Title:- Birth of a contact language did not favour simplification

Authors:-
Felicity Meakins[1,3], Xia Hua[2,3], Cassandra Algy[4], Lindell Bromham[2,3]

[1]School of Languages and Cultures
University of Queensland
Brisbane QLD 4072 Australia

[2]Macroevolution and Macroecology
Research School of Biology
Australian National University
Canberra ACT 0200 Australia

[3]Centre of Excellence for the Dynamics of Language
Australian National University
Canberra ACT 0200 Australia

[4]Karungkarni Arts
Kalkaringi
via Katherine NT 0852 Australia

Corresponding author:-
Felicity Meakins
f.meakins@uq.edu.au

Birth of a contact language did not favour simplification

Abstract:-*

We report on the rapid birth of a new language in Australia, Gurindji Kriol, from the admixture of Gurindji and Kriol. This study is the first investigation of contact-induced change within a single speaker population which uses multiple variants. It also represents an innovative modification of the Wright-Fisher population genetics model to investigating temporal change in linguistic data. We track changes in lexicon and grammar over three generations of Gurindji people, using data from 78 speakers coded for their use of Gurindji, Kriol and innovative variants across 120 variables (with 292 variants). We show that the adoption of variants into Gurindji Kriol was not random, but biased towards Kriol variants and innovations. This bias is not explained by simplification, as is often claimed for contact-induced change. There is no preferential adoption of less complex variants and, in fact, complex Kriol variants are more likely to adopted over simpler Gurindji variants.

*We thank Nick Evans, Russell Grey, Steve Levinson, Miriam Meyerhoff, Patrick McConvell, Matt Hahn and Luis Miguel Rojas-Berscia for their comments on the manuscript.

Keywords:- language contact, morphological complexity, linguistic variation, creolization, Wright-Fisher, population genetics

1. INTRODUCTION

Language contact processes such as creolisation, borrowing and shift have been shown to have a radically reductive effect on morphological complexity (Dorian 1989; Gardani, Arkadiev, & Amiridze 2014; Matras & Sakel 2007; McWhorter 2001; Parkvall 2008), although the universality of this loss of complexity has been disputed (Bakker et al. 2011; DeGraff 2005; Matras 2003; Plag 2003a, 2003b; Trudgill 2001). Simplification in contact-induced change occurs either by the preferential adoption of simpler elements (e.g. borrowing prepositions over case morphology), or by the simplification of complex language features already present in a language (e.g. syncretism of case in language shift scenarios). The case for simplification has often been limited by the relatively small number of language features studied, usually inflectional morphology, which have been chosen for their known patterns of simplification (§2). Ideally, the process of contact-induced change should be studied over a large number of language elements, chosen arbitrarily rather than for their known patterns of change, and from a wide range of speakers in the population.

In this study, we test the generality of morphological simplification in a situation of contact-induced change by examining the genesis of Gurindji Kriol. Gurindji Kriol is a mixed language, spoken in northern Australia, which formed in the last 40 years from pervasive code-switching practices in the context of a widespread shift to Kriol (McConvell & Meakins 2005; Meakins 2011b). We apply the Wright-Fisher models, derived from evolutionary biology, to 120 language features (with 292 variants) drawn from a 96 hour corpus representing 78 speakers across three generations of Gurindji people (Meakins 2011a, 2011b). Our application of a Wright-Fisher model to multi-generational language change includes novel modifications of classic population genetics methods in order to allow analysis of temporal change in linguistic data.

Gurindji Kriol draws structural and lexical elements from both Gurindji (Ngumpin-Yapa, Pama-Nyungan) and Kriol (an English-based creole language). Morphological simplification processes are thought to have less impact on mixed languages than they do in other contact situations, due to the high level of bilingualism in the communities where mixed languages develop (Matras 2003; Meakins 2013b). Nonetheless, most (if not all) mixed languages, including Gurindji Kriol, form in situations of language shift which involve extensive code-switching or borrowing. These contact scenarios are generally associated with simplification, i.e. the loss of morphological marking and distinctions, and restrictions on morphological transfer. While the mixed language literature tends to focus on what is preserved rather than lost (Bakker 2015; Matras 2003; Meakins 2013b; Thomason 1997), understanding the competing pressures of language shift and maintenance within these complex ecologies has ramifications for notions of simplification in other types of contact-induced change. Moreover, taking a large sample of language elements rather than focussing on individual features allows a broader assessment of the processes underlying mixed language development.

Because Gurindji Kriol has developed rapidly over three generations, all stages of its development are still represented in the current Gurindji population. In fact, most houses in the community contain members of three generations. These generations represent different stages in the formation of Gurindji Kriol (cf. the apparent-time hypothesis, Bailey et al. 1991; Labov 1963). The 'grandparent' generation (Generation 1) are bilingual in Gurindji and Kriol, and code-switch between the two. The 'parent' generation (Generation 2) cannot speak either language fluently, instead they use a single language, Gurindji Kriol. However, for many linguistic features, individual members of this generation differ in whether they use the Gurindji variant or Kriol variant or both. The ‘child’ generation (Generation 3) speak a single shared language where all speakers tend to share the same variants for each linguistic feature, some of which are derived from Gurindji, some from Kriol, and some of which are innovations (Meakins 2008) (§3).

We examine the usage of 292 Gurindji, Kriol and innovative variants in the expression of 120 language features for each of 78 Gurindji people, representing Generation 1 (>40 years), Generation 2 (20-40 years) and Generation 3 (8-14 years). The 120 language features are chosen for the fact that they vary rather than for their specific patterns of change (such as simplification) in order not to bias the analysis. The language features include both lexical items and morpho-syntactic variables. For example, ‘search’ may be expressed using the variants *luk-aran* (Kriol) or *warlakap* (Gurindji), or both. Similarly, the locative may be expressed using the variants *langa* (Kriol), *-ta* (Gurindji) and/or double-marking (innovation). All variants are coded for three levels of morphological complexity (H=high, M=medium, L=low). For example, bound morphemes such as case suffixes (e.g. *–ta*) are coded as more complex than free forms such as prepositions (e.g. *langa*: see Table 2). Words which are bimorphemic (e.g. *luk-aran*) are coded as more complex than words consisting of one morpheme (e.g. *warlakap*: §4.4).

This snapshot of variant use across generations give us an opportunity to test ideas about contact-induced language change. We model language change across the three generations of Gurindji people by adapting the Wright-Fisher model (Fisher 1956; Wright 1931) to evaluate whether the adoption of elements into the mixed language is biased by source language (e.g. preference for Kriol terms over Gurindji), as well as evaluating whether simpler variants are preferentially adopted over more complex forms (§4.5). While the Wright-Fisher model was originally formulated in terms of the change in frequency of genetic variants in a randomly mating biological population, more generally it describes a process of unbiased, serial sampling of variants from one time step to another. While this model has been shown to simulate plausible patterns of language change (Baxter et al. 2006; Baxter et al. 2009; Bentley, Ormerod, & Batty 2011; Blythe 2012; Kirby, Griffiths, & Smith 2014; Newberry et al. 2017; Reali & Griffiths 2010), this is the first application of Wright-Fisher models to a large sample of language variants within a single speaker community in order to model broad-scale language change over time. We show that the adoption of language elements into the mixed language Gurindji Kriol has not been random, but has been biased towards Kriol and innovative variants. Importantly, we show that this bias is not explained by simplification, because there is no preferential adoption of less complex variants into Gurindji Kriol.

2. Complexity and language contact

One of the oft-claimed results of contact-induced change is the reduction of complexity, particularly in overt morphological marking and paradigmatic distinctions (see Miestamo, Sinnemäki, & Karlsson 2008 for a collection of papers). The contact literature generally examines the *absolute* complexity of a linguistic system, which is the number of syntagmatic and paradigmatic distinctions described for a language, rather than *relative* complexity, which is a measure of user and learner difficulty or computational effort (Ackerman, Blevins, & Malouf 2009; Ackerman & Malouf 2013; Dahl 2004; Kusters 2008; Miestamo 2006, 2008; Nichols 2009).

Morphology provides the usual means of measuring absolute complexity in a language (Baerman, Brown, & Corbett 2015: 4). For example, languages which have morphological structure, i.e. polysynthetic or agglutinating languages, are considered to be more complex than languages which do not, i.e. isolating languages (Anderson 1992; Sapir 1921). Extreme cases of contact-induced language genesis such as creolisation have been claimed to have a radically reductive effect on morphology (McWhorter 1998, 2001), however this claim has not gone undisputed. There have been a number of reviews of the evidence, which have both supported the claim of morphological reduction in contact-induced language genesis or presented a counter-argument in response to this claim (e.g. Parkvall 2008; Plag 2003a; 2003b; and the commentary section in Linguistic Typology (2001) Vol 5).

Mixed languages are often considered to be the exception to the assumed morphological simplification processes associated with language contact. The perceived lack of simplification may be due to features of the speaker communities in which they form. For example, they may be due to the high level of bilingualism in the communities where they develop (Meakins 2013b). The way that the source languages combine in mixed language genesis also might affect the degree to which simplification occurs. Most mixed languages are characterised as combining the morpho-syntactic structure of one language with significant amounts of lexical content from another language; these are referred to as G(rammar)-L(exicon) mixed languages (Matras & Bakker 2003). More interestingly, some mixed languages, including Michif, Mednyj Aleut, Gurindji Kriol and Light Warlpiri, contain morphology from both languages, creating what Myers-Scotton (2002) calls a 'composite matrix language'. Matras (2003) suggests that a particular feature of these mixed languages is the seemingly unconstrained transfer of grammatical elements that are usually considered resistant to borrowing and are therefore often labelled as 'loan proof', such as inflectional morphology. For example in Michif, verbal inflections are derived from Cree and the nominal system preserves French plural morphology and adjectival agreement (Bakker 1997). Mednyj Aleut consists of many Aleut nominal inflections, including two case distinctions and various derivational suffixes, as well as Russian finite verbal inflectional morphology (Golovko 1994; Thomason 1997). And, as will be discussed in §3, Gurindji Kriol combines the nominal structure of Gurindji including case marking, with the verbal structure from Kriol. In this respect, mixed languages may be less prone to morphological simplification than other contact language types such as creoles.

Nonetheless, most (if not all) mixed languages develop in the context of language shift, where code-switching or borrowing is prevalent. In these contact situations, morphology is observed to be susceptible to simplification, loss and transfer restrictions. For example, morphological marking and paradigmatic distinctions are lost early in cases of language attrition in individuals and community-level language loss. In this situation, periphrastic constructions are often preferred over functionally-equivalent affixes or clitics (Austin & Simpson 2007; Dorian 1976; Gal 1989; Janse & Tol 2003; Sallabank & Austin 2011). For example, de Groot (2008: 194-207) notes that Hungarian spoken outside of Hungary shows a tendency for analytic modal, reflexive and causative Hungarian forms to be used rather than synthetic equivalents. Similarly, analytic forms from another language often replace synthetic forms in shift situations. For example, in another case of contact-induced language change in an Australian language, Dyirbal spatial case-markers were incrementally replaced by English prepositions (Schmidt 1985: 52-54). Affixes and clitics are also less likely to be borrowed or switched into another language's grammatical frame than free form equivalents such as prepositions or adverbs (e.g. Gardani 2008; Myers-Scotton 2002; Thomason & Kaufman 1988; Weinreich 1974 [1953]). For example, in a cross-linguistic study of 27 pairs of languages, Matras (2007: 42) finds no instances where case-markers are borrowed, although equivalent functional categories which are expressed by adpositions are susceptible to borrowing. As will be discussed in §3, Gurindji Kriol was created in a situation of language shift via an intermediate code-switching stage where one of the contributing languages is a creole language.

Where bound morphemes do remain in situations of language contact, different dimensions of complexity are affected. Morphological marking which does not rely on the broader clause, such as derivational morphology (e.g. nominalisers, diminutives) and inherent inflection (e.g. gender, TAM), appear to be more resilient than contextual inflection (e.g. case, or person, number or gender agreement) whose presence is dictated by the morpho-syntax of the clause (Gardani 2008; Roberts & Bresnan 2008). Morphological paradigms (e.g. case, person, TAM, number and gender) are also susceptible to change (Anderson 2015). Syncretism, allomorphic simplification and increased paradigmatic regularity are all observed outcomes of contact-induced change (e.g. Janse & Tol 2003). For example, the case system of Pennsylvania German shows syncretism in nominative and accusative case (articles and agreement), thereby creating a 'common' case (Huffines 1989). Finally, paradigms themselves are considered the least 'borrowable' of all morphological categories, although see Seifart (2012) for an exception.

Most of the claims of simplification (or maintenance of complexity, in the case of mixed languages) in situations of contact-induced language change have been supported by focussing on aspects of morphology where simplification (or maintenance) has occurred. Indeed, in studies of creole languages, claims of simplification have been rapidly followed by counter-claims of complexification, which suggests that the selection of morphological features for study is largely theoretically-driven (e.g. McWhorter 2001, and extensive rebuttals). We approach this question from a more neutral perspective, examining linguistic features which show variation in a language system undergoing contact-induced change, and asking whether the selection of variants is driven by an orientation to morphological simplicity.

Modelling linguistic variability and language change has long been the preoccupation of quantitative sociolinguistics (Labov 1994a, 1994b) and, more recently, probabilistic models of syntax (e.g. Bresnan 2007). These approaches use mixed models or logistic regression (e.g. (Va)rbrul) to measure the predictability of the use of a linguistic variable against a range of grammatical and social factors within a speaker population. Language change is inferred where differences in variation occur across generations of speakers, referred to as the 'apparent-time hypothesis' (cf. Bailey et al. 1991; Labov 1963). This approach models language variation and change for single variables. The study presented here scales up the object of study to examining multiple variables simultaneously which allows a broader view of variability and change in a linguistic system.

Other studies have examined multiple linguistic features within a single analysis with the question of language contact and complexity in mind (Bakker et al. 2011; Blasi, Michaelis, & Haspelmath 2017; Parkvall 2008). For example, Parkvall (2008) compared 32 creole languages with 153 non-contact languages across 46 features from the WALS database. Languages were scored for complexity, ranging from 0.62 for the most complex language (Burushaski) to 0.15 for the least complex language (Sango). Creoles fell within the range of 0.15-0.33 which was below the average range of complexity (0.38-0.39). Bakker and colleagues (2011) used the same dataset, applying different tree and network models. They found that creole languages cluster separately from the non-contact languages, but were not differentiable on the basis of complexity. These studies are interesting synchronic comparisons of the complexity of creole languages and non-contact languages, however little can be inferred about the diachronic process of language change, creolisation in this case. Moreover, it is difficult to come to firm conclusions either about the relative simplification of creoles and mixed languages compared to their source languages, or the absolute lack of complexity in such languages, since a number of non-contact languages are ranked below creole languages in complexity. Our study examines a large number of features within a single contact language at different stages of its formation. This diachronic view allows us to ask questions about the processes which contributed to its development, specifically simplification.

3. GURINDJI KRIOL

Gurindji Kriol is a newly-formed mixed language spoken by Gurindji people under the age of 40 years in northern Australia around 900 km south-west of Darwin. It combines Gurindji, a member of the Ngumpin-Yapa subgroup of the Pama-Nyungan family and the traditional Australian language of the region, and Kriol, an English-lexifier creole language spoken across much of northern Australia. In Figure 1, blue shading represents the area where Kriol is spoken, yellow shading indicates where Gurindji is spoken, i.e. in the communities of Kalkaringi and Daguragu, and purple shading shows where Gurindji Kriol is spoken, i.e. Kalkaringi and Daguragu but also Pigeon Hole and Yarralin, which are traditionally Bilinarra and Ngarinyman-speaking communities, respectively.

