Journal of Language Evolution, 2020, 75–91
doi: 10.1093/jole/lzaa002
Research article

Language endangerment: a multidimensional analysis of risk factors

L. Bromham,*,†,‡ X. Hua,†,‡,§ C. Algy and F. Meakins‡,††**

†Macroevolution and Macroecology, Research School of Biology, Australian National University, Canberra, ACT 0200, Australia, ‡Centre of Excellence for the Dynamics of Language, Australian National University, Canberra, ACT 0200, Australia, §Mathematical Sciences Institute, Australian National University, Canberra, ACT 0200, Australia, **Karungkarni Arts, via Katherine, Kalkaringi, NT 0852, Australia and ††School of Languages and Cultures, University of Queensland, Brisbane, QLD 4072, Australia

*Corresponding author: Lindell.Bromham@anu.edu.au

Abstract

The world is facing a crisis of language loss that rivals, or exceeds, the rate of loss of biodiversity. There is an increasing urgency to understand the drivers of language change in order to try and stem the catastrophic rate of language loss globally and to improve language vitality. Here we present a unique case study of language shift in an endangered Indigenous language, with a dataset of unprecedented scale. We employ a novel multidimensional analysis, which allows the strength of a quantitative approach without sacrificing the detail of individual speakers and specific language variables, to identify social, cultural, and demographic factors that influence language shift in this community. We develop the concept of the 'linguatype', a sample of an individual's language variants, analogous to the geneticists' concept of 'genotype' as a sample of an individual's genetic variants. We use multidimensional clustering to show that while family and household have significant effects on language patterns, peer group is the most significant factor for predicting language variation. Generalized linear models demonstrate that the strongest factor promoting individual use of the Indigenous language is living with members of the older generation who speak the heritage language fluently. Wright–Fisher analysis indicates that production of heritage language is lost at a significantly faster rate than perception, but there is no significant difference in rate of loss of verbs vs nouns, or lexicon vs grammar. Notably, we show that formal education has a negative relationship with Indigenous language retention in this community, with decreased use of the Indigenous language significantly associated with more years of monolingual schooling in English. These results suggest practical strategies for strengthening Indigenous language retention and demonstrate a new analytical approach to identifying risk factors for language loss in Indigenous communities that may be applicable to many languages globally.

Key words: bilingual education; Australian languages; language shift; language variation and change; Wright–Fisher

1. Introduction

Over 7,000 languages are currently recognized worldwide, but over 6% are no longer spoken, and between one-third and one-half of all languages are considered endangered (Austin and Sallabank 2011; Evans 2011; Rehg and Campbell 2018; Romaine 2017; Sutherland 2003). Estimates of the rate of ongoing language loss are roughly equivalent to the loss of one language every fortnight (Crystal 2000), with the potential for the majority of languages to cease being spoken within a century (Krauss 1992). The heaviest areas of language loss have been in Australia, Canada, and the USA, where more than 70% of languages are highly endangered or no longer spoken (Loh and Harmon 2014; Simons and Lewis 2013). Australia has suffered catastrophic rates of language loss: fewer than fifteen of the more than 300 distinct Australian Indigenous languages are still actively acquired by children (Simpson and Wigglesworth 2019). Urgent intervention is needed to maintain and revitalize the world's remaining language diversity (Crystal 2000; Krauss 1992). To best focus our efforts to stem the tide of language loss, we need to know which factors present the greatest threat to minority languages, and which aid in their retention within speech communities. In particular, we want to examine factors that are open to adjustment or manipulation in real-world situations, such as exposure to heritage language and aspects of formal education.

Studies of language endangerment have typically been either quantitative analyses focused on correlates of language endangerment at regional or global levels (e.g. Amano *et al.* 2014; Sutherland 2003), or qualitative studies of social factors influencing language shift in specific communities (Austin and Sallabank 2011; Brenzinger 1992; Crystal 2000; Dorian 1981; Hinton 2011; Krauss 1992; Meakins and O'Shannessy 2016; Rehg and Campbell 2018; Sasse 1992; Simons and Lewis 2013). Here we develop a new approach, combining the strength of a quantitative approach and a large dataset with the advantages of a sociolinguistic focus on specific language variables and individual language use, in order to examine the process of language change within an Indigenous population. This multidimensional approach allows the detection of general trends across many linguistic features. By doing so, we can identify microlevel processes within speaker communities which allow us to make predictions about risk factors at the macrolevel of language endangerment, and potentially support effective interventions to support language vitality.

To understand processes of language shift and language endangerment, we need to consider multiple levels of language use. For example, Sasse (1992) identifies three levels of factors relevant to language loss: external setting (the cultural, sociopolitical, historical processes that result in pressure on a speaker community to shift language), speech behavior (sociolinguistic aspects of changing variable use), and structural consequences (changes in phonology, morphology, syntax, and lexicon within an endangered language). A full explanation of language loss should ideally draw on research at all of these levels (Fishman 2002; Sasse 1992). More generally, there is wide agreement that patterns of language change require both synchronic studies of variation and diachronic studies of language shift (Labov 1994a,b; Meyerhoff 2009). In many ways, this is similar to the approach that biologists take to understanding the evolution of biodiversity and the patterns of species endangerment: studies of current patterns of genetic variation are used both to understand the processes of diversification and to assess population vulnerability (Allendorf 2017; Avise 2009).

While there are key differences in the way that languages and species change over time, and in the types of processes that cause vulnerability and loss, some broad concepts are similar between biology and linguistics and these shared concepts may help inform new modes of analysis (Bromham 2017). The uniformitarian principle is widely accepted in both linguistics and evolutionary biology: that changes within and between lineages (languages or species) are ultimately driven by the processes of variation and change that can be observed operating within populations (Christy 1983). Yet these processes of change are typically studied either at the 'micro' level (change in specific language variables, individual variation between speakers, or sociolinguistic patterns within populations) or at the 'macro' level (language diversification, shift, and loss). Evolutionary biology faces a similar disjunct between studies of population genetic variation and the processes of mutation (the causes of differences between individuals), selection and drift (which influence patterns of variation within populations), and speciation and extinction (which determine the fates of species and lineages). In biology, DNA sequences provide a way of linking these levels together because the same genes can be compared between individuals, among populations, and between species (Hua and Bromham 2017). But it has thusfar been difficult to develop a similar framework for a micro-to-macro analysis of language variation because the data typically recorded at individual and population level (in sociolinguistic studies) rarely corresponds directly to the language variation studied at the lineage level (exemplified by historical linguistics).

When biologists wish to characterize the variation present in a population to explain the history of change and potentially to evaluate capacity for future change, they sample many individuals from the population and identify genetic variables that show variation within the population, whether these are alternative DNA sequences of particular genes, protein variants, microsatellites (short DNA repeats sometimes used in 'DNA fingerprinting'), or any other form of genetic variant (e.g. Allendorf 2017; Avise 2009; Luikart *et al.* 2003). If many genetic markers are sampled, without specifically targeting traits likely to be actively shifting (changing under selection) in this population, they give an unbiased view of rates and patterns of change. With sufficient variable sites from across the genome, estimates of both individual variability and population levels of variation can be made and specific groups of variables can be compared to contrast their patterns of variation and change. For example, many studies of genetic variation seek to identify single nucleotide polymorphisms, variable sites in the genome at which not all individuals have identical DNA sequences (e.g. Morin, Luikart and Wayne 2004). At each of these identified variable sites, individuals in the population may carry one or two of multiple variants for that locus. The 'genotype' of an individual is the set of variants it carries at all of the sampled variable sites—it is a representation of the individual's genetic makeup, not a complete description of its genome.

Surveying variants in individuals sampled from the population provides a baseline for change across the genome against which to compare the patterns and rates of change at specific sequences in order to detect significant patterns. Examples range from identifying genes under selection in particular environments (e.g. Evans *et al.* 2014), determining the relative contributions of different lines of ancestry of individuals and populations (Rius and Darling 2014), documenting the response of populations to changing pressures (e.g. Agrawal *et al.* 2012), tracing population history (e.g. Moreno-Estrada *et al.* 2013), reconstructing large-scale patterns of change and diversification (e.g. Coulson *et al.* 2011), and predicting future evolutionary potential (e.g. Harrisson *et al.* 2014). In this way, the genotype allows patterns of variation at the microlevel to be connected to change at the macrolevel, connecting the origins of new variants to variation within populations to the diversification of species and lineages. Although genotyping nuclear markers (inherited from both parents) or mitochondrial markers (typically maternally inherited) may be used to trace parentage or ancestry, the concept of genotyping is not restricted by the mode of inheritance of the markers. For example, we could also genotype an individual's microbiome (which could be inherited from parents, siblings, unrelated individuals or via the environment) or its virome (which could even be 'inherited' from other species such as domestic animals or wildlife). These genetic variants could change over the course of an individual's lifetime, but we could still genotype the variants present at a given point in time.

