Depending on gender

The role of Gurindji women in contact-induced language change

Felicity Meakins,[1] Lindell Bromham,[2] and Xia Hua[2]
[1] University of Queensland | [2] Australian National University

Sociolinguists have grappled with how speakers of different genders use linguistic variables differentially to constitute their identities. Two seemingly-conflicting generalisations have emerged, referred to as the gender paradox. Women at once maintain standard forms which are overtly-discussed and positively-evaluated; yet lead change in non-standard forms which are below the level of consciousness. These trends are relatively robust in cases of language-internal change in Western societies but less so in non-Western societies and situations of language contact. This paper examines the role of gender in a north Australian Indigenous community where there is a shift underway from Gurindji to Kriol. The dataset consists of 185 variables used by 78 speakers across three generations. Here we examine the results of the application of the BayesVarbrul to the dataset which suggest that women tend to be more conservative than men in retaining Gurindji variants, despite a more general shift to Kriol.

Keywords: gender paradox, Gurindji Kriol, language contact, BayesVarbrul, language variation

1. Introduction

For decades now, variationist sociolinguists have been interested in gender differences in speech, specifically where speakers use variants differentially in the constitution of their gender identities (see Coates, 1998; Eckert & McConnell-Ginet, 2003; Meyerhoff, 2010 for some overviews). Two surprisingly robust generalisations have emerged across many studies: women at once maintain standard forms which are overtly-discussed and positively-evaluated; and yet lead change in non-standard forms which are below the level of consciousness. These apparently conflicting trends are referred to by Labov (2001, pp. 293–294) as the *gender paradox*.

https://doi.org/10.1075/aplv.24001.mea | Published online: 3 March 2025
Asia-Pacific Language Variation 10:2 (2024), pp. 183–208. ISSN 2215-1354 | E-ISSN 2215-1362

Nonetheless, these generalisations are not always upheld in non-Western societies (e.g., Bakir, 1986; Bucholtz, 2003), and the gender paradox becomes even more perplexing in situations of language contact where gender is not a significant driver of the uptake of particular variants or the direction of language change (Meyerhoff & Birchfield, 2019) (§2).

Here we discuss the results of a new form of sociolinguistic analysis that has been used to examine the role of gender in a situation of contact-induced language change in northern Australia, using the Gurindji Kriol dataset which consists of 185 linguistic variables (with 357 variants) used by 78 speakers (~15% of the Gurindji community).[1] The speakers in the Gurindji Kriol dataset represent three generations which means the snapshot of variant use across generations gives us an opportunity to test ideas about contact-induced language change (cf. the apparent time hypothesis (Bailey et al., 1991; Labov, 1963)). For example, the dataset has been previously used to test the claim that language contact leads to morphological simplification (Meakins et al., 2019). It has also been used to examine how social factors such as education level and living with elders affect rates of language change (Bromham et al., 2020) and to understand how variants cluster or 'cohere' to form sociolects (Hua et al., 2022) (§4).

These previous studies have identified significant associations between language variant use and social categories, including generation, number of years in formal school-based education and exposure to heritage language through living with an elder who is fluent in Gurindji. But gender remained unexamined until Hua (2022) included gender as a potential predictor of language change in the Gurindji community. Previous studies had not included gender as a factor because, while the 78 speakers in the dataset are coded for gender, the dataset has male speakers only from the youngest generation and the oldest generation while female speakers are represented across all three generations. This gap in the dataset stems from the practicalities of fieldwork. It would have been culturally inappropriate for Meakins and Algy who collected the data, both women, to work with young men (regardless of their marital status). Working with young men would have been disapproved of by the Gurindji community because it is unusual for a woman, or even a pair of women, to spend much time alone in a group of men (Meakins, Green, & Turpin, 2018, p. 249).

Despite this gap in the Gurindji Kriol dataset, Hua's (2022) new method — `BayesVarbrul`[2] — made it possible to test the potential effect of gender on language change among Gurindji people. `BayesVarbrul` is a new multivariate analysis of lan-

1. See Meakins et al. (2019) for a detailed description of the database and collection of data.

2. The code and the instructions are implemented in R and available at https://github.com/huaxia1985/BayesVarbrul.

guage change which combines Wright–Fisher models and the logic behind `Varbrul` analyses under a Bayesian framework. Wright-Fisher models are generally used in population genetics for describing a serial sampling process of genetic variants from one timestep to the next. Hua (2022) adapts Wright–Fisher models for language change, mapping the uptake of multiple linguistic variants across different generations (§ 4.1).

Hua's modelling of Gurindji Kriol dataset using `BayesVarbrul` suggests with 70% certainty that men use more Kriol variants relative to Gurindji variants than women (§ 4.3). Hua (2022) focussed on the methodological aspects of the model, but although gender was included in the model it was not a focus of the study so was not examined in any detail. Because the question of the role of gender in language change has been a subject of interest in the sociolinguistics literature, here we focus on the inclusion of gender in the model of Hua (2022) and examine it in the context of the literature on the role of gender in non-Western societies and in situations of contact-induced language change (§ 2). Additionally, this result demonstrates how `BayesVarbrul` can be applied to incomplete datasets and still yield robust results (§ 4).

2. Gender as a driver in contact-induced language change

Variationist sociolinguists have shown how speakers use variants to constitute their gender identities. Early studies identified two generalisations about the differential use of variants by men and women. Firstly, women use more standard variants (often morpho-syntactic or lexical) than men where the standard variants are those which are overtly discussed and positively-evaluated. Secondly, where variants (often phonetic) are used that are below the level of awareness in a speech community, women tend to lead men in linguistic change. These two generalisations form what Labov (2001, pp. 294–293) has dubbed as the *gender paradox*, that is, women at once maintain standard forms and yet lead change in the non-standard forms which are below the level of consciousness. Nonetheless these two generalisations are linked. The effect of gender on the maintenance of standard forms and the rise of new forms is shaped by social salience. Overt awareness of the social meaning of variants is likely to be important in women's maintenance of variants and the rejection of variants that are negatively evaluated. However, women may lead change for variants that have less social salience and are less overtly evaluated.

The first generalisation about the association between women and standard forms has been shown to be empirically robust. As Trudgill (1983, p. 162) observes, it is "the single most consistent finding to have emerged from social

dialect studies over the past twenty years". For example, in Trudgill's (1972) classic study of variation in the use of *ing* in Norwich English, he observed that women tended to use the [iŋ] variant, rather than the [in] variant more than men across class and style, for example, *singing* vs *singin'*. Trudgill (1972) also observed that Norwich women tended to overreport their usage of standard forms and men tended to underreport them. He suggested that this pattern was the result of women finding the overt prestige associated with the standard forms more important than men. Trudgill argued that women use language to achieve status denied to them by societal structures and the use of the standard is a reflection of the powerlessness of women in the public sphere. Eckert and McConnell-Ginet (2003, p. 294) suggest there might be something else at play here. Grammatical variables are highly salient which means that non-standard grammar attracts more attention in family and schools. Rebelliousness, defiance and toughness is tolerated and even valued in boys but not in girls which means that girls are less likely to show grammatical deviance.

The degree of awareness in the use of variants in the expression of gender (i.e., social salience) also drives change in the use of forms, according to the second part of Labov's gender paradox. Women are the leaders in changes that are below the level of consciousness (i.e., less salient). Labov (1972) refers to these types of variants as 'indicators'. Indicators are variables which pattern according to social groups but are below the level of awareness in the community. For example, young women are leading a merger of EAR and AIR diphthongs in onsets in New Zealand English, such that the pronunciation of 'beer' and 'bear' is indistinguishable (Batterham, 2000). Holmes (1995) also found that young women were leading a change in previously stigmatised glottalised variants of the word-final /t/ in New Zealand English.

