Language shift in Vanuatu's 2020 census: Investigation of a national dataset for 250,000 people

Guy A. Lavender Forsyth
University of Auckland

Language shift in Vanuatu's 2020 census: Investigation of a national dataset for 250,000 people

Guy A. Lavender Forsyth
University of Auckland

Abstract

Growing concern surrounds the threats faced by Vanuatu's famous diversity of Indigenous languages, with particular attention directed towards the pressure that Bislama exerts as the country's lingua franca. Here, I report the results of exploratory analyses of Vanuatu's most recent national census, leveraging individual-level data for nearly 250,000 people to gauge the magnitude of the shift away from Indigenous languages and identify the factors implicated in this process. The data reveal that 83.9% of Vanuatu's population can 'easily' speak an Indigenous language and that 74.9% learnt an Indigenous first language. Conversely, 9.6% of people cannot speak an Indigenous language (including 9.4% of Indigenous Ni-Vanuatu), and a severely conservatively-biased estimate shows that Bislama is the first language learnt for at least 14.2% of the population, including 14.3% of Ni-Vanuatu, with the true figure likely closer to 25%. The youngest generations show considerably poorer Indigenous language abilities and substantially greater learning of Bislama as a first language, a pattern consistent with a movement away from Indigenous languages over time. Further patterns that emerge indicate that Indigenous languages fare best in Vanuatu's rural areas, in the Area Councils with the smallest populations, and in the communities that are the least diverse in terms of people's island backgrounds. Focusing in on these rural areas, where Bislama likely poses the greatest threat to Indigenous languages, individuals who have a history of migration, who have access to modern communication technologies, who live in economically better-off households, and who live in households without *kastom* land tenure are less likely to speak an Indigenous language easily and are more likely to learn Bislama as a first language. While strictly correlational and non-causal, these exploratory analyses reflect both the vitality of Indigenous languages in Vanuatu today and the serious challenges they face from a suite of likely interconnected factors.

Summary in Bislama

Plante man oli wari se ol lanwis blong Vanuatu bae oli lus. Hemia, hemi from se Bislama i stap kam antap long kantri ia. Blong faenemaot wanem i stap long situesen ia, mi luk long infomesen blong nasanol sensas blong 2020 we Vanuatu Bureau Blong Statistik i bin givim long mi. Infomesen ia i inkludum klosap evri man, woman, mo pikinini we oli stap long Vanuatu tede. Klosap long 250,000 pipol i stap insaed long sensas. Mi faenemaot se ova 80% blong ol man Vanuatu i save toktok gud long aelan lanwis, mo tu, aelan lanwis hemi fes lanwis blong 75% blong ol man we oli stap long Vanuatu. Be tu, klosap long 10% blong ol man Vanuatu oli no save toktok eni aelan lanwis. Mo ova long 14% blong ol man Vanuatu oli stap toktok Bislama olsem fes lanwis blong olgeta. Ol yangfala oli yusum Bislama moa bitim ol olfala. Mo tu, ol yangfala oli no save gud ol aelan lanwis olsem ol olfala. Hemia, hemi minem se long fiuja, maet moa man i stat yusum Bislama olsem fes lanwis mo bae i no gat tumas man

we i save aelan lanwis. Mi faenemaot se ol aelan lanwis oli moa strong long sam komuniti bitim sam narafala komuniti. Ol komuniti we lanwis i strong oli stap aotsaed long taon. Ol komuniti ia oli stap long ol eria we i no gat tumas man, mo ol komuniti ia oli no gat tumas man we oli kamaot long ol difdifren aelan. Mi faenemaot se samfala man oli save gud lanwis blong olgeta mo oli no toktok Bislama olsem wan fes lanwis blong olgeta. Ol man ia oli stap long sem ples we oli bin bon, oli no gat mobael fon, mo oli no yusum intanet tumas, oli no gat tumas mane blong pem ol niufala materiel blong bildim haos, mo haos blong olgeta i stap long kastom graon blong olgeta. Olgeta samting ia we mi faenemaot, oli jes kamaot long infomesen blong sensas 2020 nomo, mo long fiuja maet bae mifala faenemaot sam difren stori bakagen. Be mi wantem mekem klia se naoia ol aelan lanwis blong Vanuatu oli gat laef yet, mo fulap man Vanuatu oli yusum olgeta lanwis ia evri dei. Be tu, ol aelan lanwis blong Vanuatu bae oli fesem hadtaem from se wol blong yumi i stap jenis.

Keywords

language shift, language endangerment, Bislama, Indigenous languages, development, migration, Vanuatu, Pacific

---

1   Introduction

Amidst the global coronavirus pandemic of 2020, as great portions of the world lived under lockdown, Vanuatu's National Statistics Office (since renamed *National Bureau of Statistics*) completed data collection for the national census, producing a dataset that constitutes the largest, most comprehensive survey yet of this South Pacific country's 300,000-strong population. With such a large amount of individual- and household-level data, the 2020 census exhibits clear potential for aiding the understanding of Vanuatu's language situation. In particular, the 2020 census data promise to provide valuable insights for ongoing discussions about the severity of, and factors implicated in, language endangerment.

Vanuatu is a roughly Y-shaped archipelago of islands, 66 of which are inhabited according to the 2020 census. Vanuatu is well known for its linguistic diversity, possessing perhaps the densest such diversity on earth with over 130 Indigenous languages (François et al., 2015). However, linguists' and ethnographers' reports provide a growing number of case studies across various communities indicating that Bislama, an English-lexified creole that is Vanuatu's national lingua franca, is playing an increasingly important role in daily life (for a review, see Lavender Forsyth, 2025). Many observers are therefore concerned that an ongoing shift towards Bislama is exerting increasing pressure on Vanuatu's famous diversity in Indigenous languages. Here, I explore the 2020 census data in relation to the question of a shift away from Indigenous languages, describe and visualise the patterns that emerge, and discuss the results in relation to existing literature and current understandings of language endangerment in Vanuatu.

National census data are a valuable resource for linguists interested in assessing language vitality and the factors implicated in language endangerment. There are many examples of census data being used for these purposes in other countries, including Australia (Kipp & Clyne, 2003), Canada (Choy et al., 2021), the USA (Stevens, 1992), the UK (Kandler et al., 2010), India (Dwivedi et al., 2020), South Africa (Posel & and Zeller, 2016), Guatemala (Hawkins, 2005), and the Soviet Union (Silver, 1975). As one review of the field of language endangerment states, census data are 'a powerful tool in revealing macro-level trends in the language dynamics of a country, state or polity' (Pauwels, 2016, p. 36). At the same time,

census data bring their own limitations, since they invariably attempt to distil complex behaviours into simple data points through surveys with strict limitations on time and cost. An important part in any analysis of census data is therefore a critical reflection on the nature and meaning of the data, and openness about what can and cannot be inferred from any particular dataset (Pauwels, 2016, pp. 35–46).

Regarding Vanuatu, existing research already makes use of earlier censuses to help understand the country's complex and changing language situation. The 1989 census was the first that took place in independent Vanuatu and was also the first to include a question about language (the two pre-independence censuses of 1967 and 1979 do not include any questions specifically addressing language). Terry Crowley (1994) used the 1989 data to estimate the speaker population for Paamese while also drawing attention to some of the difficulties in using census data to make inferences about the strength or decline of Vanuatu's languages. He notes that the 1989 census makes no attempt to record the names of the different Indigenous languages—something that remains true for the 2020 census. This is an understandable decision on the part of the National Bureau of Statistics due to the difficulties involved in asking respondents to name their own language, which include the lack of mutually agreed names for many languages and differences in individuals' perceptions of the distinction between dialects and languages (Crowley, 1994, pp. 7–8). Crowley also emphasises that while censuses are highly attractive in the size of their datasets and the extent of their coverage, caution must always be taken in interpreting the meaning of the numbers. He points out several issues in the data collection and analysis for the 1989 census that mean the figures regarding languages cannot be taken at face-value. Thus, while Lynch and Crowley rely on the 1989 census to estimate speaker populations in their landmark volume *Languages of Vanuatu* (2001), their estimates are based on geographical population statistics rather than the language-specific census question. As I discuss further below, an accurate interpretation of the figures for the 2020 census also relies on a careful consideration of the process of data generation.

The Vanuatu censuses of 1999 and 2009 each included a question addressing language shift in their 'household' surveys (meaning one response per household, rather than one per individual). This question records the main language used by each household: an Indigenous language, Bislama, English, or French. François (2012) uses these data to show that the percentage of households with an Indigenous main language decreased across the decade, while the percentage of households using Bislama as their main language increased. Significantly, this trend is not specific to urban areas but is also visible in the rural areas where some observers had assumed that Indigenous languages were not so threatened by Bislama. This finding, derived from self-report census data, represents one important piece of evidence which, taken together with the 'on the ground' reports of linguists and ethnographers (e.g., Barbour et al., 2018; Gray, 2012; Guérin, 2008; Shipman, 2008; Takau, 2016; Walworth et al., 2021), indicates that Vanuatu's Indigenous languages are threatened by the growth of Bislama as the main language of daily life (Lavender Forsyth, 2025).

Unfortunately, the question about the main household language was not included in the 2020 census and so it is not possible to test whether the trend identified between 1999 and 2009 has continued. However, the 2020 census does include two new questions relevant to language shift and endangerment: one that records individuals' competence in an Indigenous language and another that records the first language they learnt (the verbatim wordings of both questions are given below). I hope to show that the 2020 census data can be used to generate useful insights on the state of language endangerment in Vanuatu. First, the 2020 census provides information that can help quantify the number of people that can speak an Indigenous language and how many of these speak an Indigenous language as a first language compared to Bislama. Second, the dataset provides information about the kinds of people that can and cannot speak

an Indigenous language, in terms of geography, demography, and other relevant characteristics. Limitations of the census dataset must be acknowledged, such as its lack of information about individual Indigenous languages, which inhibits analysis of language shift between Indigenous languages (see Duhamel, 2025, this issue), and its self-report nature (see the Discussion for further details). Nevertheless, the 2020 census provides an unparalleled opportunity to test current understandings of the process of language endangerment in Vanuatu – in terms of potentially declining transmission of Indigenous languages – against a nation-wide sample of fine-grained individual-level data.

2   Methods

The National Bureau of Statistics shared with me the 2020 census data for the two questions relevant to language shift, alongside several other variables useful for understanding differences in individuals' responses to these two language questions. The data were collected by Vanuatu's National Bureau of Statistics, which is required by law to collect this information from everyone living in Vanuatu. The data are anonymised to help protect individuals' privacy and any potentially identifiable data cannot be shared publicly. Permission for me to access the data was granted by the Chief Statistician on 24/04/2024. For others wishing to access the same data, they are available from the National Bureau of Statistics but only with explicit permission from the Chief Statistician. This research received ethics approval from the Max Planck Institute for Evolutionary Anthropology (no. 2023_16, dated 23/11/2023) and I hold a Vanuatu Kaljoral Senta research permit (dated 20/02/2024) to conduct research on this topic in Vanuatu.

The sample of census data I use for all analyses reported below is restricted to those with at least five years of age, excluding roughly 40,000 individuals younger than this. This is because it is by the age of five that children will likely have developed the identifiable language behaviours relevant to the current analysis. Another 5,645 individuals do not have data for either of the two language questions and are therefore not included in my analyses. These exclusions do not significantly change the makeup of the basic demographic descriptors published in the official census reports. Readers interested in finding out more about the nature of the sample can therefore refer to the publications of the Vanuatu National Bureau of Statistics (Hakkert & Pontifex, 2022; VNSO, 2020).

The two language questions tap into different aspects of language shift. The first question records each individual's degree of competence in an Indigenous language. The original question wording is: 'Can [respondent] speak an indigenous (vernacular) language?' (Bislama: '[respondent] hemi save toktok lo wan indijines (lokol) lanwij?') Responses were recorded on a three-tier scale: 'No', 'Yes with difficulty' ('Yes wetem difikolti'), and 'Yes easily' ('Yes, isi nomo'). The second question records the respondent's first language learnt: 'What is the first language [respondent] learned to speak?' (Bislama: 'Wanem hemi fes lanwij we [respondent] hemi lanem blo toktok?') Responses were recorded for 'English', 'French', 'Bislama', and 'Indigenous (vernacular)'. Both questions are individual-level, meaning one data point per person. This provides an opportunity to correlate individuals' answers for the language questions with other individual-level, household-level, and area-level variables. However, it is crucial to know, for an accurate interpretation of the figures, that the second language question was not asked for everyone—only those that could speak an Indigenous language with ease or difficulty. This means the two language questions are not independent. Knowledge of whether someone's first language is Bislama or an Indigenous language is dependent on their speaking an Indigenous language (with ease or with difficulty). The census thus almost certainly underestimates the percentage of first-language speakers of Bislama,

because anyone who learnt Bislama as a first language and does not speak an Indigenous language is excluded from this estimate.

Some information about the two language variables is already provided in official publications. The Basic Tables report (VNSO, 2020) gives raw numbers for how many people's first language is Indigenous, Bislama, English, or French in each Area Council and breaks this down for men and women, but does not provide any information regarding the question about competence in an Indigenous language. The Analytical Report (Hakkert & Pontifex, 2022) presents graphs of Indigenous language competence by age for men and women, and provides a percentage breakdown of first language responses by urban-rural location. However, there remains much more that can be done with this data than has thus far been undertaken. In future I plan to publish further analyses of these data. In the meantime, I here take the opportunity to share some insights from an initial exploration and visualisation of the 2020 census data.

All analyses are exploratory and none were preregistered. All data cleaning, analysis, and visualisation was conducted in R (version 4.4.1; R Core Team, 2024) and RStudio (version 2024.09.0+375; Posit Team, 2024). Graphs were produced using the ggplot2 package (Wickham, 2016). The maps were downloaded using the geodata package (Hijmans et al., 2024) and were plotted using the packages sf and ggspatial (Dunnington, 2023; Pebesma, 2018). The complete code for data cleaning, analysis, and visualisation is publicly available on the OSF (https://doi.org/10.17605/OSF.IO/B8SUT) and does not compromise the restrictions on sharing the census data described above.

3   Results

3.1   Who speaks what?

Figure 1 presents the percentage breakdown for responses to the two census questions relevant to language shift. Figure 1 also visualises the non-independence of the two questions, with respondents' answers to the first question, about whether they speak an Indigenous language, determining whether their responses are recorded for the second question about what their first language is. The full sample for the first question about Indigenous language competence numbers 244,377 respondents. Of these, 83.9% speak an Indigenous language 'easily', 6.5% speak an Indigenous language 'with difficulty', and 9.6% do not speak an Indigenous language. The sample for the second question about respondents' first language (which is a sub-sample of that for the former question) numbers 220,913, of whom 82.8% learnt an Indigenous first language and 14.2% learnt Bislama as a first language. Importantly, however, the sample for this second question excludes everyone who cannot speak an Indigenous language (including nearly one in ten Indigenous Ni-Vanuatu: see below). If we presume that everybody who is unable to speak an Indigenous language learnt a non-Indigenous first language (rather than treating them as NA), we can calculate an overall estimate for first-language speakers of Indigenous languages in Vanuatu of 74.9%. These speakers of non-Indigenous languages most likely learnt Bislama, English, or French as a first language, as no other languages are widely spoken in the country. The true percentage of people who speak Bislama as a first language in Vanuatu is therefore almost certainly higher than the conservative figure of 14.2% and is more likely closer to 25%.