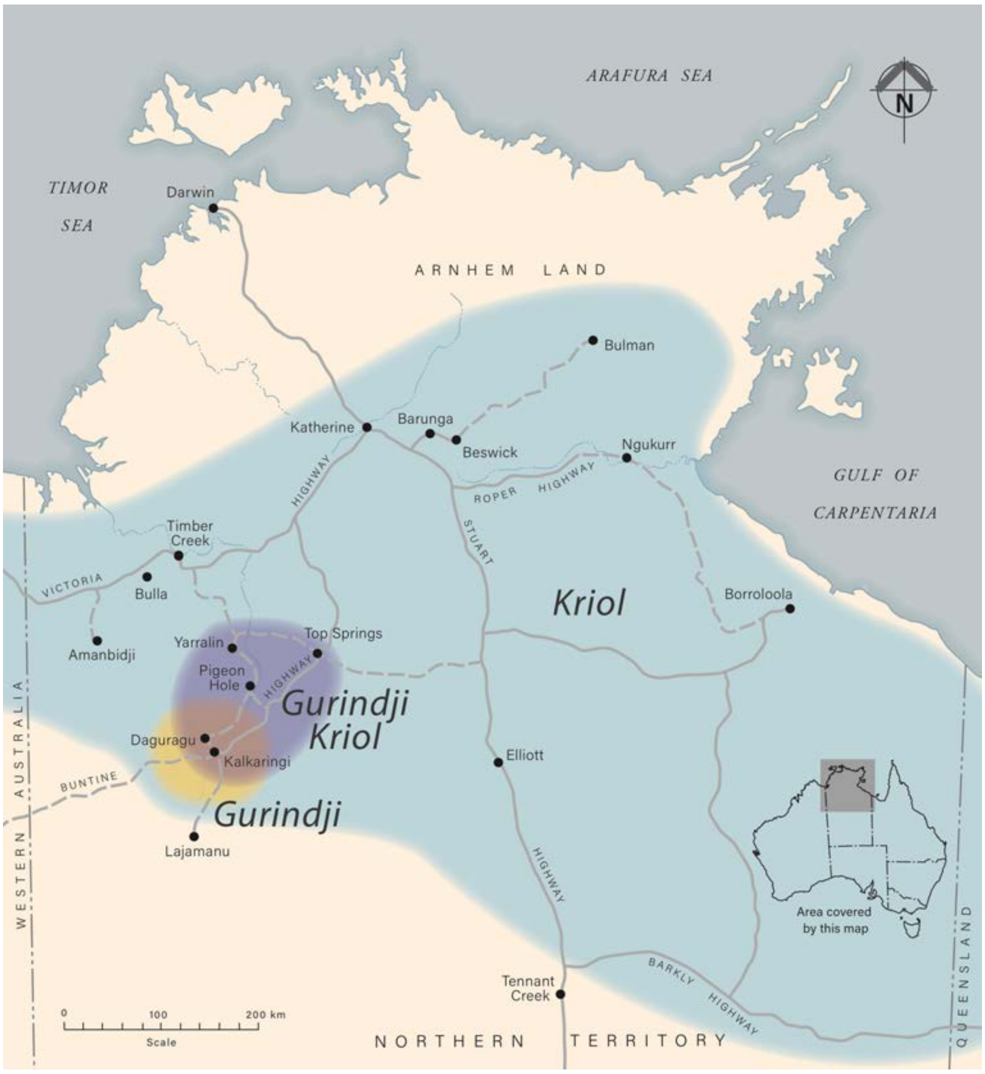

Figure 1 Areas in northern Australia where Gurindji, Kriol and Gurindji Kriol are spoken (Map: Brenda Thornley 2017)

Gurindji Kriol originated from the socio-historical circumstances associated with the colonisation of Gurindji country by European pastoralists. From the late 1800s onwards, the traditional lands of the Gurindji were seized by colonists for cattle grazing. During the establishment of the cattle stations, colonists killed many Gurindji people, with the survivors brought to work on the stations in slave-like conditions with other Aboriginal groups (Charola & Meakins 2016). The cattle station owners communicated with the Aboriginal workers in a cattle station pidgin which later developed into the creole now referred to as ‘Kriol’. The Gurindji added this language to their communicative repertoire and intra-sentential code-switching between Gurindji and Kriol became the norm and provided fertile ground for the formation of a mixed language (McConvell & Meakins 2005; Meakins 2011b). Despite the fact that Gurindji Kriol finds its origins in these code-switching practices, it can be distinguished from code-switching in a number of ways: (i) the mixing patterns in the mixed language are more consistent between speakers, (ii) the mixed language is the main language of acquisition despite the fact that children still receive monolingual Gurindji input and hear code-switching, and (iii) the mixed language has created innovative forms from Gurindji and Kriol which are not present in either source language (Meakins 2012).

As explained in §1, all stages of the development of Gurindji Kriol are still represented in the current Gurindji population. To characterise this change over time, we define three different generations of speakers. Generation 1 are the 'grandparent' generation, people over the age of 40 years who are bilingual in Gurindji and Kriol, and code-switch between these languages. They are also most likely to be the generation who creolised the cattle station pidgin. Generation 2 are the 'parent' generation, Gurindji people 20-40 years old who speak Gurindji Kriol and cannot speak either Gurindji or Kriol fluently. Members of Generation 2 use either a Gurindji, Kriol or innovative variant, or a combination of these, for many linguistic features. Generation 3 is the 'child' generation (<20 years old) who tends to share the same variants for each linguistic feature, whether they are derived from Gurindji, Kriol or are innovations. These three generations represent a diachronic snapshot of the development of Gurindji Kriol, allowing us to test the hypothesis about contact-induced change and simplification.

Typologically, Gurindji Kriol is classified as a V(erb)-N(oun) mixed language (cf. Matras & Bakker 2003) because it exhibits a structural split between the noun phrase system and the verb phrase system. The core verbal grammar is largely derived from Kriol, e.g. TAM auxiliaries such as *bin* 'past, (< been)' and cross-referencing pronouns such as *yutubala* '2DU, (< you two fellows)'. Much of the nominal grammar, such as the case marking, comes from Gurindji, while other suffixes and prepositions find their origin in Kriol. Other forms in Gurindji Kriol are innovative, for example *-tu* which takes its form from the Gurindji ergative case marker, but has transformed into a nominative case marker under the influence of the Kriol nominative-accusative case system (Meakins 2015). The lexicon of Gurindji Kriol is mixed: nouns and verbs have been adopted from both Kriol and Gurindji. An overview of the structure of Gurindji Kriol can be found in Meakins (2013a) and many aspects of the language including case, spatial relations, phonology and verbs have been described in more detail elsewhere, often in comparison with another mixed language, Light Warlpiri, which is spoken in the neighbouring community of Lajamanu (see Figure 1) (Jones, Meakins, & Muawiyath 2012; Meakins 2010, 2011a; Meakins, Jones, & Algy 2016; Meakins & O'Shannessy 2012). The example below in Figure 2 demonstrates the split schematically (Gurindji features are given in yellow, Kriol features in blue, and innovations in green).

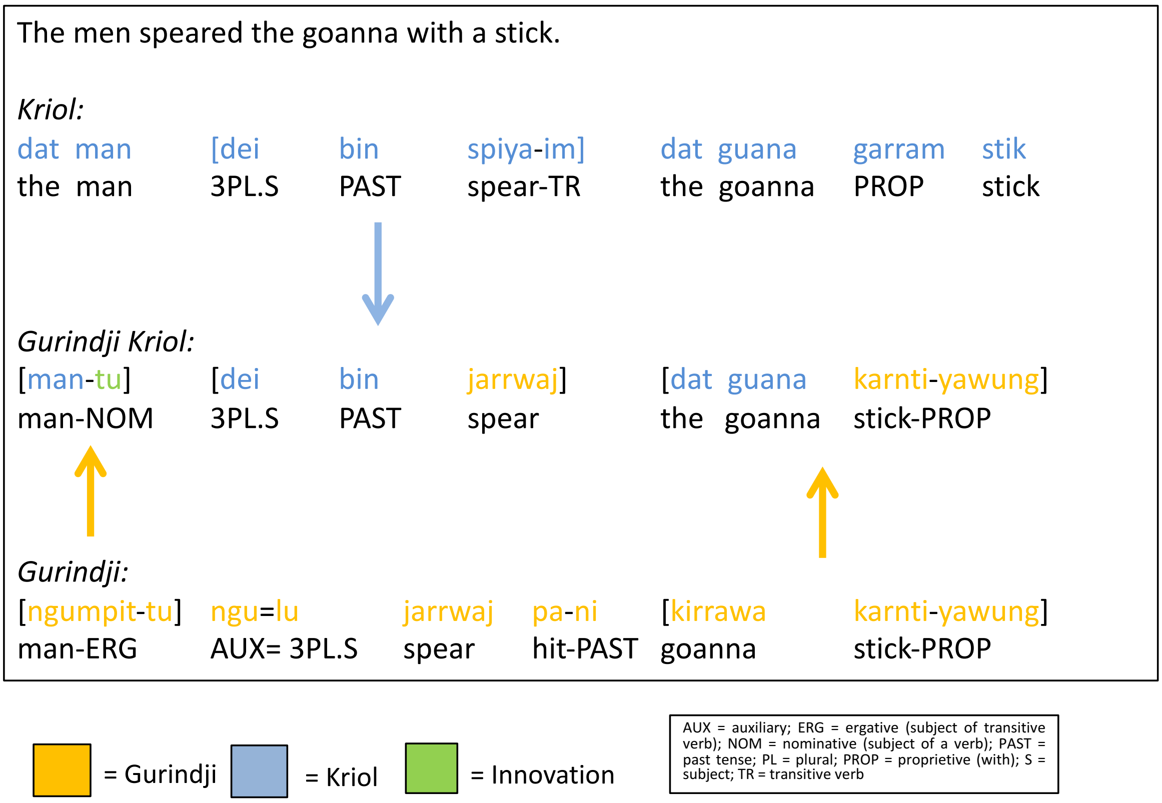

Figure 2 Annotated example of structural and lexical sources of Gurindji Kriol (Meakins & O'Shannessy, 2010: 1697)

Some simplification has certainly occurred in the formation of the mixed language. For example, syncretism has occurred in the marking of transitive and intransitive subjects which precipitated a shift from an ergative system to a nominative system (Meakins 2015) and the nominative allomorphs have reduced from seven forms to two (Meakins 2011b: 23, 30). Table 1 shows this shift from the Gurindji allomorphs used by Generation 1 (left) to the Gurindji Kriol allomorphs used by Generation 2 (right).

Table 1 Allomorphic reduction and core argument shift in Gurindji Kriol

GURINDJI (ERGATIVE)			**GURINDJI KRIOL (NOMINATIVE)**	
V-FINAL	DISYLLABIC	*-ngku*	**V-FINAL**	*-ngku*
	MULTISYLLABIC	*-rlu*		
C-FINAL	PERIPHERAL	*-kulu*	**C-FINAL**	*-tu*
	CORONAL	*-tu, -rtu*		
	LIQUID	*-u*		
	PALATAL	*-ju*		

Other areas of the grammar have complexified, for example nominative marking has become optional which means that speakers no longer simply apply case marking obligatorily to subjects, but must attend to other syntactic, semantic and pragmatic features in the clause and the broader discourse context to use the case marker appropriately (Meakins 2009; Meakins & O'Shannessy 2010). For example, in (1) the same semi-transitive verb *kuli* 'attack' is used in both clauses, but the agent in the first clause is unmarked, and in the second clause receives nominative marking. This non/use of the nominative marker contrasts the two agents (Gurindji elements in italics, Kriol elements in plain font).[i]

(1) a.

Dat	***mukmuk***	bin	*kuli*	la	im
the	owl	PST	attack	LOC	3SG

'The owl attacked (the boy).'

b.

Dem	**bi-*ngku***	*kuli*	la	im	dat	*warlaku=warla*
those	bee-NOM	attack	LOC	3SG	the	dog=CONTR

'(And) the bees really went for the dog instead.' (Meakins 2009: 82)

Meakins and Wilmoth (2019) argue that optional case marking of this kind is a common result of language contact and is an example of morphological complexification. The existence of instances of both morphological simplification and complexification in Gurindji Kriol shows that it is easy to cherry pick individual morphological features to support the argument for either simplification or complexification. Indeed the complexity and language contact literature can be characterised by the proliferation of individual examples and counter-examples to the claims of simplification in contact-induced change. The aim of this paper is to conduct a general investigation into whether the genesis of the mixed language Gurindji Kriol, which formed in a situation of contact and shift, is characterised by an overall process of simplification. In order to do this, we examine multiple features simultaneously. These features are only chosen only because they are variable, rather than because of specific patterns of variation. This unbiased selection of variable features avoids the ascertainment bias that arises from focusing on features known to be undergoing a specific pattern of change, such that they are noticeable above any general patterns of language change.

4. Did simplification shape Gurindji Kriol?

To investigate the role that simplification has played in the formation of the mixed language Gurindji Kriol, we used Wright-Fisher models to investigate patterns of change across multiple linguistic features and their variants. These features were chosen only because they show variation within the speaker community, allowing us to investigate whether subsequent generations of Gurindji people oriented to simpler variants as the new contact language emerged.

4.1 Data

The data consists of 120 linguistic features (with 292 variants) used by 78 Gurindji people across three generations. Sampling of speakers is proportional to the numbers of people in each generation in the speaker community, with 18 people from Generation 1, 38 from Generation 2 and 22 from Generation 3, so that our dataset reflects the current demographic structure of the speaker community. For each of the 78 speakers, we recorded their use of variants for each of the 120 linguistic features to determine whether the speakers used the Gurindji variant, Kriol variant or both to express particular lexical concepts (such as 'tree' or 'search') or grammatical functions (such as subject, location, tense etc). We also recorded the use of innovations (I), such as hybrid variants that fuse the form from one source language with the grammatical pattern of the other: these innovations cannot be coded as either solely Gurindji or Kriol but as new forms derived from the mixture of both (§4.3). Each of the language variants was also coded for one of three levels of morphological complexity with the criteria described in §4.4. Table 2 lists all of the 120 features and their variants. Features 1-80 are lexical, and features 81-120 are grammatical. They are also coded for their language of origin and their level of complexity (low, medium or high). For glossing conventions, see Endnote i.

