefixes, although in elicitation some speakers do accept (C)VCV- variants of Future and Prior forms with a negative prefix. The Apprehensive prefix has the form *na-* and it co-occurs with the Distant Future paradigm of the subject prefixes, except in the third-person singular where it appears independently. Vatlongos also has an extensive system of verb-initial consonant mutation (Ridge, 2018c, 2019, Chapter 5). For regular verbs, the initial consonant of the verb stem remains constant in all inflectional environments, and inflectional categories are only indicated by the verbal

prefixes. However, some verbs with onsets /h/, /v/, /t/, /x/, /k/ or a vowel, undergo lexically specified verb initial consonant mutation in certain inflectional environments, most notably the Non-future in affirmative polarity, negative polarity, Immediate Future with a first person singular subject, and Immediate Future with other singular and plural subjects (but not dual and paucal numbers, contra Crowley, 1991). We have seen examples of this process in the Non-future affirmative above: *ti* 'stay' as *di* in Example (1) and *ha* 'go' as *ba* in (2) and (3). As suggested by these examples, this process applies to many very high-frequency verb lexemes.

The most important trend in these patterns for the analysis of variation in the subject-indexing prefixes is the consistent use of the prenasalised phonemes (which are relatively rare elsewhere in the language) in the Non-future affirmative environment. For all patterns that involve only two initial consonants, this is the environment that is uniquely identified. These prenasalised variants most likely arose through accretion of a realis prefix, of which the 3SG.NFUT prefix *mi-* in Vatlongos is a reflex. It is worth discussing the possible historical developments involved in some detail as it has consequences for the presumed direction of change in the pattern of variation between the *mi-* and zero forms of the 3SG.NFUT prefix discussed in Section 6.

Blevins and Lynch (2009, pp. 118–124) reconstruct **mi-* to Proto-Paamese-Vatlongos, and discuss how it has been completely replaced by the zero morpheme in Northern Paamese. Retention of *mi-* in most phonological environments is therefore a conservative feature of Vatlongos. However, they suggest that **mi-* was lost before labials in Proto-Paamese-Vatlongos, so the use of the zero-variant of the 3SG.NFUT prefix before labials in Vatlongos is inherited. When *mi-* is used before labial consonants, which is especially common in Endu-Vatlongos (see Section 6), this is therefore an innovative analogical extension.[7]

Crowley (1991, pp. 206–215) argues that the prenasalised verb-initial consonants in Vatlongos, as well as in Paamese and other language of North-Central Vanuatu, arose from accretion of a realis prefix **mV-*, of which the 3SG.NFUT form *mi-* in Vatlongos is a reflex. Lynch (2020) suggests that this would have transpired through unstressed vowel deletion, assimilation of the preceding /m/ to the place of articulation of the following consonant voiceless stop, and finally reanalysis as a phonemic prenasalised voiced stop. This implies that the use of *mi-* in these contexts in Vatlongos is again an innovative analogical extension. If *mi-* had been retained with these verb lexemes in Vatlongos, we would not expect the verb-root to also be prenasalised. This is corroborated by data from three verb classes in Northern Paamese where fossilized reflexes of Proto-Paamese-Vatlongos **mu-* are

7. I'd like to thank an anonymous reviewer for this suggestion.

retained. These forms occur in the same positions as the prenasalised initial consonants in Central and Southern Paamese, but precede the basic verb-initial consonant rather than the mutated prenasalised consonant (Crowley, 1991, p. 190).

This means that the use of the *mi-* variant of the 3SG.NFUT prefix in contexts where the zero form is possible must be innovations motivated by analogical levelling, regularising the system so that the 3SG.NFUT prefix acts like the other Non-future subject prefixes, which do occur with both mutated prenasalised onsets and labial onsets. This has important consequences for the direction of change in different communities implied by the patterns of variation discussed in Section 0.

5. Patterns of variation across communities

We have identified four patterns of variation in the Vatlongos subject-indexing prefixes. 1) There is variation between a variant including /o/ or /a/ in all second person singular cells, and 2) between *mi-* and zero in the third person singular Non-future in certain phonological contexts. 3) A pattern of variation that recurs across the subject-indexing paradigm is variation between forms with (C)VCV or (C)VC syllable structure. 4) The final pattern is not a choice between different forms used to express a grammatical category, but instead a choice to apply different grammatical categories to a referent: the differential use of the paucal versus plural number categories.

5.1 Individual styles: 2SG /o/ vs. /a/

In the second-person singular Non-future, Immediate Future, and Prior cells there is variation between forms including /o/ or /a/.

This phonological alternation is found in some lexical items, but there are also many examples of minimal pairs requiring a distinction between these phonemes (for example *ga* '3SG.NFUT.eat' and *go* '3SG.NFUT.pass'), so this is not simply underspecification at a phonological level. There is no straightforward phonological conditioning by adjacent consonants, vowels in subsequent syllables or preceding words, and the same verbal root can occur with either form.

The *a-* form seems to be an innovation, and possibly a recent one. Only the *o-* form is mentioned in Parker's (1968, p. 28, 1970, p. v) descriptions based on data collected in the late 1960s. Blevins and Lynch (2009, p. 118) reconstruct **ko* – for Proto Paamese Vatlongos. Metalinguistic awareness of this feature has come up in conversations in the field, especially with transcribers. Some speakers suggest the /a/ versions started in Mele Maat and have spread to the Ambrym communities from there.

My first attempt to explore this feature quantitively threw up a disproportionately high number of tokens in the Endu subcorpus, and a very high proportion of /o/ forms (see Ridge, 2019, p. 204). On closer inspection, this turned out to be largely due to the extreme behaviour of a single speaker, Elder Saksak Reuben, who overwhelmingly used /o/ forms rather than /a/ forms (199 tokens of /o/ compared to only 4 tokens of /a/). This high number of tokens reflects his use of the second person singular as a generic pronoun[8] in several long procedural texts on the cultivation of breadfruit and yam, and a tendency to confirm the understanding of his audience by asking *okila?* 'you know?' For this reason, his texts have been excluded from the data in this section, leaving a smaller but less skewed Endu sub-corpus. However, that a speaker can have such an extreme personal style in relation to this variant is interesting in itself. Elder Saksak is one of the Endu speakers who professes puristic language attitudes, condemning perceived innovations by Ase-Taveak and Mele Maat speakers, so this might be a stylistic feature that indexes his commitment to Endu-Vatlongos.

