Variation Rolls the Dice

A worldwide collage in honour of Salikoko S. Mufwene

EDITED BY

Enoch Oladé Aboh
Cécile B. Vigouroux

JOHN BENJAMINS PUBLISHING COMPANY

Variation Rolls the Dice

Contact Language Library (COLL)

Contact Language Library (CoLL), formerly *Creole Language Library,* presents descriptive and theoretical studies on contact languages, such as pidgin and creole languages, mixed languages and other languages which have undergone significant contact-induced change. All CoLL publications are anonymously and internationally refereed.

For an overview of all books published in these series, please see *benjamins.com/catalog/cll* and *benjamins.com/catalog/coll*

Volume 59

Variation Rolls the Dice. A worldwide collage in honour of Salikoko S. Mufwene
Edited by Enoch Oladé Aboh and Cécile B. Vigouroux

Variation Rolls the Dice

A worldwide collage in honour of Salikoko S. Mufwene

Edited by

Enoch Oladé Aboh
University of Amsterdam

Cécile B. Vigouroux
Simon Fraser University

Table of contents

Acknowledgements

This editorial project has been both exciting and enriching, and we hope that the readers will find as much pleasure as we had in editing this book.

Acronyms and Glosses

Acronyms

ALA	Adult Language Acquisition
APEC	Asia-Pacific Economic Cooperation
APiCs	Atlas of Pidgin & Creole Language Structures
BV	Basic Variety (model)
CL	Computational Linguistics
ERC	English-Related Creole/s
ESF	European Science Foundation
FoR	Frames of Reference
FRC	French-Related Creole/s
FSL	French as a Second Language
FUO	Finite Utterance Organization
IPA	International Phonetic Alphabet
IUO	Infinite Utterance Organization
MPI	Max Planck Institute for Psycholinguistics
NSW	New South Wales
NUO	Nominal Utterance Organization
P/Cs	Pidgin and Creole languages
SCA	Sound Class Alignment
SgE	Singapore English
SLA	Second Language Acquisition
TL	Target Language
TMA	Tense Mood Aspect
TPAS	Peranakan Association of Singapore
WALS	World Atlas of Language Structures

Glosses

1/2/3/	First/second/third/person
1>23	1st person subject, 2nd or 3rd person object
ABS	Absolutive
ACC	Accusative

AF	Affirmative
AFFECT	Affective
AG	Agentive
AGR	Agreement
ALL	Allative
ANT	Anterior
AREA	Area
ASS	Associative
AUX	Auxiliary
BEN	Benefactive
CAR	Endearment
CL	Classifier
CNG	Congruence
CNTR	Contrastive
COM	Comitative
COMPL	Completive
CONJ	Conjunction
CONT	Continuative
COP	Copula
COOR	Coordinator
DAT	Dative
DCL	Declarative
DEIC	Deictic
DEL	Delimitative
DEM	Demonstrative
DET	Determiner
DIM	Diminutive
DIR	Directional
DIS	Discourse marker
DISJ	Disjunction
DS	Different subject
DU/DUAL	Dual
DUB	Dubitative
EMP	Emphatic
ERG	Ergative
EU	Euphonic
EV.SEEN	EVIDENTIAL: Seen without speaker participation
EVI	Evidential
EVIT	Evitative
EX/EXCL	Exclusive

EXIS	Existential
F	Feminine
FOC	Focus
FU/FUT	Future
GEN	Genitive
HORT	Hortative
IDEO	Ideophone
IFR	Inferential
IM	Imperative
IMP/IMPFV	Imperfective
IN/INCL	Inclusive
INC	Inchoative
INCP	Incipient
IND	Indefinite
INF	Infinitive
INT	Intentional
INTENS	Intensifier
ITER	Iterative
LOC	Locative
MOD	Modality
NEG	Negation/Negative
NFUT	Non-future
NMLZ	Nominaliser
NOM	Nominative
NPST	Non-past
NS	Non-singular
Ø	Null marking
OB	Object
OBL	Oblique
PAS/PST	Past
PAS2	Secondary past
PERF/PFV	Perfective
PERL	Perlative
PL	Plural
PLS	Plural subject
PO/POSS	Possessive
PR/PROG	Progressive
PRED	Predicator
PRES	Present
PRO	Pronoun

Q	Question
REDUP	Reduplicated
REL	Relative
RFL	Reflexive
S	Subject
SG	Singular
SGS	Singular subject
SIDE	Side
SIM	Simultaneous
SS	Same subject
SU	Subject
SUFF	Suffix
TM	Thematic vowel
TOP	Topic/ Topicalizer
TRANS/TR	Transitive
V	Verb
VOC	Vocative
WAY	Away