FEATURE	VARIANT	LANGUAGE	COMPLEXITY	COMMENTS
F1 UP	ontop	K	L	Does a Gurindji person use the Kriol or Gurindji variant to express 'up', 'foot' etc? These variants are all monomorphemic and are therefore low in complexity.
	kankula	G	L	
F2 DOWN	dan	K	L	
	kanyjurra	G	L	
F3 FOOT	fut	K	L	
	jamana	G	L	
	jina	G	L	
F4 LIGHTNING	laitning	K	L	
	janginyina	G	L	
F5 BEE	bi	K	L	
	nama	G	L	
	lanu	G	L	
F6 OWL	owl	K	L	
	mukmuk	G	L	
F7 FROG	canetoad	K	L	
	ngakparn	G	L	
F8 BOY	boi	K	L	
	malyju	G	L	
F9 HOLE	hol	K	L	
	jimpiri	G	L	
F10 TURTLE	tetul	K	L	
	narrinyjila	G	L	
	kuwarlamparla	G	L	
F11 SICK	sik-wan	K	M	This variant has morphological structure 'sick-NMLZ' i.e. is bimorphemic.
	janga	G	L	These variants are all monomorphemic.
F12 WATER	woda	K	L	
	ngawa	G	L	
F13 CAR	motika	K	L	Although 'motorcar' is a compound in English, neither 'motor' nor 'car' exists as separate stems in Kriol.
	kurrurij	G	L	This variant is monomorphemic.
	rarraj-kaji	I	M	This variant has morphological structure 'run-AGENT', i.e. it is bimorphemic and is therefore coded as medium-level complexity. Although both morphemes come from Gurindji, the derived word is a recent innovation.
F14 HAPPY	hepi-wei	K	M	This variant is a bimorphemic word using derivational morphology 'happy-ADV'.
	pleij	K	L	These variants are all monomorphemic.
	marrinyu	G	L	
F15 BIRD	bird	K	L	
	jurlaka	G	L	
F16 DOG	dog	K	L	
	warlaku	G	L	
F17 GIRL	gel-wan	K	M	This variant is a bimorphemic word using derivational morphology 'girl-NMLZ'.
	wamala	G	L	These variants are all monomorphemic.
F18 NEST	nes	K	L	
	juru	G	L	

F19 MOTHER	mami-wan	K	M	This variant is a bimorphemic word using derivational morphology ‘mother-NMLZ’.
	ngamayi	G	L	This variant is monomorphemic.
F20 FATHER	dadi-wan	K	M	This variant is a bimorphemic word using derivational morphology ‘father-NMLZ’.
	ngaji	G	L	These variants are all monomorphemic.
F21 BARRAMUNDI	barramandi	K	L	
	manyirrkila	G	L	
	yinarrwa	G	L	
F22 BUTTERFLY	badaflai	K	L	
	malimali	G	L	
F23 TEA	ti	K	L	
	nalija	G	L	
F24 BOTTLE	botul	K	L	
	murlukurn	G	L	
F25 FENCE	fens	K	L	
	partik	G	L	
F26 BEEHIVE	bihaib	K	L	Although ‘beehive’ is a compound in English, it is not recognised as a bimorphemic word by Gurindji people.
	namawurru	G	L	Although *nama-wurru* ‘bee-COM’ is bimorphemic in a neighbouring language (Jaminjung), it was borrowed as a frozen form into Gurindji.
	ngarlu-waji	G	M	This variant is a bimorphemic word using derivational morphology ‘honey-AGENT’.
	bi-yu hawuj	I	M	This variant is a construction consisting of a bimorphemic word using inherent inflection ‘bee-DAT’ and another monomorphemic word ‘house’.
F27 NOSE	nos	K	L	These variants are all monomorphemic.
	jitji	G	L	
F28 HEAD	hed	K	L	
	ngarlaka	G	L	
F29 HORSE	ojij	K	L	
	ors	K	L	
	yawarta	G	L	
	timana	G	L	
F30 SNAKE	jinek	K	L	
	wari	G	L	
	jawulwara	G	L	
F31 GOANNA	guana	K	L	
	kirrawa	G	L	
F32 ONE	wan-bala	K	M	This variant is a bimorphemic word using derivational morphology ‘one-NMLZ’.
	jintaku	G	L	These variants are all monomorphemic.
F33 TREE	tri	K	L	
	karnti	G	L	
	watiya	G	L	
F34 CLOSE	kulojap	K	L	
	wijku-pari	G	M	This variant is a bimorphemic word using derivational morphology ‘close-ANOTHER’.
F35 FAR	long-wei	K	M	This variant is a bimorphemic word using derivational morphology ‘long-ADV’.
	yikili	G	L	These variants are all monomorphemic.
F36 FAMILY	femli	K	L	
	nungkiny	G	L	
F37 OUR	bla mela	K	L	These variants are monomorphemic words ‘DAT 1PL.INC’.
	ngantipa-ny	G	M	This variant is a bimorphemic word using inherent inflection ‘1PL.INC-DAT’.
	ngantipa-nguny	G	M	This variant is a bimorphemic word using inherent inflection ‘1PL.INC-

				DAT'.
F38 WHAT	nyampa	K	L	These variants are all monomorphemic.
	wat	G	L	
F39 WHERE	weya	K	L	
	wijei	K	L	
	wanyji-ka	G	M	This variant is a bimorphemic word using inherent inflection 'which-LOC'.
F40 NEXT TO	wan-said	K	M	This variant is a bimorphemic word using derivational morphology 'one-NMLZ'.
	nekstu	K	L	This word derives from 'next to' but is monomorphemic for Gurindji people.
	kurlpap	G	L	This variant is monomorphemic.
F41 WHITE PERSON	wait-bala	K	M	This variant is a bimorphemic word using derivational morphology 'white-NMLZ'.
	kartiya	G	L	These variants are all monomorphemic.
	kartipa	G	L	
F42 LITTLE	lidul-wan	K	M	This variant is a bimorphemic word using derivational morphology 'little-NMLZ'.
	yapakayi	G	L	These variants are all monomorphemic.
F43 BIG	bigija	K	L	
	big-wan	K	M	This variant is a bimorphemic word using derivational morphology 'big-NMLZ'.
	jangkarni	G	L	This variant is monomorphemic.
F44 FIRE	faya	K	L	These variants are all monomorphemic.
	jawi	G	L	
	warlu	G	L	
F45 AHEAD	fran	K	L	
	kamparri	G	L	
F46 HUSBAND	hasben	K	L	
	ngumparna	G	L	
F47 CAT	pujikat	K	L	Although 'pussycat' is a morphologically complex word in English, it is a monomorphemic word for Gurindji people, i.e. neither 'cat' nor 'pussy' are used independently.
	nyutinyuti	G	L	While this variant originally arose though a productive process of reduplication, it is no longer bimorphemic, evidenced by the fact that it no longer has monomorphemic counterparts, i.e. *nyuti*.
F48 WHATSITSNAME	ting	K	L	These variants are all monomorphemic. This word is to as a replacement word if a speaker can't remember the actual word or to be deliberately oblique for stylistic reasons.
	nganayirla	G	L	
F49 THREE	jirri-bala	K	M	This variant is a bimorphemic word using derivational morphology 'three-NMLZ'.
	murrkun	G	L	These variants are all monomorphemic.
F50 EYE	ay	K	L	
	mila	G	L	
	ngapanyji	G	L	
F51 WOMAN	wuman	K	L	
	kajirri	G	L	
	janka	G	L	
	kirri	G	L	
F52 SHADE	sheid	K	L	
	ngantawi	G	L	
	yarti	G	L	
F53 FISH	fij	K	L	
	yawu	G	L	
F54 MOUTH	mawuj	K	L	
	kangarnta	G	L	

F55 ROCK	rok	K	L	
	stoun	K	L	
	wumara	G	L	
F56 RETURN	gu-bek	K	M	This variant is a bimorphemic word using inherent inflection 'go-DIR'.
	lurrpu	G	L	These variants are all monomorphemic.
	wart	G	L	
F57 SQUASH	squash-im	K	M	This variant is a bimorphemic word using derivational morphology 'squash-TR'.
	jampurlk	G	L	These variants are all monomorphemic.
	nurt	G	L	
F58 LOOK	luk	K	L	
	karrap	G	L	
F59 SEARCH	luk-aran	K	M	This variant is a bimorphemic word using inherent inflection 'look-DIR'.
	warlakap	G	L	This variant is monomorphemic
F60 BLEED	blid-ing	K	M	This variant is a bimorphemic word using inherent inflection 'bleed-CONT'
	kungulu	G	L	This variant is monomorphemic.
F61 FISHING	bijinbat	K	L	This word is complex in Kriol 'fish-CONT-CONT', but is used monomorphemically by Gurindji people, i.e. this form is morphologically frozen.
	bij-ing	K	M	This variant is a bimorphemic word using inherent inflection 'fish-CONT'.
	wuyurrunkarra	G	L	This word appears to be *wuyurrun-karra* 'fish-CONT' but is not. It is a morphologically frozen form. Instead the word *wuyurrun* which means 'fishing line' was derived through back formation.
F62 BITE	bait-im	K	M	This variant is a bimorphemic word which uses derivational morphology 'bite-TR'.
	katurl	G	L	This variant is monomorphemic.
F63 LICK	lik-im	K	M	This variant is a bimorphemic word using derivational morphology 'lick-TR'.
	ngalyak	G	L	These variants are all monomorphemic.
F64 JUMP	jamp	K	L	
	tipart	G	L	
F65 CLIMB	klaim-ap	K	M	This variant is a bimorphemic word using inherent inflection 'climb-TEL'.
	partaj	G	L	These variants are all monomorphemic.
F66 CRY	krai	K	L	
	lungkarra	G	L	Historically this word was 'cry-CONT', but is now monomorphemic (see Meakins & Nordlinger, 2014: 286-287 for the historical details).
F67 CHASE	jeij-im	K	M	This variant is a bimorphemic word using derivational morphology 'chase-TR'.
	kayikayi	G	L	While this variant originally arose though a productive process of reduplication, it is no longer bimorphemic, evidenced by the fact that they no longer have monomorphemic counterparts, i.e. *kayi* and *jayi*. The form *jayijayi* is a recent innovation which is the result of phonological coalescence between *jejim* and *kayikayi*.
	jayijayi	I	L	
F68 QUIET	kwait-wei	K	M	This variant is a bimorphemic word using derivational morphology 'quiet-ADV'.
	muk	G	L	These variants are all

	wampal	G	L	monomorphemic forms.
F69 SLEEP	jilip	K	L	
	makin	G	L	
F70 STEAL	steal-im	K	M	This variant is a bimorphemic word using derivational morphology 'steal-TR'.
	jawurra	G	L	These variants are all monomorphemic. Although *futwok* is derived from 'foot walk', it is not recognised as having complex morphological structure by Gurindji people.
F71 WALK	futwok	K	L	
	wok	K	L	
	kalu	G	L	
F72 HUNT	hant-ing	K	M	This variant is a bimorphemic word using inherent inflection 'hunt-CONT'.
	mutap	G	L	This variant is monomorphemic.
F73 HIT HEAD	kil-im hed	K	M	This variant is a construction consisting of a bimorphemic word using derivational morphology 'hit-TR' and another monomorphemic word 'head'.
	pangkily	G	L	These variants are all monomorphemic.
	tarl	G	L	
F74 EAT	ab-im	K	M	This variant is a bimorphemic word using derivational morphology 'eat-TR'.
	id-im	K	M	This variant is a bimorphemic word using derivational morphology 'eat-TR'.
	jaartkarra	G	L	Historically this word was 'eat-CONT' but is now monomorphemic, i.e. *jaart* does not exist on independently (see Meakins & Nordlinger, 2014: 286-287 for the historical details).
F75 SWELL	juwel-ap	K	M	This variant is a bimorphemic word using inherent inflection 'swell-TEL'.
	rimpu	G	L	This variant is monomorphemic.
F76 SIT	jidan	K	L	Historically this word was 'sit-DIR' but is now monomorphemic, i.e. 'sit' does not exist independently.
	lurlu	G	L	These variants are all monomorphemic.
	tak	G	L	
F77 SHOW	jou-im	K	M	This variant is a bimorphemic word using derivational morphology 'show-TR'
	wiyit	G	L	These variants are all monomorphemic.
F78 WAIT	weit	K	L	
	liwart	G	L	
F79 LISTEN	lijin	K	L	
	kurru	G	L	
F80 TALK	tok	K	L	
	jarrakap	G	L	
F81 ALL	holot	K	L	
	al	K	L	
	purrp-pa-rni	G	M	This variant is a multimorphemic word 'finish-EP-ONLY'.
F82 DISCOURSE	nyawa-ma	G	M	This variant is a bimorphemic word 'this-TOP'.
F83 THEN	abta dat	K	L	These variants consist of two free morphemes and are therefore not complex morphologically.
	brom deya	K	L	
	an den	K	L	
	yala-nginyi	G	H	This word 'that-ABL' belongs to a paradigm.
	nyila-nginyi	I	H	This word 'that-ABL' belongs to a paradigm. Although both morphemes are derived from Gurindji, the combination of them is a recent innovation (see Meakins & Nordlinger, 2014: 171-172 for further details).
F84 EXACTLY	rait la	K	L	This variant consists of two free

				morphemes and is therefore not complex morphologically.
	-ta=rni	G	M	This variant is a bimorphemic structure using inherent inflection and a discourse clitic (which only has scope over the word to which it attaches) '-LOC=ONLY'.
F85 ADJECTIVE	-wan	K	M	This variant is a derivational morpheme.
	zero	G	L	This variant is zero derivation.
F86 TOPIC	=ma	G	H	This variant is a discourse clitic.
F87 HERE	hiya	K	L	This variant is monomorphemic.
	murla-ngka	G	H	This word 'this-LOC' belongs to a paradigm.
	nyawa-ngka	I	H	This word 'this-LOC' belongs to a paradigm. Although both morphemes are derived from Gurindji, the combination of them is a recent innovation (see Meakins & Nordlinger, 2014: 171-172 for further details).
	hiya-ngka	I	M	This variant is a bimorphemic word using inherent inflection 'here-LOC'. The stem is derived from Kriol and the locative suffix is derived from Gurindji.
F88 OBLIQUE	la wi	K	L	This variant consists of two free morphemes and is therefore not complex morphologically.
	ngantipa-ny	G	M	This variant is a bimorphemic word using inherent inflection '1PL.INC-DAT'.
F89 DETERMINER	da	K	L	These variants are all monomorphemic.
	dat	K	L	
F90 NASAL CLUSTER DISSIMILATION (NSD)	-ka etc	G	H	This is an allomorphic process involving inflection which has been lost among younger generations. See McConvell (1988) for more information about NSD. It increases the number of allomorphs for ergative and locatives case suffixes.
F91 PLURAL	dem	K	L	This variant is a free form article.
	-rrat	G	M	This variant is inherent inflection.
	-rra	G	M	This variant is inherent inflection.
F92 ALLATIVE	la	K	L	This variant is a free form preposition.
	nothing	K	L	This variant is has no morphological structure.
	-ngkurra	G	M	This variant is inherent inflection.
	-ngkirri	I	M	This variant is inherent inflection. It is the result of phonological coalescence between the original Gurindji allomorphs *–jirri, -yirri and –ngkurra* (see Meakins, 2012 for more detail).
F93 DATIVE	-wu	G	H	This variant is contextual inflection.
	-yu	I	H	This variant is contextual inflection. See Meakins (2011b) for more information about allomorphy.
	-ngku	I	H	This variant is contextual inflection. See van den Bos, Meakins & Algy (2017) for more information about this variant.
F94 PROPRIETIVE	gat	K	L	This variant is a free form prepositions.
	with	K	L	
	-yawung	G	M	This variant is inherent inflection.
F95 DATIVE	-ku	G	H	This variant is contextual inflection.
	-tu	I	H	This variant is contextual inflection. See Meakins (2011b) for more information about allomorphy.
F96 ERGATIVE	-lu	G	H	This variant is consists of a reduction in allomorphy. See Meakins (2011b) for more information about allomorphy.