Nonetheless, Labov's generalisations — that women maintain socially salient variants but drive change in the use of variants that are not salient — are not always supported by studies of language use in non-Western societies. Indeed, variationist sociolinguists have long noted that data and therefore theorising on gendered language use and change has been skewed towards Western societies. For example, Bucholtz (2003, p. 403) has pointed out that there is a "tendency for language and gender researchers to focus on the white, English-speaking middle class". Where gender differences and language use in other cultures have been examined, opposite trends have sometimes been observed. For example, Bakir (1986) found that Iraqi men make more use of prestigious Classical Arabic variants than women in a study of the distribution of Classic Arabic variants and their local Iraqi equivalents. Similarly, Meakins et al. (2024) find that female Shawi speakers in Peru with a tertiary education use more prestigious variants (in this case ancestral forms) relative to the other variants.

Gender explains variation in most sociolinguistic studies, often in combination with class and formality, at least in monolingual communities, whether Western or non-Western. However, the strength of Labov's generalisations about gender and language change has been questioned, particularly with regard to language contact situations. For example, Meyerhoff and Birchfield (2019, p. 246) observe that for language contact situations in both Western or non-Western contexts "(i)t proves much harder to find evidence that gender is a significant predictor of variation: gender effects are very weak, inconsistent in the direction of their effect, or simply not present at all in variationist studies of language contact".

Some studies of languages in contact situations do conform to Labov's predictions. For example, Dorian (2006) found that a conservative form of the vocative in Sutherland Gaelic was more likely to be retained by women than men, even though there was a community-wide shift to English underway. Maclagan et al. (2013) also found that Māori women led non-stigmatised sound changes which are the result of contact with New Zealand English, but were more conservative when changes are salient and stigmatised. Holmquist (1985) made similar findings in his study of Castilian Spanish spoken in Ucieda which is located in the Cantabria mountain region of northern Spain. He found young women were leading a shift to standard Spanish vowels as they shifted away from rural life to factory work. These trends reflect those found in monolingual studies where women orientate to the standard variants consciously upheld by speech communities.

Nonetheless, other studies of contact-induced change go against Labov's predictions by demonstrating that women are using more innovative, stigmatised forms, even where they are salient. For example, Liu (2012) reports that, in Shanghainese (a dialect of the Chinese language), women are leading men in a shift to the non-standard pronunciation of the pronoun *wo* 'I' as [u:] instead of the standard [ŋu:]. The variant is a stigmatised alternative used by migrants from other parts of China. Other contact studies show no differences between genders. In Nagy's (2011) study of Italian influence on Faetar, an endangered Franco-provençal variety spoken in southern Italy and by diaspora groups in Toronto, she found no lexical differences in men and women's speech. Meyerhoff (2000, 2002, 2015) also found no gender differences in multiple studies of morpho-syntactic variation in Vanuatu where multilingualism between Austronesian languages and Bislama, an English-based Creole language, is the norm. A similar story is found for Bequia, an English dialect of the Caribbean island of Bequia (Meyerhoff, 2008).

Meyerhoff and Birchfield (2019) use Ochs' (1992) theory of indexicality to explain the widespread absence of gender effects in many language contact situations. Ochs (1992, p. 340) suggests that "few features of language directly and

exclusively index gender". Some examples of linguistic features which directly index gender are pronouns and kinship terms. In most cases, however, the association of linguistic features and gender is secondarily mediated through the use of linguistic features to index a stance, act or activity. In situations of language contact, Meyerhoff and Birchfield (2019) suggest that:

> Because the mapping of stances, acts, and activities onto gender ideologies is culturally mediated, situations of language contact create conditions where (a) there may be more important social identities to inscribe linguistically, and (b) there may be competing ideologies about and ways of experiencing gender that need to be negotiated. Both of these considerations mean other social identities may be quantitatively more significant.

Relatedly, Eckert and McConnell-Ginet (2003, p. 4) note that, "gender doesn't just exist, but is continually produced, reproduced, and indeed changed through people's performance of gendered acts". Language is one site of gender performance. Language then does not merely reflect existing social categories but constructs them under this view. This re-positioning of language and gender contrasts with the earlier studies which treated gender as a fixed category which was echoed in variant use. Rather, Eckert and McConnell-Ginet (2003, p. 293) suggest that linguistic resources index aspects of social practice that constitute social categories rather than the categories themselves. In situations of contact-induced language, it may well be the case that the external social dynamics of language communities which make salient different social identities (ethnicity, generation), perhaps render other within-community social identities (gender, class) as relatively less relevant such that they are only minimised activated.

We turn now to the Gurindji community of Kalkaringi in northern Australia where the socio-historical situation has led to a shift from Gurindji to Kriol. In Section 4, we examine the role of gender in this case of contact-induced change.

3. The language ecology of Kalkaringi

Gurindji people mostly live in the twin communities of Kalkaringi and Daguragu in northern Australia, although Gurindji diaspora also live in surrounding communities and the larger towns of Katherine and Darwin. Kalkaringi and Daguragu are only eight kms apart with shared kin relations and joint administration so they essentially function as a single community. The name "Kalkaringi" is used in this paper to refer to both communities.

Kalkaringi is a complex language ecology where there is much variation in the use of Gurindji (Ngumpin-Yapa, Pama-Nyungan), the traditional language of

the region, and Kriol, the English-lexifier Creole language spoken across much of northern Australia (among other languages such as Warlpiri and English) (Meakins, 2008). Code-switching between Gurindji and Kriol has also led to the emergence of the mixed language, Gurindji Kriol (McConvell & Meakins, 2005; Meakins, 2011b). Gurindji Kriol is also spoken as an L1 in the Bilinarra and Ngarinyman communities of Pigeon Hole and Yarralin, respectively (see Figure 1).

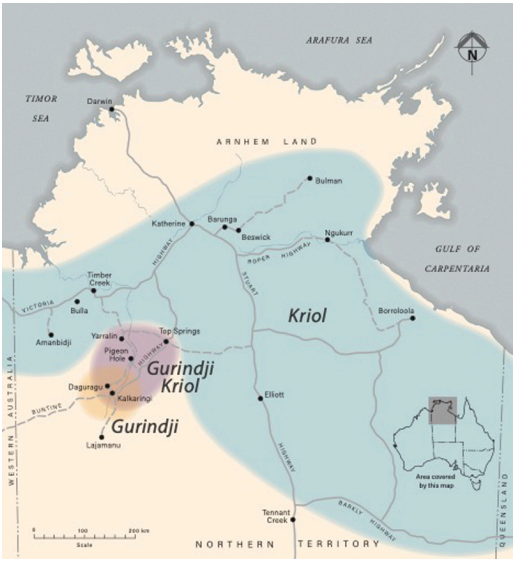

Figure 1. Areas in northern Australia where Gurindji, Kriol and Gurindji Kriol are spoken (Meakins et al., 2019)

Gurindji Kriol developed rapidly over three generations from the early 1980s onwards, with all stages of its development still represented in the current Gurindji population. Most households in the community contain members of three generations. These generations represent different stages in the formation of Gurindji Kriol. The 'grandparent' generation (Generation 1) are bilingual in Gurindji and Kriol, and they code-switch between the two. The 'parent' generation (Generation 2) cannot speak either language fluently and are the first generation of Gurindji Kriol speakers. The 'child' generation (Generation 3) also speak Gurindji Kriol but with more and lighter (i.e., acrolectal) Kriol variants and less variability both between and within speakers. Figure 2 illustrates the three generations represented in the database and their use of Gurindji, Kriol and Gurindji Kriol.