As shown in Table 6, the remaining speakers in the Endu subcorpus use the /o/ form for 64% of the 56 tokens of these prefixes, a very similar rate to Mele Maat speakers at 65% of 71 tokens. Ase-Taveak speakers on the other hand use the /o/ form for 76% of 133 tokens. These patterns could support speakers' suggestions that the /a/ forms are an innovation that originated in Mele Maat. On the other hand, it is interesting that Endu speakers pattern closely with Mele Maat speakers despite the perception that Endu-Vatlongos is a conservative dialect. However, the effect of community on this variation is not significant (Pearson's chi-squared test $p = 0.133$, $N = 260$).

Table 6. Distribution of *o-* and *a-* tokens of 2SG forms by community

Community	*o-* tokens	*o-* %	*a-* tokens	*a-* %	Word count
Endu	36	64%	20	36%	7,458
Ase-Taveak	101	76%	32	24%	23,271
Mele Maat	46	65%	25	35%	14,737
All	183	70%	77	30	45,466

Instead, both gender and age group had a significant relationship with the proportion of these variants in the corpus, suggesting a change in apparent time being led by women. In the corpus, women used /a/ tokens almost exactly half the time, while men were more likely to use /o/ tokens, for 85% of second person sin-

8. Other texts in the corpus use the third-person singular or plural for this purpose.

gular prefixes (Table 7). This was a highly significant effect (Pearson's chi-squared test with Yates' continuity correction $p = <.005$, $N = 260$).

Table 7. Distribution of *o* and *a* tokens of 2SG forms by gender

Gender	*o*- tokens	*o*- %	*a*- tokens	*a*- %	Word count
Female	56	50%	55	50%	19,044
Male	127	85%	22	15%	26,422
All	183	70%	77	30	45,466

Grouping the corpus by age group (Table 8) results in less reliable trends as there are very few tokens in some cells, and the youngest age group is only represented by a very small corpus.[9] However, there is a clear pattern whereby older speakers in the 46–65 and older than 66 age groups use the /o/ variants nearly 90% of the time, while younger speakers use them much less often. In combination with the evidence from documentation in the late 1960s and reconstruction of Proto-Paamese-Vatlongos mentioned above, this supports the hypothesis that the /a/ forms are innovative. This effect was also highly significant (Pearson's chi-squared test $p = <.005$, $N = 260$).

Table 8. Distribution of *o*- and *a*- tokens of 2SG forms by age group

Age	*o*- tokens	*o*- %	*a*- tokens	*a*- %	Word count
<31	12	63%	7	37%	5,527
31–45	49	48%	54	52%	14,938
46–65	74	88%	10	12%	12,297
66+	48	89%	6	11%	12,704
All	183	70%	77	30%	45,466

A serious limitation to this finding is that the coding of these tokens is based on the perceptions of three transcribers, checked by the researcher during data-entry but not analysed acoustically. To strengthen the observation, it would be useful to conduct acoustic analysis of the examples, which may reveal a more complex picture or phonetically intermediate tokens. Given the relatively small number of tokens and the evidence that an individual speakers' style can skew the

9. The expected frequency for the smallest cell (/a/ tokens in the <31 age group) is just over the recommended minimum of 5 at 5.63 (to 2 decimal places), so it is possible to use Pearson's chi-squared test, but the results should be treated with caution especially in relation to the youngest age group.

results, it would also be useful to revisit this pattern of variation with a larger corpus using regression modelling.

In summary, the variation in the second person singular forms suggests a pattern of change in apparent time where an innovative variant is diffusing through the community, a change which seems to be led by women. Despite speakers' impressions that the innovative variant originates in Mele Maat, there is no significant relationship with community membership of the speaker in the corpus.

5.2 Variation in phonological conditions: 3SG.NFUT *mi-* vs. zero

In the third-person singular Non-future there is variation between a zero morpheme and *mi-*. There are identifiable conditions on these allomorphs, but they do not account for all the variation. As discussed above, a morphological condition is that the zero morpheme always occurs before the negative Non-future prefix *taa-*. This is related to a lexical condition: the *mi-* prefix never occurs before the verb *tavuol, tavorel* 'be absent', which is probably a lexicalised negated verb form.

The phonological conditions on the third-person singular zero morpheme are unidirectional. The zero morpheme only occurs before prenasalised consonants (/^{m}b/ <b>, /^{n}d/ <d>, /^{ŋ}g/ <g>), bilabial consonants (/p/, /m/ and again /^{m}b/), and, for some speakers, especially in Endu and Sameo villages, /k/. However, the *mi-* variant also occurs in all these environments (compare (5), (6), (7) with (8), and (9) with (10)), although ~~mi-~~ occurs less often than the zero morpheme.

(5) *mai* *Ø-ba*
pigeon 3SG-NFUT.go
'the pigeon went' [https://doi.org/10.24397/pangloss-0000079#S19]

(6) *xi* *su* *mi-ba*
3SG first 3SG.NFUT-NFUT.go
'he went first' [https://doi.org/10.24397/pangloss-0000057#S21]

(7) *xi* *Ø-gusil* *xalu Ø-ba*
3SG 3SG-NFUT.follow 3DU 3SG-NFUT.go
'it followed them along' [https://doi.org/10.24397/pangloss-0000081#S35]

(8) *mi-gusil-i*
3SG.NFUT-NFUT.follow-3OBJ
'it followed her' [https://doi.org/10.24397/pangloss-0000037#S18]

(9) *tati nalu Ø-pus-i*
dad 3DU.POSS 3SG-see-3OBJ
'their dad saw' [https://doi.org/10.24397/pangloss-0000081#S16]

(10) *mi-pus-i*
3SG.NFUT-see-3OBJ
'I saw it' [https://doi.org/10.24397/pangloss-0000051#S101]

As discussed in Section 6, the use of the *mi-* variant before labials and prenasalised consonants is most likely an innovative analogical extension to environments where it has historically been lost through accretion to the verb root, which resulted in the mutated verb-initial prenasalised consonants in the first place. Regularisation of the paradigm so that all the Non-future subject prefixes occur before the mutated verb stems could motivate this extension. In Parker's (1968, p. 40a) description, the *mi-* variant is observed to never occur before labials (including prenasalised / ᵐb/), but to be optional before /ⁿd/ and /ᵑg/. In the corpus used here, this pattern is mostly confirmed for speakers without connections to Endu. It could be that the relationship between *mi-* and the mutated verb-initial consonants is less transparent in the non-labial prenasalised consonants, making the analogical extension of *mi-* to these contexts more likely than extension to the labials.