Ni-Vanuatu make up 98.9% of the sample for the former language question and 99.2% for the latter. The vast majority of these identify as only Ni-Vanuatu, with only 464 people in the entire sample identifying as part Ni-Vanuatu. It is thus unsurprising that responses to the two language questions show little difference when looking only at responses from Ni-Vanuatu. Of

Ni-Vanuatu specifically, 84.1% speak an Indigenous language 'easily', 6.5% speak an Indigenous language 'with difficulty', and 9.4% do not speak an Indigenous language. Of those that speak an Indigenous language, 82.9% of Ni-Vanuatu learnt it as their first language and 14.3% learnt Bislama as their first language. Given that the existence of non-Ni-Vanuatu is part of the reality of the country's language situation, all following analyses will proceed with the full sample including both Ni-Vanuatu and non-Ni-Vanuatu.
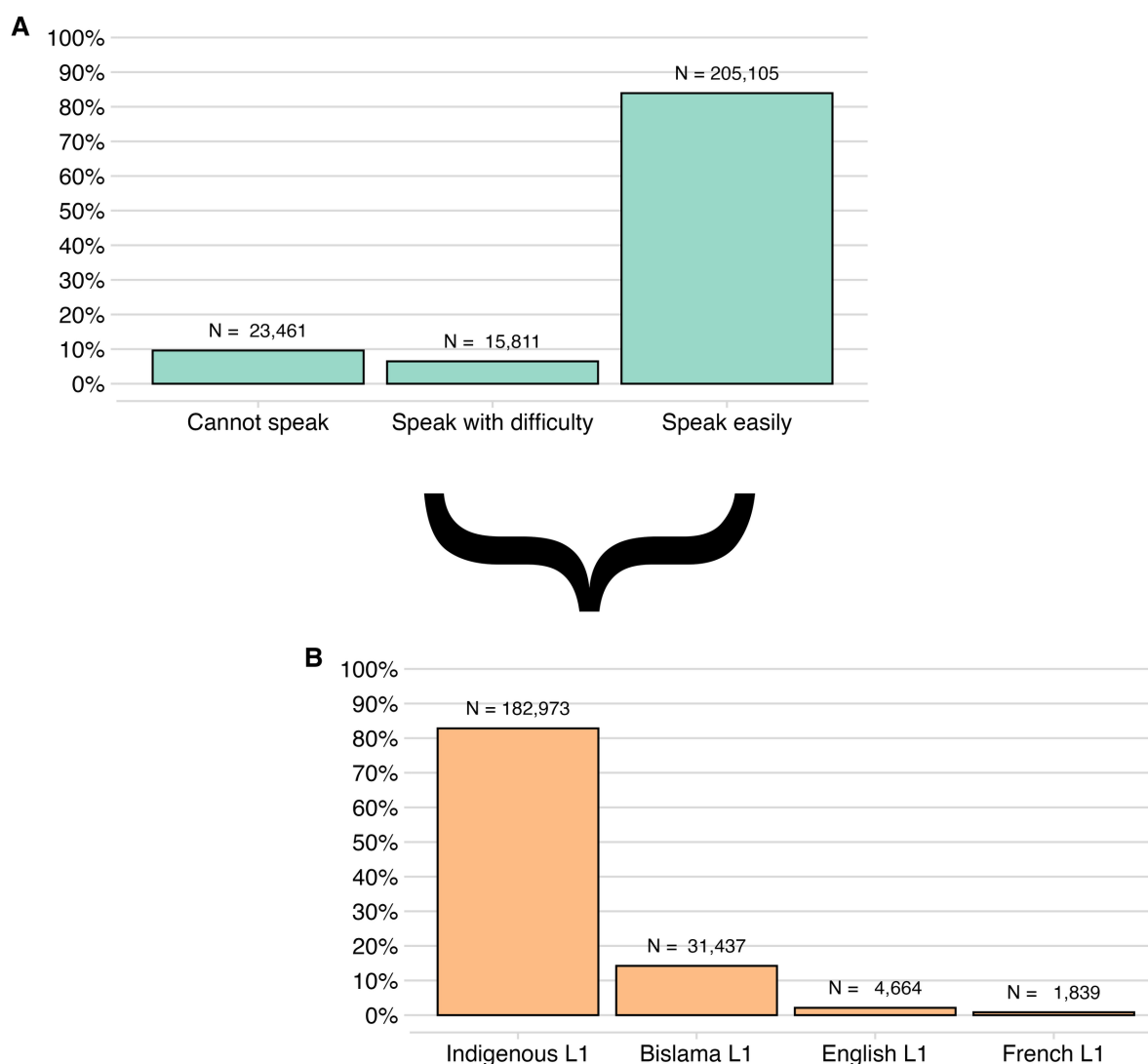

Figure 1. Two graphs showing (A) responses to the question about Indigenous language competence and (B) responses to the question about first language. Only individuals with at least 5 years of age are included in this analysis. The brace annotation highlights the dependence between the two questions, whereby data for individuals' first language are only available for those that can speak an Indigenous language at least with difficulty.

3.2   The geography of language shift in Vanuatu

Administratively, Vanuatu is divided into Provinces and these in turn are divided into Area Councils (see Figure 2-A). Figure 2-B presents each Area Council's total population over the age of four, showing that only three islands contain Area Councils with populations greater than 6,000. This includes Efate where Port Vila is located, which is the country's capital and

the only Area Council to have a population greater than 24,000, Santo which contains Luganville, Vanuatu's 'second city' and the only Area Council with a population greater than 12,000 (and less than 24,000), and Tanna which has two Area Councils with populations greater than 6,000 (and less than 12,000). All other islands' Area Councils have populations under 6,000 people, highlighting the rural nature of most of the country.
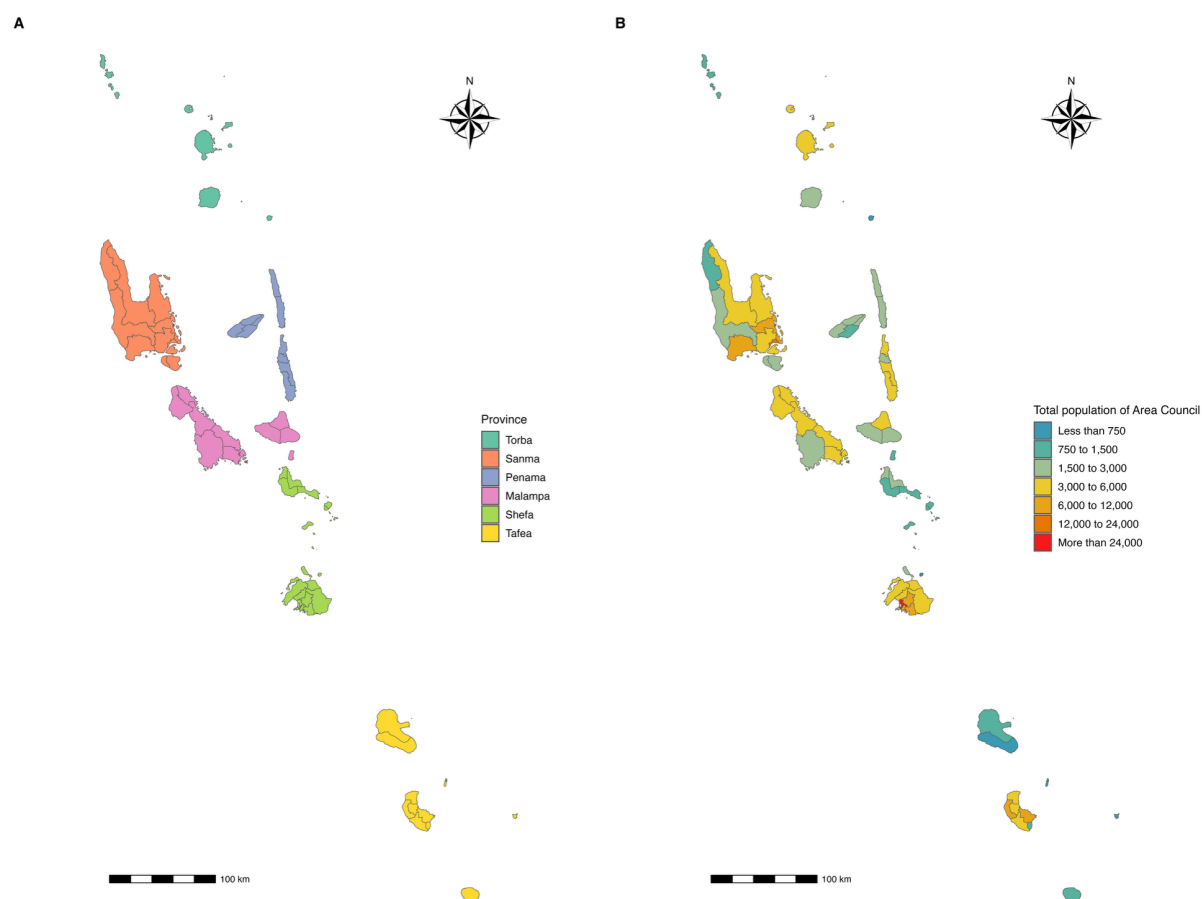

Figure 2. Maps of Vanuatu showing (A) the administrative division of the archipelago into Provinces and Area Councils, and (B) the raw number of people over the age of four living within each Area Council according to the 2020 census. Note that some smaller islands are grouped together into a single Area Council and for these the population total refers to that of entire Area Council rather than just the island. The Area Council boundaries reflect those used in the 2020 census, which in some cases (e.g., Santo) are slightly different to those currently used for government administration.

Turning to the census language data, Figure 3 presents percentage responses to the two language questions based on whether the respondent's location is urban or rural. For this I employ the National Bureau of Statistics' definition of urban and rural, which divides Port Vila and Luganville (N = 56,256; 23% of the full sample) from the rest of the country (N = 188,121; 77%). Comparing urban areas to rural, a lower percentage of people speak an Indigenous language easily (Figure 3-A), a higher percentage of people speak an Indigenous language with difficulty (3-B), a lower percentage of people learnt an Indigenous first language (3-C), and a higher percentage of people learnt Bislama as a first language (3-D).
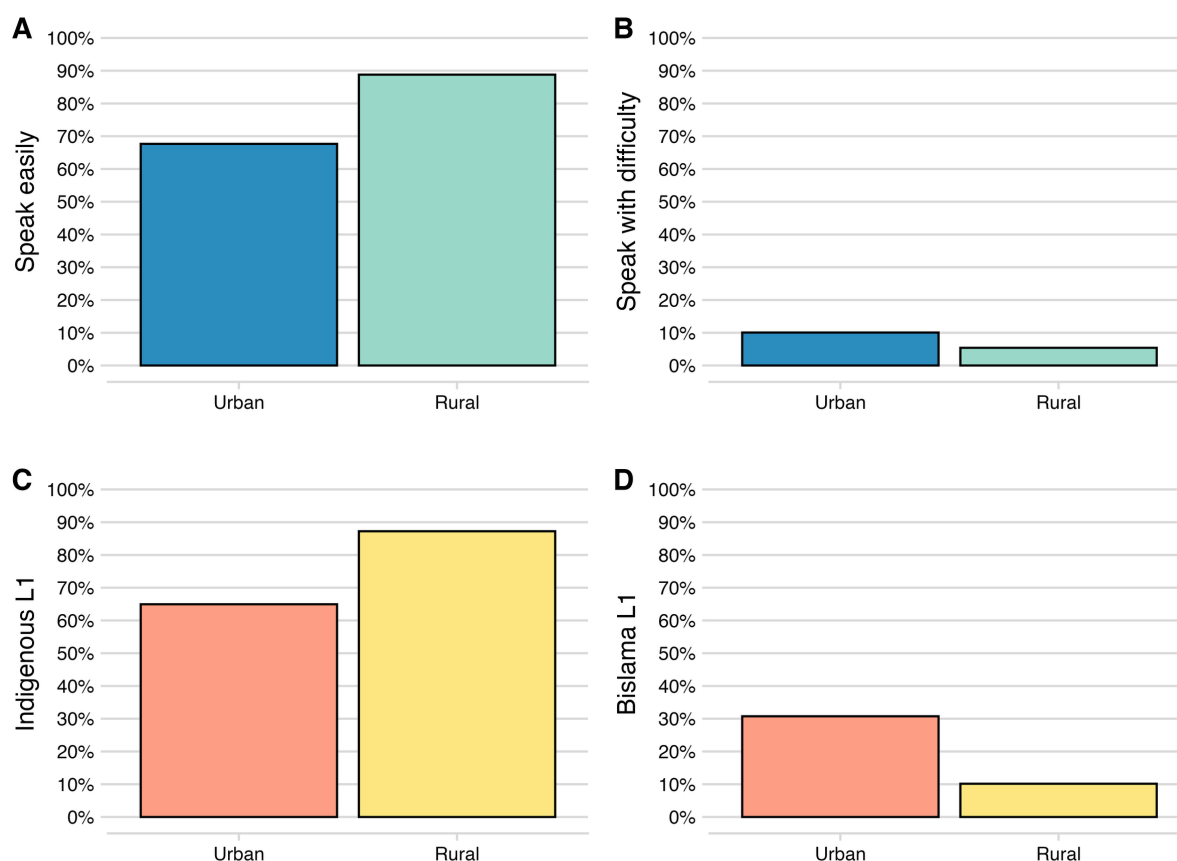
Figure 3. Responses to the language questions by urban-rural location, with separate graphs for the percentage that speak an Indigenous language easily (A) and with difficulty (B), and that learnt an Indigenous first language (C) and Bislama as a first language (D).

The patterning of responses to the two language questions by urban-rural location is clear, but this does not mean that the causal story that explains the pattern is equally straightforward. There are many aspects of life that differ between urban and rural Vanuatu, and it is not immediately obvious which factors would be most responsible for the lower Indigenous language competence and increased learning of Bislama as a first language in the urban areas.

To begin with, the diversity of responses to the language questions across different parts of the country can be visualised. Does significant geographical variation exist beyond the urban-rural binary? Figure 4 presents the percentage of people in each Area Council that speak an Indigenous language easily (4-A) and that learnt an Indigenous first language (4-B). This visualisation demonstrates a range of variation across Area Councils: from West Santo and Aniwa where 99.2% of people speak an Indigenous language easily to East Malo where only 59.9% do; and from South Erromango where 99.1% learnt an Indigenous first language to Aniwa where—despite the high proportion of people who speak an Indigenous language easily—only 56.9% apparently learnt it as a first language. Despite the intriguing exception of Aniwa, generally there seems to be a correspondence between responses to the two language questions. Areas with a higher percentage of people that speak an Indigenous language easily generally seem to have a higher percentage of people that learnt an Indigenous first language.