The approach to genotyping—recording a large number of variables that are known to show variation within the population but without regard to their specific patterns of change—has not yet been applied in linguistics. Instead, sociolinguistic studies tend to focus on one or few variables selected specifically because they vary with some factor of influence, such as age or class (Guy and Hinskens 2016; Labov 1994a,b; Meyerhoff 2010; Tagliamonte 2006). But with sufficient data, an analogous approach can be developed by selecting linguistic variables known to show variation within the population, sampling speakers across many different demographic or sociocultural sections of the society, and recording for each speaker which variants they use in their linguistic repertoire. Recording which language variants an individual uses for each language variable allows us to describe their 'linguatype', analogous to an individual's genotype. A genotype does not describe the whole genome, but uses a representative sample of genetic markers that is known to vary within the population in order to characterize genetic variation of individuals, for example, to describe the variation attributed to different ancestral populations. In the same way, the linguatype uses a broad sample of language elements that vary within the population to characterize an individual's pattern of language use, such as the use of elements from different source languages. If a sufficient number of variables are chosen only because they are known to have multiple forms in the population, not because they have a known pattern of change or association within the speaker community, then the linguatype provides an unbiased sample of the 'idiolect' (an individual's characteristic pattern of language use) at a particular point in time. This allows us to study the process of the incorporation of language elements from one language into another, such as large-scale borrowing that occurs during the ongoing attrition of a heritage language in situations of language shift. A large sample of variables also allows patterns to be contrasted between different language elements, such as nouns and verbs, or tracked between different kinds of language modalities, such as comprehension versus production.

Here we demonstrate how the 'linguatype' sampling strategy can be used to integrate analysis of language variation and change across micro- and macrolevels, using an analysis of an endangered Indigenous Australian language. The Australian situation presents an especially informative focus for studies of language shift and loss. Much of Australia's rich linguistic heritage has undergone rapid change since the arrival of English-speaking colonists from the eighteenth century onward, and the processes of language loss and endangerment are ongoing (Loh and Harmon 2014; Simpson and Wigglesworth 2019). Here we focus on one highly endangered Indigenous language (Gurindji: Ngumpin-Yapa, Pama-Nyungan: Fig. 1) which is undergoing rapid contact-induced shift. A mixed language, drawing elements from Gurindji and Kriol (an English-based creole), is now the main language of the younger generations in this community, with an ongoing pattern of replacement of Gurindji language elements with Kriol (McConvell and Meakins 2005; Meakins *et al.* 2019). School-based education is conducted entirely in English. However, English is rarely used outside the classroom, restricted to interactions with local government and in some employment settings.

We scale-up the classic Labovian paradigm of modeling language variation and change (Labov 1966) to multiple dimensions by analyzing 185 variables including lexicon and grammar, both in production and comprehension, drawn from 96 hours of language recordings of seventy-eight speakers, representing 15% of the entire speech community (Meakins *et al.* 2019). For each individual, we have information on key social factors, including age, family relationships, household, location, and education level. We analyze all language variables in a multidimensional framework, allowing us to detect general trends of language change without losing the ability to trace the fate of specific language elements or identify individual speakers.

2. Data

2.1 Data collection

This study builds on a recently published language corpus, based on recordings of people from the Gurindji communities of Kalkaringi and Daguragu (Fig. 1), including seventy-eight speakers, which represent approximately 15% of the current Gurindji community (Meakins *et al.* 2019). The Gurindji language is undergoing rapid shift: while the older generations speak both Gurindji and Kriol (an English-lexifier creole widely spoken in Indigenous communities across Northern Australia), younger generations speak neither language fluently, but instead use a mixed language which consists of both Gurindji and Kriol elements, as well as some innovative features found in neither source language. English is the language of formal education but is rarely spoken in these communities outside the school setting (Meakins 2008).

The corpus consists of 96 hours of recordings of a range of tasks including free narratives, conversation, procedural descriptions (such as how to make artifacts or prepare bush foods), picture-prompt narratives, and director-matcher tasks (Meakins *et al.* 2019). The picture-prompt narratives and director-matcher tasks are conducted between peers from the speaker community (Meakins, Green and Turpin 2018), to avoid adjustment of language usage to accommodate a non-native speaker (i.e. the linguist). Peers are matched by generation to ensure the speaker does not accommodate a Gurindji person of another generation (e.g. a member of an older generation might use more Kriol when speaking to a person from a younger generation). Elicitation tasks included narrative descriptions using picture books and director-matcher tasks (see Meakins *et al.* (2019) for details).

The dataset consists of 185 language variables, with 130 production variables (recorded use of a language variant by a speaker) and 55 comprehension variables (experimentally determined understanding of a language variant by a speaker). Within production variables, forty are grammar and ninety are lexicon, and, of these, ninety are nouns and forty are verbs. See Supplementary Table S1 for a full list of variables, and all 356 variants analyzed.

The production variants (V1-V130) are drawn from a 96-hour corpus of audio and video recordings which consist of free narratives, conversation, procedural descriptions (e.g. how to make artifacts, how to collect and prepare bush foods, etc), picture-prompt narratives (e.g. Frog stories (Mayer 1969)) and picture-prompt sentences which were elicited through a series of director-matcher tasks targeted at particular linguistic variables (Meakins 2011) (see Supplementary Video). The corpus is held in PARADISEC (the Pacific And Regional Archive for Digital Sources in Endangered Cultures), which is a digital repository for language recordings. For each variable, we record whether each speaker uses Gurindji variants and/or non-Gurindji variants (Supplementary Information Table S2). The non-Gurindji variants are typically from Kriol, but we also include some innovative forms in the 'non-Gurindji' category (Meakins *et al.* 2019). These innovations are often a novel blend of Gurindji and Kriol forms. We also

Figure 1. Map of northern Australia showing the location of the two Gurindji communities, Kalkaringi and Daguragu.

record production variables, which record whether a speaker can give the Gurindji term for a concept or not. For example, variables V120-V130 require speakers to say a word indicating a direction (e.g. North, Darwin: see Supplementary Table S1). There is no Kriol variant for 'north', so using the Gurindji word (*kayirra*) is recorded as 1, otherwise is recorded as 0 (null). In Section 3.3, we describe how we have adapted the Wright–Fisher (WF) method to analyze these kinds of production variables that have only one variant.

The comprehension variables (V131-V185) are drawn from experimental data from a number of 4AFC (alternative forced choice) tasks where participants hear a Gurindji word and choose from four pictures the one they believe matches the audio stimulus best (see Meakins and Wigglesworth (2013) for more information on methods). Spatial comprehension variables come from dead reckoning tasks (see Meakins and Algy (2016) for more information on methods). Each comprehension variable records whether a speaker understands a Gurindji form, so by definition, each comprehension variable has one Gurindji variant.

There are several important strengths of this approach. First, unlike recording free conversation, the elicitation tasks ensure that every speaker has multiple opportunities to produce each language variable. Therefore, we have information for each one of the seventy-eight speakers for every one of 185 variables, and we can be confident that failure to use a variant is indicative of its absence from the linguatype of that speaker—that is, this variant is not a feature of their current linguistic repertoire, even if they have used it in the past or might adopt that variant in future (see Supplementary Table S2). This illustrates that the linguatype is a sampling strategy that provides a description of the individual's language usage pattern at the time of sampling, rather than an attempt to characterize their lifetime repertoire or 'idiolect'. The linguatype approach does not require us to assume that the individual's idiolect was formed in childhood and has remained essentially unchanged. Instead, the linguatype represents an informative but unbiased sample of language variation from one individual at a particular point in time.