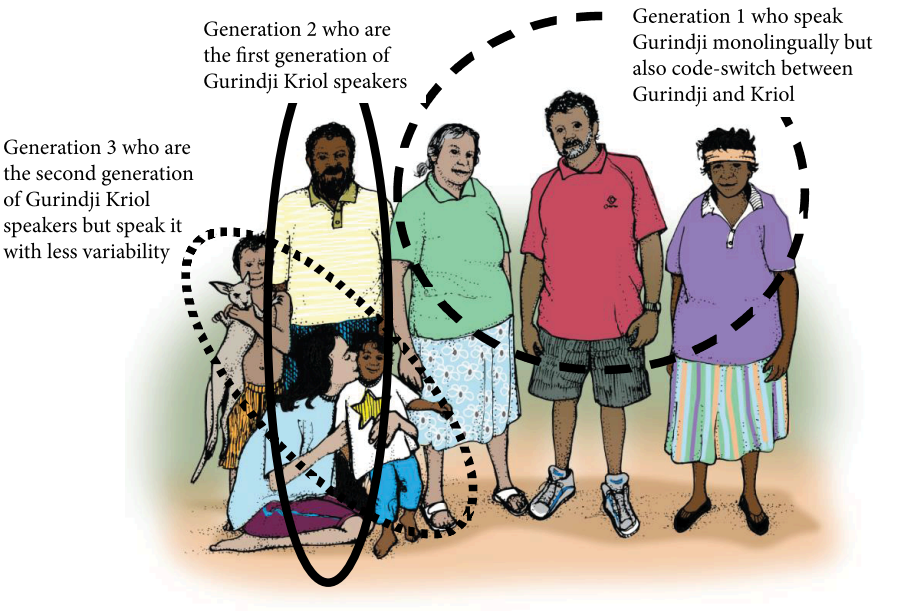

Figure 2. Three generations of Gurindji people and their language practices (Meakins, 2023, p. 77) (Figure based on drawing by Jenny Taylor AILIB-PEOPLE_019–00 (Kral, Disbray, & Green, 2022))

Gurindji Kriol is characterised by the adoption of some Gurindji features, some Kriol features and the development of some innovative elements not used in either source language (Meakins, 2012; Meakins et al., 2019). In addition, some of the Kriol features have since acrolectalised as a result of increasing contact with English. This mix is shown in the following two sentences where Kriol elements are bolded, Gurindji elements are italicised and innovative elements are underlined. The English origins of Kriol words are given in the fourth line. Example (1) shows a 'heavier' form of Gurindji Kriol more characteristic of adults and (2) shows a 'lighter' form more characteristic of younger generations.

(1)	**Dat**	**bigija**	*warlaku*	**bin**	*makin*	**wansaid**	**<u>langa</u>**	*warlu-<u>ngka</u>*.
(2)	Dat	big-wan	*warlaku*	**bin**	*makin*	**nekstu**	**langa**	**faya**
	the	big-NMLZ	dog	PST	sleep	next.to	LOC	fire-LOC
	<that	<biggest-<one		<been			<along	

The large dog slept next to the fire. (adapted from Hua et al., 2022, p. 84)

There has been a shift towards an increased use of Kriol lexicon and morpho-syntax, and a narrowing of variation from Generation 1 to Generation 3 (Bromham et al., 2020; Meakins et al., 2019). Nonetheless variation is still present. For example, younger generations are more likely to use the Kriol word *faya* 'fire' than the Gurindji equivalent *warlu*; and are more likely to use the more acrolectal Kriol word *nekstu* than the alternative basilectal Kriol word *wansaid* (<one side). Adult Gurindji people tend to use the basilectal Kriol form *bigija* (alongside the Gurindji word *jangkarni*) but children are more likely to use the acrolectal form *bigwan* (<big one). Adults are also more likely to use the innovative double-marked variant *langa … -ngka* instead of a sole Kriol preposition *langa* or a sole Gurindji locative suffix *–ngka* (Hua et al., 2022, p. 84).

There are many papers on generational differences in the speech of Gurindji people which consider *individual* features such as lexicon (Meakins & Wigglesworth, 2013; Sloan, Meakins, & Algy, 2022), grammatical relations (Meakins, 2009, 2015; Meakins & Wilmoth, 2020), static location (Meakins, 2011a, 2016), possession (Meakins & O'Shannessy, 2005; van den Bos, Meakins, & Algy, 2017), complex predicates (Meakins, 2010) and spatial relations (Dunn, Meakins, & Algy, 2021; Meakins, 2011c; Meakins & Algy, 2016; Meakins, Jones, & Algy, 2016); and *multiple* features in single analyses (Bromham et al., 2020; Hua et al., 2022; Meakins et al., 2019). This paper focuses on a new previously unexplored potential driver of language change in this community, speaker gender.

4. Gender as a driver of change among Gurindji people

As introduced in Section 1, gender, as a potential predictor of language change in the Gurindji community, was absent in studies of the Gurindji Kriol dataset until Hua (2022) which showcased `BayesVarbrul`, a new multidimensional analysis of language change using a Bayesian framework which allows a unified analysis of multiple variables. In § 4.1, we discuss some of the uses of `BayesVarbrul` in comparison with other sociolinguistics analysis tools `Varbrul`, `Rbrul`, mixed effects models (e.g., `glmm` and `mlogit` packages), random forests and Bayesian mixed models (`RStan` and `MCMCglmm` packages). We show that `BayesVarbrul` addresses a number of limitations of previous models. Its defining feature is being able to simultaneously analyse multiple dependent variables using Wright–Fisher model as the null model which is appropriate for analysing language change, as we discuss. We then summarise the results from Hua (2022) (§ 4.2), giving particular attention to gender (§ 4.3).

4.1 Introducing BayesVarbrul as a new analytic method

Varbrul (Variable Rules Analysis) has been the tool of choice for many variationist sociolinguists since it was first developed in the mid 1970s (Cedergren & Sankoff, 1974) and was then later implemented as the user-friendly Goldvarb (Sankoff, Tagliamonte, & Smith, 2005). It allows sociolinguists to understand variation in a speech community's use of a linguistic feature (lexical, grammatical or phonetic) according to social factors (e.g., speaker class, generation, ethnicity, gender, etc.) and linguistic factors (e.g., animacy, edge weight, phonological context, etc.). The linguistic feature can be the use or non-use of a feature, or variation in the use of two variants of a feature. Of particular interest is the change in the usage frequency of variants over time which has been studied in both real time and apparent time (Tagliamonte, 2006).

Essentially Varbrul is a fixed effects binary logistic regression model and, as such, has a number of limitations.[3] Firstly, the linguistic feature of interest, the dependent variable, can only be binary (i.e., present/absent or two variants). This is a problem for continuous dependent variables such as vowel formant measurements and for language variables with multiple variants. Secondly, the factors (independent variables or predictors) affecting the dependent variable can also only be categorical, not continuous, which is a problem for variables such as age, many phonetic measures, proficiency scores, etc. Thirdly, Varbrul essentially runs a series of monofactorial analyses instead of a single multifactorial analysis which means it cannot capture interactions between independent variables (Bayley, 2013, p. 19; Gorman & Johnson, 2013, p. 220).