F97 ABLATIVE	brom	K	L	This variant is a free form prepositions.
	burrum	K	L	
	-nginying	G	M	This variant is inherent inflection.
	-nginyi	G	M	This variant is inherent inflection.
F98 DATIVE	blanganta	K	L	This variant is a free form preposition.
	-yu etc	G	H	This variant is contextual inflection.
	bo	I	L	This innovation using the Kriol form, but the Gurindji dative pattern. See Meakins (2011a) for more information.
F99 FACTIVE	meik-im	K	M	This variant is a bimorphemic using derivational morphology 'show-TR'.
	-k	G	M	This variant is derivational morphology.
F100 INSTRUMENTAL	garram	K	L	This variant is a free form preposition.
	-yawung-tu	G	H	This variant is '-PROP-ERG' which involves agreement with a transitive subject.
	-yawung	I	M	This variant is inherent inflection. This is an innovation because traditional Gurindji requires agreement in transitive clauses.
F101 LOCATIVE	langa	K	L	This variant is a free form preposition.
	-ta	G	M	This variant is inherent inflection.
	langa xxx-ta	I	H	This variant is double-marking (therefore agreement) across an NP.
F102 CASE AGREEMENT	no case agreement	K	L	This variant has no morphological structure.
	case concord	G	H	This variant is an inflection requiring agreement across an NP.
	edge marking	I	M	This variant is an inflection requiring only the last word of the NP to be marked.
F103 SUBJECT MARKING	no marking	K	L	This variant is has no morphological structure.
	obligatory	G	H	This variant is contextual inflection.
	optional	I	H	This variant is optional case marking (OCM). See Meakins & Wilmoth (2018).
F104 CASE ALIGNMENT	no marking	K	L	This variant has no morphological structure.
	ergative	G	H	This variant is contextual inflection.
	nominative	I	H	This variant is contextual inflection.
F105 INSUBORDINATION	-ta	I	M	This variant is inherent inflection. See Meakins (2016) for more explanation.
F106 MARKING OF REDUCED SUBORDINATE CLAUSES	no marking	K	L	This variant has no morphological structure.
	allative	G	H	This variant is contextual inflection because it involves switch reference.
	locative	I	M	This variant is inherent inflection because it does not involve switch reference. See Meakins (2016) for more explanation.
F107 AGAIN	=du	K	M	This variant is a discourse clitic which only has scope over the noun to which it attaches.
	igin	K	L	This variant is a free form particle.
	=rningan	G	M	This variant is a discourse clitic which only has scope over the noun to which it attaches.
F108 PASSIVE	ged x-nginyi	K	H	This variant is a structure which has morphological consequences in the rest of the clause. Passives are discussed in detail in Meakins (2011b).
	nothing	G	L	This variant is has no morphological structure.

F109 PLACEMENT	-im-TEL	K	M	This variant is a multimorphemic word involving derivational morphology and inherent inflection.
	yuwa-na-na	G	H	This word 'put-IMPF-PRS' belongs to a paradigm.
	put-im	I	M	This variant is a bimorphemic word 'put-TR' involving derivational morphology.
F110 ACCOMPANIMENT	gat	K	L	This variant is a free form preposition.
	ka-nga-na	G	H	This word 'take-IMPF-PRS' belongs to a paradigm.
	teik-im	I	M	This variant is a bimorphemic word 'take-TR' involving derivational morphology.
F111 MOTION	gon	K	L	This variant is a monomorphemic word
	ya-na-na	G	H	This variant is a word 'go-IMPF-PRS' which belongs to a paradigm.
F112 CAUSATIVE	meik-im	K	M	This variant is a bimorphemic word 'make-TR' involving derivational morphology.
	ma-na-na	G	H	This word 'do-IMPF-PRS' belongs to a paradigm.
F113 Imperative LOOK	luk	K	L	This variant is a monomorphemic word.
	nya-ngka	G	H	This word 'look-IMP' belongs to a paradigm.
F114 Imperative GIVE	gib-it	K	M	This variant is a bimorphemic word 'give-TR' involving derivational morphology.
	jayi-ngka	G	H	This word 'give-IMP' belongs to a paradigm.
F115 PAST TENSE	bin	K	L	This variant is a free form auxiliary.
	=m	K	M	This variant is inherent morphology.
	-ni	G	H	This variant is inherent morphology in paradigm.
F116 NEGATION	nomo	K	L	These variants are free form particles.
	don	K	L	
	kula	G	L	
F117 COPULA	nothing	K	L	This variant has no morphological structure.
	bi	I	L	This variant is a free form auxiliary.
	karrinyana	G	H	This word 'be.IMPF.PRS' belongs to a paradigm.
F118 GO	gu	K	L	This variant is a monomorphemic word.
	ya-na-na	G	H	This word 'go-IMPF-PRS' belongs to a paradigm.
F119 HIT	kil-im	K	M	This variant is bimorphemic word 'hit-TR' involving derivational morphology.
	pa-na-na	G	H	This word 'hit-IMPF-PRS' belongs to a paradigm.
F120 SUBORDINATOR	to	I	M	This word has morphological consequences in rest of clause.

Table 2 120 linguistic features (with 292 variants) and their complexity

The linguistic features were chosen for their variability and comparability across the corpus. By variability, we mean that features could be expressed by one or more Kriol and Gurindji variants and sometimes an additional innovative variant. By comparability, we refer to the fact that each speaker had a number of opportunities in the process of constructing the corpus to express the linguistic feature of interest (e.g. a lexeme such as SEARCH or a grammatical function such as LOCATION) using a Kriol, Gurindji or an innovative variant. The advantage of this approach over recording less structured conversation is that we are able to test each speaker for the use of each feature of interest (see §4.2 for details). This means that the dataset is complete for each speaker with no missing data. Thus 'zero' in our dataset is meaningful, indicating the absence of that feature from an individual's usual language repertoire, rather than being 'missing data' (failure to record variant usage).

4.2 DATA COLLECTION METHODS

The variants are drawn from 96 hours of recordings which were made by Meakins and Algy and are sound-linked, transcribed and translated into English using the software Computerized Language Analysis (CLAN) which links transcriptions of recordings to media (sound or video) and allows morpheme-by-morpheme annotation of these transcripts (MacWhinney 2000). The majority of the recordings (80 hours; 57,179 clauses) are annotated for a number of values such as the gloss (e.g. locative, nominaliser, past tense), part of speech (e.g. verb, noun), transitivity (transitive, intransitive), animacy (human, inanimate, animal, kin, body part), language of origin (Gurindji or Kriol) and other values which enabled the use of CLAN searches on the usage of the linguistic variants to construct the dataset. An example of an annotated clause extracted from a CLAN transcript of a recording is given in (2) (Meakins & Wilmoth 2019) (For annotation conventions, see Endnote i).

(2) Dat *yawarta=ma* leg i bin kil-im fens-*tu*
'The fence hit the horse's leg.'

Dat	*yawarta*	*=ma*
the	horse	=TOP
dem|dat&k=the	n:animal|yawarta&g=horse	suf:top|ma&g=TOP

leg	i	bin
leg	3SG.S	PST
n:bp|leg&k=leg	pro|i&3SG/S&k=he/she/it\S	v:aux|bin&PST&k=PST

kil	-im	fens
hit	-TR	fence
v:tran|kilim&k=hit	suf|im&TR&k=TR	n:inanimate|fens&k=fence\S

-*tu*
-NOM
case:erg|tu&g=NOM

The remaining 16 hours of the corpus are not (yet) morphologically annotated so that searches for linguistic variants in this part of the corpus were undertaken using a text editor.

The recordings consist of free narratives, conversation, procedural descriptions (e.g. how to make artefacts, how to collect and prepare bush foods etc), picture-prompt narratives (e.g. frog stories (Mayer 1994 (1969)) and the monster stories series (O'Shannessy 2004)), and picture-prompt sentences which come from a series of director-matcher tasks designed to target particular linguistic features (Meakins 2011b). Around 60 hours of the corpus consists of free narratives, conversation and procedural descriptions, and the remaining 36 hours consists of picture-prompt tasks.

The picture-prompt tasks were particularly crucial to the development of the dataset. Although naturalistic data remains the gold standard for the development of variationist corpora (e.g. Poplack 2015) and formed the basis for identifying variability among Gurindji people in the expression of particular linguistic features, a reliance on conversation or free narrative data would have resulted in a dataset with many missing data points (see §4.1 for a discussion of the interpretation of 'zero'). In conversations and free narratives, speakers choose the topic and therefore you cannot guarantee the expression of particular linguistic features. In this respect, it is difficult to produce a comparable dataset for large numbers of features and speakers. Picture-prompt tasks supplement the data by ensuring that speakers have the opportunity to express particular linguistic features a number of times using their variant(s) 'of choice'.

The picture-prompt tasks consist of narrative descriptions using picture books or single image descriptions using director-matcher tasks. These tasks require speakers to express the characters, objects and actions depicted by the pictures. All characters, objects and actions occur at least three times (in most cases many more) in the pictures, guaranteeing the multiple use of a Kriol, Gurindji and/or innovative variant(s) by a speaker to express a particular linguistic feature. For example, the use of picture books such as the frog story meant that every participant had a number of opportunities to express the verb SEARCH as either *lukaran* (<Kriol) or *warlakap* (<Gurindji).

The director-matcher tasks were conducted as a series of games (picture-match, bingo or 'go fish') (see Meakins 2009; Meakins 2011b: 50-54; 2011c, 2016; Meakins & Algy 2016; Meakins & O'Shannessy 2010; Meakins & Wilmoth 2019 for a detailed description of the tasks). In total, 148 pictures were used throughout the different tasks. The director and matcher sat near each other with matching sets of pictures in front of them. The director described a picture, which differed only minimally from another picture, in such a way that the matcher could identify the correct picture. The director and matcher's pictures were screened off from each other to ensure that the director was required to produce detailed descriptions. For example, one set of minimally different pictures showed a woman putting a drink on a chair and the same woman putting the same drink on a table instead. The differing locations guaranteed the use of a Gurindji case suffix *-ta* (or *-ngka*), a Kriol preposition *la* (or *langa*), or a double-marked innovative variant, for example *langa teibul-ta* 'LOC table-LOC'. Director-matcher tasks have been used extensively in semantic typology research, such as studies of spatial relations and reciprocals (Evans et al. 2011; Levinson & Wilkins 2006), to produce comparable data across languages. We use this method in our study to produce comparable data across speakers within a single speech community.

More generally, we also adopted the peer elicitation approach (cf. Meakins, Green, & Turpin 2018: 253) to ensure that the data captured were as close to everyday speech as possible. Peer elicitation involves replacing the usual linguist-speaker dyad with peer-peer dyads or groups. In the case of elicited narrative descriptions using picture books, the person giving the instructions came from the same generation. In the case of director-matcher tasks, both the director and matcher came from the same generation. This ensured that the data was not skewed by adjustment of language use to accommodate a Gurindji person of another generation (or, more importantly, not an English-dominant researcher). For example, a Gurindji speaker might accommodate for the presence of children by using more Kriol.

4.3 CODING OF LANGUAGE FEATURES

For each language feature, a speaker is recorded as using either only a Gurindji variant, only a Kriol variant, only an innovative variant, or different combinations of these, because the speaker is given multiple opportunities to express each language feature, i.e. there are multiple picture stimuli which require description and therefore the use of a particular linguistic feature. For example, during the description of the frog story, the word SEARCH requires expression a number of times because there are multiple pictures of a boy looking for his pet frog. During the frog story description, a speaker could express SEARCH using only the Gurindji variant *warlakap*, or only the Kriol variant *lukaran,* or both. Thus there are three possible usage patterns produced by any one speaker for this particular language feature. Usage pattern 1 involves a speaker only using the Kriol variant *lukaran* for all expressions of SEARCH. Usage pattern 2 involves a speaker only using the Gurindji variant *warlakap*. Usage pattern 3 involves a speaker using both the Kriol and Gurindji variants across multiple expressions of SEARCH. Figure 3 shows these three usage patterns for the expression of SEARCH.

Figure 3 Usage patterns for the expression of SEARCH

	KRIOL	GURINDJI
F59 SEARCH	*lukaran*	*warlakap*
Usage pattern 1	1	0
Usage pattern 2	0	1
Usage pattern 3	1	1

An example of the usage of a grammatical feature is the expression of LOCATION. LOCATION could be expressed by a Gurindji case suffix, Kriol preposition or an innovative double-marked variant, or a combination of these. Figure 4 shows the five usage patterns, used by speakers in our study, which are found in the corpus.

Figure 4 Observed usage patterns for the expression of LOCATIVE

	KRIOL	GURINDJI	INNOVATION
F102 LOCATIVE	*preposition*	*case*	*double-marking*
Usage pattern 1	0	1	0
Usage pattern 2	1	0	0
Usage pattern 3	1	1	0
Usage pattern 4	1	0	1
Usage pattern 5	1	1	1

Speakers were coded for whether they used the variant or not (binary) rather than the frequency of use (for further discussion see §5.2). We did not record the usage frequency of a variant by a speaker because this measure is not reliable given that we only have 96 hours of recordings. Instead, we use a modelling approach to account for uncertainties in the usage frequency (see §4.5).

We record a variant of a language feature as Gurindji (G) or Kriol (K), if it is identified as such by the Gurindji speech community. In a very few cases, this coding does not reflect the historical origin of a variant. For example, the Gurindji community identify *partik* 'fence' and *kururij* 'car' as Gurindji words, despite their

English origin (*paddock* and *carriage*, or perhaps *Goodrich tyres*). The equivalent Kriol words are *fens* and *motika*, which are also in the repertoire of Gurindji people. It is likely that these words were borrowed very early in the contact between Kriol/English and Gurindji such that knowledge about their source has been lost.

We record a variant of a language feature as an innovation (I) if it consists of a hybrid Gurindji-Kriol form. Innovations take a number of forms. In many cases, they fuse the form (phonological material) of one language with the grammatical distribution of another language. For example, the Kriol dative preposition *bo* is used by Gurindji Kriol speakers to mark direct objects where Kriol speakers used the locative preposition *langa.* Gurindji Kriol speakers use the dative preposition *bo* rather than the locative case marker *-ngka* because the dative case marker *-wu* is used by Gurindji speakers to mark indirect objects (Meakins 2011a, 2011b). Other types of innovations include phonological coalescence where two forms are fused into one. For example, the verb *jayijayi* 'chase' is the result of the coalescence of the Kriol verb *jeijim* and the Gurindji equivalent *kayikayi* (Meakins 2012).

In a few cases, speakers used neither the Gurindji or Kriol or innovative variants. These cases are represented as black bars in Figure 9 (F82, F86, F90, F96, F105, F120). All of these cases are Gurindji, Kriol, or innovative variants where there are no equivalents in the other language. For example, feature 85 is the Kriol suffix -*wan* which derives adjectives from nouns and verbs and does not have a Gurindji equivalent because Gurindji has no adjectives (instead nouns have the same descriptive function as adjectives).

4.4 CODING OF COMPLEXITY

Each linguistic variant was coded according to three levels of morphological complexity (H=high, M=medium, L=low). Here we are measuring *absolute* complexity, i.e. the number of elements contained in each subsystem of a grammar and the number of distinctions each of these elements makes (e.g. Nichols 2009), since this is how complexity has been approached in the literature about morphological simplification and language contact (§2). For example, in terms of inflectional morphology, absolute complexity (termed '(E)numerative complexity' by Ackerman & Malouf (2013)), refers to the number of morpho-syntactic distinctions that a language makes both in terms of the internal structure of words and their arrangement into inflectional classes.

We code the three levels of complexity based on measures of complexity from the literature, summarised in Table 3 and discussed in detail below (with more information about the features and their variants available in Table 2 and the complete appended dataset in Appendix 1). Note that all large quantitative analyses are somewhat reductive by nature, and the coding systems does not necessarily capture some of the nuances of the morphological structures, but the strength of a quantitative approach is in the ability to detect general patterns and distinguish them from selectivity, co-incidence and background noise.