Second, the 185 features are chosen because they are known to have multiple variants in the speaker community, but not because of any specific pattern of change in those variables. This is equivalent to the identification of polymorphic loci in population genetic studies, which aim to find genetic traits known to vary within communities in order to track variation and change over space and time (Bromham 2016), but differs from the classic sociolinguistic approach of focusing on variables known to be undergoing a specific trajectory of change (Labov 1994a,b). A speaker's linguatype can change over their life span, so the linguatype represents a snapshot of a speaker's idiolect at the time of sampling. Note that linguatype sampling strategy is agnostic with respect to the mode of transmission of linguistic variants, for example, whether they are learned from parents or shared among unrelated peers. Indeed, this approach can be used to contrast the relative strengths of 'vertical transmission' (from parents and grandparents) to 'horizontal transmission' (from peers or unrelated speakers).

Third, by sampling a large number of language variables, we have sufficient statistical power to track influences of peers, family, household, and external cultural factors (such as education and language exposure) on individual patterns of language use, and to examine differential rates of change in grammar and lexicon, nouns and verbs, and production and comprehension variables.

2.2 Social, demographic, and cultural factors

For each individual in this study, we record the following social, demographic, and cultural factors:

- *Generation*: Age is used to divide participants into three generations: Generation 1 ('Grandparents': >40 years old), Generation 2 ('Parents': 20–40 years old); and Generation 3 ('Children': 8–14 years old). Note that these labels do not imply that everyone in Generation 1 has grandchildren, or that all 20- to 40-year olds are biological parents, but we use these as culturally appropriate labels that reflect social and classificatory relationships within this community.
- *Family*: All participants are members of ten extended families, identified by a participant's lineage across the three generations. Two participants could be included in more than one family, but for the purposes of the analysis they were assigned to the family they most closely identify with.
- *Household*: All participants are divided into thirty-one residential households according to their primary place of residence at the time of data collection. Note that membership of household and family have a degree of overlap, but they are not identical: family is defined by pedigree and household by co-habitation.
- *Community*: Households are clustered into two communities, Daguragu and Kalkaringi, which are the only permanent Gurindji settlements (8 km apart: Fig. 1). Community is a higher-level category of spatial clustering, used in analyses where we do not wish to include all households as separate factors.
- *Level of education*: We define three levels, according to years of formal school-based education (which has always been conducted in English in these communities): Low = up to third grade of primary school; Medium = completion of primary school and up to fourth year of high school; High = senior school or tertiary. The binning is based on common educational experience in this community. The Low category is common for people 40+ years old (Generation 1), as this was the extent of primary education available for many older people. Many members of Generation 2 have completed primary school and up to fourth year of high school. Most Generation 2 individuals have not completed all 6 years of high school, although some have tertiary education. Because many members of the Generation 3 (children) have not yet completed their schooling, we analyze generations separately when we are looking for correlates of education level. We use these three levels for the discriminant correspondence analysis (DCA), but we combine these into two levels for the generalized linear mixed model (GLMM), in order to avoid having too

many groupings in the analysis with zero speakers. In the GLMM we combine medium and high, referring to this category as 'high'.

- *Exposure to heritage language*: We categorize participants by their level of exposure to a fluent Gurindji speaker. All members of Generation 1 (Grandparent generation) are fluent Gurindji speakers, but members of Generations 2 and 3 speak the mixed language and are not fluent in Gurindji. Therefore we rank exposure by the amount of time each speaker has spent co-habiting with a fluent Gurindji speaker from Generation 1: Low = has never lived with a Generation 1 person; Medium = lived with a Generation 1 person most of their childhood; High = has lived with a member of Generation 1 most of their life. Members of Generation 1 do not vary in their exposure level so these analyses are only conducted on Generations 2 (Parents) and 3 (Children). As for education level, we use three levels for the DCA (low, medium, high), but two levels for the GLMM (low, high + medium), again to avoid having too many groups in the analysis with zero speakers.

The distribution of speakers between age classes, exposure levels, and education levels is summarized in Fig. 2.

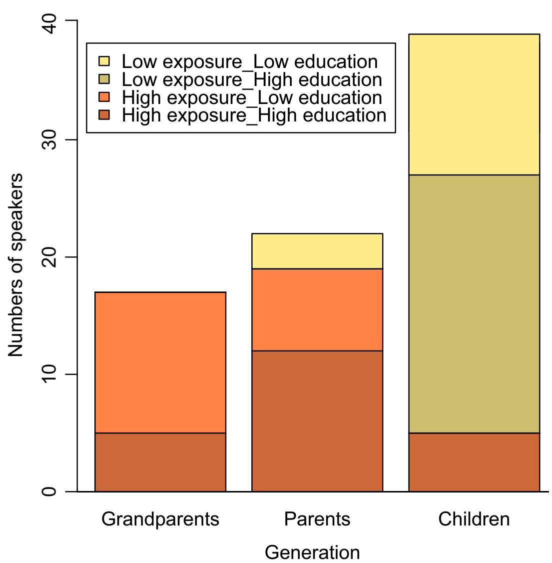

Figure 2. Number of speakers in each generation with each level of education and exposure to heritage language. The levels are as defined for the GLMM analysis, high (= medium + high) and low: see Section 2.2 for details.

3. Analysis and results

We perform three different forms of analysis, each of which asks a different, complementary question about language variation and change in the Gurindji speech community.

3.1 Discriminant Correspondence Analysis (DCA)

We use DCA (Williams *et al*. 2010) to ask if linguatypes (individual language patterns) cluster by social, demographic, and cultural factors. We calculate the *R*-squared as the percentage of the variation between linguatypes that can be explained by the factors of interest: generation, family, household, education, exposure (see Section 2.2). As the *R*-squared increases from 0 to 1, more variation in linguatypes between individuals is explained by the factor that defines group membership, indicating that linguatypes are clustered together into groups defined by shared social, demographic, and/or cultural factors. To test the statistical significance of clustering in linguatypes, we generate a null distribution by randomly assigning speakers to factors (1,000 permutations). All the factors have higher R-squared values than the 1,000 permutations, suggesting that they have a statistical significance of clustering in linguatypes at significance level 0.001. We also apply *F*-tests on each factor to approximate P-values. Because different generations have different average levels of key factors such as education and exposure, we perform DCA on education and exposure separately for each generation (Table 1).

The DCA analyses show that the factor responsible for the strongest clustering effect in grouping similar linguatypes is 'Generation': the first principal component (PC) separates speakers by generation, rather than education level and level of exposure to heritage language (Fig. 3). Within each generation, speakers tend to cluster with other individuals with the same level of education or exposure to heritage language. When we group speakers by both generation and education, 92% of the language variation between those groups is accounted for by different generations and 8% of the variation by levels of education. When we group speakers by both generation and exposure to Gurindji, 85% of the variation between groups is accounted for by generation and 15% of the variation by exposure.

3.2 Generalized linear mixed model (GLMM)

The DCA analysis asks which social, demographic, and cultural factors capture overall variation in language use between individuals, clustering linguatypes into groups. Next we use GLMM (Hadfield 2010) to look at the

Table 1. Factors influencing individual language patterns (linguatype)

	Generation	Family	Household	Education level			Language exposure	
				Grandparents	Parents	Children	Parents	Children
R^2_{Adj}	0.89	0.41	0.46	0.84	0.58	0.63	0.73	0.70
P	9×10^{-38}	5×10^{-7}	4×10^{-4}	7×10^{-7}	1×10^{-5}	4×10^{-9}	3×10^{-7}	7×10^{-12}

Notes: DCA, with *R*-squared value for each factor adjusted by the number of levels in each factor, for three generations within the speaker population: Grandparents (>40 years old), Parents (20–40 years old), and Children (8–14 years old). P-values are based on *F*-tests against a null distribution generated by 1,000 permutations of randomizing individuals to factors.