Rbrul, developed by Johnson (2019), solves some of these problems by allowing the sociolinguist to model both continuous as well as categorical variables. Underlyingly, it is a mixed model which uses the lme4 package containing the *glmer* function. As such, it allows the use of two types of independent variables or predictors: fixed effects and random variables. The additional of random variables takes into consideration important characteristics of many datasets, for example an uneven representation of tokens from different people, the fact that an individual is more likely to behave like themselves than other people and the fact that particular types of variables, such as high frequency lexemes, might skew results (Tagliamonte & Baayen, 2012). Rbrul can also model interactions between fac-

3. We note that problems have been identified with Varbrul including the automated step-wise model selection (Harrell, 2021) and whether the difference between weightings is significant, but do not focus on this here. For smaller datasets, conditional inference trees and random forests are becoming popular. For datasets with more than two variants in competition, multinomial logistic regression can be used.

tors (more easily than Varbrul) (Johnson, 2009). More recently, many variationist sociolinguists have moved from Rbrul to directly use mixed models from the lme4 package (Bates, Maechler, & Bolker, 2012; Pinheiro & Bates, 2000). Mixed-effects *logistic* regression models have become commonplace in variationist research for *binary* categorical variables (i.e, present/absent or two variants), for example, Travis, Grama and Purser's (2023) study of *ing*/*in* alternation in Sydney English. Mixed-effects *linear* regression models are used for *continuous* dependent variables such as phonetic variables, for example, Walker and Meyerhoff's (2020) study of low-back vowels in eastern Caribbean English.

One of the appeals of Rbrul is that it reports on 'factor weight' (the relative importance of predictors in an analysis) which makes it more familiar to users of Varbrul. Nonetheless there is no way of measuring whether the differences between factor weights are significant. More recently, some variationist sociolinguists have used random forests and conditional inference trees which have the advantage of reporting on the relative ranking of predictors (much like factor weight). Random forests also have advantages for smaller datasets or unbalanced datasets, for example, where tokens are unevenly contributed by different speakers, or where there is more of one kind of variant than another, or where one age group is better represented than another etc. Since random forests make no assumptions about the underlying distribution of the data set, they are appropriate for these kinds of datasets (Breiman, 2001; Tagliamonte & Baayen, 2012). Some examples have been reflexive constructions in north Australian Kriol (Dickson & Durantin, 2019), dative alternation in Asian Englishes (Bernaisch, Gries, & Mukherjee, 2014), verb complementation constructions across New Englishes (Deshors & Gries, 2016) and objects in Vera'a (Schnell & Barth, 2018).

While Rbrul, other mixed models and random forests are more sophisticated analyses of variation in datasets, none of the forms of analyses are appropriate for variables which have more than two variants, for example, the comparative in English which has three forms *bigger, more big* and *more bigger*. The solution has been to group variants together or to assume two stages of binary choice (Sankoff & Rousseau, 1989; Szmrecsanyi et al., 2016). For example, in the literature on /t/ variation in regional British English varieties there are three variants of /t/: [t], [tʔ] and [ʔ], and various grouping strategies have been used (Drummond, 2011; Foulkes, Docherty, & Watt, 2005; Straw & Patrick, 2007). Nonetheless this style of analysis requires making various assumptions about the data before the analysis, for example, concerning the functional equivalence of some variants. Multinominal analyses are a solution to this problem in that they allow the analysis of more than two variants of a single variable. This solution was identified early on by Sankoff and Labov (1979, pp. 218–219) but was not implemented in linguistic analyses until recently. Multivariant analyses are implemented in a multinomial

logistic regression model using the *mlogit* function from the `mlogit` package. For example, Levshina (2015, pp. 277–288) demonstrates this form of analysis in the English permissive construction which uses the three verbs *allow, let* and *permit*. Levshina (2016) went on to re-implement this analysis in a Bayesian mixed-effects multinomial regression analysis using the R package `RStan`. This departure from frequentist statistics follows a recent trend in the social sciences to move beyond simple hypothesis testing. Bayesian statistics estimates the probability of a hypothesis given the data, rather than the probability of data given a (null) hypothesis.

> While frequentist statistics only allows one to test whether the null hypothesis can be rejected, Bayesian statistics enables one both to test the null hypothesis and to estimate the probability of specific parameter values given the data.
>
> (Levshina, 2016, p. 251)

Other sociolinguists have used the R package `MCMCglmm` (Hadfield, 2010), which is another method for implementing a Bayesian mixed-effects regression analysis. For example, it has been used to examine constraints on subject and non-subject English relative clauses (Grafmiller, Szmrecsanyi, & Hinrichs, 2018), and *is* and *has* contractions in English (MacKenzie, 2020).

While multinomial regression models (either frequentist or Bayesian versions) can be used to analyse multiple *variants* of single variables, they cannot analyse multiple *variables* with multiple variants:

Multinominal regression:
VARIABLE
VARIANT 1
VARIANT 2
VARIANT 3
....

Multivariable regression:
VARIABLE I
VARIANT 1
VARIANT 2
VARIANT 3
....
VARIABLE II
VARIANT 1
VARIANT 2
VARIANT 3
....

VARIABLE III
VARIANT 1
VARIANT 2
VARIANT 3
....

The covariance of multiple linguistic variables is important to theories of language evolution but has been previously difficult to implement. Quite early on, Labov (1972) observed that multiple variables often co-vary according to social categories. In his study of New York City English, Labov observed that five language variables all separately indexed similar categories of social status and levels of formality. The observed similarity in the indexical nature of multiple language variables led Guy (2013, p. 63) to question whether "multiple variables cohere in forming sociolects", a hypothesis which he referred to as 'linguistic coherence'. Some studies of linguistic coherence extend to regional-level clustering of variables, characteristic of dialects (Gregersen & Pharao, 2016; Thelander, 1982; van Hout, Kruijsen, & Gerritsen, 2014). The ability to map from individual-level variation and change over time (micro-level variation) to variation between speech communities over time (macro-level variation) is an important step in understanding language evolution.

Nonetheless, these studies were constrained by the limitations of the type of analysis they used. In all of the studies, each of the variables was examined independently rather than combined within a single analysis. Gregersen and Pharao (2016) use multiple regressions for individual variables + correlations between each pair of variables. Thelander (1982) uses cohesion-coefficients (KK) between each pair of variables. Van Hout, Kruijsen and Gerritsen (2014) use ANOVAs for individual variables + factor analysis on variables, testing the same predictor variables against each of the dependent variables one-by-one. These approaches are not flawed but become untenable when the number of linguistic variables increases. These studies used between three and 12 variables which made a series of single analyses possible. On the other hand, it would have made the Gurindji Kriol study almost impossible given that 185 separate analyses would have been required for each of 185 variables (see below). Using a large number of linguistic variables also increases the chance of identifying incidental associations as clusters, requiring an adjustment of the *p*-value down for multiple tests to reduce the chance of false positives. The solution is either a single analysis that tests co-variation between all the variables and all the predictors simultaneously, or a more efficient approach to summarising the results of independent tests for each

linguistic variable, so that we can detect general patterns in their co-variation (Wright, 1992).

This is the motivation for designing `BayesVarbrul` which provides the tools for sociolinguists to map the flow and clustering of multiple variables (Hua et al., 2022). `BayesVarbrul` can model the effect of social factors on the uptake of variants from multiple variables in a single analysis, showing how some social groups use these variants differently from other groups. It can also be used to model linguistic predictors if they apply to all of the dependent variants in a multivariable analysis. The variables can be categorical (i.e., use or non-use of a variant) or frequency-based, or a combination of both. Variables can also have more than two variants. The Gurindji Kriol dataset only contains categorical variables but a more recent application of `BayesVarbrul` to a Shawi dataset contains both categorical and frequency-based variables (Meakins et al., 2024). `BayesVarbrul` is flexible because it uses a Bayesian framework rather than a frequentist framework. For highly complex models, the use of a Bayesian framework makes it more feasible to fit the model to data. The underlying model in `BayesVarbrul` is a complex hierarchical model, where the lowest hierarchy fits expected usage frequency of each word variant by each speaker to the data. The middle hierarchy fits a regression model with fixed effects on speakers and/or language variables to the expected usage frequencies. The highest hierarchy fits the residuals in the regression model (including random effects) to the Wright-Fisher model that allows social interactions among speakers.