VARIANT MORPHOLOGICAL PROFILE	COMPLEXITY SCORE	EXAMPLES AND FEATURE NUMBER
Monomorphemic content words	L	*jamana* 'foot' (F3), *hol* 'hole' (F9), *lungkarra* 'cry' (F66), *jidan* 'sit' (F76)
Function words	L	*gat* 'PROP' (F94), *brom* 'ABL' (F97), *bo* 'DAT' (F98), *kula* 'NEG' (F116), *bi* 'copula' (F117), *gu* 'serial motion verb' (F118)

Multimorphemic content words	M	*mami-wan* 'mother-NMLZ' (F19), *wijku-pari* 'near-ANOTHER' (F34), *wan-said* 'next to, one-SIDE' (F40), *jirri-bala* 'three-NMLZ' (F49), bait-im 'bite-TR' (F62), *klaim-ap* 'climb-TEL' (F65)
Derivational morpheme	M	*-rra* 'this-PL' (F91), *-k* 'FACT' (F99)
Inherent inflection	M	*-ta=rni* 'bang on, LOC-ONLY' (F84), *-yawung* 'PROP' (F94), *ngantipa-ny* 'we-OBL' (F88), *-nginyi* 'ABL' (F97), *-ngka* 'subordinator' (F106), *=m* 'PST' (F116)
Contextual inflection	H	*-yu* 'DAT' (F93), *-lu* etc 'ERG' (F96), *-ngku* 'NOM' (F104), *-jirri* 'switch reference' (F106), *ged-nginyi* 'passive' (F108)
Inflection in paradigm	H	*-ka* 'nasal cluster dissimilation' (90), *-ni* 'PST' (F115), y*a-na-na* 'go' (F118), *pa-na-na* 'hit' (F119)
Discourse	H	*=ma* 'TOP' (F86), optional case (F104), *=du* 'AGAIN' (F107)

Table 3 Measures of morphological complexity

Monomorphemic words were coded as simpler (L) than bimorphemic or multimorphemic words, which were coded as M. This reflects the fundamental claim that languages which possess more structure at the level of the word (agglutinating and polysynthetic languages such as Turkish or Mohawk) are more complex than isolating languages (such as Vietnamese) (Anderson 1992, 2015). For example, the expression of HOLE in both Gurindji and Kriol involves monomorphemic words *jimpiri* and *hol*, which are therefore both coded L (Figure 5). Similarly the expression of ONE and SEARCH in Gurindji involves monomorphemic words *jintaku* and *warlakap*, whereas in Kriol, bimorphemic words are used *wan-bala* 'one-NMLZ, (< one-fellow)' and *luk-aran* 'look-TEL, (< look around)' (further examples are given in Table 2, Table 3 and Appendix 1).

Figure 5 Examples of complexity coding for monomorphemic and bimorphemic words

	KRIOL	GURINDJI
F6 HOLE	*hol*	*jimpiri*
Complexity	L	L
	KRIOL	GURINDJI
F32 ONE	*wan-bala*	*jintaku*
Complexity	M	L
	KRIOL	GURINDJI
F59 SEARCH	*luk-aran*	*warlakap*
Complexity	M	L

In a similar vein, bound morphemes, i.e. affixes and enclitics, were coded as more complex than free form equivalents, based on the claims in the literature about the inherent complexity of bound morphemes. For example, case-marking is considered more complex than functionally-equivalent prepositions by virtue of being bound and therefore contributing to the internal structure of words, rather than being a free form (e.g. Baerman, Brown, & Corbett 2015: 4). The replacement of bound morphemes (synthetic forms) with periphrastic constructions (analytic forms) is labeled a process of simplification in the language shift and attrition literature (de Groot 2008; Dorian 1978; Schmidt 1985), and the reduction of inflectional morphology in creole languages is also considered simplification by some (e.g. McWhorter 1998). An example of this process in the Gurindji Kriol dataset is variation between synthetic forms such as Gurindji case suffixes, e.g. the locative or proprietive, and Kriol analytical equivalents which are free form prepositions. The Gurindji analytical forms were coded as more complex than the Kriol synthetic forms. Examples are given in Figure 6.

Figure 6 Examples of complexity coding for bound vs free morphemes

	KRIOL	GURINDJI
F103 LOCATIVE	*langa*	*-ta*
Complexity	L	M

	KRIOL	GURINDJI
F94 PROPRIETIVE	*gat*	*-yawung*
Complexity	L	M

Different types of bound morphemes were coded as having different levels of complexity. Contextual inflection (e.g. core case) is coded as highly complex (H) in contrast with inherent inflection (e.g. non-core case) or derivational suffixes (e.g. number, nominalisation, telic) (M) because contextual inflection is determined by the broader syntactic context (cf. Booij 1996). The role of contextual inflection is to mark the relationship between a head and a dependent in a relationship of either government or agreement whereas inherent inflection is determined by the information a speaker wishes to convey, and derivation simply creates new lexemes. This distinction has ramifications for languages in contact. For example, Gardani (2008) and Myers-Scotton (2002) have shown that contextual inflection (called 'outsider late-system morphemes' by Myers-Scotton) is less likely to transfer than inherent inflection and derivational morphology. Contextual inflection is also commonly lost in situations of language shift and attrition, as discussed above. In the dataset, the Gurindji dative, which is an example of contextual inflection, is coded as more complex (H) than the Gurindji ablative (M) or Gurindji plural (M), which are examples of inherent inflection and a derivational suffix, respectively. All of these inflections are ranked more complex than their Kriol free form counterparts (Figure 7).

Figure 7 Examples of complexity coding for different types of bound morphemes

	KRIOL	GURINDJI	INNOVATION
F98 DATIVE	*bla*	*-yu*	*bo*
Complexity	L	H	L

	KRIOL	KRIOL	GURINDJI	GURINDJI
F97 ABLATIVE	*burrum*	*brom*	*-nginyi*	*-nginying*
Complexity	L	L	M	M

	KRIOL	GURINDJI	GURINDJI
F91 PLURAL	*dem*	*-rra*	*-rrat*
Complexity	L	M	M

Words with affixes that form part of a paradigm were coded as highly complex, reflecting the observation that paradigms are (almost) never borrowed (referred to as the 'Principle of Morphosyntactic Subsystem Integrity') and fall out of wholesale use in language shift and attrition situations (and are replaced by analytic equivalents) (Evans 2016a; Seifart 2012). For example, *ya-* 'go' in Gurindji belongs to a paradigm which distinguishes tense (past, present), mood (potential, indicative, imperative, hortative), and perfectivity (imperfective, perfective), whereas the Kriol variant *gu* marks these categories by combining with free form auxiliaries such as *bin* 'past', *ledim* 'hortative', *garra* 'potential', *yusta* 'past imperfective', rather than suffixes. Similarly, variation exists in the expression of the causative. The Gurindji causative is coded as highly complex (H) because it makes use of the 'make' paradigm, whereas the Kriol equivalent involves a serial verb consisting of a bimorphemic minor verb *meik-im* and a major verb (Meakins 2010). Similarly, the expression of past tense in conjunction with third person pronoun varies in complexity from a Gurindji past tense morpheme which is a part of the verb conjugation, a Kriol clitic *=in* and the Kriol free form *bin*. These complexity codings are summarized in Figure 8.

Figure 8 Examples of complexity coding for morphemes that form part of a paradigm

	KRIOL	GURINDJI
F118 GO	*gu*	*ya-na-na*
Complexity	L	H

	KRIOL	GURINDJI
F112 CAUSATIVE	*meik-im V*	*ma-na-na*
Complexity	M	H

	KRIOL	KRIOL	GURINDJI
F115 PAST	*im=in*	*i bin*	*-ni*
Complexity	M	L	H

Morphemes which require attention to the broader morpho-syntactic and discourse context such as optional case marking (see §3) and some discourse clitics are coded as highly complex (H). In the former case, Anderson (2015: 22) observes the complexity of multiple word forms corresponding to the same morpho-syntactic representation, for example 'dove' and 'dived', which are different word forms of the past tense form of DIVE in English. Campbell & Muntzel (1989: 189) note this type of variation is common in language shift, and Meakins & Wilmoth (2019) argue that it represents morphological complexification.

4.5 WRIGHT-FISHER MODEL

Given this language corpus sampled from a range of ages within the speaker population, we can track the formation of the new mixed language, Gurindji Kriol, from the two source languages, Gurindji and Kriol. We wish to ask whether the mixed language is an unbiased mix of elements from the two source languages, or whether some classes of language variants had a greater rate of incorporation into the mixed language than others - for example, we wish to know whether the formation of the mixed language has favoured Gurindji or Kriol terms, or whether there has been a preferential adoption of simpler language elements over more complex alternatives. This kind of problem is familiar to evolutionary biologists, who often wish to ask whether particular heritable variants have increased in frequency over time more than would be expected from random sampling processes alone.

We adapted the Wright-Fisher model (Fisher 1956; Wright 1931) to allow us to detect non-random patterns of incorporation of elements from the two source languages into the mixed language. The Wright-Fisher model is a classic model in population genetics that describes changes in variant frequencies over time. While it has typically been used to describe changes in the frequency of alleles (genetic variants) in an interbreeding population, it can be used to represent any process where each new group of variants at a time step is formed by sampling from the group of variants at the previous time step. The reason we need to use a formal model such as the Wright-Fisher rather than simply tracking the relative frequency of each variant over time is that we need to account for the effects of random sampling on frequencies. If the sample size is not infinite, then any sample is unlikely to exactly match the frequencies of variants at the previous time step. In other words, due to random sampling error, the frequencies of language variants are expected to fluctuate from one time step to the next. In a serial sampling procedure, the sampling errors can accumulate over time, so a random overrepresentation of one variant at one time step may be followed by an increase or decrease in the following time step, and by chance an element may undergo a series of changes in frequency in the same direction, resulting in a rise or fall in frequency over time. It would be possible to use a simpler statistical test to ask whether there is a greater increase in one class of language variant over another – for example, more increase in Kriol language variants than we would expect by chance. But using a formal model of the process of language change over time provides us with a general and flexible approach to generating a distribution of possible variant frequencies under a serial random sampling process. This model can take into account different starting frequencies or number of available variants, and can be adapted to different kinds of data, or modes of change.

The Wright-Fisher provides us with a way of generating a null distribution, which is the range of expected outcomes we might get from this random sampling process alone, in the absence of any bias or selection. In an unbiased sampling procedure, every variant has the same chance of being sampled and included in the next time step. Therefore, any difference in frequency from one time step to the next is just stochastic variation due to the sampling error. We can use the expected distribution from an unbiased sampling procedure to detect departures from random sampling. For example, we might detect that one class of variants has significantly higher frequency in the next time step, and that this increase in frequency is more than we would expect from sampling error alone. In this case, we would say that sampling is biased with respect to that class of variant. We can then ask how likely it is that the observed frequencies fall within the distribution of expected outcomes under unbiased sampling, and compare that to a model where classes have different rates of being sampled over time (a biased sampling model). If a biased sampling model fits the observed frequencies significantly better than an unbiased sampling model, then we conclude that some variants are preferentially adopted over others.

This use of the Wright-Fisher model to generate a null distribution of variant frequencies under a serial random sampling of variants does not rely on a particular model of language change. Indeed, the Wright-Fisher model has been used as a convenient tool for many different models of language change. Several theoretical studies have examined the mathematical link between the Wright-Fisher model and models of language or cultural evolution (Baxter et al. 2006; Baxter et al. 2009; Bentley, Hahn, & Shennan 2004; Bentley et al. 2007; Bentley, Ormerod, & Batty 2011; Kirby, Griffiths, & Smith 2014; Newberry et al. 2017; Reali & Griffiths 2010). These studies use the Wright-Fisher model to mathematically formulate different models of language change, using different definitions of the sample size in the model. For example, the Wright-Fisher model can be used to describe an individual's learning process of a new variant of a language feature (Reali & Griffiths 2010), where the person updates his/her usage frequency of the variant after he/she heard other speakers using the language feature. Under this model, the sample size is how many times the person heard other speakers using the language feature. Another example is more similar to gene transmission, where a speaker copies a variant of a language feature from the variant used by another random speaker in the previous time step (Bentley, Ormerod, & Batty 2011). Under this concept, the Wright-Fisher model is used to predict changes in the number of speakers who use the variant in the population, so the sample size in this particular model is the number of speakers in the population.

So the use of Wright-Fisher to model random sampling of language elements is flexible with respect to the precise nature of available data or specific assumptions about the process of language change. The flexibility of the Wright-Fisher model is illustrated in the different ways that we could interpret the parameter N. This represents the size of the sample of variants made from the previous time step to create the set of variants in the subsequent time step. In our application of the Wright-Fisher, we are modeling a sampling process from one generation of speakers to the next. By using this model, we treat the sample size as a hidden variable (see below), so that we are not making assumptions about the way that language is acquired in this population. But as the fundamental assumption of the Wright-Fisher model, we assume that the observed language changes are the result of a serial sampling process, such that there are discrete time steps and the group of variants at one time step is sampled only from the previous time steps, not from any earlier time steps. Here, we are interested in the distinct language usage in three age groups of the Gurindji Kriol community. The 'grandparent' generation (>40 years) are those who speak both source languages fluently and use features from both; the 'parent' generation (20-40 years) use features of both source languages but speak neither fluently; and the 'child' generation (<20 years) share a common mixed language that has incorporated features from both source languages. In our model, we define three time steps, corresponding to the three generations. Defining finer time steps within each generation will result in larger variance in language usage across the three generations, making any difference in the adoption rate of different variants less significant. If we find no evidence for biased sampling, such as preferential adoption of simpler language elements, when using three time steps, then we will also find no evidence using finer time steps. Our null hypothesis is that the distinct language usage in the three generations is generated in an unbiased way such that variants used in each generation are randomly sampled from those used by speakers in the previous generation. The null hypothesis may be rejected, either because some variants are more likely to be adopted than others, or because each generation samples not only from the previous generation but also from other generations or from peers (Hua et al in prep).

We made several modifications to the Wright-Fisher model to allow its extension from population genetic studies to this linguistic dataset. First, the Wright-Fisher model typically assumes only two possible variants of each feature (i.e. two alleles for each gene). But in language data there may be more than two variants of a given language feature. For example, in our data some of the features have more than one Gurindji variant, more than one Kriol variant, and may also have an innovative variant, thus some features have up to six available variants. For unbiased sampling and for biased sampling during the first several time steps, changes in variant frequencies of features with two variants can be well approximated with Beta distribution (Tataru et al. 2017). We mathematically demonstrate that Dirichlet distribution, an extension of Beta distribution to multiple dimensions, also provides a good approximation to features with more than two variants (see Appendix 2 for the proof). Here we briefly describe the mathematical outcome.