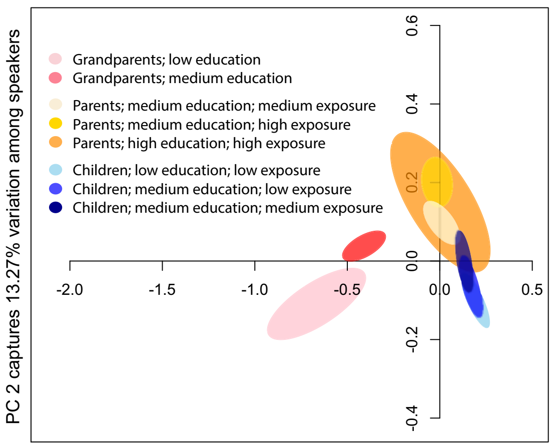

Figure 3. DCA for groups defined by generation, language exposure, and education level. We plot the range that contains 95% speakers from each group in PC space according to their linguatypes. PC1, which separates different generations, produces the greatest separation between groups, suggesting that there is much larger variation between generations than among levels of education or levels of exposure to heritage language in the speaker community. Note that there is no variation in exposure in the Grandparent generation who are all fluent Gurindji speakers.

effect of the social, demographic, and cultural factors on changes in language use across all language variables, analyzed at the level of patterns of variants used for each language variable. We ask whether these factors predict relative frequencies of patterns of language variants. To do this, we first describe the patterns of variant usage that are observed in the linguatypes of the population. Each speaker must have at least one variant of each language variable, but they may have more than one. For example, an individual may use Gurindji and Kriol variants for a variable, or multiple Gurindji variants, or multiple Kriol variants, or innovative forms, or any combination of the available variants.

In this study, we are specifically interested in identifying factors that are associated with inclusion of Gurindji variants in the linguatype. To do this, we express the frequency of patterns of variant use with respect to a reference pattern, so that we can identify which patterns of use have more or fewer Gurindji variants than the reference pattern. Comparison to a reference pattern is also required to account for the dependency of frequencies of patterns: since each individual is characterized by one pattern of variants, then, by necessity, a change in frequency of one pattern in the population must cause a change in frequency of other patterns. The reference pattern is used to scale the frequency of patterns to allow us to detect relative changes in frequency. To compare patterns of language usage among groups defined by the factor of interest, we need to ensure that the reference pattern has a nonzero frequency in each of the groups, so we select the reference pattern as the most common pattern in the group with the fewest incidences (e.g. only five speakers from Generation 3 have both high education level and high exposure level to heritage language, so the most common pattern in this group is set as the reference pattern).

Now we can use GLMM to estimate the effect of each factor on the frequency of patterns of language use relative to the reference pattern, specifically focusing on patterns with more or fewer Gurindji variants than the reference pattern. In each GLMM, the dependent variable is the language variable, and family is a random effect (because we have no expected direction of influence of family on linguatype). More specifically, the dependent variable is a vector of the frequencies of all patterns of variant usage in the language variable relative to the reference pattern and each pattern of usage is marked as having more or fewer Gurindji variants than the reference pattern. To avoid too few incidences in each level of a factor, we reduced the level of education to two levels (high + medium and low) and the level of exposure to Gurindji to two levels (high + medium and low), and, rather than including each household in the analysis, households are grouped into the two communities (Kalkaringi and Daguragu). We first ask how speakers from different generations use variants differently for each variable, by defining generation and community as fixed effects (because there are only two communities and therefore only a single coefficient to describe community effect on a pattern of usage in a language variable, it does not matter whether community is treated as a fixed effect or a random effect). We then ask how speakers from the same generation, but with different education levels and exposure to Gurindji, use variants differently for each variable, by defining education level, exposure to Gurindji, and community as fixed effects.

For each fixed effect, we calculated a z-value from the posterior distribution of its coefficient, so the effect of each factor is comparable across different language variables. If a factor has no significant influence on the language variant patterns for a particular language variable, so that each pattern is just as likely to increase or decrease in frequency relative to the reference pattern, we then expect the z-value of the factor on each pattern to follow a standard normal distribution, therefore the significance test on the factor is whether its z-value is larger than 1.96 or smaller than -1.96. This allows us to test for a significant impact of factors on each variable. But it also allows us to look for significant trends across all variables. For all the patterns over all the language variables, the sum of their z-values will follow a normal distribution with mean equal to zero and variance equal to the number of patterns. Using this normal distribution as the null distribution, we can test if a factor has significant influence on the frequency of patterns of use of Gurindji variants over all the language variables. Figure 4 shows that individuals in the Parent and Children generations tend to use patterns with fewer Gurindji variants than individuals of the Grandparent generation. This result indicates decreasing inclusion of heritage language elements over generations.

Figure 5 shows that, for the majority of variables, patterns of use with relatively more Gurindji variants are associated with lower levels of formal education and greater exposure to the older generation of fluent Gurindji speakers. Most speakers in the Parent generation have high exposure to Gurindji, and most Children have low exposure to Gurindji (Fig. 2), so our estimated effects of education on members of the Parent and Children generations should be independent to the exposure effect. In addition, all Parents with low exposure to Gurindji also have a low education level, and all children speakers with a high exposure to Gurindji also have a high education level. This suggests that the estimated associations between education and exposure and the retention of heritage language variants are not driven by collinearity between education and exposure levels, because if they were, then we would expect education and exposure to be associated with the retention of heritage variants in the same direction.

For the family factor, we calculate the intraclass correlation of the random effect that describes how often speakers from the same family have the same frequency of each pattern compared with speakers from the other families. The significance test on the random effect is based on a randomization test with 1,000 permutations, where speakers are randomly assigned to families, to generate a null distribution for comparison. We find that language variant patterns that are significantly associated with generation (Fig. 4) are often not significantly influenced by family. This result suggests that the family effect does not overwhelm similarity among peers in patterns of language variants.

3.3 Wright-Fisher (WF) analysis

The first analysis (DCA) tells us that demographic, social, and cultural factors have significant effects on individual patterns of language use in the Gurindji speaker community. The second analysis (GLMM) tells us, more specifically, that inclusion of relatively more Gurindji variants in an individual's language usage patterns (as represented by the linguatype) is significantly associated with more years living with elders and fewer years of formal education in English, across all generations. Now we wish to ask which aspects of language have greater rates of retention or loss.

We extend our previous work in applying the WF model to language change (Meakins *et al.* 2019). The WF model describes a serial sampling process from one timestep to the next (Fisher 1956; Wright 1931). In this case, speakers in one generation sample the language variants they hear in the previous generation's usage. Modeling this series of sampling processes over generations allows us to describe the expected degree of fluctuation in the frequency of patterns due to sampling error alone. The value of this model is that it allows us to generate an expected variation in frequencies of variants over time just due to random sampling processes; that is, if the frequency of different language variants fluctuates randomly, without preference to any particular variant. This provides a null distribution to compare the observed change in frequencies over generations to identify significant trends that are unlikely to be due to random sampling effects. Now we can consider which kinds of variants have a greater rate of retention (maintaining their frequency in the population over generations) or adoption (rising in frequency in the population over generations).

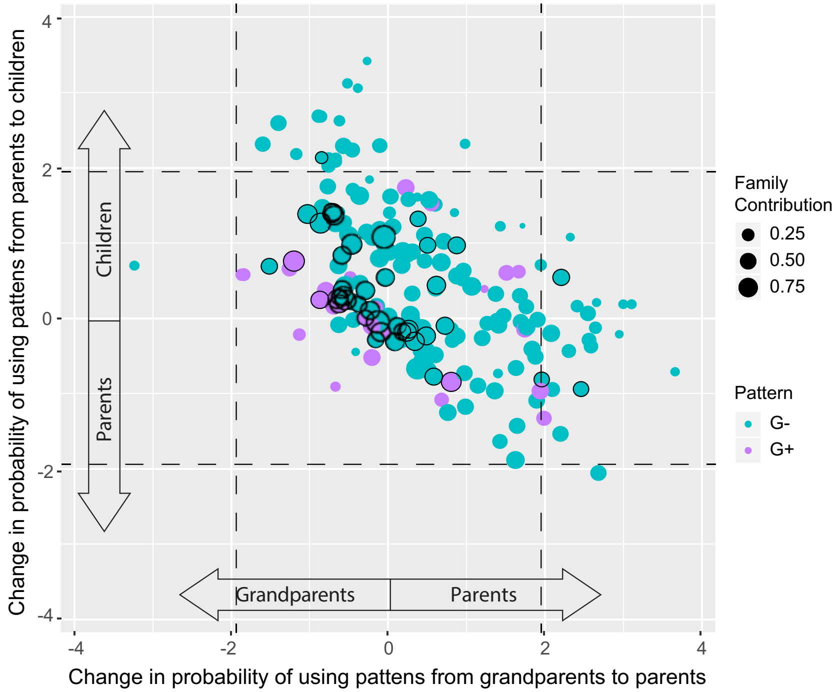

Figure 4. Results of GLMM on frequency of usage of Gurindji variants across generations. Each data point is a pattern of language use that has more (G+) or fewer (G−) Gurindji variants than the reference pattern. For data points with *x*-axis ≥1.96 (the right vertical dashed line), parents use the pattern significantly more often than grandparents. For datapoints with *y*-axis ≥1.96 (the top horizontal dashed line), children use the pattern significantly more often than parents. The size of each data point represents the degree of family contribution to the pattern. Family has significant effect on data points with a black outline. Most of these data points are in the centre of the plot, not significantly influenced by generation. In general, there is an increase in patterns of use with fewer Gurindji variants and so a decrease in inclusion of Gurindji variants over generations.