Another motivation for designing `BayesVarbrul` is to model language change in a more realistic way. For example, all the regression models we reviewed above compare patterns in the data to the expected pattern under a logistic distribution (a probability distribution like the normal distribution but for categorical data). This distribution is "chosen for statistical convenience not because theories of language evolution predict that a speaker's usage frequency of a variant would follow a logistic distribution if there were no social effects" (Hua, 2022, p. 41). In contrast, Wright-Fisher models can be used to generate a null model for language data using a process of serial sampling of variants from one time step to another (e.g., Baxter et al., 2009; Baxter et al., 2006; Bentley, Ormerod, & Batty, 2011; Meakins et al., 2019; Reali & Griffiths, 2010), which is conceptually similar to the exemplar theory for language change (Pierrehumbert, 2001). Like `Varbrul` and `Rbrul`, `BayesVarbrul` is essentially a regression model but instead of assuming a logistic distribution as the null model, it uses Wright-Fisher to generate the expected distribution of usage frequencies for a variant, which also allows us to fit residual bias towards using any particular variants or residual covariance among speakers that have not been fully explained by the regression model. In short, because it avoids

choosing the underlying distribution of dependent variables for statistical convenience, this makes BayesVarbrul a better method for modelling language variation and change.

4.2 Summary of Hua (2022) results

Hua (2022) applied BayesVarbrul to the Gurindji Kriol dataset which has 185 linguistic variables (357 variants) used by 78 speakers across three generations.[4] The dataset codes seven social factors for each of the 78 speakers — generation, age, gender, family, household, level of formal school-based education and level of exposure to Gurindji[5] — and asks whether these social factors are significantly associated with differences in the relative use of Gurindji and Kriol variants. The significant effects are summarised in Figure 3.

Hua (2022, p. 46) finds that there is nearly 100% certainty that each generation uses Kriol variants more often than the previous generation (Figure 3A). This generational effect is independent of an age effect which shows that younger speakers use more Kriol relative to Gurindji. Thus the generational effect does not simply represent a gradual loss of the Gurindji variants over time but characterises distinct patterns of language use within particular generations (see also Hua et al., 2022). Hua (2022, p. 48) also finds there is over 70% certainty that speakers with more years of formal school-based education in English use more Kriol variants (Figure 3B) and that there is about 60% certainty that speakers who lived with a fluent Gurindji speaker during childhood use Gurindji variants more often than those who did not (Figure 3C). These results agree with an earlier study that used a multivariate generalisation of the Varbrul analysis (Bromham et al., 2020).

4. See Meakins et al. (2019) for a detailed description of the database and collection of data.

5. See Hua et al. (2022, pp. 86–87) and Bromham et al. (2020, p. 80) for a detailed description of the social categories. In brief, there are three generations, each having a social reality in the historical events which led to the establishment of the Gurindji communities. All sampled speakers are members of 10 extended families, identified by a speaker's pedigree. They are divided into thirty-one residential households according to their principal place of residence. There are two levels of education, according to years of attendance of formal school-based education (conducted in English). There are two levels of exposure to Gurindji, according to whether the speaker has lived with a fluent Gurindji speaker during childhood (Hua, 2022, pp. 45–46).

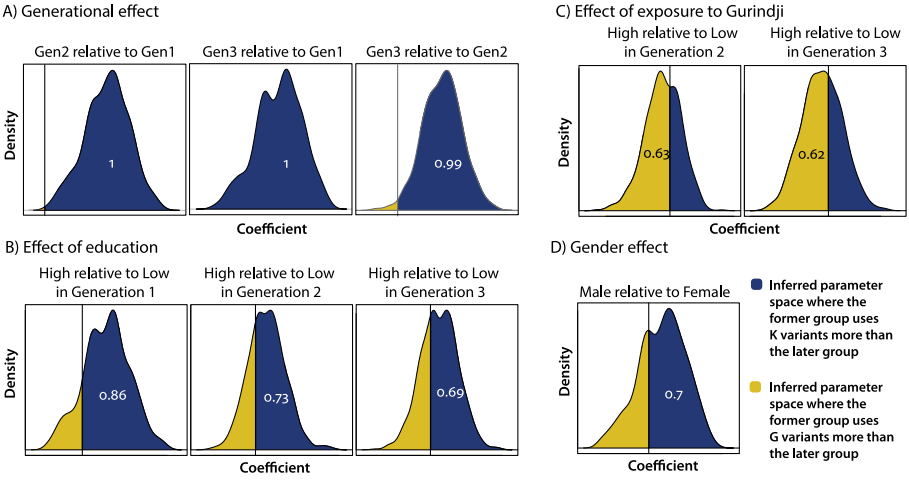

Figure 3. The effect of each social factor on language use in the Gurindji community. The number in the blue area gives the posterior probability, or the proportion of inferred parameter space where the factor increases or decreases the usage of Kriol variants (K) relative to the Gurindji variants (G) (adapted from Hua, 2022, p. 47)

4.3 Discussion of gender results

While Hua (2022) included a number of social factors, here we focus on the effect of gender, which was not explored in detail in the original publication. The `BayesVarbrul` analysis suggests with 70% certainty that in this community, men use more Kriol variants relative to Gurindji variants than women. As Hua (2022, p. 48) notes, this result does not necessarily suggest that gender is a causal factor on language usage. It only suggests that the female speakers in the dataset tend to use more Gurindji variants than male speakers. There could be other factors that correlate with gender and affect language usage, but these factors were not included in the model. Regardless, these results align with the gender literature which suggests that women use more standard variants (often morpho-syntactic or lexical) than men where the standard variants are those which are overtly discussed and positively-evaluated, in this case Gurindji variants. Note that there are no phonetic variables in the Gurindji Kriol dataset to test the other part of the gender paradox which is that women lead change in the non-standard forms (often phonetic variants) which are below the level of awareness.

To put these results in their sociohistorical context, the greater frequency of use of Kriol variants among Gurindji men is likely attributable to their work situation in early 1900s up until the 1980s. Gurindji Country was invaded in the 1860s by British settlers looking for good pasture lands. The latter part of the 1800s was

a battle front with unknown but large numbers of Aboriginal people massacred, including Gurindji people. In the early 1900s, the British invaders realised that Aboriginal people were a potential source of free labour and forced Gurindji and other groups to move to the cattle stations including Wave Hill Station. Life on the cattle stations afforded some peace for Gurindji people but the working and living conditions were extremely harsh. In protest against these conditions, Gurindji people walked off Wave Hill Station in 1966, establishing their own settlement at Daguragu. Their resistance eventually culminated in the transfer of some of Wave Hill Station back to the Gurindji people in 1986 (Berndt & Berndt, 1987; Charola & Meakins, 2016; Hardy, 1968; Lewis, 2012; Ward, 2016).