But there are factors that could inhibit the extension of the *mi-* variant into these environments. Firstly, for verbs which undergo verb-initial consonant mutation, the prenasalised initial consonant already disambiguates the un-prefixed verb from the other use of the bare stem in Vatlongos, for second-person singular Imperative (which takes the basic initial consonant). Secondly, economy might favour the use of the zero-morpheme in the positions where it is phonologically licensed, especially given that for many verb lexemes there is no cost of ambiguity. Economy is a factor discussed by Koch (1995) in relation to the creation of zero morphemes.

Although the *mi-* form is possible in more phonological environments, many very frequent verbs undergo verb-initial consonant mutation, and so usually occur with the zero morpheme. The zero morpheme is therefore more frequent than the *mi-* form, accounting for 79% of 7270 tokens in the corpus.

The phonological conditions on these variants are a salient marker of Endu-Vatlongos. Speakers from Endu use the *mi-* variant before the prenasalised and, especially, the bilabial consonants, more frequently than other speakers (as in Examples (6), (8), and (10)). All the tokens in the corpus where *mi-* is used before /m/ are from Endu speakers. In this respect, Endu therefore seems to be a relatively innovative dialect, despite the community's perception that their dialect is "older" than South-Vatlongos.

Apart from the trends in the corpus, there is also observational evidence that speakers in Endu and elsewhere are aware of this pattern in Endu-Vatlongos. In preparing texts for distribution, Endu speakers tried to make this pattern more consistent in Endu-Vatlongos stories by adding the prefix in these environments, and transcribers from other communities often commented on the use of *mi-* in these environments as an amusing feature of Endu-Vatlongos.

On the other hand, speakers from Endu and Sameo (one of the villages closest to Endu) can use the zero morpheme before /k/. There are only 16 tokens of this in the corpus, but they are all in texts by speakers from Endu or Sameo, or speakers with connections to these communities. Many of the forms involved seem to be variants of verbs which elsewhere in the corpus show verb-initial consonant mutation between basic /h/ and Non-future affirmative /ŋg/.

The observed tendency to use the *mi-* form in more phonological environments is surprisingly not reflected in the overall frequency of these variants in the community-level subcorpora (Table 9). Both Ase-Taveak and Endu speakers use the zero morpheme at a very similar rate in the corpus: 76% of 3884 tokens for Ase-Taveak, and 77% of 1669 tokens in Endu. Although speakers in Mele Maat use the zero form under the same phonological conditions as speakers in Ase-Taveak, they are using it more often: the zero morpheme accounts for 84% of 1859 tokens. Community has a highly significant effect on the proportion of these two variants in the corpus (Pearson's chi-squared test $p = <.005$, $N = 7412$).

Table 9. Distribution of Ø- and *mi-* tokens of 3SG.NFUT forms by community

Community	Ø- tokens	Ø- %	*mi-* tokens	*mi-* %	Word count
Endu	1,293	77%	376	23%	10,501
Ase-Taveak	2,967	76%	917	24%	23,271
Mele Maat	1,569	84%	290	16%	14,737
All	5,829	79%	1,583	21%	48,509

Speakers in urban Mele Maat are under more pressure to shift to Bislama and English than speakers on Ambrym, and this could lead to language attrition in Vatlongos. There is a perception among Vatlongos speakers that speakers from Mele Maat speak more limited Vatlongos than speakers on Ambrym, and in a language attitudes survey respondents in Mele Maat were the only group to self-assess their competence in Bislama more highly than Vatlongos (Ridge, 2018a, 2019, Chapter 2). It may be that the higher rate of the zero-morpheme in the Mele Maat subcorpus could be because speakers are using a more restricted range of more frequent verb lexemes, which are more likely to undergo verb-initial conso-

nant mutation. The higher rate of use for the zero variant in Mele Maat is probably also partly because of the higher frequency of loan verbs in this subcorpus. Loan verbs in Vatlongos are often introduced by the copular verb (Ridge, 2017, 2019, Chapter 3). The copular verb *he* undergoes verb-initial consonant mutation to give the Non-future form *be,* which only ever occurs with the zero form of the third person singular Non-future in the corpus.

A good proxy for exposure to English and Bislama speaking environments, and removal from Vatlongos-speaking environments, is level of education. Education beyond seven years requires interaction with students who speak other languages (at a minimum, with students at secondary schools in Northern Ambrym), and study beyond Year 10 usually involves studying in an urban centre or abroad. Mele Maat speakers who contributed to the corpus are on average more educated than speakers in the island communities, with easier access to schools and universities in Port Vila (Ridge, 2018a, 2019, Chapter 2). The relationship between this pattern of variation and level of education was therefore also examined, to try to tease apart whether individual language attrition or dialectal conventions at the community level might explain the higher use of the zero variant in Mele Maat.

The results showed very little difference in the use of the variants by speakers at different levels of education (Table 10). Speakers with less than seven years of education use the zero variant for 78% of 4,049 tokens; speakers with seven to ten years of education use zero for 79% of 1971 tokens; and speakers with more than ten years education use zero for 80% of 1392 tokens. The relationship is not significant (Pearson's chi-squared test p =.1385, $N = 7412$).