A key parameter of this model is the rate of retention of a class of language variants, representing the relative rise or drop in frequency of a class of variants from one generation to the next (Meakins *et al.* 2019). While we might naturally think of maintenance of Gurindji variants as 'retention' and rise in use of Kriol or Innovative variants as 'adoption', these patterns are both modeled as the rate of change in frequency of variation over generations, so represent the same parameter in the WF analysis, which we will refer to as retention rate.

We improved our previous application of the WF model in two ways. First, our previous analysis described changes in the frequency of Gurindji (G), Kriol (K) and innovative (I) variants over generations, so we fitted the probability of speakers from each generation using each variant of a language variable to a Dirichlet distribution (Meakins *et al.* 2019). But here we are specifically testing three hypotheses: two hypotheses contrast change in frequency of Gurindji variants in different features, and one hypothesis concerns the uptake of contact-driven language variants (K or I). In each case, we need to only consider changes in frequency of one type of variant (either G variants or K + I variants). This simplifies our analysis because we only need to calculate the probability of using one type of variant, rather than calculate the probability of using each variant in a language variable. So we can make our analysis more efficient by using the marginal distribution of the Dirichlet distribution for the type of variants of interest.

Our second improvement is to address a limitation of the previous model, which could not be applied to language variables for which some speakers have zeros for all available variants. This is particularly a problem for comprehension variables, because each comprehension variable records whether a speaker knows the Gurindji form of a language variable (1) or not (0). So unlike most production variables, where a participant must use at least one variant, the comprehension variables can have a null variant (0). This issue also applies to some production variables which, instead of using alternative G or K variants, can shift to a different part of the communicative repertoire. For example, production variable

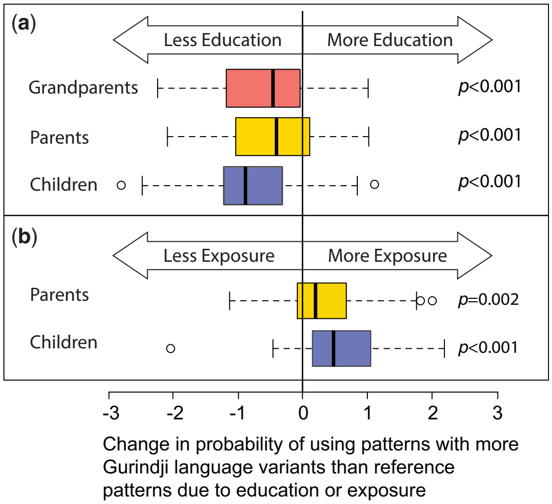

Figure 5. Effect of education and exposure on retention of Indigenous language elements in the linguatype. Boxplots of *z*-values from GLMM on 185 language variables, showing changes in the frequency of patterns of usage that have more Gurindji variants than a reference pattern (see Supplementary Information for details). Patterns with positive z-values suggest that speakers with a higher education level or a higher exposure level use the patterns more often, and vice versa. Here the z-values are positive for exposure showing that speakers with a high exposure level use G+ pattern more often (see Fig. 4). Conversely, the *z*-values are negative for education level showing that speakers with a high education level use G+ pattern less often. P-value is the probability that the sum of *z*-values follows a normal distribution with mean zero and variance equal to the number of *z*-values, which is expected if there is no directional influence of (**a**) number of years of formal schooling or (**b**) living with a member of grandparent generation on frequency of patterns with more Gurindji variants. Patterns of use with more Gurindji elements are significantly more common in individuals with lower education levels across all three generations (a), and in individuals with higher exposure to fluent speakers (living with a member of the grandparent generation) in the parent and child generations (b). Out of the patterns that have different number of Gurindji variants to the reference patterns, there are 50 (47%) patterns with more Gurindji variants than the reference patterns in the Grandparent generation, 83 (59%) in Parents, and 58 (44%) in Children.

SAY NORTH only has one variant (*kayirra*, a Gurindji variant: see Supplementary Table S1), but a speaker can express this meaning by either using the variant or simply pointing north—so again the state for this variant can be 0 if they do not use the Gurindji word.

To solve this problem, we model the state of not having a language variant as an additional variant. The probability of a speaker using the variant of the variable then follows a beta distribution (a Dirichlet distribution with two dimensions) determined by the adoption rates of both the variant and the null variant (e.g. pointing or a lack of comprehension). Following our previous application, the final likelihood of a speaker knowing a particular variant of a variable is the cumulative probability of the beta distribution under an arbitrary threshold, which we set to 5%. Our previous application demonstrates that the analysis is robust to the choice of the threshold (Meakins *et al.* 2019).

We use a maximum likelihood method to estimate the retention rates of three types of language variants: Gurindji, Kriol + Innovation and null (using no variant). The retention rate of a variant is the rate that the variant will be used in preference to other variants after each sampling process (we could use the term 'adoption rate' for the same parameter when we are considering new variants that start at lower frequency). A likelihood ratio test is performed to compare a null model, where the type of variants under test has the same retention rate in different language features, to an alternative model, where the type of variant has a significantly different adoption rates in different language features. For example, to test the hypothesis that the production of Gurindji is lost more rapidly than comprehension, we compare a null model, in which G variants have equal retention rate between production and comprehension, to a biased model, where G variants have a different retention rates depending on whether they are production or comprehension variables. The hypothesis is supported if G variants in production have a significantly lower retention rate and so faster rate of loss than G variants in comprehension.

We can use this model to compare the rates of retention of Gurindji variants in different language features. To do this, we perform a likelihood ratio test to compare a null model, where different language features have the same rates of retention of Gurindji variants, to an alternative model, where different rates of retention are fitted to different classes of language features. We found that Gurindji variants in production variables decreased in frequency at a greater rate than comprehension variables, but there was no statistical difference in rate of change of frequency of Gurindji grammar vs lexicon, or nouns vs verbs (Table 2).

4. Discussion

We recognize multiple influences on language variation and change, ranging from the use or disuse of specific variants, to different rates of change across elements of language such as frequently used lexicon or grammatical structures, to membership of different social or demographic groups, to aspects of language evolution or contact induced-shift that are manifest at the level of whole

Table 2. Relative rate of retention of Indigenous (Gurindji) language variants for different language features.

Proposed rate difference	Relative rate	Likelihood ratio	P-value
Perception > Production	0.09	262.06	<0.001
Lexicon > Grammar	0.91	0.42	0.52
Verb > Noun	1.36	0.30	0.58

Notes: WF analysis, with different retention rates of Gurindji variants fitted to different categories of language features (production vs perception, grammar vs lexicon, and noun vs verb) and a likelihood ratio test against the null hypothesis that there is no difference in the retention rates of Gurindji variants between the two language features. The proposed rate differences represent hypotheses of rate differences in retention rates of the different categories. A significant result of the likelihood ratio test and a relative rate of retention lower than 1 support the hypothesis of significantly different rates of retention.

speaker populations. We can consider all of these factors together when we use an individual's linguatype to place them in a multidimensional space, defined by variation in the use of sampled language features, each of which has at least two variants. We can then ask which demographic, social or cultural factors cluster linguatypes (i.e. explain significant amounts of variation in patterns of language usage between individuals).

We performed discriminant correspondence analysis to contrast the relative explanatory power of peers, family, education and co-habitants on shaping language variation between individual speakers across 185 language variables. We found that all of these factors were significant in clustering patterns of language usage (Table 1), but that generation is the strongest clustering factor, suggesting that a speaker's linguatype is most similar to that of their peers rather than their family, household or community (Fig. 3). The second strongest clustering factor is exposure to Gurindji as a result of living with a member of the older generation in childhood. Number of years of formal school-based education also has a strong influence on language variation. Family and household have significant effects on clustering language variation, but their effects are not as strong as generation, exposure, and education. Within each generation, speakers cluster with those with similar levels of formal education or Gurindji language exposure (Fig. 3).