On the cattle stations, Gurindji men spent extended periods of time with non-Indigenous stockmen in remote stock camps which were less segregated than the 'blacks camps' in the station homestead areas. In the stock camps, the *lingua franca* was the Cattle Station Pidgin which eventually creolised. Gurindji women also spent time interacting with non-Indigenous women in the homestead areas as domestic servants and nannies but there were fewer Gurindji women employed and fewer non-Indigenous women than men present on the stations in general. Gurindji women used Cattle Station Pidgin with non-Indigenous women but spoke Gurindji to their non-Indigenous child charges. There are many non-Indigenous people, now adults, who still have a knowledge of Gurindji as a result of being raised by these Gurindji women. The 'blacks camps' in the station homestead areas where Gurindji women spent much of their time were segregated from the non-Indigenous houses which would have also contributed to the lower exposure to and use of English. In the 1970s, Gurindji boys and girls attended school at the neighbouring Welfare Settlement (now Kalkaringi) which would have levelled the access by boys and girls to English somewhat. Gurindji children continue to be schooled in Kalkaringi to this day. The differential access to and use of English by men and women has also probably levelled in the communities of Daguragu and Kalkaringi, although more men than women work in jobs with non-Indigenous people which require at least functional levels of English. Women continue to be the main caregivers of Gurindji children and the elderly and, in this respect, spend more time around older people who speak Gurindji and in households where Gurindji Kriol is the main language. This demographic situation is a likely explanation for the higher use of Kriol variants by Gurindji boys and men.

The importance of Hua's (2022) modelling of gender in the Gurindji Kriol database is also its handling of missing data, specifically the generation gap of Gurindji men in the dataset. As noted in Section 1, there are no Generation 2 men represented in the dataset, only women. In contrast, there are women represented across all generations. In general, there are also more girls and women ($n = 63$,

81%) represented in the dataset than boys and men ($n = 15$, 19%), as shown in Table 1.

Table 1. Number of female and male participants represented in each generation in the Gurindji Kriol dataset

	Female	Male	Total
Generation 1	11	6	17
Generation 2	22	0	22
Generation 3	30	9	39
Total	**63**	**15**	**78**

Hua (2022) handled the issue of the uneven representation of Gurindji female and male participants by not including an interaction between gender and generation. In other words, the model includes gender and generation effects separately, with the gender effect assumed to be the same across generations and the generation effect the same for both men and women. The significant gender effect of men using Kriol variants more often than women is unlikely due to the generation effect that younger generation uses Kriol variants more often because the ratio of female to male participants in the young generation (Generation 3) is nearly double that of the oldest generation (Generation 1: see Table 1). Of course, we cannot rule out the possibility that the significant gender effect is driven by interaction between generation and gender, for example, if Generation 1 males use Gurindji variants more often than Generation 1 females and Generation 3 males use Kriol variants more often than Generation 3 females. In that case, we could not conclude that Gurindji males generally use more Kriol than Gurindji females. However, because there are roughly the same number of male participants in both Generation 1 and Generation 3 this possibility is small (Table 1). It is more likely that men of at least one generation use more Kriol than women in the same generation.

Hua's (2022) approach is a common practice in regression models to handle missing data. However, for a regression which models a categorical predictor on the use or non-use of language variants, the response variable in a regression model is the proportion of participants in a category that use a particular variant. In this case, there is greater uncertainty in the response variable in categories which have relatively fewer participants. `BayesVarbrul` differs from a regression model because it applies hierarchal modelling that first estimates the usage frequency of each variant by each individual speaker, and then fits a regression model to the estimated usage frequencies. As a result, the response variable in the

regression model is no longer the proportion of speakers using particular variants, but the usage frequencies of each speaker. Thus, the power to detect the effect of a categorical predictor depends less on the sample size and is therefore less affected by uneven numbers of participants in different social categories.

The gender analysis of the Gurindji Kriol dataset shows the power of `BayesVarbrul` to produce meaningful results in datasets that record use or non-use of a variant and includes participants unevenly distributed in different social factors, even in cases where there are no sampled participants in a particular social category (in this case missing participants from the male Generation 2 category). This is an advantage for the development of datasets from speaker communities where social factors are already highly unevenly represented or from older corpora which were not originally designed for quantitative analysis. Note that to develop datasets from older corpora, it is better to record frequency data than use/non-use data. It is important to distinguish two interpretations of the value of 0 (zero). In use/non-use data, zero means that a speaker does not use that variable and does not simply represent missing data (i.e., zero cannot be used to denote that this variable did not appear in the corpus since this is not a 'true zero' but simply missing data).

5. Conclusion

This paper has examined the gender results from Hua (2022) in the context of the literature on gender and language change. Hua (2022) introduced a new method for conducting multivariate analyses of language change, using the Gurindji Kriol dataset as a proof of concept for `BayesVarbrul`. In doing so, Hua (2022) produced new results for understanding language change in the Gurindji community in northern Australia. She showed that gender is significantly associated with different use of language variants, with females using relatively fewer Kriol variants than males, retaining relatively more Gurindji lexical and morpho-syntactic variants, despite a more general shift to Kriol in this community. This result aligns with the Labov's generalisation that women maintain standard forms which are overtly discussed and evaluated. Given that other studies of language use in contact situations have either found no support for Labov's generalisation (e.g., Meyerhoff & Birchfield, 2019), or have found the opposite pattern (e.g., Bakir, 1986), this case study adds another observation of the role of gender in contact-induced language change to the debate. However, this case study also highlights the complexities in separating the within-community effects of gender (e.g., women maintain conservative language variants) to gender-based differences in contact outside the community (e.g., men participate in work that brings them

into more contact with English speakers). Thus, this result may reflect a combination of both fixed social categories (e.g., Labov's view of gender) and more dynamic environmental effects (e.g., Eckert's view of gender).

Funding

Research funded by ARC Centre of Excellence for the Dynamics of Language (CE140100041) to Felicity Meakins.

Abbreviations

NMLZ nominaliser
PST past
LOC locative case

References

Bailey, Guy, Wikle, Tom, Tillery, Jan, & Sand, Lori (1991). The apparent time construct. *Language Variation and Change*, *3*, 241–264.

Bakir, Murtadha (1986). Sex differences in the approximation to Standard Arabic: A case study. *Anthropological Linguistics*, *28*(1), 3–9. https://www.jstor.org/stable/30027941

Bates, Douglas, Maechler, Martin, & Bolker, Ben (2012). *lme4: Linear mixed-effects models using 'Eigen' and S4 classes*. In http://cran.r-project.org/web/packages/lme4/index.html

Batterham, Margaret (2000). The apparent merger of the front centring diphthongs — EAR and AIR — in New Zealand English. In Allan Bell & Koenraad Kuiper (Eds.), *New Zealand English* (pp. 111–145). Amsterdam: John Benjamins.

Baxter, Gareth, Blythe, Richard, Croft, William, & McKane, Allan (2006). Utterance selection model of language change. *Physical Review E*, *73*, 046118.

Baxter, Gareth, Blythe, Richard, Croft, William, & McKane, Allan (2009). Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change*, *21*(2), 257–296.

Bayley, Robert (2013). Variationist sociolinguistics. In Robert Bayley, Richard Cameron, & Ceil Lucas (Eds.), *The Oxford handbook of sociolinguistics* (pp. 11–30). Oxford: Oxford University Press.

Bentley, R. Alexander, Ormerod, Paul, & Batty, Michael (2011). Evolving social influence in large populations. *Behavioral Ecology and Sociobiology*, *65*, 537–546.

Bernaisch, Tobias, Gries, Stefan, & Mukherjee, Joybrato (2014). The dative alternation in South Asian English(es): Modelling predictors and predicting prototypes. *English World-Wide*, *35*(1), 7–31.

Berndt, Ronald, & Berndt, Catherine (1987). *End of an era: Aboriginal labour in the Northern Territory*. AIAS.

Breiman, Leo (2001). Random forests. *Machine Learning*, *45*, 5–32.

Bromham, Lindell, Hua, Xia, Algy, Cassandra, & Meakins, Felicity (2020). Language endangerment: A multidimensional analysis of risk factors. *Journal of Language Evolution*, *5*(1), 75–91.