Notably, the two urban Area Councils—Port Vila and Luganville—are not the areas with the lowest percentages for Indigenous language competence and Indigenous first language. They place near the bottom, but Central Malekula places below Port Vila on both metrics and, below both of these, East Malo places below Luganville on both metrics too. Correspondingly,

while more remote parts of the country such as West Santo and the northerly Torres and Banks regions do appear to be strongholds where many people speak an Indigenous language easily and as a first language, rural Vanuatu is not an uninterrupted haven for Indigenous languages. Even within the same island variation can exist, as is clearly the case for Malekula and Epi. I thus find variation in the extent of language shift both between islands and within them.
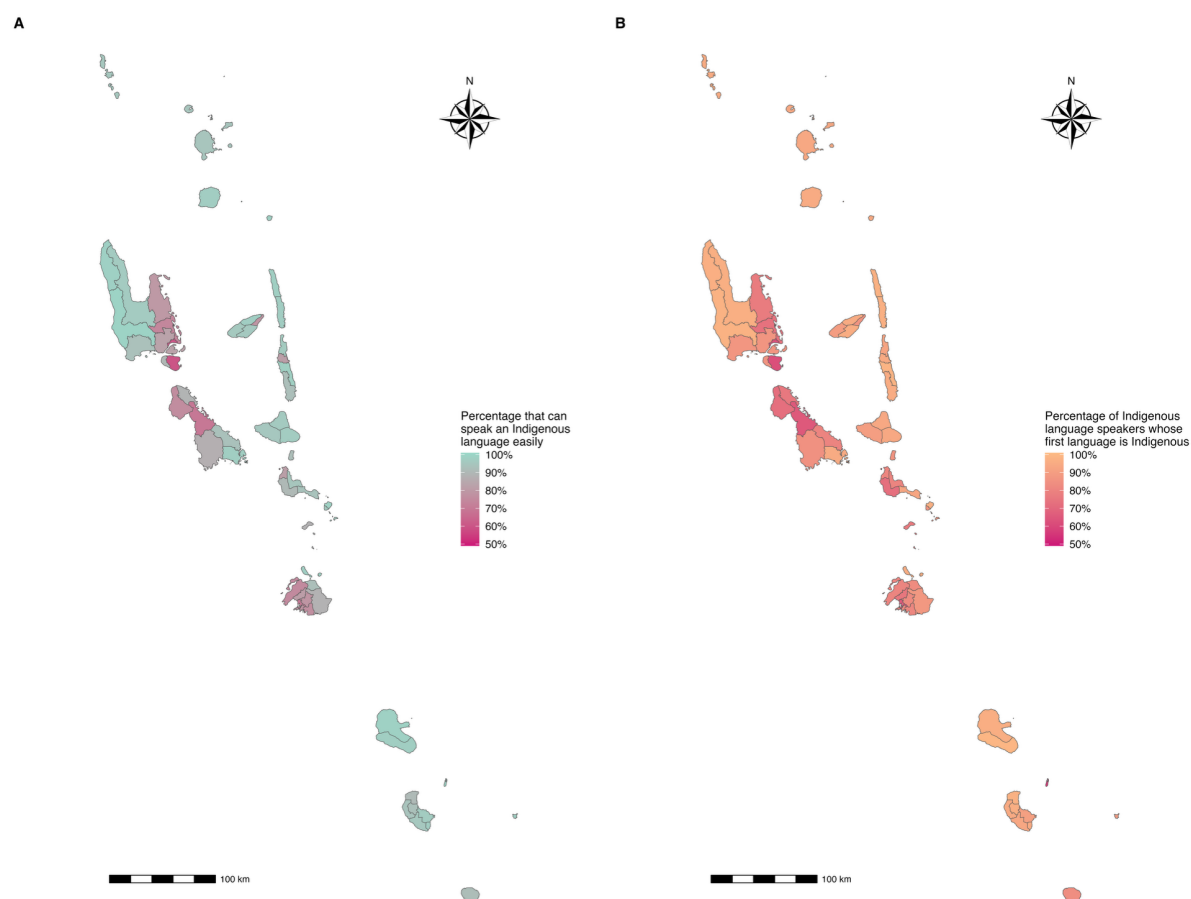

Figure 4. Maps of Vanuatu showing (A) the percentage of people in each Area Council who speak an Indigenous language easily, and (B) the percentage of people in each Area Council who both speak an Indigenous language *and* whose first language is an Indigenous language. Only individuals over the age of four are included in this analysis.

What might account for this geographical variation? Here I explore three non-exhaustive possibilities: the total population of an area, the amount of 'home island diversity' within a community, and the number of Indigenous languages traditionally spoken on an island.

Figure 5 presents each Area Council's total population (those with at least five years of age) against percentage responses to the two language questions. Population size is represented on a log base-10 scale to make it easier to read. Figure 5-A plots the percentage of people in each Area Council that speak an Indigenous language easily and 5-B plots the percentage of people in each Area Council that learnt an Indigenous first language. This visualisation reveals negative associations between total population and these responses to the two language questions. While it is possible to have a large population and high percentages of people who speak an Indigenous language easily and who learnt an Indigenous first language—as is the case for West Tanna—the overall trend seems to be for Area Councils with larger populations

to have lower percentages of people who speak an Indigenous language easily and who learnt an Indigenous first language.
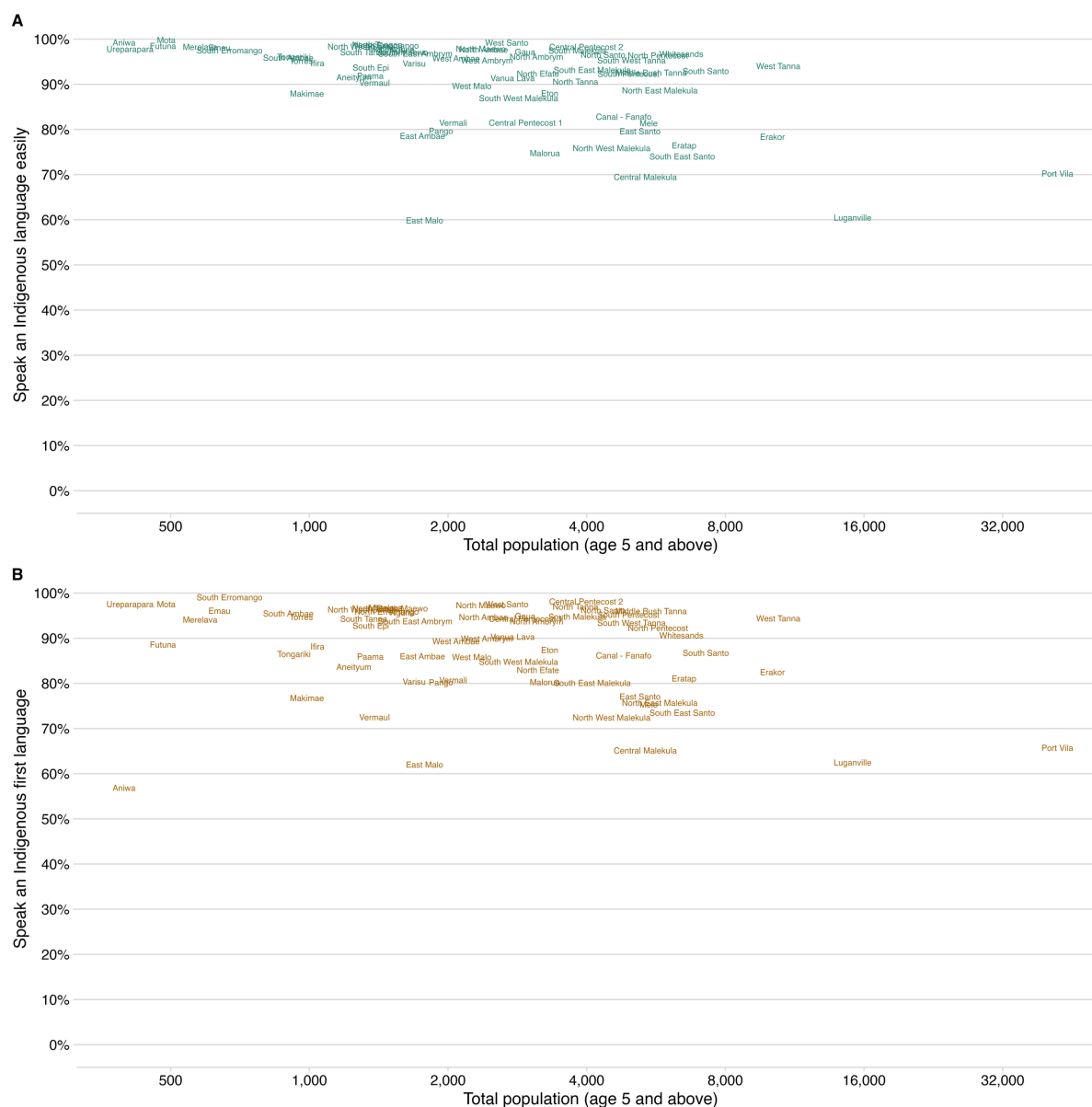

Figure 5. Graphs showing each Area Council's total population and (A) the percentage of people in the Area Council that speak an Indigenous language easily and (B) the percentage of people in the Area Council that learnt an Indigenous first language. Note that population, represented on the x-axis, is on a log base-10 scale to aid readability.

Figure 6 presents responses to the two language questions against the amount of 'home island diversity' within each Enumeration Area. An Enumeration Area is a geographical unit the National Bureau of Statistics uses which is below that of the Area Council. There are 881 Enumeration Areas in the 2020 census, of which the current analysis uses 865 (16 are excluded for having fewer than ten respondents for each language question). The median population size of these Enumeration Areas is 251. In Figure 6, each dot represents an Enumeration Area, the colour of which indicates whether the National Bureau of Statistics classes it as urban or rural. Enumeration Areas are ranked along the x-axis by the amount of 'home island diversity' represented amongst their residents. I calculate this by taking the total number of unique home

islands claimed by the residents of an Enumeration Area and dividing it by the Enumeration Area's total population. 'Home island diversity' thus represents a population size-adjusted measure of the number of different home islands that people in a community affiliate to. To take a couple of worked examples, one Enumeration Area has a population of 134 and all these people have the same home island. Dividing the number of home islands by the population total thus creates a 'home island diversity' score of 0.007. Contrastingly, another Enumeration Area has a population of 231 who share a total of 16 different home islands, resulting in a 'home island diversity' score of 0.069.

One pattern observable in Figure 6 is a clear difference in the amount of home island diversity between urban and rural Enumeration Areas, with urban Enumeration Areas showing the greater diversity. Figure 6-A plots this information against the percentage of people in each Enumeration Area that speak an Indigenous language easily, while Figure 6-B plots it against the percentage of people in each Enumeration Area that speak an Indigenous first language. Each figure includes a trend line produced by regressing home island diversity onto the language question. The data indicate that Enumeration Areas with greater home island diversity generally have smaller percentages of people who speak an Indigenous language easily and who learnt an Indigenous first language. However, there are comparatively fewer Enumeration Areas with home island diversity scores above 0.2, so the extrapolation of simple linear regression lines for these regions should be taken with caution. Furthermore, the plotted lines take no account of whether Enumeration Areas are rural or urban. It may be the case, for example, that the trend only applies to the comparison of urban and rural areas or amongst rural Enumeration Areas. Indeed, looking only at the urban Enumeration Areas, this visualisation suggests there might be less of a relationship between home island diversity and responses to the language questions. The current analysis should therefore be taken only as an initial description of the data. Further analyses will be necessary to test the nature and robustness of the observed associations in greater detail.
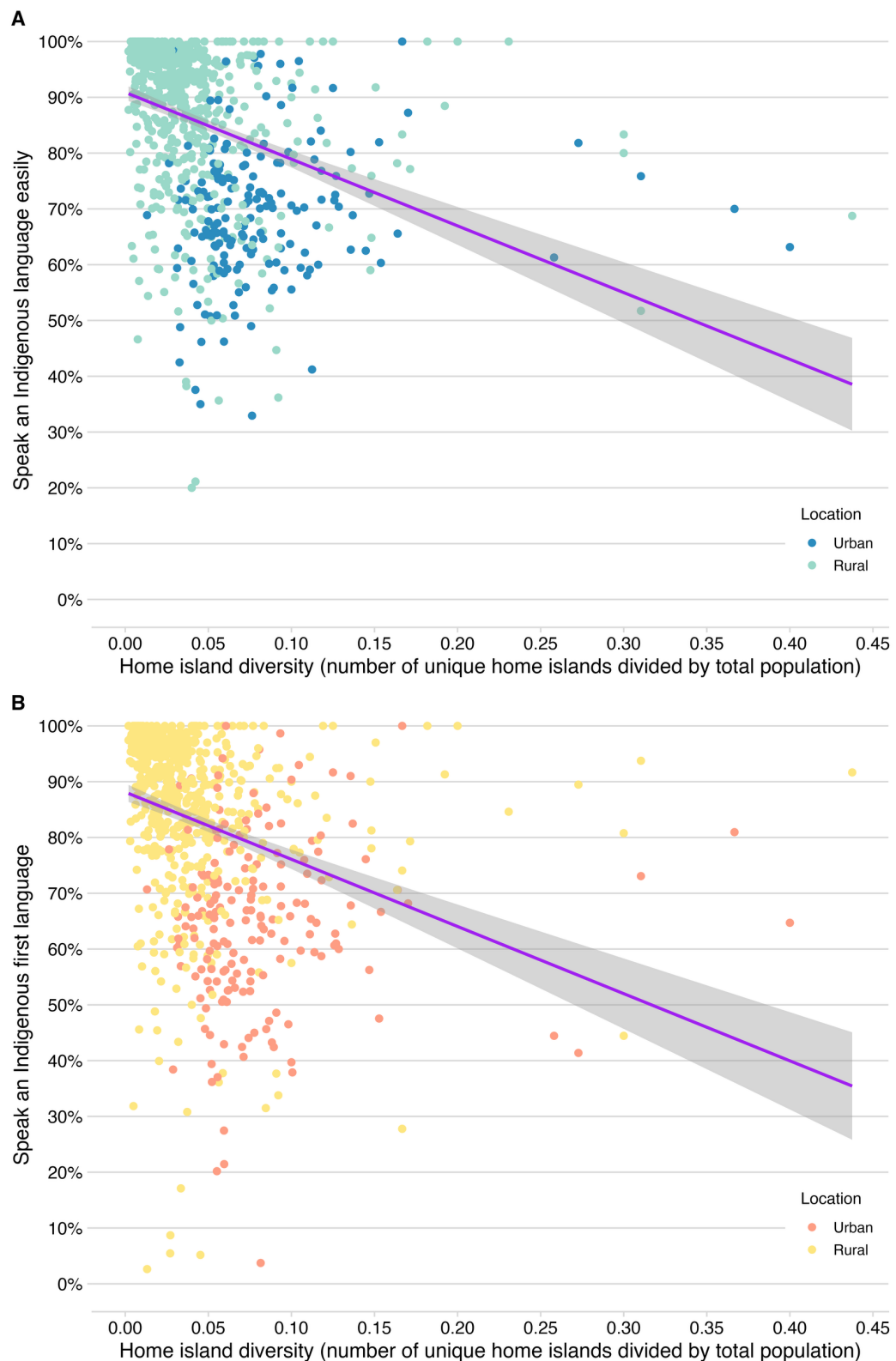
Figure 6. Graphs showing each Enumeration Area's amount of 'home island diversity', whether it is urban or rural, and (A) its percentage of people who speak an Indigenous language easily and (B) its percentage of people who learnt an Indigenous first language. Lines are produced by simple linear regressions of home island diversity onto each language variable, displayed with 95% confidence intervals. Enumeration Areas with fewer than ten respondents for the language questions are excluded from this analysis.