Having identified factors that shape linguatypes in this speaker community, we used a GLMM to examine how these social factors predict patterns of use of specific language elements. We did this by comparing the relative frequencies of patterns of use that include Gurindji variants for each language variable between groups defined by these significant factors. Inclusion of heritage language elements in individuals' linguatypes decreased across generations of speakers, reflecting the ongoing process of language shift in this community (Fig. 4). Within each generation, cultural factors shape the use of Gurindji. In particular, individuals with more years of formal education are less likely to use Gurindji variants (Fig. 5). However, adults and children who live with a member of the 'Grandparent' generation (i.e. fluent Gurindji speakers) in childhood are more likely to use variants from their heritage language than those growing up in households without elders (Fig. 5). While families do shape linguatype, these effects do not overwhelm the effect of peers on language variation (Fig. 4).

Which aspects of Gurindji are most vulnerable to loss? Because our database contains a large number of language variants, we can contrast rates of loss between grammar and lexicon, nouns and verbs, production and comprehension. We compare the patterns of change over three generations using a Wright-Fisher (WF) analysis, which compares observed rates of change to the range of outcomes expected under unbiased sampling (Meakins *et al.* 2019). We show that there is a significantly lower rate of retention in Gurindji variants of production variables (requiring active use of a language feature) than comprehension variables (testing understanding) (Table 2). This result suggests that Gurindji production is lost faster than comprehension, a pattern supported by previous observations (Dorian 1981; Fishman 1991; Meakins and Wigglesworth 2013; Tsunoda 2005; UNESCO 2003). However, we find no evidence that Gurindji grammar is lost faster than lexicon, or that noun variants are lost at a greater rate than verbs (Table 2). Therefore, our results do not support the widely held view that nouns are replaced at a greater rate than verbs (Haugen 1950; Moravcsik 1978; Muysken 1981; Singh 1982), or that grammar of heritage language suffers greater rates of loss under language shift than lexicon (Matras 2009; Myers-Scotton 2002; Thomason and Kaufman 1988; Weinreich 1974 [1953]; Zenner and Kristiansen 2014).

4.1 The effects of education in English and exposure to Gurindji

In summary, we find that family, household, peers and community all shape individual patterns of language use, but the effect of peers is the strongest of these, as individuals cluster most strongly with others of the same generation. We also identify aspects of the 'external setting' within this community that are having a significant impact on the vitality of this highly endangered Indigenous language (Sasse 1992). A notable result of our analysis is that formal schooling appears to have a significant negative association with retention of heritage language variants, across all generations in the speaker community. This result does not imply that learning English is harmful, but that monolingual education can have unintended negative consequences on Indigenous language competency. While monolingual education appears to be associated with erosion of Indigenous language competency in this community, it has not resulted in high competency in English: most students in this community are not fluent in English, and the school performs substantially below national averages in all aspects of standard literacy and numeracy tests, which are conducted entirely in English (Australian Curriculum Assessment and Reporting Authority 2018). The association between formal education and Indigenous language attrition in this community is not simply the result of perceived economic or social advantages of competency in the dominant language (Mufwene 2003), because schooling is not leading to widespread use of English. Instead, the pattern of shift is leading to the formation of a mixed language, characterized by loss of elements of Gurindji grammar and lexicon and their replacement by Kriol. While Kriol is widely spoken in Indigenous communities in northern Australia, it is rarely used in mainstream economic activities or educational settings, nor is it used in wider social contexts for communicating with non-Indigenous people. The preferred language of the younger generations in this community, the mixed language Gurindji Kriol, is not commonly spoken outside the communities of Kalkaringi and Dagaragu, and so does not have a broader socioeconomic reach than Gurindji.

Our findings are consistent with many remote Indigenous communities in Australia in which English is not the first language of most students (Macqueen *et al.* 2018). In Australia, education in Indigenous communities is almost always conducted in English, even when English not commonly spoken in the day-to-day interactions in the community. Furthermore, in common with many other Anglophone countries, high-stakes educational testing in literacy and numeracy has been introduced, partly justified as a means to reduce Indigenous disadvantage by ensuring competency in English (Allendorf 2017; Macqueen *et al.* 2018; McCarty 2009; Ryan and Whitman 2013; Wyman *et al.* 2010). In Australia, USA, and Canada, where rates of Indigenous language loss have been the most dramatic, the results of nationwide tests for educational attainment, conducted in English, have been used to reduce support for bilingual education (Combs and Nicholas 2012; Devlin 2011; Wyman *et al.* 2010). But development of these policies focused on English competency has seen less emphasis on the effect of formal education on competency in Indigenous languages, nor on appropriate ways to evaluate such an effect (Hinton 2011; Macqueen *et al.* 2018; O'Grady 2018), leading to the 'squandering of the personal, community, and national linguistic and intellectual resources within the mainstream classroom' (Cummins 2005). Conversely, promoting Indigenous language development in a bilingual setting can have flow-on effects to improve educational performance (Cummins 2005; McCarty 2003; Usborne *et al.* 2009).

Our analysis highlights positive influences on retention of Indigenous language elements within the speaker community. While we confirm that the strong effect of peers on shaping individual language patterns is a powerful force in language shift, we also demonstrate that growing up in a household with a member of the older generation is a significant factor in the retention of the Indigenous language. Language loss ultimately results from the disruption of the transmission of a heritage language across generations (Forrest 2018). While bilingual education may be the gold standard for developing and maintaining language vitality (Devlin, Disbray and Devlin 2017), it is not available in all communities with endangered languages. In such situations, Indigenous language vitality can be encouraged through other means, in partnership with schools and community members. In many Indigenous communities, where school teachers speak the nationally dominant language and cannot provide language instruction in an Indigenous language, the focus should be on developing an education system that supports existing language transmission pathways or at least does not disrupt them. For example, while promoting book-reading in the home has a positive influence on English competency in Indigenous children, oral storytelling can have a stronger effect on Indigenous language competency (Farrant *et al.* 2014). Fluent speakers from older generations can be included in school activities, as they are in Kohanga Reo ('language nest') settings in New Zealand (May 1999; Reedy 2000).

4.2 Patterns of loss cross-linguistically

Once language shift is underway, as it is in the Gurindji community, how does it affect different linguistic systems? Fishman and others note that one indication of language shift is generational differences in speaking versus understanding the heritage language, that is, younger generations will be able to understand their parents or grandparents, but not necessarily have fluency in the language itself (Dorian 1981; Fishman 1991; Meakins and Wigglesworth 2013; Tsunoda 2005;

UNESCO 2003). The WF analysis confirms this observation that production is lost faster than comprehension. In this case study, the use of Gurindji language is being lost eleven times faster than the comprehension of Gurindji.

The WF model also produced some surprising results. The first unexpected result is that the WF did not confirm the cross-linguistic observation that nouns are generally considered to be more easily adopted than verbs (Haugen 1950; Moravcsik 1978; Muysken 1981; Singh 1982). In our dataset, the rate of adoption of Kriol nouns is not significantly higher than the rate of adoption of Kriol verbs. This result is in contrast to broadscale studies of language borrowing, such as the Leipzig Loanword Typology project which includes forty-one languages, which found that nouns were twice as likely to be borrowed as verbs (Haspelmath and Tadmor 2009). One of the reasons that verbs are transferred between languages less often than nouns relates to the fact that they generally carry more inflection than nouns, for example, TAM marking vs number or gender marking (Weinreich 1974 [1953]; Wohlgemuth 2009). Kriol verbs are largely uninflected and are therefore easier to borrow or code-switch, which probably explains why they behave in a similar manner to nouns (Meakins and O'Shannessy 2012). Indeed, equal rates of noun and verb borrowing have also found been found for older borrowings into Gurindji from neighboring Indigenous languages, with the lack of inflection again being an explanation (McConvell 2009).