Denoting the initial usage frequency of variant *i* as $\theta_i(0)$, the possible number of usages of variant *i* in the first time step is described by a multinomial distribution. The mean of the multinomial distribution (denoted $\theta_i(1)$) describes the usage frequency of the variant as a function of the frequency in the previous time step ($\theta_i(0)$). The mean of the multinomial distribution describes any deterministic change in the usage frequency of variant *i.* We use the general mathematical formula for the deterministic change (Blythe 2012; Reali & Griffiths 2010):

$$\theta_i(1) = \frac{N\theta_i(0)+r_i}{N+\sum r_i} \qquad \text{eqn. 1.}$$

where parameter *N* is the sample size (the number of variants drawn from the previous time step) and parameter r_i is the adoption rate of the ith variant. This equation describes the proportional increase in each variant relative to other variants in the samples, so we need to divide the adoption rate of each variant by the sum of adoption rate of all variants in the sample. The adoption rate is included simply to allow increasing or decreasing trend in the usage frequency of some variants relative to others, under a biased sampling model. Variants with higher adoption rates than others tend to increase in usage frequency over time. Under an unbiased sampling model, where no class of variant is more or less likely to be adopted than any other, the values of r_i are equal for different variants. Repeating this sampling process for *t* time steps, when *t/N* is small, the usage frequency of variant *i* after *t* time steps can be well approximated by a Dirichlet distribution with parameters:

$$\alpha_i = \left\{e^{-(1-\frac{N}{N+\sum r_i})t}\left(\theta_i(0)-\frac{r_i}{\sum r_i}\right)+\frac{r_i}{\sum r_i}\right\}\left(\frac{N}{t}-1\right) \qquad \text{eqn. 2.}$$

In addition, unlike population genetic data, some speakers can also have zero variants for a particular language feature: for some language features, some speakers did not use a Gurindji, Kriol or innovative variant in the elicitation tasks, despite being given multiple opportunities to do so. Rather than missing data, these are typically features where there is only one variant available (either from Gurindji, Kriol, or an innovation), yet speakers have not used that variant and express the function in an alternative way. For example, younger generations do not use the Gurindji nasal cluster dissimilation (NCD) rule which changes the locative or ergative suffix from *-ngka* or *-ngku* into *-ka and -ku*, respectively (McConvell 1988). Failure to use variants of NCM and other language features is indicated by black bars in F82, F86, F90, F96, F105, F120 (Figure 9). We do not include these features in our Wright-Fisher model because they give little information on the relative adoption rates of different variants.

Another modification that we must make to the Wright-Fisher model to allow its extension from population genetic studies to datasets typical in linguistic studies is to distinguish the usage frequency of each variant of a language feature across the whole speaker population versus the usage frequency of a single speaker. For population genetics, an individual either carries an allele or does not, so the Wright-Fisher is used to model change in allele frequencies in a biological population. But for language evolution, a speaker has a chance of using a variant relative to the other variants. So the Wright-Fisher model cannot be used directly to model change in the number of speakers who use a variant in a speaker community, but to model change in variant usage frequencies of each speaker first, which then results in a change in variant frequencies of a speaker community. However, our data on the variant usage frequencies of each speaker are binary - each speaker either uses or does not use a particular variant (G, K and I) of each language feature. Because our data are derived from elicitation tasks, we are unable to calculate the frequency with which each variant is used in everyday speech of each speaker for all of the language features. So we link the data to the distribution of variant usage frequencies from the Wright-Fisher model by assuming that each binary variable represents an arbitrary threshold imposed on the distribution of variant usage frequencies: each individual has a frequency of usage for each language variant, but when that frequency of usage drops below 5% of the distribution, then that variant will have been recorded as 'absent' in the language elicitation tasks. Mathematically, the probability of a speaker not using a variant is the cumulative distribution function of the Dirichlet distribution described in equation 2. The 5% threshold is an arbitrary value typically used in analyses of distributions, however we test whether the selection of the threshold value has any impact on the results by rerunning all analyses with higher and lower thresholds, and demonstrate that it has little effect on the outcome of the analysis. This approach allows us to model all possible trajectories of changes in usage frequency of each speaker over generations, and calculate the likelihood of each speaker's presence/absence data given these trajectories. This is a strength of our approach, because it allows the Wright-Fisher model to be extended to a broader range of language change datasets. Instead of requiring usage frequency data, we modify the method to take binary (presence/absence) data, but model change along a frequency spectrum.

Finally, when applying a Wright-Fisher model to changes in a speaker's usage frequencies of different variants, we do not want to make assumptions about the way that language is acquired by the speaker. Because N is not the number of speakers, but the number of variants sampled from one time step to the next, it is not a known quantity because we do not know exactly how this sampling process occurs. For example, we don't know whether the speakers in one time step sample their language variants equally from all speakers from the previous time step, or whether this sampling process is biased toward some speakers. Nor do we know how often their usage frequencies are updated. Therefore, we treat the sample size (N) in the Wright-Fisher model as a hidden variable. To do this, we integrate out N by applying a uniform prior to N because N is a discrete variable and a uniform distribution is an uninformative prior for a discrete variable. The prior has upper boundary at 1000. Variance of usage frequencies is an inverse function of N. The conditional likelihood on even larger N is close to 0 and so increasing the higher bound of the prior does not influence the final estimates. We also need to account for the usage frequencies in the previous unrecorded generation ($\theta_i(0)$) in our likelihood calculation, because the usage frequencies of each speaker in Generation 1 depends on the usage frequencies in the previous generation (i.e. the 'great-grandparent' generation). For this study, given the recent language change due to colonisation, we can make a valid assumption that two generations before Generation 1, all speakers used only Gurindji variants, with no Kriol or innovations. When there is more than one Gurindji variant, we need to account for the relative frequency of different Gurindji variants in this Gurindji-speaking ancestral population. Since we cannot know these frequencies, we uniformly integrate over the simplex of different Gurindji variants in two generations before Generation 1 by a deterministic iterative adaptive algorithm implemented in the 'cuhre' function in R package 'R2Cuba' (Hahn, 2015).

As a result of these modifications to the Wright-Fisher model, our likelihood calculation starts with three Dirichlet distributions, each of which models the speaker's usage frequencies of different variants in each of the three generations. These distributions depend on a value of sample size (*N*) and a set of usage frequencies of Gurindji variants in two generations before Generation 1 ($\theta_i(0)$). For example, the usage frequencies of a language feature in a speaker in generation 1 follow the Dirichlet distribution described in equation 2 with $t = 2$ because there are two generations before Generation 1 (back to the earlier generation that spoke only Gurindji and no Kriol). The usage frequencies of a language feature in a speaker in generation 2 follow the Dirichlet distribution with $t = 3$. The cumulative distribution function of these Dirichlet distributions are then used to calculate the probability of a speaker using or not using particular variants in the elicitation task with an arbitrary threshold. For example, the likelihood of a speaker using variant G, but not variant K, is the probability from the Dirichlet distribution of the speaker's generation that the usage frequency of variant G is higher than the threshold and the usage frequency of variant K is lower than the threshold. The likelihood is then integrated over values of sample size and usage frequencies of Gurindji variants in two generations before Generation 1. As a result, there is a likelihood for each speaker using each particular language variant in the elicitation task. The product of likelihoods over language features and over speakers is the final likelihood of the presence/absence data given our model. A likelihood ratio test is performed to compare an unbiased sampling model (where all variants have the same adoption rates) to the biased sampling model (where some classes of variants have a greater or lower adoption rates than others). Depending on the alternative hypothesis, the biased sampling model is designed with different adoption rates for G, K, I variants, or different adoption rates for different complexity levels within G, K, I variants. Adoption rates of different variants are estimated using maximum likelihood. A significant result of the likelihood ratio test suggests that the adoption rates for different categories of variants are significantly different from each other.

4.6 RESULTS

We find the mixed language Gurindji Kriol did not form by a random sampling of Gurindji and Kriol variants, but was significantly biased towards the inclusion of Kriol variants (Figure 9; Table 4). Figure 9 shows the proportion of variants used for the language feature in the speaker community (represented by bar height). The speakers are divided into three generations: Generation 1 ('grandparent', >40 years of age), Generation 2 ('parent', 20-40 years) and Generation 3 ('child', in our dataset represented by children aged 8-14 years). The numbers refer to the 120 linguistic features (Table 2). Black bars represent the proportion of speakers who did not use any of the identified variants for a given language feature, although given multiple opportunities to do so in elicitation tasks (§4.3). Features 1-55 represent noun lexicon and features 81-106 are from the noun grammar. Similarly, features 66-80 are from the verb lexicon and features 107-120 are from the verb grammar. However, in practice, we have coded the complexity of the internal morphological structure of the nouns and verbs, which makes a clear differentiation between lexical and grammatical features unrealistic (§4.4).

Figure 9 Observed patterns of Gurindji (G), Kriol (K) and innovative (I) variants across three generations of Gurindji people

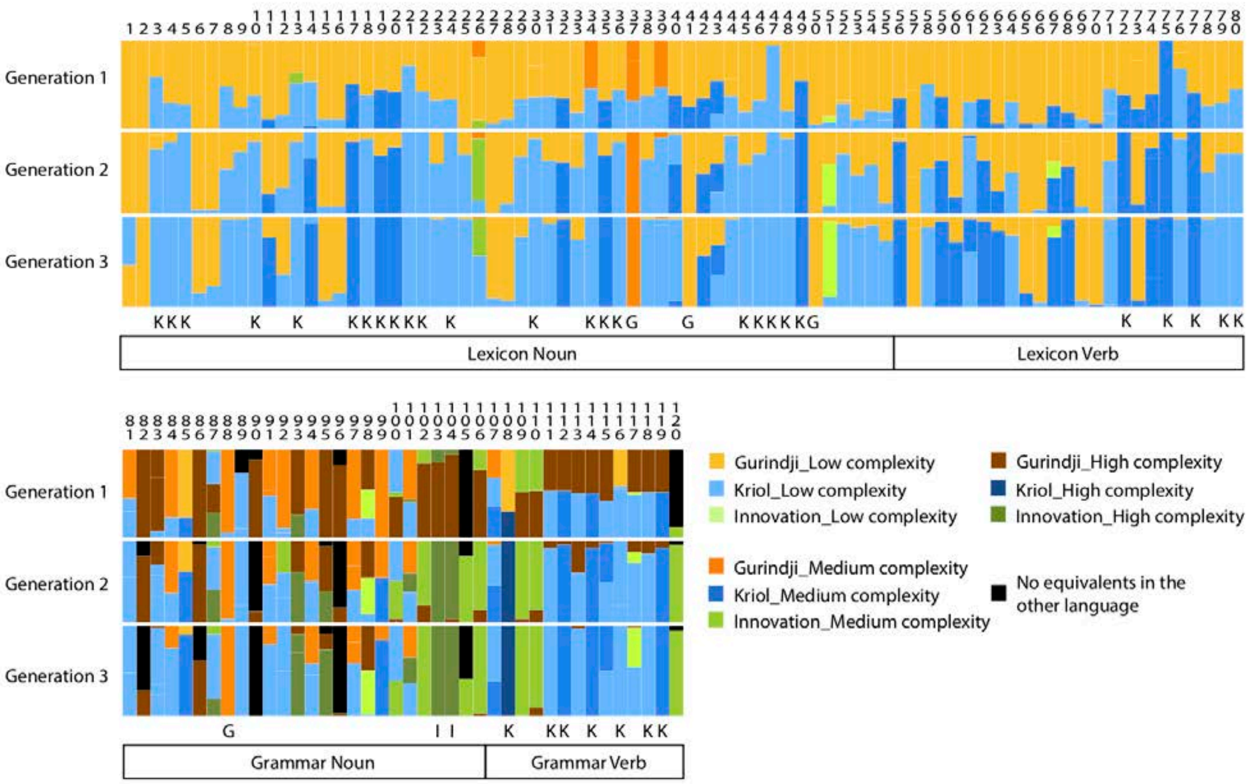

Figure 9 also shows that the formation of the mixed language is not simply the gradual replacement of Gurindji language elements with Kriol. In some cases, the Kriol variant used in Generation 1 is completely replaced by the Gurindji variant by Generation 3, and some innovative variants also become fixed by Generation 3. The letters ‘K(riol)’, ‘G(urindji)’ and ‘I(nnovation)’ along the bottom of the Generation 3 bar in Figure 9 show the language of the linguistic variants which have gone to fixation and replaced all other variants for that feature. For 33 features, the Kriol (K) variant has replaced the Gurindji (G) variant. For three features, G replaces K, and for two features, innovations (I) replace G (Figure 9).

Table 4 gives the results of the Wright-Fisher analysis of Gurindji (G), Kriol (K) and innovative (I) variants for all language features having three (G, K, I), or two (G, K) types of variants, and also separately on noun features and verb features which have G and K variants (combining both lexical and grammatical features: see Figure 9). For each set, we estimated adoption rates of each of the variants and performed a likelihood ratio test (LRT) against the null model that all variants have same adoption rate. Note that some sets have no innovative variants (see Table 2). There are only 16 features that include G, K and I, which are not enough to get reliable estimates for the separate categories nouns and verbs. For each analysis, we report the number of features (*n*), the estimated adoption rate, the negative log likelihood value (-lnL), and the likelihood ratio test (LRT) and its *p*-value. The analyses are repeated for three different thresholds to express binary data as usage frequencies (§4.5).

Table 4 gives the results of the likelihood ratio tests. Each line in the table reports the result of one model test, giving the set of language features being analysed (e.g. the set of all features that include G, K and I elements), the number of features that fall into that set (e.g. 16), the relative adoption rate of each class of elements, whether G, K or I, the likelihood of the model and the result of the likelihood ratio test (LRT) of this model against a null model in which all classes have the same rate of adoption, and the p-value that indicates whether this model provides a significantly better fit to the data than an equal rates model. Because the rates for G and K adoption are always greater than the rates of G adoption, and because the p value in all cases is well below 0.05, we conclude that K and I variants have a greater rate of adoption across the entire dataset, and its subdivisions into nouns and verbs, at all levels of the threshold for converting binary presence/absence into continuous frequencies. These tests allow us to identify a case of biased transmission, where one class of variants has a significantly greater rate of adoption over the three generations. We find that Kriol variants have a greater rate of adoption into the Gurindji Kriol language when considered over all language features (Table 4). Kriol variants also have a greater rate of adoption for noun features, and verb features. Innovations that generate novel variants of linguistic features through the hybridisation of Gurindji and Kriol variants also have a greater rate of adoption into the mixed language than Gurindji variants (Table 4).

Table 4 Wright-Fisher analysis of the relative adoption rates of Gurindji (G), Kriol (K), and innovative (I) variants

Language features	*n*	Adoption rate			-lnL	LRT	*p* - value
		Gurindji	Kriol	Innovation			
Threshold = 0.05							
All (G, K, I)	16	<0.001	2.765	3.437	2502	983	<0.001
All (G, K)	94	<0.001	5.926	--	9059	8266	<0.001
Noun (G, K)	59	<0.001	5.921	--	5756	5079	<0.001
Verb (G, K)	35	0.002	6.313	--	3301	3208	<0.001
Threshold = 0.01							
All (G, K, I)	16	0.001	2.278	3.460	4130	1363	<0.001
All (G, K)	94	<0.001	6.597	--	16283	12770	<0.001
Noun (G, K)	59	<0.001	6.249	--	10477	7447	<0.001
Verb (G, K)	35	0.310	8.593	--	5794	5000	<0.001
Threshold = 0.1							
All (G, K, I)	16	<0.001	3.206	3.821	1854	837	<0.001
All (G, K)	94	<0.001	5.586	--	6590	6483	<0.001
Noun (G, K)	59	<0.001	5.432	--	4172	3951	<0.001
Verb (G, K)	35	0.001	5.592	--	2417	2463	<0.001

Table 5 shows that the bias towards Kriol variants is not driven by preferential adoption of simpler language variants. For each set of language features, we start with a null model where G, K, I variants have different adoption rates, but there is no difference in rate between complexity levels within Gurindji, Kriol, and innovation (see Table 4). We first test if adoption rates differ among complexity levels within Gurindji by asking if allowing differences in adoption rate among complexity levels within Gurindji increases model fit. We then test if adoption rates differ among complexity levels within Kriol by asking if allowing difference in adoption rate among complexity levels within both Gurindji and Kriol increases model fit compared to the model in the previous step. Lastly, for the dataset with G, K, and I variants, we test if adoption rates differ among complexity levels within the class of innovative variants by asking if allowing differences in the adoption rate among complexity levels within Gurindji, Kriol, and innovations further increases model fit (compared to the model in the previous step). For each analysis, we report the number of features (*n*), the estimated adoption rate, the negative log likelihood value (-lnL), and the likelihood ratio test (LRT) and its *p*-value. Taking, for example, the 94 features (all = noun+verb) which have a Gurindji and Kriol variant (G, K), we give the relative adoption rates for high medium and low complexity features within each category (e.g. near zero rate of adoption of G elements of all complexity levels, but greater rates of adoption of high complexity K variants over medium and low), the likelihood of that model (-lnL), the null model it is compared to (e.g. H, M and L have same rate of adoption within each category), the likelihood ratio of the comparison between the model and the null model. A p-value less that 0.05 suggests that the model parameters are a significantly better description of the data than the null model of equal rates The analyses are repeated for three different thresholds to express binary data as usage frequencies (§4.5). We demonstrate that the threshold chosen to convert binary data to frequency data has little effect on the conclusions of our analysis (Table 4 and 5). Our approach, therefore, demonstrates not only how modelling can be adapted to the nature of the data at hand, but also how to test whether such modelling decisions affect the results of the analysis.