Bucholtz, Margaret (2003). Sociolinguistic nostalgia and the authentication of identity. *Journal of Sociolinguistics*, *7*(3), 398–416.

Cedergren, Henrietta, & Sankoff, David (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, *50*, 333–355.

Charola, Erika, & Meakins, Felicity (Eds.). (2016). *Yijarni: True stories from Gurindji country*. Aboriginal Studies Press.

Coates, Jennifer (1998). *Language and gender: A reader*. London: Blackwell.

Deshors, Sandra, & Gries, Stefan (2016). Profiling verb complementation constructions across New Englishes: A two-step random forests analysis of ing vs. to complements. *International Journal of Corpus Linguistics*, *21*(2), 192–218.

Dickson, Greg, & Durantin, Gautier (2019). Variation in the reflexive in Australian Kriol. *Asia-Pacific Language Variation*, *5*(2), 171–207.

Dorian, Nancy (2006). Negative borrowing in an indigenous language shift to the dominant national language. *International Journal of Bilingual Education and Bilingualism*, *9*(5), 557–577.

Drummond, Rob (2011). Glottal variation in /t/ in non-native English speech: Patterns of acquisition. *English World-Wide*, *32*(3), 280–308.

Dunn, Vivien, Meakins, Felicity, & Algy, Cassandra (2021). Acquisition or shift: Interpreting variation in Gurindji children's expression of spatial relations. In Enoch Aboh & Cécile Vigouroux (Eds.), *Variation rolls the Dice* (pp. 105–131). Amsterdam: John Benjamins.

Eckert, Penelope, & McConnell-Ginet, Sally (2003). *Language and gender*. Cambridge University Press.

Foulkes, Paul, Docherty, Gerry, & Watt, Dominic (2005). Phonological variation in child-directed speech. *Language*, *81*(1), 177–206.

Gorman, Kyle, & Johnson, Daniel Ezra (2013). Quantitative analysis. In Robert Bayley, Richard Cameron, & Ceil Lucas (Eds.), *The Oxford handbook of sociolinguistics* (pp. 214–240). Oxford: Oxford University Press.

Grafmiller, Jason, Szmrecsanyi, Benedikt, & Hinrichs, Lars (2018). Restricting the restrictive relativizer: Constraints on subject and non-subject English relative clauses. *Corpus Linguistics and Linguistic Theory*, *14*(2), 309–355.

Gregersen, Frans, & Pharao, Nicolai (2016). Lects are perceptually invariant, productively variable: A coherent claim about Danish lects. *Lingua*, *172–173*, 26–44.

Guy, Gregory (2013). The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables. *Journal of Pragmatics*, *52*, 63–71.

Hadfield, Jarrod (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, *33*(2), 1–22.

Hardy, Frank (1968). *The unlucky Australians*. Sydney: Nelson.

Harrell, Frank (2021). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer.

Holmes, Janet (1995). Glottal stops in New Zealand English: An analysis of variants of word final /t/. *Linguistics*, *33*, 433–463.

Holmquist, Jonathan (1985). Social correlates of a linguistic variable: A study in a Spanish village. *Language in Society*, *14*, 191–203.

Hua, Xia (2022). BayesVarbrul: a unified multidimensional analysis of language change in a speaker community. *Journal of Language Evolution*, *7*(1), 40–52.

Hua, Xia, Meakins, Felicity, Algy, Cassandra, & Bromham, Lindell (2022). Language change in multidimensional space: New methods for modelling linguistic coherence. *Language Dynamics and Change*, *12*(1), 78–123.

Johnson, Daniel Ezra (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, *3*(1), 359–383.

Johnson, Daniel Ezra (2019). *RBrul Manual*. Retrieved 4 June 2019 from http://www.danielezrajohnson.com/Rbrul_manual.html

Kral, Inga, Disbray, Samantha, & Green, Jennifer (Eds.). (2022). *Arlkeny map-akert — Australian Indigenous Languages Image Bank*. PARADISEC. http://catalog.paradisec.org.au/repository/AILIB

Labov, William (1963). The social motivation of a sound change. *Word*, *19*, 273–309.

Labov, William (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

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

Levshina, Natalia (2015). *How to do Linguistics with R*. Amsterdam: Benjamins.

Levshina, Natalia (2016). When variables align: a Bayesian multinomial mixed-effects model of English permissive constructions. *Cognitive Linguistics*, *27*, 235–268.

Lewis, Darrell (2012). *A wild history: Life and death on the Victoria River Frontier*. Melbourne: Monash University.

Lui, Guo-Qiang (2012). Social identity and sound change: The case of WO in Shanghainese. *Australian Review of Applied Linguistics*, *35*(2), 203–214.

MacKenzie, Laurel (2020). Comparing constraints on contraction using Bayesian regression modeling. *Frontiers in Artifical Intelligence*, *3*(58), 1–18.

Maclagan, Margaret, Harlow, Ray, King, Jeanette, Keegan, Paul, & Watson, Catherine (2013). The role of women in Māori sound change. In Yousef Elhindi & Theresa McGarry (Eds.), *Gender-linked variation across languages* (pp. 5–21). Champaign: Common Ground Publishing LLC.

McConvell, Patrick, & Meakins, Felicity (2005). Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics*, *25*(1), 9–30.

Meakins, Felicity (2008). Unravelling languages: Multilingualism and language contact in Kalkaringi. In Jane Simpson & Gillian Wigglesworth (Eds.), *Children's language and multilingualism: Indigenous language use at home and school* (pp. 247–264). New York: Continuum.

Meakins, Felicity (2009). The case of the shifty ergative marker: A pragmatic shift in the ergative marker in one Australian mixed language. In Jóhanna Barðdal & Shobhana Chelliah (Eds.), *The role of semantics and pragmatics in the development of case* (pp. 59–91). Amsterdam: John Benjamins.

Meakins, Felicity (2010). The development of asymmetrical serial verb constructions in an Australian mixed language. *Linguistic Typology*, *14*(1), 1–38.

Meakins, Felicity (2011a). Borrowing contextual inflection: Evidence from northern Australia. *Morphology*, *21*(1), 57–87.

Meakins, Felicity (2011b). *Case marking in contact: The development and function of case morphology in Gurindji Kriol*. Amsterdam: John Benjamins.

Meakins, Felicity (2011c). Spaced out: Inter-generational changes in the expression of spatial relations by Gurindji people. *Australian Journal of Linguistics*, *31*(1), 43–77.

Meakins, Felicity (2012). Which Mix? — Code-switching or a mixed language — Gurindji Kriol. *Journal of Pidgin and Creole Languages*, *27*(1), 105–140.

Meakins, Felicity (2015). From absolutely optional to only nominally ergative: The life cycle of the Gurindji ergative suffix. In Francesco Gardani, Peter Arkadiev, & Nino Amiridze (Eds.), *Borrowed morphology* (pp. 189–218). Berlin: Mouton de Gruyter.

Meakins, Felicity (2016). No fixed address: The grammaticalisation of the Gurindji locative as a progressive suffix. In Felicity Meakins & Carmel O'Shannessy (Eds.), *Loss and renewal: Australian languages since colonisation* (pp. 367–396). Berlin: Mouton de Gruyter.

Meakins, Felicity (2023). Quantifying the language dynamics of bilingual communities. *Linguistic Approaches to Bilingualism*, *13*(1), 76–82.

Meakins, Felicity, & O'Shannessy, Carmel (2005). Possessing variation: Age and inalienability related variables in the possessive constructions of two Australian mixed languages. *Monash University Linguistics Papers*, *4*(2), 43–63.