Second, the WF analysis did not support the hypothesis that grammar is more vulnerable to loss than lexicon in language shift scenarios, because we found no significant difference in the rate of loss of grammar and lexicon. In situations of language shift, it has been observed that speakers will maintain words in their speech for longer than grammatical structures. Similarly, in code-switching and borrowing scenarios, grammar is rarely transferred between languages, whereas words are commonly borrowed and switched into another language's grammatical frame (Greenhill *et al.* 2017; Matras 2009; Myers-Scotton 2002; Thomason and Kaufman 1988; Weinreich 1974 [1953]; Zenner and Kristiansen 2014). Given these observations, it is not clear why the loss of grammar does not proceed significantly faster in our dataset. However, this observation may relate to differences in methods and the selection of language features. Most studies that show *grammar* > *lexicon* and *noun* > *verb* patterns of loss or transfer examine small numbers of variables which are chosen often for their known patterns of change (see Greenhill *et al.* (2017) for an exception). This method of selection can result in an ascertainment bias that can skew results by focusing on features known to be undergoing a specific pattern of change. In our study, we examine 185 features that were chosen only because they are variable, rather than because of their specific patterns of variation. Ideally, studies of language contact and change should be extended over a large representative sample of language features which show variation, in order to detect general patterns of change. Using a multivariate approach as we have done avoids any risk of cherry-picking features that are undergoing common patterns of change.

5. Conclusion

We have shown that, in this community, there is a rapid shift in language usage away from using language elements from the Indigenous language and toward more inclusion of Kriol (Meakins *et al.* 2019). While family and household have an important effect on individual language patterns, peers exert the strongest influence on shaping 'linguatypes'. However, the identification of significant external factors that impact on Indigenous language retention—more years of formal schooling results in greater loss of Indigenous language, and exposure to older generations results in retention of Indigenous language elements—points to the potential for interventions that increase vitality in endangered Indigenous languages. Measures of language endangerment typically focus on both usage (e.g. whether used in formal schooling) and transmission (e.g. whether being learned by children at home) (Lee and Van Way 2016; Lewis and Simons 2010; Moseley 2010). Our results have relevance for both of these aspects of language vitality, suggesting that practices that promote the contact between children and elders will enhance Indigenous language retention, and that measures that limit the spheres of usage of the language (e.g. not being used in the classroom) can accelerate erosion of Indigenous language. In this way, our study on a single language community provides an informative view of larger scale processes of language vitality and endangerment.

The idea that monolingual education and exposure to heritage language at home are important factors influencing language vitality is widely accepted (which is why these factors are included in our study). Yet there have been relatively few quantitative studies of the effect of such factors on language shift due to the difficulty of collecting appropriate data and having an effective analytical framework. Most existing studies use self-reported information for relatively few participants, and focus on broad-brush measures of language competency.

Our study is unique in the quantity of data we bring to this question, both in terms of number of speakers and number of language variables, and in the novel analytical approach we take, which provides confirmation of qualitative impressions or more limited quantitative analyses. We hope this study adds weight to the conviction that these key cultural factors are important in language maintenance, providing focus for useful interventions in the maintenance of minority language vitality.

Acknowledgements

Funding for the data collection comes from an Australian Research Council (ARC) Future Fellowship awarded to Felicity Meakins (FT170100042) and the ARC Centre of Excellence for the Dynamics of Language (Project ID: CE140100041). We thank the Gurindji participants for their contributions to this study.

Data and materials availability

Data is available in the Supplementary Material, but any information that might identify individuals, such as year of birth, gender, household, and family, has been removed.

Supplementary data

Supplementary data is available at *Journal of Language Evolution* online.

Conflict of interest statement. None declared.

References

Agrawal, A. A. et al. (2012) 'Insect Herbivores Drive Real-Time Ecological and Evolutionary Change in Plant Populations', *Science*, 338: 113–16.

Allendorf, F. W. (2017) 'Genetics and the Conservation of Natural Populations: Allozymes to Genomes', *Molecular Ecology*, 26: 420–30.

Amano, T. et al. (2014) 'Global Distribution and Drivers of Language Extinction Risk', *Proceedings of the Royal Society B: Biological Sciences*, 281: 20141574.

Austin, P. K., and Sallabank, J. (2011) *The Cambridge Handbook of Endangered Languages*. Cambridge University Press, Cambridge.

Australian Curriculum Assessment and Reporting Authority (2018). Kalkaringi School, Kalkarindji, NT <myschool.edu.au>.

Avise, J. C. (2009) 'Phylogeography: Retrospect and Prospect', *Journal of Biogeography*, 36: 3–15.

Brenzinger, M. (1992) *Language Death: Factual and Theoretical Explorations with Special Reference to East Africa*. Berlin: De Gruyter Mouton.

Bromham, L. (2016) *An Introduction to Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford.

—— (2017) 'Curiously the Same: Swapping Tools between Linguistics and Evolutionary Biology', *Biology & Philosophy*, 32: 855–86.

Christy, T. C. (1983) *Uniformitarianism in Linguistics*. John Benjamins Publishing.

Combs, M. C., and Nicholas, S. E. (2012) 'The Effect of Arizona Language Policies on Arizona Indigenous Students', *Language Policy*, 11: 101–18.

Coulson, T. et al. (2011) 'Modeling Effects of Environmental Change on Wolf Population Dynamics, Trait Evolution, and Life History', *Science*, 334: 1275–8.

Crystal, D. (2000) *Language Death*. Ernst Klett Sprachen.

Cummins, J. (2005) 'A Proposal for Action: Strategies for Recognizing Heritage Language Competence as a Learning Resource within the Mainstream Classroom', *Modern Language Journal*, 585–92.

Devlin, B. (2011) 'The Status and Future of Bilingual Education for Remote Indigenous Students in the Northern Territory', *Australian Review of Applied Linguistics*, 34: 260–79.

——, Disbray, S., and Devlin, N. R. F. (2017) *History of Bilingual Education in the Northern Territory*. New York: Springer.

Dorian, N. (1981) *Language Death: The Life-Cycle of a Scottish Gaelic Dialect*. Philadelphia: University of Pennsylvania Press.

Evans, L. M. et al. (2014) 'Population Genomics of Populus Trichocarpa Identifies Signatures of Selection and Adaptive Trait Associations', *Nature Genetics*, 46: 1089.

Evans, N. (2011) *Dying Words: Endangered Languages and What They Have to Tell us*. John Wiley & Sons.

Farrant, B. M. et al. (2014) 'Early Vocabulary Development of Australian Indigenous Children: Identifying Strengths', *Child Development Research*, 2014: 1–7.

Fisher, R. A. (1956) *The Genetical Theory of Natural Selection*. New York: Wiley.

Fishman, J. (1991) *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages*. Clevedon: Multilingual Matters.

Fishman, J. A. (2002) 'Endangered Minority Languages: Prospects for Sociolinguistic Research', *International Journal on Multicultural Societies*, 4: 270–5.

Forrest, W. (2018) 'The Intergenerational Transmission of Australian Indigenous Languages: Why Language Maintenance Programmes Should Be Family-Focused', *Ethnic and Racial Studies*, 41: 303–23

Greenhill, S. et al. (2017) 'Evolutionary Dynamics of Language Systems', *Proceedings of the Royal Society*, 114: E8822–E8829.

Guy, G., and Hinskens, F. (2016) 'Linguistic Coherence: Systems, Repertoires and Speech Communities', *Lingua*, 172-173: 1–9

Hadfield, J. D. (2010) 'MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package', *Journal of Statistical Software*, 33: 1–22

Harrisson, K. A. et al. (2014) 'Using Genomics to Characterize Evolutionary Potential for Conservation of Wild Populations', *Evolutionary Applications*, 7: 1008–25

Haspelmath, M., and Tadmor, U. (2009) 'The Loanword Typology Project and the World Loanword Database', *Loanwords in the World's Languages: A Comparative Handbook*, Berlin: De Gruyter Mouton, 1–34.

Haugen, E. (1950) 'The Analysis of Linguistic Borrowing', *Language*, 26: 210–31

Hinton, L. (2011) 'Language Revitalization and Language Pedagogy: New Teaching and Learning Strategies AU - Hinton, Leanne', *Language and Education*, 25: 307–18

Hua, X., and Bromham, L. (2017) 'Darwinism for the Genomic Age: Connecting Mutation to Diversification', *Frontiers in Genetics*, 8.

Krauss, M. (1992) 'The World's Languages in Crisis', *Language*, 68: 4–10

Labov, W. (1966) *Social Stratification of English in New York City*. Washington: Center for Applied Linguistics.

—— (1994a) *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.

—— (1994b) *Principles of Linguistic Change: Social Factors*. Oxford: Blackwell.