Table 5 Wright-Fisher analyses of the relative adoption rates of Gurindji (G), Kriol (K), and innovative (I) variants with High, Medium and Low complexity levels

Language features	*n*	Variant	Adoption rate			-lnL	Null model	LRT	*p*-value
			H	M	L				
Threshold = 0.05									
All (G, K, I)	16	G	<0.001	<0.001	<0.001	2502	No difference in rate among H, M, L within G, K, I	0.002	0.999
		K	--	2.538	2.786	2499	No difference in rate among H, M, L within K, I	5.516	0.063
		I	4.968	3.647	2.328	2433	No difference in rate among H, M, L within I	132	<0.001
All (G, K)	94	G	<0.001	<0.001	<0.001	9059	No difference in rate among H, M, L within G, K	0.014	0.993

		K	17.162	5.847	5.924	9039	No difference in rate among H, M, L within K	41	<0.001
Noun (G, K)	59	G	<0.001	<0.001	<0.001	5756	No difference in rate among H, M, L within G, K	0.010	0.995
		K	--	6.426	5.681	5745	No difference in rate among H, M, L within K	22	<0.001
Verb (G, K)	35	G	<0.001	<0.001	0.001	3259	No difference in rate among H, M, L within G, K	85	<0.001
		K	16.897	5.724	7.028	3249	No difference in rate among H, M, L within K	20	<0.001
Threshold = 0.01									
All (G, K, I)	16	G	<0.001	2.197	0.001	2502	No difference in rate among H, M, L within G, K, I	0.370	0.831
		K	--	2.419	2.488	4118	No difference in rate among H, M, L within K, I	23	<0.001
		I	7.988	4.291	2.194	3993	No difference in rate among H, M, L within I	251	<0.001
All (G, K)	94	G	0.001	0.708	0.001	16282	No difference in rate among H, M, L within G, K	1.460	0.482
		K	23.916	6.604	6.604	16282	No difference in rate among H, M, L within K	122	<0.001
Noun (G, K)	59	G	<0.001	<0.001	<0.001	10477	No difference in rate among H, M, L within G, K	<0.001	1.000
		K	--	6.951	6.090	10461	No difference in rate among H, M, L within K	30	<0.001
Verb (G, K)	35	G	<0.001	0.002	0.237	3259	No difference in rate among H, M, L within G, K	125	<0.001
		K	30.810	6.847	9.121	5655	No difference in rate among H, M, L within K	151	<0.001
Threshold = 0.1									

All (G, K, I)	16	G	<0.001	0.001	<0.001	1854	No difference in rate among H, M, L within G, K, I	0.002	0.999
		K	--	2.809	3.169	1850	No difference in rate among H, M, L within K, I	6.758	0.034
		I	7.301	3.787	2.527	1787	No difference in rate among H, M, L within I	126	<0.001
All (G, K)	94	G	<0.001	<0.001	<0.001	6590	No difference in rate among H, M, L within G, K	0.056	0.972
		K	13.343	5.535	5.592	6576	No difference in rate among H, M, L within K	29	<0.001
Noun (G, K)	59	G	<0.001	<0.001	<0.001	4172	No difference in rate among H, M, L within G, K	<0.001	1.000
		K	--	5.688	4.924	4156	No difference in rate among H, M, L within K	32	<0.001
Verb (G, K)	35	G	<0.001	0.002	0.686	2397	No difference in rate among H, M, L within G, K	85	<0.001
		K	18.435	6.335	7.212	2367	No difference in rate among H, M, L within K	60	<0.001

Our results give no support to the hypothesis that contact-induced language change will favour simplification, because we find no evidence that simpler variants have a higher rate of adoption than more complex variants, for any class of features. Table 5 shows that Kriol variants are more likely to be adopted into the mixed language than Gurindji variants, regardless of their complexity levels: the estimated adoption rates of Kriol variants with different complexity levels are always higher than the adoption rates of Gurindji variants. Medium-complexity Kriol variants are frequently adopted over their low-complexity Gurindji alternatives (Figure 9). For example, many of the nouns and verbs from Gurindji have a low-complexity morphological structure compared with Kriol nouns and verbs e.g. *jintaku* (G) versus *wan-bala* (K) 'one' or *warlakap* (G) versus *luk-aran* (K) 'search'. Feature 108 is the only feature that has high-complexity Kriol variant and this high-complexity Kriol variant has replaced the low-complexity Gurindji variant over the three generations. In addition, Table 5 shows that medium-complexity Kriol variants of nouns have a significantly greater adoption rate than low-complexity Kriol variants of nouns in the Noun (G,K) dataset. High-complexity innovative variants also have a greater adoption rate than medium- and low-complexity innovative variants, and medium-complexity innovative variants have a greater adoption rate than low-complexity innovative variants in the All (G,K,I) dataset. However, medium-complexity Kriol variants of verbs have lower adoption rate than low-complexity Kriol variants in the Verb (G,K) dataset. Similarly, medium-complexity Gurindji variants of verbs have lower adoption rate than low-complexity Gurindji variants in the Verb (G,K) dataset. This pattern is mostly driven by the adoption of Kriol verb grammar into the mixed language and the fact that there are many low-complexity Kriol variants and medium-complexity Gurindji variants in verb grammar (Figure 9).

5. DISCUSSION

5.1 CONTACT-INDUCED CHANGE AND SIMPLIFICATION

Our results show that the formation of a contact language does not necessarily involve a general preference for the adoption of simpler language variants. The lack of bias towards simpler variants may be because Gurindji Kriol is the primary language of the younger generation in the Gurindji community. While morphologically complex languages may be more challenging for adults to learn, they are not more difficult for children to learn (Miestamo 2008; Stoll, Mazara, & Bickel 2017). Therefore, a critical question for understanding contact languages is whether their formation is driven primarily by adult learners or whether it is a child acquisition phenomenon (Trudgill 2011). Indeed the cause of contact-induced simplification is often attributed to large populations of adult second language learning (Lupyan & Dale 2010; Trudgill 2011). As Trudgill (2001, p. 372) suggests, ‘complexity disappears as a result of the lousy language-learning abilities of the human adult. Adult language contact means adult language learning; and adult language learning means simplification.’

O’Shannessy (2012) suggests that the neighbouring mixed language, Light Warlpiri (§3), is the result of child acquisition rather than adult L2 learning, which may be why many inflectional features from Warlpiri are retained. Certainly morphology has been noted as being striking in its presence in some mixed languages in comparison to other contact languages such as creoles. Matras (2003) suggests it is not unusual in mixed languages to find the transfer of grammatical elements, which are often considered ‘loan proof’, for example inflectional morphology. Indeed, the most extreme cases of mixed languages, Michif (Bakker 1997), Mednyj Aleut (Golovko & Vakhtin 1990), Gurindji Kriol (Meakins 2011a, 2011b) and Light Warlpiri (O'Shannessy 2005), retain inflectional morphology from both source languages.

5.2 MODELLING MULTIPLE VARIABLES IN SINGLE POPULATIONS

This paper also demonstrates the power of a large linguistic dataset for testing hypotheses and detecting general trends. Modelling language change has traditionally been the preserve of Variationist Sociolinguistics (Labov 1994b) and, to a lesser extent, Probabilistic Syntax (Bresnan 2007). These approaches use mixed models and variable rules analysis to gauge the relative use of a linguistic variant against a range of grammatical and social factors. But this methodology is often restricted to a single language variable, and makes simplifying assumptions that are analytically tractable. For example, such methods may assume that residual follows a normal distribution, which may not adequately capture the random variation in language usage generated by the way that languages are transmitted over generations. By allowing simultaneous analysis of a large number of language variables and by explicitly modelling the way that languages are transmitted over generations, our analysis allows us to adapt classical evolutionary models, providing a broader view of variability and change in a linguistic system.

We have modelled language change by adapting the Wright-Fisher population genetic model to evaluate whether the adoption of elements into the mixed language is unbiased or biased with respect to source language and level of complexity. While this model has been shown to generate plausible patterns of language change, and has been used to generate null predictions (§2), this is the first application to studying variation and change for a population-level sample of a large number of language features within a single speech community. The Wright-Fisher model describes the expected outcomes of a process where the variants present in one step are sampled to generate the variation in the next step. We integrated out uncertainties in sampling error and usage frequency in previous generations when calculating the likelihood of the language data given our model. The model allows us to detect cases where the usage of linguistic variants have increased or decreased at a significantly greater rate than other variants with respect to the previous generation.

Our modifications to the Wright-Fisher model allow a more general applicability to language change, by permitting application to a wider range of types of data. In order to analyse presence/absence data, where we record only whether a particular speaker uses a given variant of a linguistic feature or not, we modified the Wright-Fisher equations to analyse binary language data rather than continuous frequency data, and to allow loss of linguistic features. We do this by modelling the underlying frequency distribution, and demonstrate that the threshold chosen to convert binary data to frequency data has little effect on the conclusions of our analysis. Our approach, therefore, demonstrates not only how modelling can be adapted to the nature of the data at hand, but also how to test whether such modelling decisions affect the results of the analysis. However, this analysis could be extended to the use of frequency data where available. In our dataset, speakers were just coded for whether they used the variant or not (binary) rather than the frequency of use. So where two or more variants are used, they are encoded as '1' and '1', which could represent '50%' and '50%' but equally it could represent '10%' and '90%' or '90%' and '10%' etc. The relative frequency of use is not used in this study but is known to be relevant to language change (e.g. Backus 2014). When data on relative frequency of use are available, these data can be readily analysed by our model, by using the probability density function of the Dirichlet distribution we derived in this study as the likelihood function, rather than setting an arbitrary threshold, and using the cumulative distribution function of the Dirichlet distribution as the likelihood function.

6. Conclusion

In this paper, we have shown that the analysis of a single, well-documented contact language can be used to test the generality of linguistic processes. Through a large dataset consisting of multiple variables from three generations of speakers, we were able to show that in this particular case, the formation of a contact language was not driven by a preference for simple linguistic variants. Extension of this approach to other contact languages would allow further investigation of the link between morphological simplification and contact, for example by comparing languages with different agents of change (e.g. children versus adults).

This paper has also demonstrated an innovative application of biological modelling to linguistic data. While adoption of methods from evolutionary biology is now commonplace in historical linguistics, particularly the use of phylogenetic methods, this is a novel application of population genetics methods to community-level language change, encompassing a large and unbiased sample of language features. Models of language change have been variously interpreted as the uptake of variants from a feature pool (Mufwene 2001), or the propagation of 'linguemes' i.e. linguistic variants (Croft 2000). The Wright-Fisher model is agnostic to the actual mechanisms whereby variants are sampled from one point in a time series to the next, so we can report the changing patterns of variant use without inferring specific processes. The flexibility of this approach may be useful in addressing general aspects of language change over different generations of the same speech community, given appropriate data. For example, it might be applied to the changes in phonetic patterns in different groups of speakers of a language such as English which are undergoing less rapid change which may in turn provide clues are to how linguistic diversification or 'speciation' occurs (Evans 2016b).

7. REFERENCES

ACKERMAN, FARRELL, JAMES BLEVINS and ROBERT MALOUF. 2009. Parts and wholes: Patterns of relateness in complex morphological systems and why they matter. *Analogy in grammar: Form and acquisition*, ed. by J. Blevins and J. Blevins, 54-82. Oxford: Oxford University Press.

ACKERMAN, FARRELL and ROBERT MALOUF. 2013. Morphological organization: The low conditional entropy conjecture. *Language* 89.429-65.

ANDERSON, STEPHEN. 1992. *A-Morphous morphology.* Cambridge: Cambridge University Press.

ANDERSON, STEPHEN. 2015. Dimensions of morphological complexity. *Understanding and measuring morphological complexity*, ed. by M. Baerman, D. Brown and G. Corbett, 11-26. Oxford: Oxford University Press.

AUSTIN, PETER and ANDREW SIMPSON. 2007. *Endangered languages*. Hamburg: Helmut Buske.

BACKUS, AD. 2014. Towards a usage-based account of language change: Implications of contact linguistics for linguistic theory. *Questioning language contact: Limits of contact, contact at its limits*, ed. by R. Nicolai, 91-118. Leiden: Brill.

BAERMAN, MATTHEW, DUNSTAN BROWN and GREVILLE CORBETT. 2015. What is morphological complexity? *Understanding and measuring morphological complexity*, ed. by M. Baerman, D. Brown and G. Corbett, 3-10. Oxford: Oxford University Press.

BAILEY, GUY, TOM WIKLE, JAN TILLERY and LORI Sand. 1991. The apparent time construct. *Language Variation and Change* 3.241-64.

BAKKER, PETER. 1997. *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Métis.* New York: Oxford University Press.

BAKKER, PETER. 2015. Typology of mixed languages. *The Cambridge handbook of linguistic typology*, ed. by A.Y. Aikhenvald and R.M.W. Dixon. Cambridge: Cambridge University Press.

BAKKER, PETER, AYMERIC DAVAL-MARKUSSEN, MIKAEL PARKVALL and INGO PLAG. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26.5-42.

BAXTER, G. J., R. A. BLYTHE, W. CROFT and A. J. MCKANE. 2006. Utterance selection model of language change. *Physical Review* E 73.046118.

BAXTER, G. J., R. A. BLYTHE, W. CROFT and A. J. MCKANE. 2009. Modeling language change: an evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change* 21.257-96.

BENTLEY, R. A., C. P. LIPO, H. A. HERZOG and M. W. HAHN. 2007. Regular rates of popular culture change reflect random copying. *Evolution and Human Behavior* 28.151-58.

BENTLEY, R. A., P. ORMEROD and M. BATTY. 2011. Evolving social influence in large populations. *Behavioral Ecology Sociobiology* 65.537–46.

BENTLEY, R.A., M.W. HAHN and S.J. SHENNAN. 2004. Random drift and culture change. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271.1443-50.

BLASI, DAMIÁN, SUSANNE MICHAELIS and MARTIN HASPELMATH. 2017. Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behaviour (Letters)* 1.723-29.

BLYTHE, R. A. 2012. Neutral evolution: a null model for language dynamics. *Advances in Complex Systems* 15.1150015.

BOOIJ, GEERT. 1996. Inherent versus contextual inflection and the split morphology hypothesis. *Yearbook of Morphology* 1995.1-16.