Table 10. Distribution of Ø- and *mi-* tokens of 3SG.NFUT forms by years of education

Years of education	Ø- tokens	Ø- %	*mi-* tokens	*mi-* %	Word count
<7	3,152	78%	897	22%	25,044
7–10	1,560	79%	411	21%	13,195
11+	1,117	80%	275	20%	10,270
All	5,829	79%	1,583	21%	48,509

Gender[10] and age[11] had significant effects in the corpus, but with very small effect sizes suggesting that these effects were more likely the result of imbalances

10. Female speakers used zero for 77% of tokens, male speakers 80%, $p = 0.008588$

11. Speakers aged <31 used zero for 72% of tokens, 31–45 78%, 46–65 79%, over 66 82%, $p < .005$. It is interesting that the overall trend here does match the suggested direction of change in apparent time, as zero might be being extended to more phonological contexts. However, the

in the corpus, such as the lower representation of younger and female speakers in the Mele Maat corpus.

To summarise, in the third person singular Non-future, the innovative zero form is more established, accounting for 79% of all tokens in the corpus. In terms of frequency of use, there is a clear divide between the two island communities, Ase-Taveak and Endu, who use the variants at very similar rates; and Mele Maat where zero is used more frequently at 84%. This could therefore be evidence of an emerging dialectal difference in Mele Maat, as it cannot be accounted for by higher levels of education in Mele Maat. However, this obscures the variation in the phonological conditioning of these variants, where Mele-Maat and Ase-Taveak pattern together, with Endu speakers more likely to use *mi-* before pre-nasalised consonants, and speakers associated with Endu and Sameo (a village in Ase-Taveak) using zero before /k/.

5.3 Phonological erosion resulting in new patterns of syncretism: (C)VCV- versus (C)VC- forms

In the dual Future, and paucal Non-future and Future, as well as all Prior cells except the third-person singular, there is variation between a form of the prefix with a syllable shape of (C)VCV- and a form with (C)VC- shape (Table). The (C)VC- shape results from phonological reduction, dropping the final vowel from the (C)VCV- variant.

While this pattern is consistent across different parts of the subject-indexing paradigm, it is potentially difficult to formulate the exact unit that undergoes reduction. Although each form is a single prefix in the descriptive approach I have taken here, it could consist of prefixes in two or three different slots in the systems proposed by Parker (1968, 1970) or Crowley (1991, 2002), as mentioned in Section 0. It is also difficult to delineate in purely phonological terms, as the process does not affect vowels in the verb root. In prosodic terms it would be difficult to formulate the correct domain as word stress is assigned from the end of the word, so in frameworks like Hayes's (1995) parametric approach, the initial syllable or two syllables may or may not form a foot, depending on the length of the verb root.

This is an interesting pattern of variation because the (C)VC- variants have different patterns of syncretism than their (C)VCV- equivalents, obscuring the distinction between the two Futures in dual numbers, and the distinction between all TAM categories in paucal numbers. In some subject number envi-

rate for the youngest age group especially is the least trustworthy as this is a much smaller sub-corpus, and the differences between the other age groups are very small.

ronments, speakers therefore have a choice whether to identify a specific TAM category, or not.

Across all communities, (C)VC- forms account for more than half of tokens of the prefixes that show this pattern of variation (Table 11). In Ase-Taveak the (C)VC- forms represent 57% of 633 tokens, and the rate is similar in Endu at 54% of 315 tokens. Mele Maat uses (C)VC- forms for 66% of 201 tokens. This could show that Mele Maat speakers are leading an innovation. The effect of community on the variation between (C)VC- and (C)VCV- forms is significant, but less significant than on the use of the different forms of the third-person singular Non-future prefixes (Pearson's chi-squared test $p = .02469$, $N = 1149$).

Table 11. Distribution of (C)VC- and (C)VCV- tokens by community

Community	(C)VC-tokens	(C)VC- %	(C)VCV-tokens	(C)VCV- %	Word count
Endu	171	54%	144	46%	10,501
Ase-Taveak	363	57%	270	43%	23,271
Mele Maat	133	66%	68	34%	14,737
All	667	58%	482	42%	48,509

The different rate of use in Mele Maat is reflected in the rate of use by the most highly-educated speakers (Table 12). Speakers who have received less than seven, or seven-to-ten years of education used the (C)VC- forms at similar rates (56% of 624 tokens and 57% of 375 tokens respectively), while speakers who have received education beyond Year 10 used the (C)VC- variants more often (69% of 150 tokens). The effect of education level on this variation was also significant (Pearson's chi-squared test $p = .01819$, $N = 1149$).

Table 12. Distribution of (C)CV- and (C)VCV- tokens by years of education

Years of education	(C)VC-tokens	(C)VC-%	(C)VCV-tokens	(C)VCV-%	Word count
<7	351	56%	273	44%	25,044
7–10	213	57%	162	43%	13,195
11+	103	69%	47	31%	10,270
All	667	58%	482	42%	48,509

This pattern could be an interim stage of language change moving towards a system with only the phonologically reduced forms. The negative prefixes only occur with the (C)VC- variants of the subject prefixes in the corpus, although

speakers usually do accept the (C)VCV- variants as grammatical in elicitation. The loss of the longer forms in negative polarity could be a first stage towards their loss in other environments. Other evidence from elicitation is that a younger speaker in Ase-Taveak (aged 12) always offered the (C)VC- variant first, before being prompted to offer the (C)VCV- version, unlike their parents who usually did the opposite. However, these observations were not tested with other younger speakers and this could simply be an individual difference; this is an important area for further research. There was no consistent chronological trend or significant effect of age group on this pattern of variation, unlike the 2SG variants.

These differences could be a stable pattern of variation, at least in the island communities. The same variation existed in the 1960s according to Parker (1968). Based on data from Italian, Thornton (2012b) argues that overabundance can be stable diachronically. Perhaps the Vatlongos system is sustainable, especially as the different patterns of syncretism mean that sometimes there is a functional difference between two variants. The phonologically heavier variant is also more semantically specific, and might be maintained by the functional need to point to a specific TAM category when it is not specified by other parts of the verbal morphology. It would be interesting to break down the variation between (C)VC- and (C)VCV- forms in different cells, to investigate whether the longer forms are more likely when they have a higher functional load. However, this avenue for future research would require a larger corpus, and perhaps targeted story prompts to encourage the use of some of the rarer cells in the paradigms.