Lee, N. H., and Van Way, J. (2016) 'Assessing Levels of Endangerment in the Catalogue of Endangered Languages (ELCat) Using the Language Endangerment Index (LEI)', *Language in Society*, 45: 271

Lewis, M. P., and Simons, G. F. (2010) 'Assessing Endangerment: Expanding Fishman's GIDS', *Revue Roumaine de Linguistique, 55*: 103–20.

Loh, J., and Harmon, D. (2014) 'Biocultural Diversity: Threatened Species, Endangered Languages', *WWF Netherlands, Zeist, The Netherlands*, 1.

Luikart, G. et al. (2003) 'The Power and Promise of Population Genomics: From Genotyping to Genome Typing', *Nature Reviews Genetics*, 4: 981.

Macqueen, S. et al. (2018) 'The Impact of National Standardized Literacy and Numeracy Testing on Children and Teaching Staff in Remote Australian Indigenous Communities', *Language Testing*. https://doi.org/10.1177/0265532218775758

Matras, Y. (2009) *Language Contact*. Cambridge: Cambridge University Press.

May, S. (1999) *Indigenous Community-based Education*. Multilingual Matters.

Mayer, M. (1969*) Frog, Where Are You*? New York: Dial Press.

McCarty, T. L. (2003) 'Revitalising Indigenous Languages in Homogenising Times', *Comparative Education*, 39: 147–63

—— (2009) 'The Impact of High-Stakes Accountability Policies on Native American Learners: Evidence from Research', *Teaching Education*, 20: 7–29

McConvell, P. (2009) 'Loanwords in Gurindji, a Pama-Nyungan Language of Australia', in Haspelmath M. and Tadmor U. (eds) *Loanwords in the World's Languages: A Comparative Handbook*, pp. 790–822. Berlin: De Gruyter Mouton.

——, and Meakins, F. (2005) 'Gurindji Kriol: A Mixed Language Emerges from Code-Switching', *Australian Journal of Linguistics*, 25: 9–30

Meakins, F. (2008) 'Unravelling Languages: Multilingualism and Language Contact in Kalkaringi', in Simpson J. and Wigglesworth G. (eds) *Children's Language and Multilingualism: Indigenous Language Use at Home and School*, pp. 247–64. New York: Continuum.

—— (2011) *Case Marking in Contact: The Development and Function of Case Morphology in Gurindji Kriol*. Amsterdam: John Benjamins.

——, and Algy, C. (2016) 'Deadly Reckoning: Changes in Gurindji Children's Knowledge of Cardinals', *Australian Journal of Linguistics*, 36: 479–501.

——, Green, J., and Turpin, M. (2018) *Understanding Linguistic Fieldwork*. Routledge.

—— et al. (2019) 'Birth of a Contact Language Did Not Favour Simplification', *Language*, 95: 294–332.

——, and O'Shannessy, C. (2012) 'Typological Constraints on Verb Integration in Two Australian Mixed Languages', *Journal of Language Contact*, 5: 216–46.

——, and —— (2016) *Loss and Renewal: Australian Languages Since Colonisation*. Berlin: De Gruyter Mouton.

——, and Wigglesworth, G. (2013) 'How Much Input is Enough? Correlating Comprehension and Child Language Input in an Endangered Language', *Journal of Multilingual and Multicultural Development*, 34: 171–88.

Meyerhoff, M. (2009) 'Sociolinguistic Variation and Change', *Linguistics-Encyclopedia of Life Support Systems*, 202–24.

—— (2010) *Introducing Sociolinguistics*. Routledge, London.

Moravcsik, E. (1978) 'Universals of Language Contact', in Greenberg J. (ed.) *Universals of Human Language: Vol. 1 Method and Theory*, pp. 95–122. Stanford, CA: Stanford University Press.

Moreno-Estrada, A. et al. (2013) 'Reconstructing the Population Genetic History of the Caribbean', *PLOS Genetics*, 9: e1003925.

Morin, P. A., Luikart, G., and Wayne, R. K. (2004) 'SNPs in Ecology, Evolution and Conservation', *Trends in Ecology & Evolution*, 19: 208–16.

Moseley, C. (2010) *Atlas of the World's Languages in Danger*, Online version. Paris: UNESCO Publishing. <http://www.unesco.org/culture/en/endangeredlanguages/atlas>.

Mufwene, S. S. (2003) 'Language Endangerment: What Have Pride and Prestige Got to Do with It', *When Languages Collide: Perspectives on Language Conflict, Language Competition, and Language Coexistence*, Ohio Stat University Press. pp. 324–46.

Muysken, P. (1981) 'Halfway Between Quechua and Spanish: The Case for Relexification', in Highfield A. and Valdman A. (eds) *Historicity and Variation in Creole Studies*, pp. 52–78. Ann Arbor: Karoma.

Myers-Scotton, C. (2002) *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. New York: Oxford University Press.

O'Grady, W. (2018) 'Assessing Language Revitalization: Methods and Priorities', *Annual Review of Linguistics*, 4: 317–36.

Reedy, T. (2000) 'Te Reo Māori: The past 20 Years and Looking Forward', *Oceanic Linguistics*, 46: 157–69.

Rehg, K. L., and Campbell, L. (2018) *The Oxford Handbook of Endangered Languages*. Oxford University Press.

Rius, M., and Darling, J. A. (2014) 'How Important is Intraspecific Genetic Admixture to the Success of Colonising Populations? ', *Trends in Ecology & Evolution*, 29: 233–42

Romaine, S. (2017) 'Language Endangerment and Language Death', in Fill A. F. and Penz H. *(eds) The Routledge Handbook of Ecolinguistics*. Routledge.

Ryan, T. G., and Whitman, A. (2013) 'The Inequity and Effect of Standardized Literacy Testing for First Nations Students-an Ontario (Canadian) Outlook', *Journal of Educational Thought/Revue de la Pensée Educative*, 163–81.

Sasse, H.-J. (1992) 'Theory of Language Death', in Brenzinger M. (ed.) *Language Death: Factual and Theoretical Explorations with Special Reference to East Africa*, pp. 7–30. Berlin: De Gruyter Mouton.

Simons, G. F., and Lewis, M. P. (2013) 'The World's Languages in Crisis. *Responses to Language Endangerment*', in *Honor of Mickey Noonan. New Directions in Language Documentation and Language Revitalization*, Amsterdam: John Benjamins. vol. 3, p. 20.

Simpson, J., and Wigglesworth, G. (2019) 'Language Diversity in Indigenous Australia in the 21st Century', *Current Issues in Language Planning*, 20: 67–80.

Singh, R. (1982) '"On Some 'Redundant Compounds" in Modern Hindi', *Lingua*, 56: 345–51.

Sutherland, W. J. (2003) 'Parallel Extinction Risk and Global Distribution of Languages and Species', *Nature*, 423: 276–9.

Tagliamonte, S. A. (2006) *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Thomason, S. G., and Kaufman, T. (1988) *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.

Tsunoda, T. (2005) *Language Endangerment and Language Revitalization*. Berlin: De Gruyter Mouton.

UNESCO (2003) *Language Vitality and Endangerment: International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages*. Paris: UNESCO.

Usborne, E. et al. (2009) 'Bilingual Education in an Aboriginal Context: Examining the Transfer of Language Skills from Inuktitut to English or French', *International Journal of Bilingual Education and Bilingualism*, 12: 667–84.

Weinreich, U. (1974 [1953]). *Languages in Contact: Findings and Problems*. The Hague: Mouton.

Williams, L. J. et al. (2010) 'A Tutorial on Multiblock Discriminant Correspondence Analysis (MUDICA): A New Method for Analyzing Discourse Data from Clinical Populations', *Journal of Speech, Language, and Hearing Research, 53: 1372–1393*.

Wohlgemuth, J. (2009) *A Typology of Verbal Borrowings*. Berlin: De Gruyter Mouton.

Wright, S. (1931) 'Evolution in Mendelian Populations', *Genetics*, 16: 97–159

Wyman, L. et al. (2010) 'High Stakes Testing, Bilingual Education and Language Endangerment: A Yup'ik Example', *International Journal of Bilingual Education and Bilingualism*, 13: 701–21

Zenner, E., and Kristiansen, G. (2014) *New Perspectives on Lexical Borrowing Onomasiological, Methodological and Phraseological Innovations*. Berlin: De Gruyter Mouton.