In Figure 7, each dot represents an island, and each dot's position represents the number of traditionally-spoken languages (x-axis) and the percentage of responses to one of the two language questions (y-axis). My count of the number of Indigenous languages per island relies on information provided by François et al. (2015). My focus on the number of languages traditionally spoken means I include the languages listed by François et al. that are moribund or that have no current speakers and I do not include languages spoken by recent migrants, for example those of people from other islands that work in Vila on Efate. My coding of the number of languages for each island is publicly available and can be checked in the uploaded R code (https://doi.org/10.17605/OSF.IO/B8SUT). Of Vanuatu's 66 inhabited islands, 60 are plotted, as six have fewer than ten respondents for each language question. Of these 60, 45 (75%) are recorded as having only one traditionally-spoken language. Two islands are clear outliers: Santo with 38 languages and Malekula with 42. The remaining 13 islands are recorded as having between two and seven languages.

Figure 7 shows no convincing relationship between the number of traditionally-spoken Indigenous languages an island has and the percentage of people that can speak an Indigenous language easily (7-A) or that learnt an Indigenous first language (7-B). While the regression lines plotted on each graph are somewhat negative, the 95% confidence intervals are wide. Across islands with just one Indigenous language there are a wide variety of responses to the language questions. Compared to these islands, those with more than one Indigenous language do not seem to have markedly different responses to the language questions. The two outlier islands, Malekula and Santo, also do not demonstrate severely lower percentage responses to the two questions. This analysis therefore does not indicate that islands with a greater number of Indigenous languages are more prone to shift away from Indigenous languages and towards Bislama.
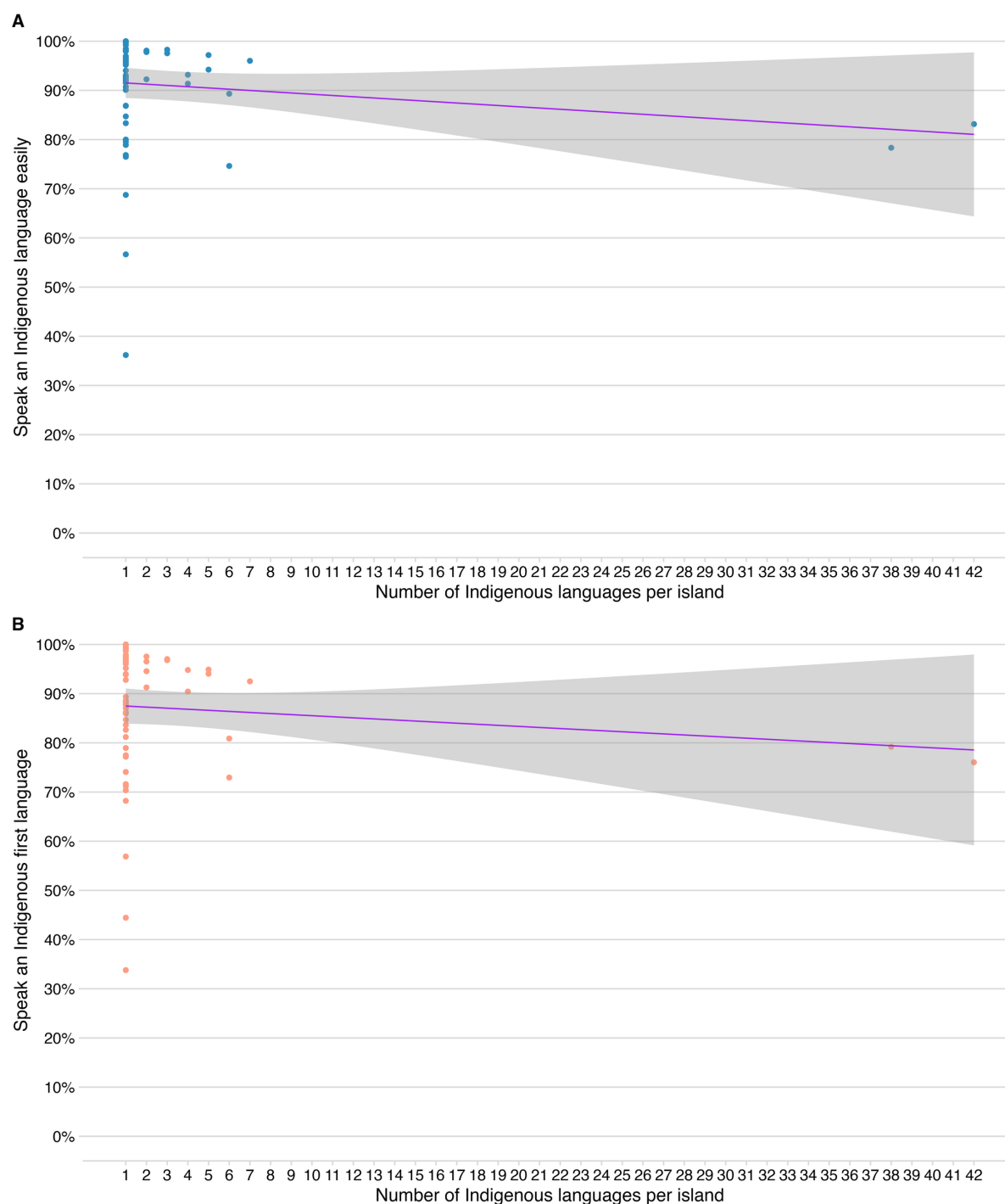
Figure 7. Graphs showing the number of Indigenous languages traditionally spoken on each island and (A) the percentage of people who speak an Indigenous language easily and (B) the percentage of people who learnt an Indigenous first language. Lines are produced by simple linear regressions of the number of Indigenous languages per island onto each language variable, displayed with 95% confidence intervals. Islands with fewer than ten respondents for the language questions are excluded from this analysis.

3.3   Individual demographic differences in language shift

Too much attention might be paid to urbanisation in discussing language endangerment in Vanuatu. This is not to deny that urban living is clearly associated with individuals' language

shift, as I show above. But it is somewhat unsurprising that people in the cities make greater use of a lingua franca. By contrast, more surprising is the growing evidence, both qualitative and statistical, that a shift towards Bislama and away from Indigenous languages is taking place in many rural communities (François, 2012; Lavender Forsyth, 2025). The factors causing this rural shift may or may not be different to those in the urban areas. In this section, therefore, where I analyse responses to the census' two language questions in relation to individual-level and household-level factors, I compare each factor across urban and rural locations. I can thereby check whether these factors are associated with less use of Indigenous languages and greater use of Bislama specifically in Vanuatu's rural areas, where Bislama's influence arguably poses the greater threat to Indigenous languages.

Figure 8 presents percentage responses for whether someone's first language is Indigenous (8-A) or Bislama (8-B) according to whether they speak an Indigenous language easily or with difficulty. This visualisation reveals a strong relationship between the two language questions, in both urban and rural areas. Nevertheless, there is still variation. Across urban and rural areas combined, 88% of those that speak an Indigenous language easily are first-language speakers, but so are 16% of those that speak an Indigenous language with difficulty (and Figure 8-A shows this to be more than 20% in rural areas). Correspondingly, 80.8% of those that speak an Indigenous language with difficulty are first-language Bislama speakers, but so are 9.1% of those that speak an Indigenous language easily (and Figure 8-B shows this to be more than 20% in urban areas). This pattern of variation indicates that the two language questions measure related but different things, as we would expect from their different foci on current language abilities and history of language learning in childhood.
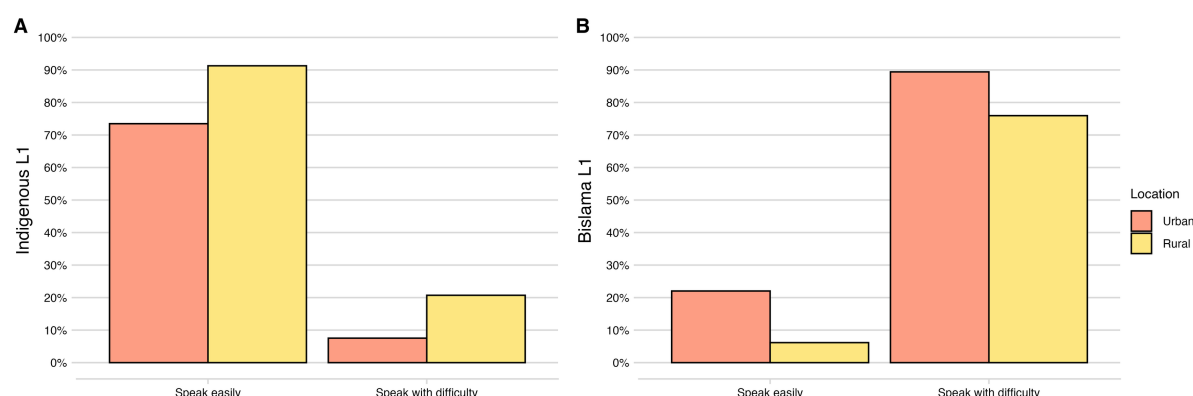

Figure 8. Percentages of people that learnt an Indigenous first language (A) and that learnt Bislama as a first language (B), according to whether someone speaks an Indigenous language easily or with difficulty, and by urban versus rural location.

Figure 9 presents the percentage of men and women's responses to the language questions, separated by urban versus rural locations. In total, 49.6% of this sample are women. The data indicates that gender is not an important factor in variation in people's responses to the language questions, with men and women having very similar responses in percentage terms.
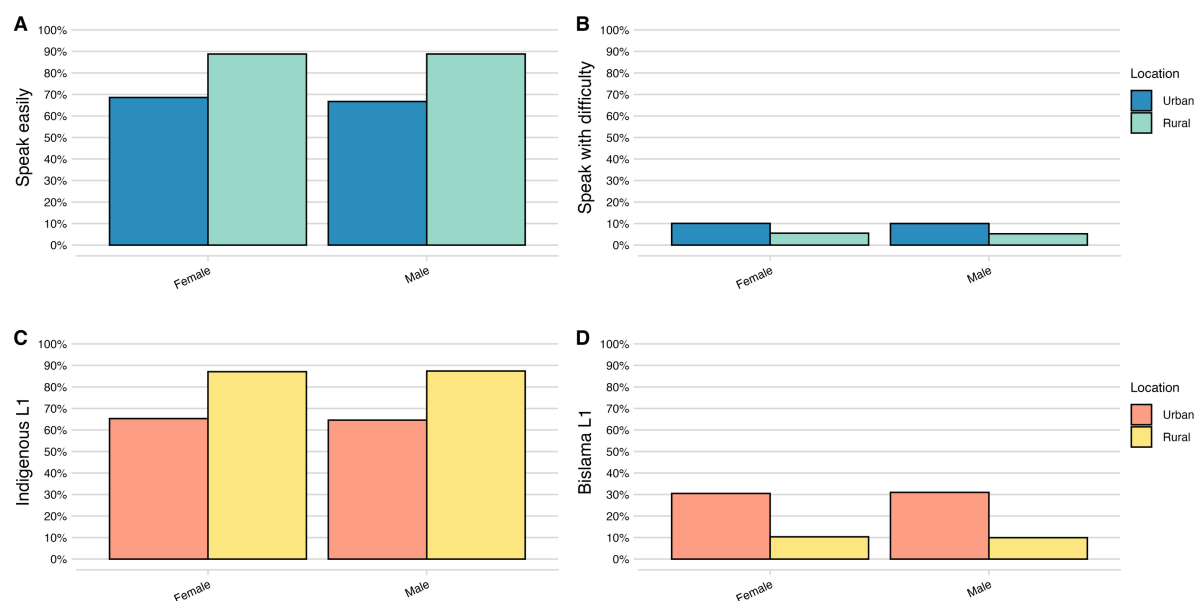
Figure 9. Responses to the language questions by gender, separated by urban versus rural location and with separate graphs for the percentage that speak an Indigenous language easily (A) and with difficulty (B), that learnt an Indigenous first language (C), and that learnt Bislama as a first language (D).

Figure 10 presents variation in responses to the language questions by age. The median age group in this sample is 25-29. The data show a clear trend whereby younger people are less likely to speak an Indigenous language easily (10-A), are more likely to speak an Indigenous language with difficulty (10-B), are less likely to have learnt an Indigenous first language (10-C), and are more likely to have learnt Bislama as a first language (10-D). These trends are more pronounced in urban areas, but they hold too for rural areas. Another insight from this visualisation is that the urban-rural difference in responses to the language questions is much greater amongst younger than older people. For the percentage of people that speak an Indigenous language easily, there is comparatively little urban-rural difference amongst those older than 45. For the percentage that speak an Indigenous language with difficulty, there is essentially no urban-rural difference amongst those over 35. Similarly striking reductions in the importance of the urban-rural divide are apparent for the percentage of people who learnt an Indigenous first language and the percentage of people who learnt Bislama as a first language.
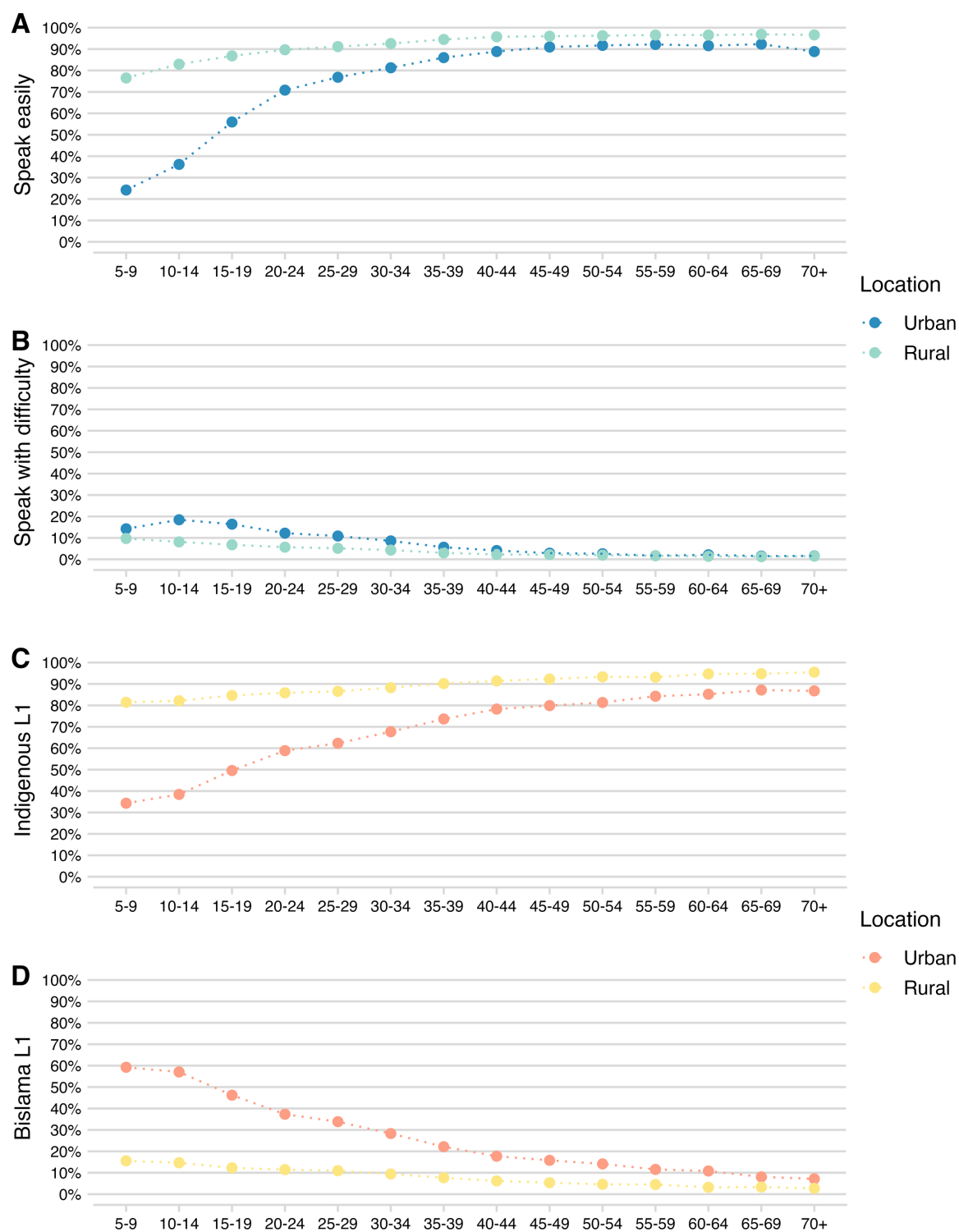
Figure 10. Responses to the language questions across the age range, separated by urban versus rural location and with separate graphs for the percentage that speak an Indigenous language easily (A) and with difficulty (B), that learnt an Indigenous first language (C), and that learnt Bislama as a first language (D).