5.4 Choice of number category: Paucal versus plural

Looking at the distribution of (C)VC- and (C)VCV- prefixes in Table 11 and Table 12, the overall number of these person-number combinations seems to be lower than expected in texts by speakers from Mele Maat and speakers educated more than ten years, compared to the overall word count for these sections of the corpus. Impressionistically during fieldwork, speakers in Mele Maat seemed to use the paucal number category less often than speakers on the island, so this was investigated as a possible explanation for the pattern observed in relation to the (C)VC- and (C)VCV- tokens.

As discussed in Section 0, the paucal can be used for quite large groups of people, especially to refer to family or village groups, so there are situations where there is a choice to use a paucal or plural form to refer to the same referent. Bislama makes a distinction between dual and plural number categories, and some speakers also use trial forms, but there is no equivalent of the Vatlongos paucal (Crowley, 2004, pp. 46–47). Speakers in Mele Maat could therefore be moving towards a system more similar to Bislama.

Within the corpus, speakers in Endu and Ase-Taveak use paucal rather than plural forms of the subject-indexing prefixes at very similar rates, for 84% of 141 tokens, and 85% of 252 tokens respectively (Table 13). Speakers from Mele Maat instead only use the paucal for 40% of 87 tokens. This difference is highly significant (Pearson's chi-squared test $p = <.005$, $N = 480$).

Table 13. Distribution of paucal and plural subject prefix tokens by community

Community	Paucal tokens	Paucal %	Plural tokens	Plural %	Word count
Endu	119	84%	22	16%	10,501
Ase-Taveak	215	85%	37	15%	23,271
Mele Maat	35	40%	52	60%	14,737
All	369	77%	111	23%	48,509

A similar pattern emerges across education levels (Table 14). Speakers with less than seven years or seven-to-ten years of education use the paucal at similar rates, for 84% of 249 tokens and 82% of 164 tokens respectively. Speakers with more than ten years of education only use the paucal for 37% of 67 tokens. Again, this difference is highly significant (Pearson's chi-squared test $p = <.005$, $N = 480$).

Table 14. Distribution of paucal and plural subject prefix tokens by years of education

Years of education	Paucal tokens	Paucal %	Plural tokens	Plural %	Word count
<7	210	84%	39	16%	25,044
7–10	134	82%	30	18%	13,195
11+	25	37%	42	63%	10,270
All	369	77%	111	42%	48,509

The differences in the use of these number categories within the corpus are very striking, but we need to be cautious in how we interpret them. The content of texts in the corpus is largely determined by individual and community preferences, so the semantic number of major protagonists in narratives is not controlled across subsections of the corpus. Similarly, we already saw in relation to the second person singular that some speakers chose to use the third person plural as a generic in procedural texts, while others used the third or second person singular. However, the very similar rates between the Endu and Ase-Taveak sections of the corpus, and the <7 and 7–10 years of education sections, suggests that the very different ratio in the Mele Maat and 11+ texts does reflect a real difference at least within the corpus.

One possibility is that highly educated and Mele Maat speakers are using the plural to refer to referents that would be indexed as paucal by other speakers in these texts. The other possibility is that the texts told by Mele Maat speakers are more likely to refer to larger groups of people, and groups that are not determined by familial relationships. For example, texts recounting employment histories were a popular choice in Mele Maat, and are likely to refer to groups of colleagues. In either case, if the corpus contents to some extent reflect input for speakers in Mele Maat, the lower incidence of the paucal could explain why the free variation between (C)VCV- and (C)VC- variants is not maintained to the same extent in Mele Maat, where there is a stronger preference for the (C)VC- variants.

6. Discussion

The patterns of variation found in the Vatlongos subject-indexing prefixes could shed light on wider questions that arise in the study of variation in under-described languages, especially in contexts of migration, urbanisation and language shift. Table 15 summarises the patterns of variation described in Section 0, along with the significant social factors.

The most straightforward pattern of variation was that in the second person singular: the distribution of two phonological variants is changing in apparent time so that younger generations use an innovative variant more often than older generations. As is often the case cross-linguistically, this change appears to be led by women. One individual speaker who was overrepresented in relation to this pattern had a consistent, conservative individual style. Once his tokens were removed, this variation did not show a significant relationship with education level or community membership, so did not seem to be affected by exposure to national languages or an urbanised environment. As the overall Vatlongos speaking community tends to be patrilocal, it is possible that changes led by women are more likely to be consistent across all regions, as women marrying into communities may propagate the changes.

Variation in the third person singular Non-future prefixes is complicated by differential phonological conditioning by speakers of Endu-Vatlongos, which, based on the historical evidence, is probably an innovative analogical extension into contexts where the *mi-* variant had previously been lost through accretion. But perhaps surprisingly, speakers in Endu and Ase-Taveak, the two rural communities, still showed very similar overall rates of use of the two variants, while speakers in Mele Maat used the zero variant more often. This difference may be indicative of community-level language attrition, as two possible explanations

Table 15. Summary of patterns of variation described in Section 0

Locus of variation	Variants	Description of patterns	Significant social factors
2SG prefixes	forms based on /o/ and /a/	The innovative /a/ variant is being used more over apparent time, especially by speakers in the 31–45 age group, and is used more often by women. Individual speakers can exhibit distinctive behaviour.	Gender Age group *Individual speaker*
3SG.NFUT prefix	*mi-* vs Ø-	Speakers in Endu use the *mi* – form more often in phonological contexts where Ø- is possible. However, speakers in Endu and Ase-Taveak use these variants at similar rates, while Mele Maat speakers use Ø more often.	Community *Gender and age group but with very small effect sizes*
All paucal, Prior (except 3SG), and dual Future prefixes.	forms with the syllable structure (C)VCV- and (C)VC-	Speakers educated for more than ten years, and speakers in Mele Maat, use the (C)VCV- forms less often.	Years of education Community
More than two subject referents	paucal or plural number marking	Speakers educated for more than ten years, and speakers in Mele Maat, use the paucal less often.	Years of education Community

of this difference could be that 1) speakers in Mele Maat are using a more limited range of highly frequent verbs which tend to undergo verb-initial consonant mutation and occur with the zero variant, and 2) speakers in Mele Maat are more likely to use loan verbs from Bislama which are introduced by the copular verb which always occurs with the zero variant in the corpus. However, this pattern does not show a significant relationship with years of education, so it does not seem to relate to individual differences in exposure to national languages. If the higher use of the zero morpheme is related to language attrition, this is at a community rather than individual level. The difference could be maintained by the relative separation of the Mele Maat community from the communities on the island.