Meakins, Felicity, & Wigglesworth, Gillian (2013). How much input is enough? Correlating comprehension and child language input in an endangered language. *Journal of Multilingual and Multicultural Development*, *34*(2), 171–188.

Meakins, Felicity, & Algy, Cassandra (2016). Deadly reckoning: Changes in Gurindji children's knowledge of cardinals. *Australian Journal of Linguistics*, *36*(4), 479–501.

Meakins, Felicity, Jones, Caroline, & Algy, Cassandra (2016). Bilingualism, language shift and the corresponding expansion of spatial cognitive systems. *Language Sciences*, *54*, 1–13.

Meakins, Felicity, Green, Jennifer, & Turpin, Myfany (2018). *Understanding linguistic fieldwork*. London: Routledge.

Meakins, Felicity, Hua, Xia, Algy, Cassandra, & Bromham, Lindell (2019). Birth of a contact language did not favour simplification. *Language*, *95*(2), 294–332.

Meakins, Felicity, Rojas Bercia, Luis Miguel, Bromham, Lindell, & Hua, Xia (2024). Modelling the emergence of new dialects in time and space. [Manuscript submitted for publication].

Meakins, Felicity, & Wilmoth, Sasha (2020). Overabundance resulting from language contact: Complex cell-mates in Gurindji Kriol. In Peter Arkadiev & Francesco Gardani (Eds.), *The complexities of morphology* (pp. 81–104). Oxford: Oxford University Press.

Meyerhoff, Miriam (2000). The emergence of creole subject-verb agreement and the licensing of null subjects. *Language Variation and Change*, *12*(2), 203–230.

Meyerhoff, Miriam (2002). Formal and cultural constraints on optional objects in Bislama. *Language Variation and Change*, *14*(3), 323–346.

Meyerhoff, Miriam (2008). Men argue, but the women doz trace: Gender and language variation in Bequia (St Vincent and the Grenadines). *Sargasso*, *1*, 115–132.

Meyerhoff, Miriam (2010). *Introducing sociolinguistics*. London: Routledge.

Meyerhoff, Miriam (2015). Turning variation on its head: Analysing subject prefixes in Nkep (Vanuatu) for language documentation. *Asia-Pacific Language Variation*, *1*(1), 78–108.

Meyerhoff, Miriam, & Birchfield, Alexandra (2019). Gender and language contact: How gender is/isn't marked in language contact. In Jeroen Darquennes, Joe Salmons, & Wim Vandenbussche (Eds.), *Language contact: An international handbook* (pp. 246–256). Mouton de Gruyter.

Nagy, Naomi (2011). Lexical change and language contact: Faetar in Italy and Canada. *Journal of Sociolinguistics*, *15*(3), 366–382.

Ochs, Elinor (1992). Indexing gender. In Alessandro Duranti & Charles Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 335–358). Cambridge University Press.

Pierrehumbert, Janet (2001). Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.

Pinheiro, José, & Bates, Douglas (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Reali, Florancia, & Griffiths, Thomas (2010). Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. *Proceedings of Royal Society B*, *277*, 429–436.

Sankoff, David, & Labov, William (1979). On the uses of variable rules. *Language in Society*, *8*(2–3), 189–222.

Sankoff, David, & Rousseau, Pascale (1989). Statistical evidence for rule ordering. *Language Variation and Change*, *1*, 1–18.

Sankoff, David, Tagliamonte, Sali, & Smith, Eric (2005). GoldVarb X: A variable rule application for Macintosh and Windows. http://individual.utoronto.ca/tagliamonte/goldvarb.html

Schnell, Stefan, & Barth, Danielle (2018). Discourse motivations for pronominal and zero objects across registers in Vera'a. *Language Variation and Change*, *30*(1), 51–81.

Sloan, Bodean, Meakins, Felicity, & Algy, Cassandra (2022). Intergenerational changes in Gurindji Kriol: Comparing apparent-time and real-time data. *Asia-Pacific Language Variation*, *8*(1), 1–31.

Straw, Michelle, & Patrick, Peter (2007). Dialect acquisition of glottal variation in /t/: Barbadians in Ipswich. *Language Sciences*, *29*(2–3), 385–407.

Szmrecsanyi, Benedikt, Biber, Douglas, Egbert, Jesse, & Franco, Karlien (2016). Toward more accountability: Modeling ternary genitive variation in Late Modern English. *Language Variation and Change*, *28*(1), 1–29.

Tagliamonte, Sali (2006). *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.

Tagliamonte, Sali, & Baayen, Harald (2012). Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, *24*, 135–178.

Thelander, Mats (1982). A qualitative approach to quantitative data of speech variation. In Suzanne Romaine (Ed.), *Sociolinguistic variation in speech communities* (pp. 65–83). London: Edward Arnold.

Travis, Catherine, Grama, James, & Purser, Ben (2023). Stability and change in (ing): Ethnic and grammatical variation over time in Australian English. *English World-Wide*, *44*(3), 435–469.

Trudgill, Peter (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, *1*, 179–195.

Trudgill, Peter (1983). *On dialect: Social and geographical perspectives*. London: Blackwell.

van den Bos, Jackie, Meakins, Felicity, & Algy, Cassandra (2017). Searching for "Agent Zero": The origins of a relative case system. *Language Ecology*, *1*(1), 4–24.

van Hout, Roeland, Kruijsen, Joep, & Gerritsen, Marinel (2014). Exosmosis along the Romance-Germanic language border in Belgium. The diffusion of French borrowings in the Dutch dialects of Haspengouw. In Raquel Casesnoves-Ferrer, Montserrat Forcadell Guinjoan, & Nuria Gavaldà-Ferré (Eds.), *Ens queda la paraula: Estudis de lingüística aplicada en honor a M. Teresa Turell* (pp. 197–223). Institut universitari de lingüistica aplicada Universitat Pompeu Fabra.

Walker, James, & Meyerhoff, Miriam (2020). Pivots of the Caribbean? Low-back vowels in eastern Caribbean English. *Linguistics*, *58*(1), 109–130.

Ward, Charlie (2016). *A handful of sand*. Melbourne: Monash University Press.

Wright, S. Paul (1992). Adjusted p-values for simultaneous inference. *Biometrics*, *48*, 1005–1013.

抽象的 (Chinese)

社会语言学家一直在努力解决不同性别的人如何以不同的使用语言的方式来构建他们的身份。他们研究发现了两种看似相互矛盾的现象，称为性别悖论。一方面女性更愿意用公认的标准形式来使用语言，另一方面她们又无意识地引领了非标准的形式来使用语言。这个性别悖论的现象在西方社会语言内部变化的时候很显著，但在非西方社会语言内部变化或在与其他语言的接触变化中并不那么显著。本文探讨了性别在北澳大利亚土著社区中语言变化的作用。该社区正在从使用古林吉语转向使用克里奥尔语。我们的数据由三代 78 位参与者使用的185 个语言变量组成。我们演示了如何应用 BayesVarbrul 来分析这个数据。结果表明，尽管整个社区普遍地转向使用克里奥尔语，但女性比男性更愿意保留使用古林吉语。

Address for correspondence

Felicity Meakins
School of Languages and Cultures
University of Queensland
Brisbane QLD 4072
Australia

f.meakins@uq.edu.au
https://orcid.org/0000-0003-4487-4351

Co-author information

Lindell Bromham
Macroevolution and Macroecology
Research School of Biology
Australian National University
Lindell.bromham@anu.edu.au

Xia Hua
Mathematical Sciences Institute
Australian National University
Xia.hua@anu.edu.au

Publication history

Date received: 14 December 2023
Date accepted: 17 December 2024
Published online: 3 March 2025