3.4   Differences in language shift by education and migration

Figure 11 plots responses to the language questions by the highest level of education an individual has achieved. This analysis excludes all those under the age of 18 and those in education at the time of the census. The median level of education for this sample is completion of Year 6, the final year of primary school. Interestingly, there are much clearer relationships between educational attainment and responses to the question about respondents' first language (11-C and 11-D) than to the question about Indigenous language competence (11-A and 11-B). While there is a slight negative relationship between education and ability to speak an Indigenous language easily, and a slight positive relationship between education and ability to speak an Indigenous language with difficulty, these relationships are surprisingly subtle. By contrast, the associations between levels of education and the percentage of people who learnt an Indigenous first language (11-C; negative direction) and the percentage of people who learnt Bislama as a first language (11-D; positive direction) are more obvious. Formal statistical modelling of these relationships will be useful to quantify their strength and test their robusticity to controls.

Figure 12 presents responses to the language questions by individuals' migration histories: whether someone lives in the same place they were born (68.2% of this sample), in the same place they have lived for a year or more (27.8%), or in the same place for less than a year (4.1%). Focusing first on people in rural locations, there is a consistent if relatively modest relationship whereby people who have lived a shorter time in their current place of residence are less likely to speak an Indigenous language easily (12-A), are (modestly) more likely to speak an Indigenous language with difficulty (12-B), are less likely to have learnt an Indigenous first language (12-C), and are more likely to have learnt Bislama as a first language (12-D). Additionally, the results show that people who were born in urban areas have highly marked differences in their responses to the language questions compared to others. While people who have lived in an urban place of residence for less than a year or more than a year have responses to the language questions that are more similar to people in rural locations, people living in urban areas who were born in the same place are markedly less likely to speak an Indigenous language easily (12-A), more likely to speak an Indigenous language with difficulty (12-B), less likely to have learnt an Indigenous first language (12-C), and more likely to have learnt Bislama as a first language (12-D).
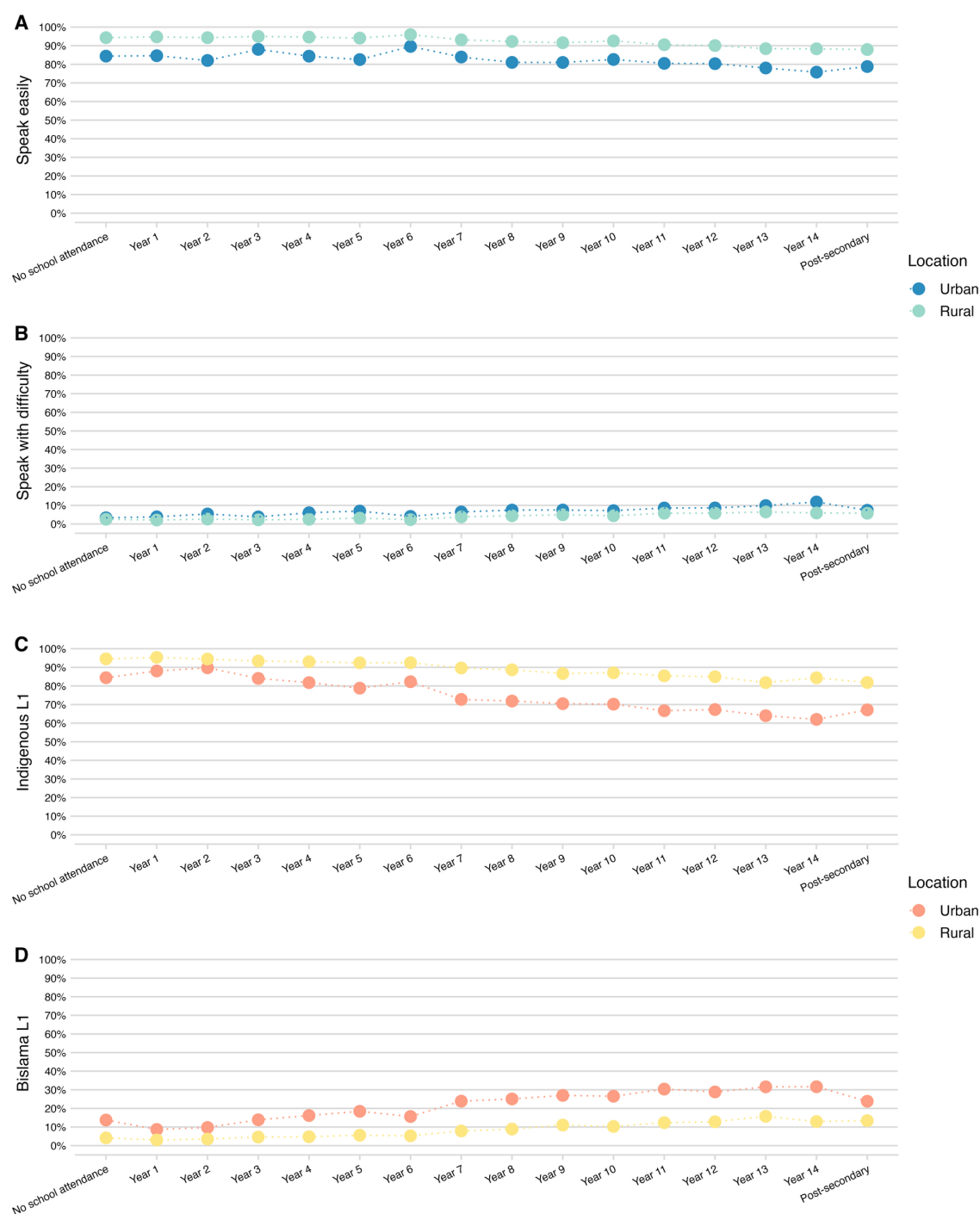
Figure 11. Responses to the language questions across levels of education, with responses separated by urban versus rural location and separate plots for the percentage that speak an Indigenous language easily (A) and with difficulty (B), that learnt an Indigenous first language (C), and that learnt Bislama as a first language (D). Individuals under the age of 18 or in education at the time of the census are not included in this analysis.
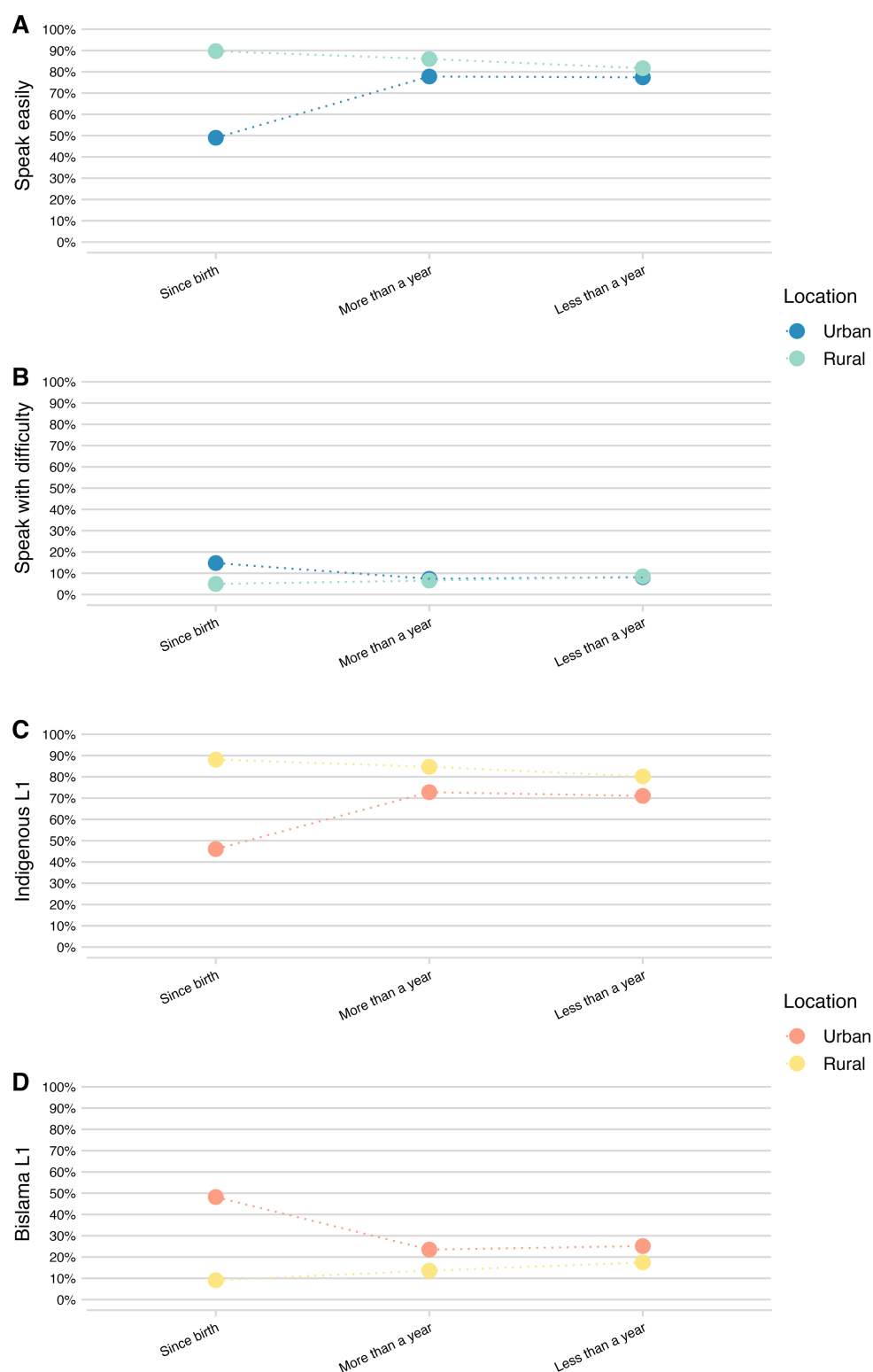
Figure 12. Responses to the language questions by residence history: people who live in the same place as they were born, who have lived in the same place for a year or more, or who have lived in the same place for less than a year. Responses are separated by urban versus rural location and there are separate plots for the percentage that speak an Indigenous language easily (A) and with difficulty (B), and that speak an Indigenous first language (C) and Bislama as a first language (D).

3.5   Differences in language shift by access to communication technologies

The next two figures present analyses of responses to the language questions by access to two forms of communication technologies: individuals' mobile phone ownership (Figure 13) and use of the internet (Figure 14). Both account for urban-rural differences. Only those with at least 15 years of age are included in these analyses. In this sample, 54.8% of people own a mobile phone in working condition and 35.2% accessed the internet in the week previous. The pattern of results for responses to the two language questions is consistent for both mobile phone ownership and internet access, and across urban and rural settings. People who own a mobile phone or who accessed the internet the week prior are slightly less likely to speak an Indigenous language easily (13-A and 14-A), slightly more likely to speak an Indigenous language with difficulty (13-B and 14-B), slightly less likely to have learnt an Indigenous first language (13-C and 14-C), and slightly more likely to have learnt Bislama as a first language (13-D and 14-D). While the trend is consistent, the sizes of the percentage differences are quite small, especially in the case of mobile phone ownership, while some of the differences in the case of internet access are somewhat larger.

Figure 13. Responses to the language questions by mobile phone ownership, with responses separated by urban versus rural location and separate plots for the percentage that speak an Indigenous language easily (A) and with difficulty (B), that learnt an Indigenous first language (C), and that learnt Bislama as a first language (D). Only people with at least 15 years of age are included in this analysis.
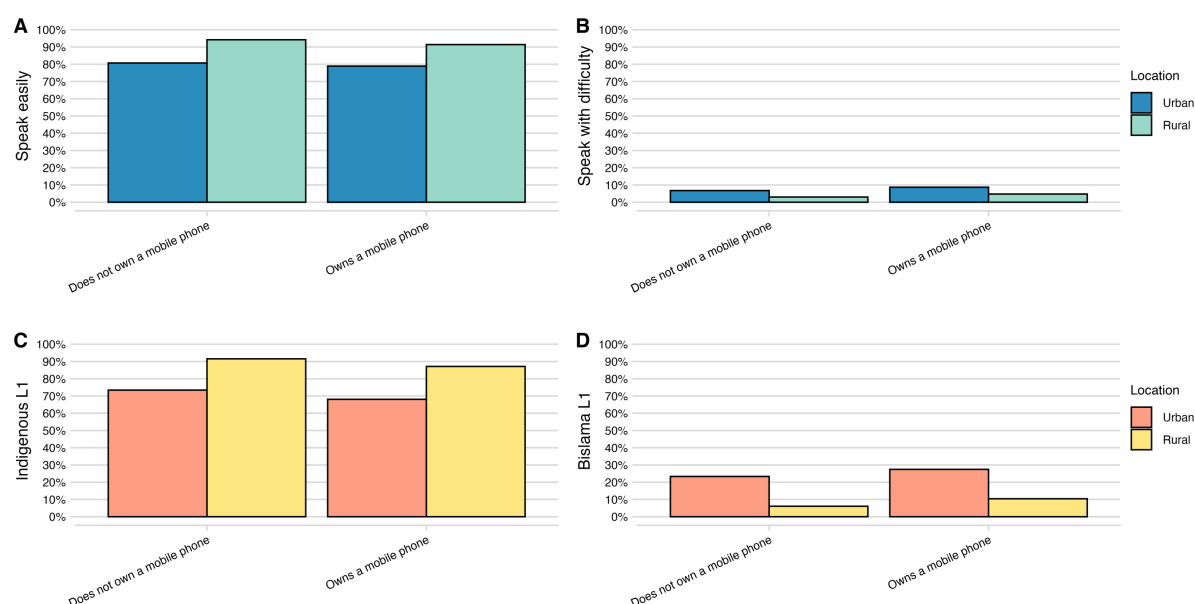
Figure 14. Responses to the language questions by internet use, with responses separated by urban versus rural location and separate plots for the percentage that speak an Indigenous language easily (A) and with difficulty (B), that learnt an Indigenous first language (C), and that learnt Bislama as a first language (D). Only people with at least 15 years of age are included in this analysis.
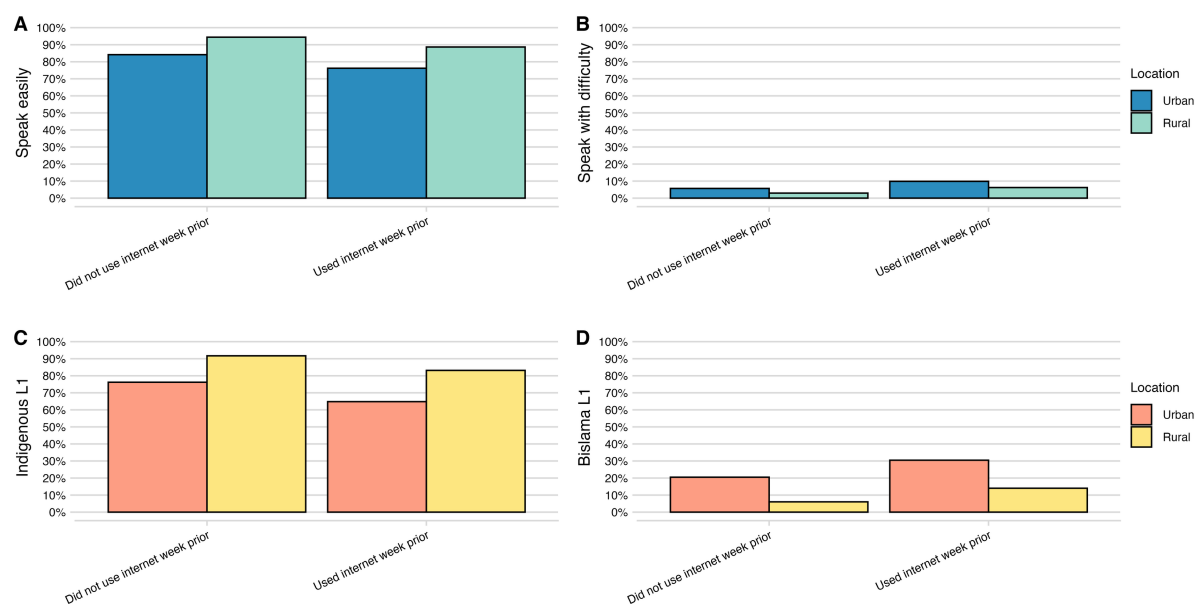
3.6   Differences in language shift by religious affiliation