The last two patterns of variation both, in different ways, relate to the expression of semantic categories. The shorter syllabic variants collapse semantic distinctions between different TAM categories in different parts of the paradigm. The choice between paucal and plural marking is a decision about how, and whether, to apply this distinction between number categories. In both cases, highly edu-

cated speakers and speakers in urban Mele Maat favour the variant that involves less frequent marking of semantic categories that are not present in the national language Bislama (e.g., Immediate vs Distant Future; paucal number). If these variants prevailed, it could lead to a smaller set of semantic distinctions and a simpler verbal paradigm. Both of these patterns of variation therefore show highly-educated speakers in urban environments favouring variants that could lead to a system with fewer semantic distinctions, or with semantic categories that are more similar to contact languages. Comparing the social conditioning of these two patterns of variation with the first two could suggest that innovations with these semantic consequences are more likely to be accelerated in contexts of language endangerment than changes which involve only phonological innovations.

An alternative explanation for the role of education in conditioning each of these patterns of variation could lie in the frequency of the different cells in the paradigm involved. The cells involved in the variation between the syllabic variants are much less frequent in discourse than the third person singular Non-future, as shown by the overall token counts (compare Table 9 and Table 11). It could be that variation in these less frequent cells is less easy to maintain for speakers who are removed from Vatlongos-speaking communities. Lower frequency items could be expected to be more prone to regularisation in contexts of language shift or individual attrition than high-frequency cells. This would parallel the way that low-frequency irregular lexical items are more likely to be regularised than the high-frequency ones, a phenomenon that Bybee describes as the Conserving Effect of frequency (Bybee, 2007). In this case the lower frequency of the dual and paucal cells across all communities is further exaggerated in the paucal cells for highly educated speakers who already mark paucal less often than other speakers, perhaps making it less likely for multiple variants to be maintained.

7. Conclusion

This article has examined four patterns of variation in the subject-indexing paradigm of Vatlongos, a language spoken in a relocated peri-urban community, Mele Maat, as well as rural communities in Southeast Ambrym. The grammatical system of Vatlongos could be undergoing simplification in Mele Maat, a possible sign of language attrition (Schmidt, 1985, pp. 212–213, 1991; Simpson, 2015). The three distinctive patterns of use in Mele Maat could all be described as trends of formal simplification: replacing a prefix with zero in the third person singular Non-future; using phonologically shorter and semantically less precise forms in the syllabic patterns; and moving from four to three number categories.

However, the very complex synchronic patterns of variation described here do not suggest a simple system, and instead demonstrate a paradox in processes described as simplification. Simplification has been characterised as both the loss of allomorphs or forms, and the extension of a form into more contexts, but in the short term this extension can lead to more variants being available to speakers (Maher, 1991, p. 68; Silva-Corvalán, 1991, p. 152). If the more specific forms (like the (C)VCV- subject-indexing prefixes) do indeed go on to be lost altogether, these developments would lead to a simpler verbal morphology in Vatlongos, and perhaps even the loss of some TAM categories. However, the synchronic situation described here is highly complex and involves a proliferation of variants within paradigmatic environments, each showing complex relationships with different combinations of social factors.

Funding

This research is based on data gathered as part of a UK Arts and Humanities Research Council Doctoral Studentship (AH/K5032771).

Acknowledgements

I would like to thank all members of the Vatlongos communities who have contributed their stories, and especially Bell Mansen, Simeon Ben, Madleen Ben for their expertise, and their work transcribing and translating recordings. I would also like to thank Miriam Meyerhoff and two anonymous reviewers for their feedback on an earlier version of this article. All remaining errors are my own.

References

Aikhenvald, Alexandra Y. (2012). Language contact in language obsolescence. In Claudine Chamoreau & Isabelle Léglise (Eds.), *Dynamics of contact-induced language change* (pp. 77–110). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110271430.77

Bermel, Neil, Knittl, Luděk, & Russell, Jean (2018). Frequency data from corpora partially explain native-speaker ratings and choices in overabundant paradigm cells. *Corpus Linguistics and Linguistic Theory*, *14*(2), 197–231. https://doi.org/10.1515/cllt-2016-0032

Blevins, Juliette, & Lynch, John (2009). Morphological conditions on regular sound change?: A reanalysis of * l-loss in Paamese and Southeast Ambrym. *Oceanic Linguistics*, *48*(1), 111–129. https://doi.org/10.1353/ol.0.0030

Bonami, Olivier, & Stump, Gregory T. (2016). Paradigm function morphology. In Andrew Hippisley & Gregory T. Stump (Eds.), *The Cambridge handbook of morphology* (pp. 449–481). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139814720.017

Bybee, Joan (2007). *Frequency of use and the organization of language*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195301571.001.0001

Chamoreau, Claudine, & Léglise, Isabelle (2012). A multi-model approach to contact-induced language change. In *Dynamics of contact-induced language change* (pp. 1–15). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110271430.1

Corbett, Greville G. (2000). *Number*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139164344

Crowley, Terry (1982). *The Paamese language of Vanuatu*. Australian National University.

Crowley, Terry (1991). Parallel development and shared innovation: Some developments in Central Vanuatu inflectional morphology. *Oceanic Linguistics*, *30*(2), 179–222. https://doi.org/10.2307/3623087

Crowley, Terry (2002). Southeast Ambrym. In John Lynch, Malcolm D. Ross, & Terry Crowley (Eds.), *The Oceanic languages* (pp. 660–670). Richmond, Surrey: Curzon Press.

Crowley, Terry (2004). *Bislama reference grammar*. Honolulu: University of Hawai'i Press.