BRESNAN, JOAN. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. *Roots: Linguistics in search of its evidential base*, ed. by S. Featherston and W. Sternefeld, 77-96. Berlin: Mouton de Gruyter.

CAMPBELL, LYLE and MARTHA C. MUNTZELL. 1989. The structural consequences of language death. *Investigating obsolescence: Studies in language contraction and death*, ed. by N. Dorian, 181-96. Cambridge: Cambridge University Press.

CHAROLA, ERIKA and FELICITY MEAKINS (eds) 2016. *Yijarni: True stories from Gurindji country*. Canberra: Aboriginal Studies Press.

CROFT, WILLIAM. 2000. *Explaining language change: An evolutionary approach.* Harlow, England: Longman.

DAHL, ÖSTEN. 2004. *The growth and maintenance of linguistic complexity.* Amsterdam: John Benjamins.

DE GROOT, CASPER. 2008. Morphological complexity as a parameter of linguistic typology: Hungarian as a contact language. *Language complexity: Typology, contact, change*, ed. by M. Miestamo, K. Sinnemäki and F. Karlsson, 191-214. Amsterdam: John Benjamins.

DEGRAFF, MICHEL. 2005. Linguists' most dangerous myth: The fallacy of Creole Exceptionalism. *Language in Society* 34.533-91.

DORIAN, NANCY. 1976. Gender in a terminal Gaelic dialect. *Scottish Gaelic Studies* 12.279-82.

DORIAN, NANCY. 1978. The fate of morphological complexity in language death: Evidence from East Sutherland Gaelic. *Language* 54.590-609.

DORIAN, NANCY. 1989. *Investigating obsolescence: Studies in language contraction and death.* Cambridge: Cambridge University Press.

EVANS, NICHOLAS. 2016a. As intimate as it gets? Paradigm borrowing in Marrku and its implications for the emergence of mixed languages. *Loss and renewal: Australian languages since contact*, ed. by F. Meakins and C. O'Shannessy, 29-56. Berlin: Mouton de Gruyter.

EVANS, NICHOLAS. 2016b. Typology and coevolutionary linguistics. *Linguistic Typology* 20.505-20.

EVANS, NICHOLAS, ALICE GABY, STEPHEN LEVINSON and ASIFID MAJID. 2011. *Reciprocals and semantic typology.* Amsterdam: John Benjamins.

FISHER, ROLand A. 1956. *The genetical theory of natural selection.* New York: Wiley.

GAL, SUSAN. 1989. Lexical innovation and loss: Restricted Hungarian. *Investigating obsolescence: Studies in language contraction and death*, ed. by N. Dorian, 313-31. Cambridge: Cambridge University Press.

GARDANI, FRANCESCO. 2008. *Borrowing of inflectional morphemes in language contact.* Frankfurt: Peter Lang.

GARDANI, FRANCESCO, PETER ARKADIEV and NINO AMIRIDZE. 2014. *Borrowed morphology.* Berlin: Mouton de Gruyter.

GOLOVKO, EVGENIY V. and NIKOLAI VAKHTIN. 1990. Aleut in contact: The CIA enigma. *Acta Linguistica Hafniensia* 72.97-125.

HAHN, THOMAS. 2015. *R2Cuba: Multidimensional Numerical Integration.* https://cran.r-project.org/package=R2Cuba.

HUFFINES, MARION LOIS. 1989. Case usage among the Pennsylvania German sectarians and nonsectarians. *Investigating obsolescence: Studies in language contraction and death*, ed. by N. Dorian, p. 211-26. Cambridge: Cambridge University Press.

JANSE, MARK and SIJMEN TOL (eds) 2003. *Language death and language maintenance:*

Theoretical, practical and descriptive approaches. Amsterdam: John Benjamins.

JONES, CAROLINE, FELICITY MEAKINS and SHUJAU MUAWIYATH. 2012. Learning vowel categories from maternal speech in Gurindji Kriol. *Language Learning* 62.997–1260.

KIRBY, S., T. GRIFFITHS and K. SMITH. 2014. Iterated learning and the evolution of language. *Current Opinion in Neurobiology* 28.108-14.

KUSTERS, WOUTER. 2008. Complexity in linguistic theory, language learning and language change. *Language complexity, typology, contact, change*, ed. by M. Miestamo, K. Sinnemäki and F. Karlsson, 3-22. Amsterdam: John Benjamins.

LABOV, WILLIAM. 1963. The social motivation of a sound change. *Word* 19.273-309.

LABOV, WILLIAM. 1994a. *Principles of linguistic change: Internal factors*. Oxford: Blackwell.

LABOV, WILLIAM. 1994b. *Principles of linguistic change: Social factors*. Oxford: Blackwell.

LEVINSON, STEPHEN and DAVID WILKINS. 2006. *Grammars of space*. Cambridge: Cambridge University Press.

LUPYAN, GARY and RICK DALE. 2010. Language structure is partly determined by social structure. *PLoS ONE* 5.

MACWHINNEY, BRIAN. 2000. *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

MATRAS, YARON. 2003. Mixed languages: Re-examining the structural prototype. *The mixed language debate: Theoretical and empirical advances*, ed. by Y. Matras and P. Bakker, 151-76. Berlin: Mouton de Gruyter.

MATRAS, YARON. 2007. The borrowability of structural categories. *Grammatical borrowing in cross-linguistic perspective*, ed. by Y. Matras and J. Sakel, 31-73. Berlin: Mouton de Gruyter.

MATRAS, YARON and PETER BAKKER. 2003. The study of mixed languages. *The mixed language debate: Theoretical and empirical advances*, ed. by Y. Matras and P. Bakker, 1-20. Berlin: Mouton de Gruyter.

MATRAS, YARON and JEANETTE SAKEL. 2007. *Grammatical borrowing in cross-linguistic perspective*. Berlin: Mouton de Gruyter.

MAYER, MERCER. 1994 (1969). *Frog, where are you?* China: Puffin.

MCCONVELL, PATRICK. 1988. Nasal cluster dissimilation and constraints on phonological variables in Gurindji and related languages. *Aboriginal Linguistics* 1.135-65.

MCCONVELL, PATRICK and FELICITY MEAKINS. 2005. Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics* 25.9-30.

MCWHORTER, JOHN. 1998. Identifying the creole prototype: Vindicating a typological claim. *Language* 74.788-818.

MCWHORTER, JOHN. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5.125-66.

MEAKINS, FELICITY. 2008. Unravelling languages: Multilingualism and language contact in Kalkaringi. *Children’s language and multilingualism: Indigenous language use at home and school*, ed. by J. Simpson and G. Wigglesworth, 247-64. 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. *The Role of semantics and pragmatics in the development of case.*, ed. by J. Barddal and S. Chelliah, 59-91. Amsterdam: John Benjamins.

MEAKINS, FELICITY. 2010. The development of asymmetrical serial verb constructions in an Australian mixed language. *Linguistic Typology* 14.1-38.

MEAKINS, FELICITY. 2011a. Borrowing contextual inflection: Evidence from northern

Australia. *Morphology* 21.57-87.

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* 31.43-77.

MEAKINS, FELICITY. 2012. Which Mix? - Code-switching or a mixed language - Gurindji Kriol. *Journal of Pidgin and Creole Languages* 27.105-40.

MEAKINS, FELICITY. 2013a. Gurindji Kriol. *The survey of pidgin and creole languages, Vol III*, ed. by S. Michaelis, P. Maurer, M. Haspelmath and M. Huber, 131-39. Oxford: Oxford University Press.

MEAKINS, FELICITY. 2013b. Mixed languages. *Contact languages: A comprehensive guide*, ed. by Y. Matras and P. Bakker, 159-228. Berlin: Mouton.

MEAKINS, FELICITY. 2015. From absolutely optional to only nominally ergative: The life cycle of the Gurindji Kriol ergative suffix. *Borrowed morphology*, ed. by F. Gardani, P. Arkadiev and N. Amiridze, 189-218. Berlin: Mouton de Gruyter.

MEAKINS, FELICITY. 2016. No fixed address: The grammaticalisation of the Gurindji locative as a progressive suffix. *Loss and renewal: Australian languages since colonisation*, ed. by F. Meakins and C. O'Shannessy, 367-96. Berlin: Mouton de Gruyter.

MEAKINS, FELICITY and CASSANDRA ALGY. 2016. Deadly reckoning: Changes in Gurindji children's knowledge of cardinals. *Australian Journal of Linguistics* 36.479-501.

MEAKINS, FELICITY, JENNIFER GREEN and MYFANY TURPIN. 2018. *Understanding linguistic fieldwork.* London: Routledge.

MEAKINS, FELICITY, CAROLINE JONES and CASSANDRA ALGY. 2016. Bilingualism, language shift and the corresponding expansion of spatial cognitive systems. *Language Sciences* 54.1-13.

MEAKINS, FELICITY and RACHEL NORDLINGER. 2014. *A grammar of Bilinarra: An Australian Aboriginal language of the Northern Territory.* Berlin: Mouton de Gruyter.

MEAKINS, FELICITY and CARMEL O'SHANNESSY. 2010. Ordering arguments about: Word order and discourse motivations in the development and use of the ergative marker in two Australian mixed languages. *Lingua* 120.1693–713.

MEAKINS, FELICITY and CARMEL O'SHANNESSY. 2012. Typological constraints on verb integration in two Australian mixed languages. *Journal of Language Contact* 5.216-46.

MEAKINS, FELICITY and SASHA WILMOTH. 2019. Complex cell-mates: Morphological overabundance resulting from language contact. *Morphological complexity*, ed. by P. Arkadiev and F. Gardani. Oxford: Oxford University Press.

MIESTAMO, MATTI. 2006. On the feasibility of complexity metrics. FinEst linguistics: *Proceedings of the annual Finnish and Estonian conference of linguistics*, ed. by K. Kerge and M.-M. Sepper, 11-26. Tallinn: TLU.

MIESTAMO, MATTI. 2008. Grammatical complexity in a cross-linguistic perspective. *Language complexity: typology, contact, change*, ed. by M. Miestamo, K. Sinnemäki and F. Karlsson, 23–41. Amsterdam: John Benjamins.

MIESTAMO, MATTI, KAIUS SINNEMÄKI and FAIUS KARLSSON (eds) 2008. *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins.

MUFWENE, SALIKOKO. 2001. *The ecology of language evolution.* Cambridge: Cambridge University Press.

MYERS-SCOTTON, CAROL. 2002. *Contact linguistics: Bilingual encounters and grammatical outcomes.* Oxford: Oxford University Press.

NEWBERRY, MITCHELL, CHRISTOPHER AHERN, ROBIN CLARK and JOSHUA PLOTKIN. 2017.

Detecting evolutionary forces in language change. *Nature* 551.223-26.

NICHOLS, JOHANNA. 2009. Linguistic complexity: A comprehensive definition and survey. *Language complexity as an evolving variable*, ed. by G. Sampson, D. Gil and P. Trudgill, 110-25. Oxford: Oxford University Press.

O'SHANNESSY, CARMEL. 2004. *The monster stories: Picture stimuli to elicit lexical subject NPs.* Nijmegen, Holland: Max-Planck-Institut-für-Psycholinguistik.

O'SHANNESSY, CARMEL. 2005. Light Warlpiri: A new language. *Australian Journal of Linguistics* 25.31-57.

O'SHANNESSY, CARMEL. 2012. The role of code-switched input to children in the origin of a new mixed language. *Linguistics* 50.305-40.

PARKVALL, MIKAEL. 2008. The simplicity of creoles in cross-linguistic perspective. *Language complexity: Typology, contact, change*, ed. by M. Miestamo, K. Sinnemäki and F. Karlsson, 265-85. Amsterdam: John Benjamins.

PLAG, INGO. 2003a. Introduction: The morphology of creole languages. *Yearbook of Morphology 2002*, ed. by G. Booij and J. van Marle, 1-2. Alphen aan den Rijn, Netherlands: Kluwer.

PLAG, INGO. 2003b. *Yearbook of Morphology 2002.* Great Britain: Kluwer.

POPLACK, SHANA. 2015. Code-switching (Linguistic). *International encyclopedia of the social and behavioral sciences*, ed. by J. Wright, 918-25. United States: Elsevier.

REALI, F. and T. L. GRIFFITHS. 2010. Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. *Proceedings of Royal Society B* 277.429 – 36.

ROBERTS, SARAH and JOAN BRESNAN. 2008. Retained inflectional morphology in pidgins: A typological study. *Linguistic Typology* 12.269-302.

SALLABANK, JULIA and PETER AUSTIN. 2011. *The Cambridge handbook of endangered languages.* Cambridge: Cambridge University Press.

SAPIR, EDWARD. 1921. *Language: An introduction to the study of speech.* New York: Harcourt Brace and company.

SCHMIDT, ANNETTE. 1985. *Young people's Dyirbal: An example of language death from Australia.* Cambridge: Cambridge University Press.

SEIFART, FRANK. 2012. The Principle of Morphosyntactic Subsystem Integrity in language contact: Evidence from morphological borrowing in Resígaro (Arawakan). *Diachronica* 29.471-504.

STOLL, SABINE, JEKATERINA MAZARA and BALTHASAR BICKEL. 2017. The acquisition of polysynthetic verb forms in Chintang. *The Oxford handbook of polysynthesis*, ed. by M. Fortescue, M. Mithun and N. Evans. Oxford: Oxford University Press.

TATARU, P., M. SIMONSEN, T. BATAILLON and A. HOBOLTH. 2017. Statistical inference in the Wright-Fisher model using allele frequency data. *Systematic Biology* 66.e30-e46.

THOMASON, SARAH GREY (ed.) 1997. *Contact languages: A wider perspective.* Amsterdam: John Benjamins.

THOMASON, SARAH GREY and TERRENCE KAUFMAN. 1988. *Language contact, creolization, and genetic linguistics.* Berkeley: University of California Press.

TRUDGILL, PETER. 2001. Contact and simplification: Historical baggage and directionality in linguistic change. *Linguistic Typology* 5.371–74.

TRUDGILL, PETER. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity.* Oxford: Oxford University Press.

VAN DEN BOS, JACKIE, FELICITY MEAKINS and CASSANDRA ALGY. 2017. Searching for "Agent Zero": The origins of a relative case. *Language Ecology* 1.4-24.

WEINREICH, URIEL. 1974 [1953]. *Languages in contact: Findings and problems.* The Hague:

Mouton.
WRIGHT, SEWELL. 1931. Evolution in Mendelian populations. *Genetics* 16.97–159.

---

[i] Abbreviations: ABL – ablative, ADV – adverbalizer, AGENT – agent, ALL – allative, COM – comitative, CONT – continuative, CONTR – contrastive focus, DAT – dative, dem – demonstrative, DIR – directional, EP – epenthetic, ERG – ergative, EXC – exclusive, FOC – focus, IMP – imperative, IMPF – imperfective, INC – inclusive, LOC – locative, n – noun, n:bp – body part noun, n:inanimate – inanimate noun, NEG – negative, NMLZ – nominaliser, NOM – nominative, NOW – focus, NPST – noon-past, O – object, OBL – oblique, ONLY – restrictive, PL – plural, pro – pronoun, PROP – proprietive, PRS – present tense, PST – past tense, REDUP – reduplication, S – subject, SG – singular, suf – suffix, TEL – telic, TOP – topic, TR – transitive, TOP – topic, v:aux – auxiliary verb, v:tran – transitive verb, 1 – first person, 2 – second person, 3 – third person, - – morpheme break, = – clitic break, &g – Gurindji origin, &k – Kriol origin.