Figure 15 presents responses to the language questions broken down by religious affiliation. The most popular religious group is Presbyterianism (27.5%), followed by Seventh-day Adventism (14.8%), the 'other' category (13.9%), Catholicism (12.2%), and Anglicanism (12.1%). The most striking pattern is for those professing a *kastom* religion and living in an urban area, but this represents a very small group of less than 40 people. Differences between the Christian denominations exist but consistent patterns are difficult to discern. There is possibly a tendency, at least in rural areas, for members of the Anglican and Catholic churches to speak Indigenous languages more easily and to have learnt an Indigenous first language rather than Bislama, compared to members of other churches. Followers of Seventh-day Adventism are also perhaps less likely than members of other denominations to speak an Indigenous language easily, again at least in rural areas. However, these patterns seem relatively subtle and require formal statistical testing to substantiate.
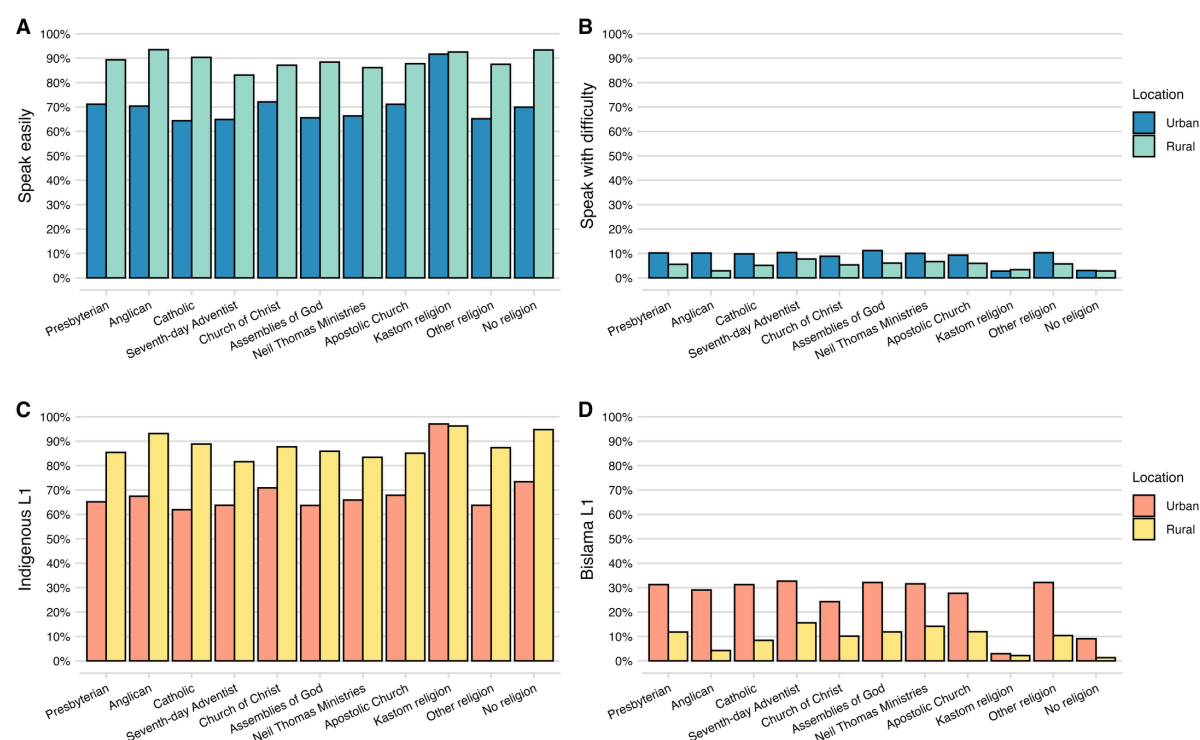
Figure 15. Responses to the language questions by religious group, with responses separated by urban versus rural location and separate plots for the percentage that speak an Indigenous language easily (A) and with difficulty (B), that learnt an Indigenous first language (C), and that learnt Bislama as a first language (D).

3.7   Differences in language shift by economic status

Figure 16 presents responses to the two language questions according to whether an individual lives in a household of 'lower', 'middle', or 'higher' economic status. I define economic status here by how many items of material wealth a household possesses, out of: (1) house walls made out of a non-'bush'/traditional material, (2) a house floor made out of a non-'bush'/traditional material, (3) a house roof made out of a non-'bush'/traditional material, (4) an improved private source of drinking water, and (5) an improved private toilet facility. The category 'lower economic status' represents households that have none or one of these items and 22.3% of individuals in this sample live in such households. 'Middle economic status' represents households that have two or three of the items, with 38.8% of individuals living in these households. 'Higher economic status' represents households that have four or five of the items and another 38.8% of individuals live in these households. This approximation of economic status corresponds with other variables that economic status should be associated with. Whereas only 7.5% of people whose main household income is from wages or business profits belong to households of 'lower economic status', 32.8% of these belong to 'middle economic status' households and 59.7% belong to 'higher economic status' households. This lends some external validity to this constructed measure of household economic status.

Figure 16 shows there is a consistent trend for people living in households of higher economic status to be less likely to speak an Indigenous language easily (16-A), (slightly) more likely to speak an Indigenous language with difficulty (16-B), less likely to have learnt an Indigenous first language (16-C), and more likely to have learnt Bislama as a first language (16-D). This is consistent in both rural and urban settings.

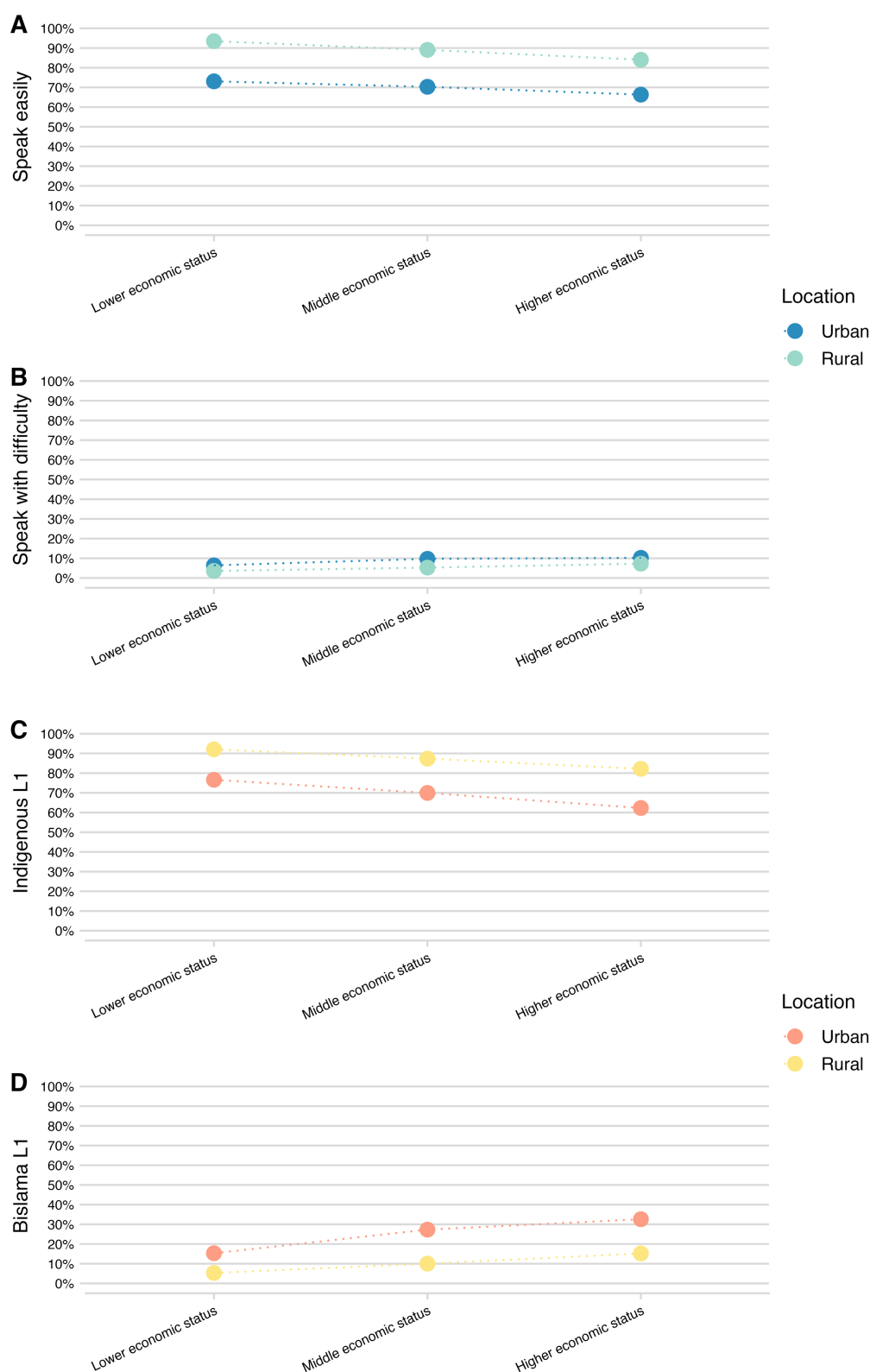**Figure 16.** Responses to the language questions by household economic status. The coding of household economic status is defined in the main text. Responses are separated by urban versus rural location, with separate graphs for the percentage that speak an Indigenous language easily (A) or with difficulty (B), that learnt an Indigenous first language (C), and that learnt Bislama as a first language (D).

3.8 *Differences in language shift by* kastom *land tenure*

Lastly, Figure 17 presents responses to the two language questions according to whether an individual lives in a household with *kastom* land tenure or not. The census classifies *kastom* land tenure based on whether one of the members of a household possesses the customary right to the land the house is built on. In total, 73% of individuals in this sample live in households with *kastom* land tenure.

Figure 17 presents interesting results that seem to run in opposite directions depending on whether someone lives in an urban or rural area. In rural areas, people who live in households with *kastom* land tenure are more likely to speak an Indigenous language easily compared to those who do not, whereas in urban areas, people who live in households with *kastom* land tenure are slightly less likely to speak an Indigenous language easily compared to those who do not (17-A). Responses to the other language questions show consistent patterns. Although the size of the percentage difference is small, people in rural areas who live in households with *kastom* land tenure are slightly less likely to speak an Indigenous language with difficulty than those who do not, whereas in urban areas those who do live in households with *kastom* land tenure are slightly more likely to speak an Indigenous language with difficulty (17-B). Rural respondents who live in households with *kastom* land tenure are more likely to have learnt an Indigenous first language than those who do not, whereas urban respondents who live in households with *kastom* land tenure are less likely to have learnt an Indigenous first language (17-C). Rural respondents who live in households with *kastom* land tenure are less likely to have learnt Bislama as a first language than those who do not, whereas urban respondents who live in households with *kastom* land tenure are more likely to have learnt Bislama as their first language (17-D).
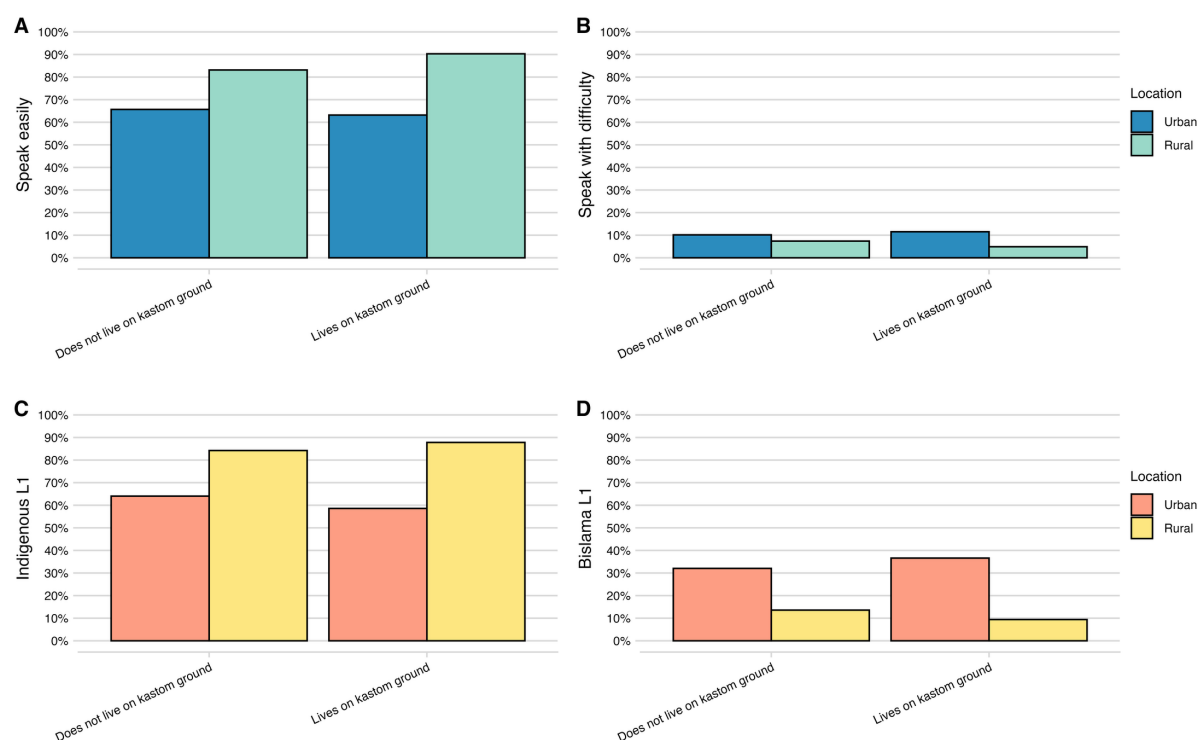

Figure 17. Responses to the language questions by whether a person lives in a household that has *kastom* land tenure, with responses separated by urban versus rural location and separate graphs for the percentages that speak an Indigenous language easily (A) and with difficulty (B), that learnt an Indigenous first language (C), and that learnt Bislama as a first language (D).