FLEx Development Team (2018). FieldWorks Language Explorer (FLEx) (8.3.11) [Computer software]. SIL International. https://software.sil.org/fieldworks/

Haspelmath, Martin (2013). Argument indexing: A conceptual framework for the syntactic status of bound person forms. In Dik Bakker & Martin Haspelmath (Eds.), *Languages across boundaries: Studies in memory of Anna Siewierska* (pp. 197–226). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110331127.197

Hayes, Bruce (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hopperdietzel, Jens (2018, May 5). *The morphosyntactic status of the subject (marker) in Daakaka (West Ambrym)* [Oral Presentation]. Austronesian and Papuan Languages and Linguistics 10, University of Surrey.

Kahle, David, & Wickham, Hadley (2013). ggmap: Spatial visualizaiton with ggplot2. *The R Journal*, *5*(1), 144–161. https://doi.org/10.32614/RJ-2013-014

Koch, Harold (1995). The creation of morphological zeroes. In *Yearbook of morphology 1994* (pp. 31–71). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-017-3714-2_2

Krifka, Manfred (2018, July 26). *Social meaning of number: Honorific duals and affiliative paucals* [Oral Presentation]. Vanuatu Languages Conference, USP Emalus Campus, Port Vila; Vanuatu. https://sites.google.com/view/vanlangconf2018

Krifka, Manfred (2011). Notes on Daakie (Port Vato): Sounds and modality. *Proceedings of the Eighteenth Meeting of the Austronesian Formal Linguistics Association (AFLA)*, 46–65.

Labov, William (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, *2*(2), 205–254. https://doi.org/10.1017/S0954394500000338

Labov, William (2001). *Principles of linguistic change, volume 2: Social factors*. Malden: Blackwell.

Labov, William (2015). The discovery of the unexpected. *Asia-Pacific Language Variation*, *1*(1), 7–22. https://doi.org/10.1075/aplv.1.1.01lab

Lim, Lisa (2016). The art of losing: Beyond java, patois and postvernacular vitality – Repositioning the periphery in global Asian ecologies. In Luna Filipović & Martin Pütz (Eds.), *Endangered languages and languages in danger: Issues of documentation, policy, and language rights* (pp. 283–312). Amsterdam: John Benjamins. https://doi.org/10.1075/impact.42.12lim

Loiseau, Sylvain (2018). *interlineaR: Importing interlinearized corpora and dictionaries as produced by descriptive linguistics software* (R package version 1.0) [R]. Paris 13 University / UMR 7107 Lacito. https://CRAN.R-project.org/package=interlineaR (Original work published 2018)

Lynch, John (2020). Consonant mutation in Southern Oceanic. *Oceanic Linguistics*, *59* (1/2), 232–268. https://doi.org/10.1353/ol.2020.0012

Lynch, John, Ross, Malcolm D., & Crowley, Terry (Eds.). (2002). *The Oceanic languages, volume 1*. Richmond, Surrey: Curzon Press.

Maher, Julianne (1991). A crosslinguistic study of language contact and language attrition. In Herbert W. Seliger & Robert M. Vago (Eds.), *First language attrition* (pp. 67–84). https://doi.org/10.1017/CBO9780511620720.005

Meyerhoff, Miriam (2003). Claiming a Place: Gender, knowledge, and authority as emergent properties. In Janet Holmes & Miriam Meyerhoff (Eds.), *The handbook of language and gender*, *volume 25*, (pp. 302–326). Oxford: Blackwell. https://doi.org/10.1002/9780470756942.ch13

Meyerhoff, Miriam (2014). Variation and gender. In Susan Ehrlich, Miriam Meyerhoff, & Janet Holmes (Eds.), *Handbook of language, gender and sexuality* (pp. 87–102). Chichester, West Sussex: Wiley Blackwell. https://doi.org/10.1002/9781118584248.ch4

Meyerhoff, Miriam (2017). Writing a linguistic symphony: Analyzing variation while doing language documentation. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, *62*(4), 525–549. https://doi.org/10.1017/cnj.2017.28

Moeller, Sarah R. (2014). SayMore, a tool for language documentation productivity. *Language Documentation and Conservation*, *8*, 66–74.

MPI Nijmegen. (2018). ELAN (5.1) [Computer software]. Max Planck Institute for Psycholinguistics, The Language Archive.

Nagy, Naomi (2015). A sociolinguistic view of null subjects and VOT in Toronto heritage languages. *Lingua*, *164*, 309–327. https://doi.org/10.1016/j.lingua.2014.04.012

Nagy, Naomi (2017). Documenting variation in (endangered) heritage languages: How and why? In Kristine A. Hildebrandt, Carmen Jany, & Wilson Silva (Eds.), *Documenting variation in endangered languages* (pp. 33–64). University of Hawai'i Press.

Nagy, Noami (2009). The challenges of less commonly studied languages: Writing a sociogrammar of Faetar. In James N. Stanford & Dennis R. Preston (Eds.), *Variation in indigenous minority languages* (pp. 397–418). Amsterdam: John Benjamins. https://doi.org/10.1075/impact.25.20nag

Parker, G. J. (1968). Southeast Ambrym verb inflection and morphophonemics. In A. Capell, G. J. Parker, & A. J. Schütz (Eds.), *Papers in linguistics of Melanesia* (pp. 27–40). Canberra, A.C.T., Australia: Australian National University.

Parker, Gary J. (1970). *Southeast Ambrym dictionary*. Canberra, A.C.T., Australia: Australian National University.

Pearce, Elizabeth (2012). Dual and plural in languages of Vanuatu. *Linguistic Typology*, *16*(2), 191–232. https://doi.org/10.1515/lity-2012-0007

R Core Team (2018). *R: A language and environment for statistical computing* (3.5.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Rassool, Romola (2013). Issues of power and privilege in the maintenance of Sri Lanka Malay: A sociolinguistic analysis. In Sebastian Nordhoff (Ed.), *The genesis of Sri Lanka Malay* (pp. 121–145). Leiden: Brill.

Ridge, Eleanor (2017). *Use of loan verbs in Vatlongos (Southeast Ambrym), Vanuatu* [Oral presentation]. Cambridge Conference on Language Endangerment 7 (CCLE7), Cambridge University.

Ridge, Eleanor (2018a). Language contexts: Vatlongos, Southeast Ambrym (Vanuatu). *Language Documentation and Description*, *15*, 87–122.

Ridge, Eleanor (2018b). *Vatlongos.* Pangloss Collection. https://pangloss.cnrs.fr/corpus/Vatlongos?lang=en&mode=pro

Ridge, Eleanor (2018c, June 15). *Developments in verb-initial consonant mutation in Vatlongos (Southeast Ambrym)* [Oral presentation]. Vanuatu Languages Conference, USP Emalus Campus, Port Vila; Vanuatu. https://sites.google.com/view/vanlangconf2018

Ridge, Eleanor (2019). Variation in Vatlongos verbal morphosyntax: Speaker communities in Southeast Ambrym and Mele Maat. Doctoral thesis, SOAS, University of London. https://eprints.soas.ac.uk/id/eprint/32205

RStudio Team (2018). *RStudio: Integrated development for R* (1.1.456) [Desktop]. RStudio Inc. www.rstudio.com

Sankoff, Gillian (2006). Age: Apparent time and real time. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., pp. 110–116). Elsevier. https://doi.org/10.1016/B0-08-044854-2/01479-6

Schmidt, Annette (1985). *Young people's Dyirbal: An example of language death from Australia*. Cambridge: Cambridge University Press.

Schmidt, Annette (1991). Language attrition in Boumaa Fijian and Dyirbal. In Herbert W. Seliger & Robert M. Vago (Eds.), *First language attrition* (pp. 113–124). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620720.008

Schnell, Schnell (2019). Variable number marking in Vera'a: Animacy and beyond. *Asia-Pacific Language Variation*, *5*(2), 208–243. https://doi.org/10.1075/aplv.00006.sch

Silva-Corvalán, Carmen (1991). Spanish language attrition in a contact situation with English. In Herbert W. Seliger & Robert M. Vago (Eds.), *First language attrition* (pp. 151–171). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620720.011

Simpson, Jane (2015). Language attrition and language change. In Claire Bowern & Bethwyn Evans (Eds.), *The Routledge handbook of historical linguistics* (pp. 537–554). London: Routledge Taylor & Francis Group.

Slowikowski, Kamil (2018). ggrepel: Automatically position non-overlapping text labels with ggplot2 (R package version 0.8.0) [Computer software].

Stamen Design (2018). *Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL*. Stamen Design.

Stanford, James N. (2016). A call for more diverse sources of data: Variationist approaches in non-English contexts. *Journal of Sociolinguistics*, *20*(4), 525–541. https://doi.org/10.1111/josl.12190

Thornton, Anna M. (2012a). Overabundance in Italian verb morphology and its interactions with other non-canonical phenomena. In Thomas Stolz, Otsuka Hitomi, Aina Urdze, & Johan van der Auwera (Eds.), *Irregularity in Morphology (and beyond)* (pp. 251–269). Berlin: Akademie Verlag.

Thornton, Anna M. (2012b). Reduction and maintenance of overabundance. A case study on Italian verb paradigms. *Word Structure*, *5*(2), 183–207. https://doi.org/10.3366/word.2012.0026

Thornton, Anna M. (2018). Troubles with flexemes. In Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo, & Fiammetta Namer (Eds.), *The lexeme in descriptive and theoretical morphology*, (pp. 303–321). Berlin: Language Science Press.

Thornton, Anna M. (2019). Overabundance in morphology. In Mark Aronoff (Ed.), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.554

Tonkinson, Robert (1968). *Maat village, Efate: A relocated community in the New Hebrides.* Eugene, OR: University of Oregon.

Tonkinson, Robert (1979). The paradox of permanency in a resettled New Hebridean Community. *Mass Emergencies*, *4*, 105–116.

Tonkinson, Robert (1981). Sorcery and social change in Southeast Ambrym, Vanuatu. *Social Analysis*, *8*, 77–88.

Tonkinson, Robert (1985). Forever Ambrymese? Identity in a relocated community. *Pacific Viewpoint*, *26*(1), 139–159. https://doi.org/10.1111/apv.261008

Vanuatu Ministry of Education (2012). *Vanuatu national language policy*. https://moet.gov.vu/docs/policies

von Prince, Kilu (2015). *A grammar of Daakaka*. Berlin: De Gruyter Mouton.

Wickham, Hadley (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. http://ggplot2.org. https://doi.org/10.1007/978-3-319-24277-4

Wycliffe Bible Translators (2015). *The New Testament in the Southeast Ambrym language of Vanuatu: Vanuvei Eo e sepinien Vatlongos na mol-Vatimol xil* [The New Testament in the Vatlongos language of the people of Southeast Ambrym]. Orlando, FL: Wycliffe.

Abstract (Bislama)

Pepa ia hemi tokabaot fo rod blo tantanem toktok lo saed blo hao blo soemaot subjek long lanwis ia Vatlongos (Saot-Is Ambrym, Vanuatu). Hemi lukluk lo hamas taem ol difren rod i stap insaed lo ol stori blo manples blo trifala komuniti: Mele Maat, we ol man Saot-Is Ambrym i bin go stap klosap lo Port Vila; Endu, we oli gat wan difren tun; mo Ase-Taveak, we hemi ol narafala vilij. Ol man Mele Maat i stap iusum rod blo soemaot nambatri person we i nogat prefiks lo hem, mo oli katkatem sam prefiks i go lo wan silabol nomo, i antap bitim ol narafala komuniti. Be oli no stap soemaot se sam man nomo i stap olsem ol narafala komuniti. Ating trifala difrens ia i minim se rod blo toktok lo Mele Maat ating i kam mo simpol smol. Blo nambatu mo nambatri rod ia, hemi no man Mele Maat nomo, be tu ol man we oli skul longtaem, mekem se oli stap toktok Bislama, Inglis mo Franis oltaem. Ol difrens ia ating i soemaot se taem we lanwis i stap deinja, i save spidemap ol kaen jenis we i mekem se yu no soemaot tumas difren mining. Be kaen jenis we i hemi jensem o katkatem saon nomo, hemi jas gohed sipos lanwis i stap deinja, or hemi sef nomo.

Address for correspondence

Eleanor Ridge
School of Humanities
Media and Creative Communication
Massey University Manawatū
Private Bag 11222
Palmerston North 4442
New Zealand
e.ridge@massey.ac.nz

Publication history

Date received: 26 July 2019
Date accepted: May 2021