4   Discussion

The results of the 2020 census provide reasons for both reassurance and caution about the status of Vanuatu's Indigenous languages. Around 84% of the population can speak an Indigenous language easily and approximately 75% learnt an Indigenous language as their first language. These represent large majorities and demonstrate that Indigenous languages are active and important in Vanuatu today. However, the census data also provide warnings for the state of Indigenous languages: around 16% of the total population either cannot speak an Indigenous language at all or can only do so with difficulty, approximately 25% of the population learnt a non-Indigenous first language, and even amongst those that can speak an Indigenous language, 14.2% are first-language Bislama speakers. This shift away from Indigenous languages and towards Bislama is strongest in the youngest age groups, a finding that aligns with prior survey data (VNSO, 2021, pp. 111–112). Though it is possible for young people to gain fluency in an Indigenous language later in life, this pattern is consistent with a population-level trend towards decreased use of Indigenous languages over time (Krauss, 2008). This pattern also aligns with evidence from previous censuses that a growing proportion of families are using Bislama rather than an Indigenous language as their main household language, and with linguists' and ethnographers' testimonies that Bislama is playing a larger role in daily life across various parts of the country (see Lavender Forsyth, 2025). If a language shift is ongoing in Vanuatu, what factors do the census data suggest are implicated?

One factor that my analysis of the census data clearly implicates in language shift is a difference between urban and rural locations. A notably smaller percentage of people in urban areas than rural areas can speak an Indigenous language easily and learnt an Indigenous first language, whereas a greater proportion of people in urban areas than rural areas speak an Indigenous language with difficulty and learnt Bislama as a first language. Growing up in an urban area seems particularly significant. Youth living in urban areas are much less likely to speak an Indigenous language easily and are much more likely to have learnt Bislama as their first language compared to others. People who have lived in an urban area since birth are especially less likely to speak an Indigenous language easily and are significantly more likely to be first-language Bislama speakers compared to people who live in an urban area but who were not born there. Although striking, patterns like these should perhaps not be surprising, as Vanuatu's urban spaces are well known as strongholds for Bislama (Tryon, 2006, p. 107) and urbanisation is implicated in processes of language shift in other contexts (Fishman, 1971, pp. 315–320; though see Bromham et al., 2022).

The census data also show that factors associated with urbanism—total population size and the degree of social diversity—also relate to individuals' decreased competence in and first language-learning of Indigenous languages. More populated Area Councils, and Enumeration Areas with greater 'home island diversity', each tend to have lower proportions of people who speak an Indigenous language easily and who learnt an Indigenous first language. The urban areas are the two most populated Area Councils and they have a higher-than-average degree of diversity in 'home islands'. It is thus understandable that they should show lower percentages of people using Indigenous languages as this is in line with wider trends. Of course, other factors are also likely to differ between urban and rural settings, including access to communication technologies and educational institutions. Future work should aim to statistically compare population size, home island diversity, and other factors to identify which best accounts for the greater shift away from Indigenous languages and towards Bislama in Vanuatu's urban areas.

Although urban living is associated with a shift away from Indigenous languages, this language shift is not constrained to urban areas. Indeed, the Area Councils with the absolute

lowest percentages of people that speak an Indigenous language easily and that learnt an Indigenous first language are rural and not urban. This is not to deny that urban living is generally associated with a weaker role for Indigenous languages, as this is clearly shown above. However, there is a striking amount of variation in shift away from Indigenous languages across rural areas. The islands of Malo and Malekula contain areas with some of the lowest percentages in the entire country for people who can speak an Indigenous language easily and people who learnt an Indigenous first language. Further work could target these areas more closely to understand why they might be outliers in Indigenous language endangerment. Moreover, the census data demonstrate how critical it is not to overlook the evidence that shifts towards Bislama and away from Indigenous languages are taking place in rural communities (François, 2012; Lavender Forsyth, 2025). The possibility of Bislama becoming the language of daily life in rural Vanuatu poses great threat to Indigenous languages and the factors causing this rural shift may or may not be different to those in the urban areas.

In a trend that is similar to, though not quite as strong as the one in urban areas, younger people in rural settings are also less likely to speak an Indigenous language easily and are more likely to be first-language Bislama speakers than older people. As noted above, this trend is consistent with an ongoing shift away from Indigenous languages and towards Bislama. What factors, then, do the census data suggest are implicated in this rural language shift?

Migration seems likely to play an important role in language shift in rural as well as urban areas. People living in rural areas tend to report lower competence in an Indigenous language and greater use of Bislama as a first language if they were not born in their current place of residence, and especially if they have lived in their current residence less than a year. Amongst those living in rural areas, people living in a household with *kastom* land tenure are slightly more likely to speak an Indigenous language easily and to have learnt an Indigenous first language. Additionally, as noted above, the degree of 'home island diversity' in a community is related to variation in the percentage of people that speak an Indigenous language easily and that learnt an Indigenous first language. This convergent evidence aligns well with many linguists' and ethnographers' contentions that migration is a key cause of the shift away from Indigenous languages and towards Bislama in rural areas, particularly through increasing numbers of marriages between people of different language backgrounds (Crowley, 2000, p. 66; Daly & Barbour, 2021, pp. 1426–1427; Hess, 2009, p. 70; Miles, 1998, p. 143; Rangelov et al., 2019, pp. 104, 115; Shipman, 2008, pp. 97–98). It would be ideal for future work to more specifically address the role of linguistically 'mixed' marriages and compare this to the overall incidence of migration, but unfortunately this is not possible with the census data alone. A separate dataset of marriage practices for different areas would be needed, which could be appended to the census data and used to assess the role that such marriages play in promoting Bislama over Indigenous languages.

Further factors that seem related to Indigenous language use in rural areas include economic status, access to communication technologies, and population size. Taken together and with the factors already mentioned, this may suggest that Indigenous languages in rural Vanuatu are under pressure from a suite of factors related to economic development. While examples of economic development are implicated in language shift in other contexts (Amano et al., 2014; Thomason, 2015, pp. 21–23; though see Bromham et al., 2022), researchers in Vanuatu have been divided on whether greater access to development would be associated with increased use of Bislama over Indigenous languages. Crowley (1995, p. 340), for instance, argues that communities that experience being 'left behind' by development might be most likely to eschew Indigenous languages and adopt Bislama in order to 'catch up' with other communities. However, the census data show that it is generally people from economically higher status households who speak Indigenous languages less easily and who are more likely

to have learnt Bislama as a first language, while the opposite is true for individuals from economically lower status households. This aligns with Bislama's long association with economic development, having been introduced to Vanuatu by plantation labourers returning from Australia and developed further in the first half of the 20th century on plantations within Vanuatu (Crowley, 1990; Tryon & Charpentier, 2004), and with linguists' observations that Bislama is perceived as important for getting ahead in business and for upward social mobility (Guérin, 2008, p. 53; Rangelov et al., 2019, p. 118). A task for future work will be to test whether the associations observed here between decreased use of Indigenous languages and factors like access to communication technologies, population size, and migration 'wash out' or remain robust after controlling for economic status. For example, it is possible that by facilitating greater contact between communities, the arrival of novel communication technologies such as mobile phones and the internet directly contributes to language shift by making Bislama more important in daily life (Riehl, 2019; Vandeputte-Tavo, 2013b). Alternatively, factors including access to communication technologies and migration could be consequences of an economic change that represents the 'true' driver of language change. Further work is thus needed to elucidate the degree of robustness of the observed relationships to statistical controls.

Another interesting set of findings shows that some factors may not relate to language shift as they might be expected to. Given previous findings about the role of education in language shift in other contexts (Bromham et al., 2020, 2022), and the strong emphasis on the influence of education in linguistic and ethnographic reports on language endangerment in Vanuatu (Lavender Forsyth, 2025), I report surprisingly mixed results for the association between responses to the language questions and respondents' level of education. A clear association is present between someone's highest level of education and their learning of Bislama as a first language rather than an Indigenous language, a finding that accords with previous survey work which shows that more educated people use Indigenous languages less frequently in daily life and are more likely to have learnt Bislama as a first language (VNSO, 2021, p. 113). However, the relationship between someone's highest level of education and their competence in an Indigenous language is noticeably weaker. One interpretation of this pattern might be that access to education is linked to having learnt Bislama as a first language (as many observers would expect: Barbour, 2009, p. 232; Jauncey, 2011, p. 7; Nako, 2004; Tryon, 2006, pp. 107–108) but that it does not significantly erode competence in Indigenous languages. It is worth noting that this finding, if it withstands further scrutiny, might temper concerns that Vanuatu's education institutions are facilitating a widespread deterioration in Indigenous languages (e.g., Crowley, 2000, pp. 85–86; Daly & Barbour, 2021; Regenvanu, 2004). However, formal statistical tests are required to draw firm conclusions about the validity of this result. For example, further analyses should try controlling for respondent age, since I show above that this has a countervailing relationship with Indigenous language abilities. It seems unlikely that this pattern would be explained away by the introduction of government reforms to the role of Indigenous languages in the first three years of education from around 2012 onwards (see Early, 2023), given the median age for this sample is 36. It is harder to rule out the possibility that those with greater education are disinclined to admit they have lower competence at a skill like languages than those with less education, perhaps for status reasons. However, the fact that people with more education clearly do admit to having learnt Bislama as a first language more than others, despite longstanding perceptions that Bislama is 'broken English' and 'not a proper language' (Lynch, 1996; Vandeputte-Tavo, 2013a; Willans, 2017), goes some way to assuage this concern.

One of the most intriguing results is that people living in households with *kastom* land tenure show different responses to the language questions depending on whether they live in

an urban or rural area. As noted above, people in rural areas who live in households with *kastom* land tenure are slightly more likely to speak Indigenous languages easily and are slightly less likely to have learnt Bislama as a first language, compared to others that do not. An interesting connection here is that prior survey work finds that both these factors—Indigenous language competence and access to *kastom* land—positively contribute to subjective wellbeing (VNSO, 2021). However, in urban areas the opposite pattern in responses to the language questions exists, with people living in households with *kastom* land tenure being slightly less likely to speak an Indigenous language easily and slightly more likely to have learnt Bislama as a first language compared to others. One possibility is that in urban areas, living on one's *kastom* ground is confounded with higher economic status, since *kastom* landowners in Port Vila tend to be quite well off (Lindstrom et al., 2024). Future work employing multiple regression modelling in which various factors can be tested together would help to clarify whether this is the case.

Consistent relationships between religious groups and responses to the language questions are difficult to make out. A possible pattern may exist whereby, at least in rural areas, members of the Anglican and Catholic churches are slightly more likely to speak an Indigenous language easily and to have learnt it as a first language compared to members of other denominations, while members of the Seventh-day Adventist Church are slightly less likely to do the same. This trend, if it were to hold up under more stringent statistical testing, would not appear to be very strong, but it would align with observations by linguists and ethnographers that the 'evangelising' denominations, such as the Seventh-day Adventists, use Bislama more than established denominations like the Anglicans and Catholics (Crowley, 2000, pp. 65–66; François, 2012, p. 108; Hyslop, 2001, p. 6). Ultimately, further analysis is needed to confirm whether this pattern does withstand greater scrutiny, particularly when controlling for potentially confounding factors like geography, population size, economic development, and migration.

I do not find convincing evidence for a relationship between responses to the census' language questions and the number of Indigenous languages traditionally spoken on each island. This null finding could challenge arguments like those made by Crowley (2006) that linguistically homogenous islands, such as Paama, might be more resilient to processes of language endangerment than islands with more complex language situations, like Malekula. However, Crowley's argument primarily addresses the process by which smaller Indigenous languages are replaced by larger ones, with secondary attention to the part played by Bislama. Because the census data do not identify which Indigenous languages individuals speak, my analysis provides no evidence against the possibility that language endangerment is worse for more linguistically diverse islands through the mechanism of larger Indigenous languages displacing others. My findings are limited to the shift from Indigenous languages towards Bislama (or another language like English or French, which is not common). Analyses that address the question of language endangerment by means of a shift from one Indigenous language to another will thus have to wait for future work (see Duhamel, 2025, this issue).

One key lesson from prior work using census data to understand language endangerment, both in Vanuatu (Crowley, 1994) and across the world (Pauwels, 2016, pp. 35–46), is that careful consideration of the data-generating process is central to the interpretation of the statistics. To understand the 2020 data, two potential limitations are particularly important to consider. First, although the data are individual-level, it is not universally the case that each individual self-reports their own responses. Instead, if an individual is under the age of 15 or is unable to provide an answer for some reason, an answer is solicited from the main household respondent who is likely to be the household head. Future regression-based analyses should therefore control for the non-independence of responses at the household level. Second, it is

possible that individuals' language behaviours depart from the perceptions of the person that provides the data. These perceptions may differ from reality in non-random ways if there are social motives for giving certain answers. A common perception that Bislama is 'broken English' (Vandeputte-Tavo, 2013a; Willans, 2017) may make people less likely to report that they are first-language Bislama speakers and they may over-state their competence in Indigenous languages (see the discussion of education above). Conversely, if Indigenous languages are regarded as 'old fashioned' or less useful than Bislama in 'getting ahead' in the modern world (Rangelov et al., 2019; Vandeputte-Tavo, 2013b), this may encourage greater reporting of Bislama use over Indigenous languages. Responses to the language questions may thus reflect differing perceptions of these languages, rather than language behaviours directly. It is difficult to rule out this possibility with the current data alone. Future analyses could try to relate the census data to more 'objective' measures, such as linguists' assessments of the vitality of Indigenous languages in different areas of the country, and to direct measures of language attitudes, such as perceptions of the importance of speaking Bislama for being 'truly Ni-Vanuatu', as recorded in the nationally-representative *Pacific Attitudes Survey: Vanuatu* (Mudaliar et al., 2024). Nevertheless, the patterns in the data I report here go some way to support the interpretation that they reflect something real about language behaviours. That individuals' responses to the two language questions are related but distinct indicates they do not simply reflect a general preference towards Indigenous languages or Bislama. It is also probably more parsimonious that factors like a community's degree of home island diversity would be related to reduced use of Indigenous languages rather than merely more negative attitudes towards them. While remaining a possibility, therefore, I do not believe that the findings reported here should be explained away purely as differences in perception unrelated to real-world language behaviours.

As I have tried to emphasise, the patterns, and absence of patterns, that I present here should be taken as provisional descriptions of the data. Other trends may exist that are statistically reliable despite being small, or that only become apparent when statistically controlling for other factors. Correspondingly, it is possible that some trends described here would be altered or diminished by more sophisticated modelling techniques, or that the patterns change or 'wash out' after controlling for other variables. Future work could thus refine or overturn current findings. Nevertheless, merely exploring and describing the patterns within this expansive and versatile dataset, as done here, is a useful and worthwhile first step towards more focused statistical tests aimed at answering particular hypotheses about the role of specific factors in language shift in Vanuatu.

In particular, techniques of multiple regression would be useful to compare between different factors and quantify the relative strength of their associations with responses to the language questions. Various hypotheses put forward by linguists and ethnographers about the factors involved in language shift in Vanuatu would thus be amenable to testing with the 2020 census dataset. Furthermore, the existence of two different questions about language shift in the census provides an opportunity to test whether the factors that best predict reduced competence in Indigenous languages are the same as those that best predict increased learning of Bislama as a first language.

Yet, even employing regression-based analyses, correlation must not be mistaken for causation. We must therefore be careful not to over-interpret the correlational evidence I present here in making strong claims about the factors that cause people to shift from Indigenous languages towards Bislama. Nevertheless, in lieu of an idealised experimental setup in which people could be assigned to different conditions and the outcomes measured, longitudinal data do provide an avenue for testing causal hypotheses because they allow researchers to test whether change in one variable at a prior time step predicts change in another

variable at a subsequent time step (see Bulbulia, 2023). Future work could thus try to create longitudinal datasets of language shift in Vanuatu by linking area-level summaries of individuals' competence in an Indigenous language, or numbers of first-language Bislama speakers, to area-level summaries of the household-level language measures in the two previous censuses. There remains, therefore, a lot to learn from the census data about language shift and endangerment in Vanuatu.

5   Conclusion

Overall, the large and representative 2020 census dataset indicates that most people in Vanuatu can speak an Indigenous language easily and most learnt an Indigenous language as their first language. However, sizeable segments of the population do not speak an Indigenous language and learnt Bislama as their first language. These segments are considerably larger amongst younger people than amongst older people, consistent with a shift away from Indigenous languages over time. Patterns that emerge from the data indicate that Indigenous languages are strongest in Vanuatu's rural areas, in Area Councils with the smallest populations, and in the communities that are the least diverse in terms of people's island backgrounds. Individuals who have a history of migration (albeit only in rural areas), who live in economically better-off households, who have access to modern communication technologies, and who live in households without *kastom* land tenure (again, only in rural areas) are less likely to speak an Indigenous language easily and are more likely to have learnt Bislama as a first language. The 2020 census data thus show that Indigenous languages are vital and important in Vanuatu today but that they face serious challenges from a range of likely interconnected factors. Future work should build on the picture developed here by leveraging the census data to test explicit hypotheses – informed by these exploratory analyses and by the work of linguists and ethnographers – about which factors are most strongly implicated in the ongoing shift away from Indigenous languages and towards Bislama.

References

Amano, T., Sandel, B., Eager, H., Bulteau, E., Svenning, J.-C., Dalsgaard, B., Rahbek, C., Davies, R. G., & Sutherland, W. J. (2014). Global distribution and drivers of language extinction risk. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1793), 20141574. https://doi.org/10.1098/rspb.2014.1574

Barbour, J. (2009). Neverver: A study of language vitality and community initiatives. In M. Florey (Ed.), *Endangered languages of Austronesia* (pp. 225–244). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199544547.003.0013

Barbour, J., Wessels, K., & McCarter, J. (2018). Language contexts: Malua (Malekula Island, Vanuatu). *Language Documentation and Description*, *15*, 151-178. https://doi.org/10.25894/ldd142

Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., & Hua, X. (2022). Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution*, *6*(2), 163–173. https://doi.org/10.1038/s41559-021-01604-y

Bromham, L., Hua, X., Algy, C., & Meakins, F. (2020). Language endangerment: A multidimensional analysis of risk factors. *Journal of Language Evolution*, *5*(1), 75–91. https://doi.org/10.1093/jole/lzaa002

Bulbulia, J. (2023). *Causal diagrams (DAGS) for evolutionary human acience: A practical guide* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/b23k7

Choy, C. M., Co, M. K., Fogel, M. J., Garrioch, C. D., Leung, C. K., & Martchenko, E. (2021). Natural sciences meet social sciences: Census data analytics for detecting home language shifts. *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 1–8. https://doi.org/10.1109/IMCOM51814.2021.9377412

Crowley, T. (1990). *Beach-la-Mar to Bislama: The emergence of a national language in Vanuatu*. Clarendon Press.

Crowley, T. (1994). Linguistic demography: Interpreting the 1989 census results in Vanuatu. *Journal of Multilingual and Multicultural Development*, *15*(1), 1–16. https://doi.org/10.1080/01434632.1994.9994553

Crowley, T. (1995). Melanesian languages: Do they have a future? *Oceanic Linguistics*, *34*(2), 327. https://doi.org/10.2307/3623047

Crowley, T. (2000). The language situation in Vanuatu. *Current Issues in Language Planning*, *1*(1), 47–132. https://doi.org/10.1080/14664200008668005

Crowley, T. (2006). *Naman: A vanishing language of Malakula (Vanuatu)* (J. D. Lynch, Ed.). Pacific Linguistics.

Daly, N., & Barbour, J. (2021). 'Because, they are from here. It is their identity, and it is important': Teachers' understanding of the role of translation in vernacular language maintenance in Malekula, Vanuatu. *International Journal of Bilingual Education and Bilingualism*, *24*(9), 1414–1430. https://doi.org/10.1080/13670050.2019.1604625

Duhamel, M.-F., (2025). Do dominant languages affect linguistic diversity in Vanuatu? A call for further research. *Te Reo*, *68*(2), 114–132.

Dunnington, D. (2023). *ggspatial: Spatial Data Framework for ggplot2* (Version R package version 1.1.9) [Computer software]. https://doi.org/10.32614/CRAN.package.ggspatial

Dwivedi, P., Shraddha, C., Mathews, S., Majumder, S., Madhumathi, R., & Vasundhara, M. R. (2020). Predicting language endangerment: A machine learning approach. *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–7. https://doi.org/10.1109/ICCCNT49239.2020.9225576

Early, R. (2023). Social dynamics, public policy and language endangerment in Melanesia: An enduring enigma. In P. Brown & N. Gaertner-Mazouni (Eds), *Small islands, big issues: Pacific perspectives on the ecosystem of knowledge* (pp. 165–185). Peter Lang Verlag. https://doi.org/10.3726/b16720

Fishman, J. A. (1971). The sociology of language: An interdisciplinary social science approach to language in society. In J. A. Fishman (Ed.), *Advances in the sociology of language* (Vol. 1, pp. 215–404). Mouton.

François, A. (2012). The dynamics of linguistic diversity: Egalitarian multilingualism and power imbalance among northern Vanuatu languages. *International Journal of the Sociology of Language*, *214*, 85–110. https://doi.org/10.1515/ijsl-2012-0022

François, A., Franjieh, M., Lacrampe, S., & Schnell, S. (2015). The exceptional linguistic density of Vanuatu: Introduction to the volume. In A. François, S. Lacrampe, M.

Franjieh & S. Schnell (Eds), *The languages of Vanuatu: Unity and diversity* (pp. 1–21). Asia-Pacific Linguistics. https://halshs.archives-ouvertes.fr/halshs-01186007

Gray, A. (2012). *The languages of Pentecost Island*. Manples (BFOV) Publishing.

Guérin, V. (2008). Writing an endangered language. *Language documentation*, *2*(1), 47–67.

Hakkert, R., & Pontifex, S. (2022). *Vanuatu 2020 National Population and Housing Census: Analytical Report—Volume 2*. Vanuatu Bureau of Statistics and the Pacific Community. https://sdd.spc.int/digital_library/vanuatu-2020-national-population-and-housing-census-analytical-report-volume-2

Hawkins, R. J. (2005). Language loss in Guatemala: A statistical analysis of the 1994 population census. *Journal of Sociolinguistics*, *9*(1), 53–73. https://doi.org/10.1111/j.1360-6441.2005.00281.x

Hess, S. C. (2009). *Person and place: Ideas, ideals and the practice of sociality on Vanua Lava, Vanuatu*. Berghahn Books.

Hijmans, R. J., Barbosa, M., Ghosh, A., & Mandel, A. (2024). *Geodata: Download Geographic Data* (Version R package version 0.6-2) [Computer software]. rspatial. https://github.com/rspatial/geodata

Hyslop, C. (2001). *The Lolovoli dialect of the north-east Ambae language, Vanuatu*. Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.

Jauncey, D. G. (2011). *Tamambo: The language of west Malo, Vanuatu*. ANU Pacific Linguistics Press.

Kandler, A., Unger, R., & Steele, J. (2010). Language shift, bilingualism and the future of Britain's Celtic languages. *Philosophical Transactions: Biological Sciences*, *365*(1559), 3855–3864.

Kipp, S., & Clyne, M. (2003). Trends in the shift from community languages: Insights from the 2001 Census. *People and Place*, *11*(1), 33–41.

Krauss, M. (2008). Classification and terminology for degrees of language endangerment. In M. Brenzinger (Ed.), *Language diversity endangered* (pp. 1–8). De Gruyter Mouton. https://doi.org/10.1515/9783110197129.1

Lavender Forsyth, G. A. (2025). Language endangerment in Vanuatu: Bislama likely does pose a threat in the world's most language-diverse country. *Humanities and Social Sciences Communications*, *12*(1), 1462. https://doi.org/10.1057/s41599-025-05866-w

Lindstrom, L., Day, J., & Rousseau, B. (2024). Usufruct in the City: Tanna Migrants in Port Vila. *Cities*, *152*, 105203. https://doi.org/10.1016/j.cities.2024.105203

Lynch, J. D. (1996). The banned national language: Bislama and formal education in Vanuatu. In F. Mugler & J. D. Lynch (Eds), *Pacific Languages in Education* (pp. 245–257). Institute of Pacific Studies, University of the South Pacific.

Lynch, J. D., & Crowley, T. (2001). *Languages of Vanuatu: A new survey and bibliography*. Pacific linguistics, Research School of Pacific and Asian Studies, Australian National University.

Miles, W. F. S. (1998). *Bridging mental boundaries in a postcolonial microcosm: Identity and development in Vanuatu*. University of Hawai'i Press.

Mudaliar, C., Leach, M., Baker, K., Barbara, J., Vaha, M., Kotra, K. K., & Leodoro, T. (2024). *Pacific attitudes survey: Vanuatu (DPA Research Reports)*. Department of Pacific Affairs, Coral Bell School of Asia Pacific Affairs, The Australian National University. https://doi.org/10.25911/57VG-Q831

Nako, A. (2004). Vernacular language policy in Vanuatu. In K. Sanga, J. Niroa, K. Matai & L. Crowl (Eds), *Re-thinking Vanuatu education together* (pp. 40–43). Vanuatu Ministry of Education; Institute of Pacific Studies, University of the South Pacific.

Pauwels, A. (2016). *Language maintenance and shift*. Cambridge University Press. https://doi.org/10.1017/CBO9781107338869

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *The R Journal*, *10*(1), 439–446.

Posel, D., & and Zeller, J. (2016). Language shift or increased bilingualism in South Africa: Evidence from census data. *Journal of Multilingual and Multicultural Development*, *37*(4), 357–370. https://doi.org/10.1080/01434632.2015.1072206

Posit Team. (2024). *RStudio: Integrated Development Environment for R* [Computer software]. Posit Software. http://www.posit.co/

R Core Team. (2024). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rangelov, T., Bratrud, T., & Barbour, J. (2019). Ahamb (Malekula, Vanuatu)—Language contexts. *Language Documentation and Description*, *16*, 86–126. https://doi.org/10.25894/LDD114

Regenvanu, R. (2004). Educational initiatives of the Vanuatu Cultural Centre. In K. Sanga, J. Niroa, K. Matai & L. Crowl (Eds), *Re-thinking Vanuatu education together* (pp. 194–200). Vanuatu Ministry of Education; Institute of Pacific Studies, University of the South Pacific.

Riehl, A. (2019). Vanuatu: Indigenous language loss and the multiplying effects of climate change. In P. Grant & Minority Rights Group International (Eds), *Minority and indigenous trends 2019: Focus on climate justice* (pp. 128–132). Minority Rights Group International. https://minorityrights.org/programmes/library/trends/trends2019/vanuatu/

Shipman, T. S. (2008). *Wanem we mifala I wantem [what we want]: A community perspective of vernacular education in Vanuatu* [Master's thesis, University of Hawai'i at Manoa]. ScholarSpace. http://hdl.handle.net/10125/20824

Silver, B. D. (1975). Methods of deriving data on bilingualism from the 1970 Soviet census. *Soviet Studies*, *27*(4), 574–597. https://doi.org/10.1080/09668137508411026

Stevens, G. (1992). The social and demographic context of language use in the United States. *American Sociological Review*, *57*(2), 171–185. https://doi.org/10.2307/2096203

Takau, L. (2016). *A grammar of Nese* [Doctoral dissertation, University of Newcastle]. Open Research Newcastle. http://hdl.handle.net/1959.13/1322479

Thomason, S. G. (2015). *Endangered languages: An introduction*. University Press.

Tryon, D. T. (2006). Language endangerment and globalisation in the Pacific. In D. Cunningham, D. E. Ingram & K. Sumbuk (Eds), *Language diversity in the Pacific: Endangerment and survival* (pp. 97–111). Multilingual Matters. https://doi.org/10.21832/9781853598685-010

Tryon, D. T., & Charpentier, J.-M. (2004). *Pacific pidgins and creoles: Origins, growth and development*. Mouton de Gruyter.

Vandeputte-Tavo, L. (2013a). Bislama in the educational system? Debate around the legitimacy of a creole at school in a post-colonial country. *Current Issues in Language Planning*, *14*(2), 254–269. https://doi.org/10.1080/14664208.2013.837217

Vandeputte-Tavo, L. (2013b). New technologies and language shifting in Vanuatu. *Pragmatics*, *23*(1), 169–179. https://doi.org/10.1075/prag.23.1.08van

Vanuatu National Statistics Office. (2020). *Vanuatu 2020 national population and housing census: Basic tables report—Volume 1 (Version 2)*. https://sdd.spc.int/digital_library/vanuatu-2020-national-population-and-housing-census-basic-tables-report-volume-1

Vanuatu National Statistics Office. (2021). *Well-Being in Vanuatu: 2019–2020 NSDP Baseline Survey*. https://vnso.gov.vu/index.php/en/nsdp-baseline-survey

Walworth, M., Dewar, A., Ennever, T., Takau, L., & Rodriguez, I. (2021). Multilingualism in Vanuatu: Four case studies. *International Journal of Bilingualism*, *25*(4), 1120–1141. https://doi.org/10.1177/13670069211023132

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed). Springer International. https://doi.org/10.1007/978-3-319-24277-4

Willans, F. (2017). Grassroots talk back on social media: An analysis of public engagement in Vanuatu's language-in-education policy. *Current Issues in Language Planning*, *18*(4), 371–387. https://doi.org/10.1080/14664208.2017.1340407