CHAPTER 1
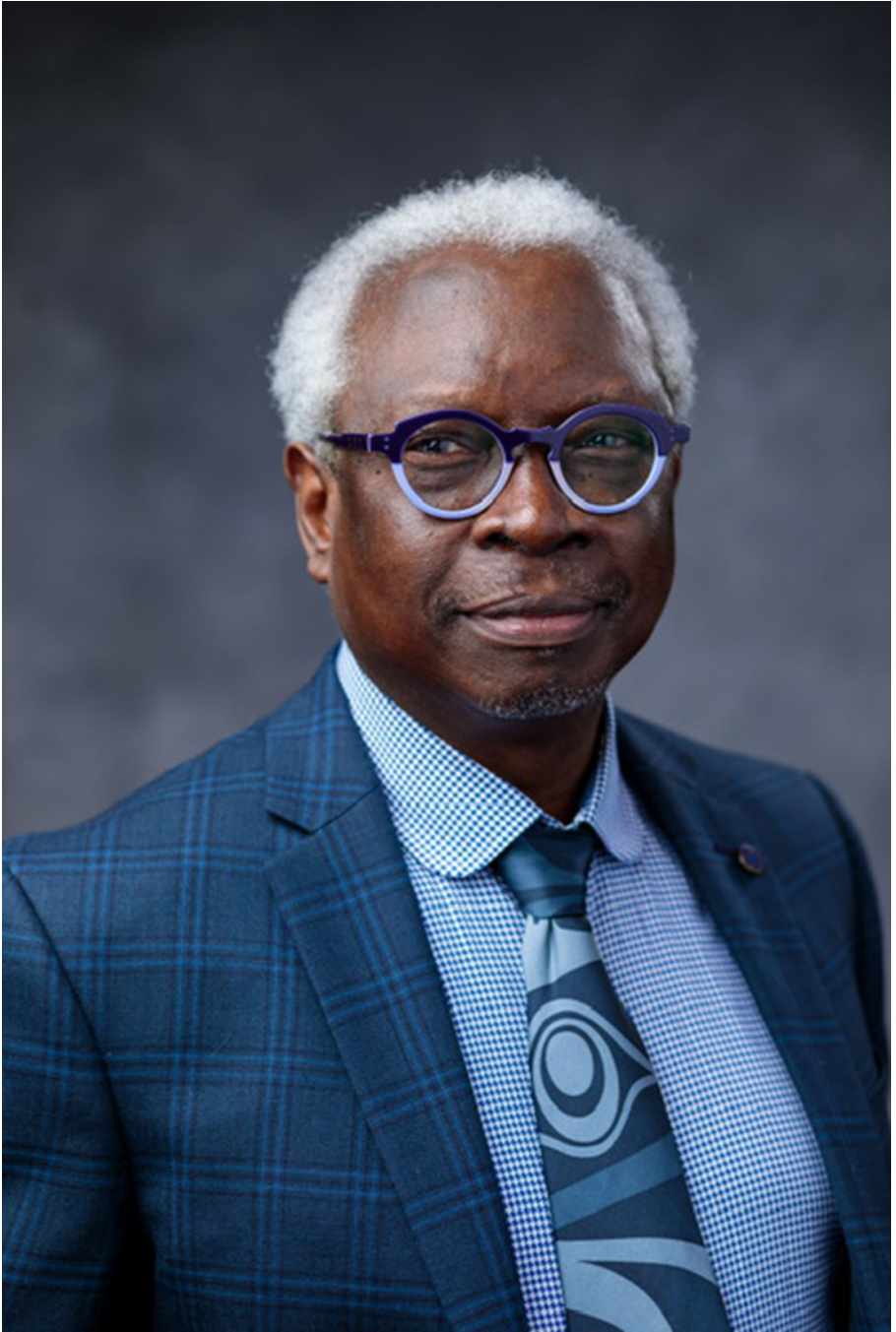
Introduction

Ecology rolls the dice

Enoch O. Aboh & Cécile B. Vigouroux
University of Amsterdam / Simon Fraser University

1.   An ecological approach to Salikoko S. Mufwene's intellectual journey

This volume honours Salikoko S. Mufwene's influential contribution to the study of language, and in particular, to creolistics. The title, *Variation Rolls the Dice*, echoes Mufwene's stance that "ecology rolls the dice" in accounting for the competition-and-selection of linguistic variants. This title is also a way to encapsulate what we may learn from Mufwene's life trajectory, his intellectual endeavour, and how his specific life and academic choices contributed to making him the iconoclast scholar he has become.[1] Indeed, Mufwene's academic path as an undergraduate in his native Zaire (i.e., present-day Democratic Republic of the Congo), a graduate student at the University of Chicago, and a faculty member in different universities in the Americas has been one of competition-and-selection. His experience as a sub-Saharan African multilingual speaker-learner of typologically distinct languages with varying degrees of social prestige has profoundly shaped the questions at the core of his research programme. For instance, how can the study of language and its contexts of use help to uncover underlying properties of speaker/signer-learners' (henceforth S-learners') linguistic repertoires, and their instinct for language? How can we account for S-learners' language capacity based on their ecology? How does a comprehensive formal account inform us about the emergence of new languages (e.g., creoles) or shed light on speakers' choice of (or

---

1.   See more on this in the postscript to this volume, "The restructuring of Salikoko Mufwene through competition and selection: A conversation between Salikoko Mufwene and Michel DeGraff".

shift to) a language in particular ecologies? How do all these questions, examined from the ecological perspective, shed light on the human mind/brain and on how humans evolved such a unique trait: language?

Sali, as he likes to be called by his students, colleagues, and friends, grew up speaking Kiyansi, the language he spoke at home in his village in the Bandundu province of Zaire where he was born, and Kikongo, the dominant language of the region. Like all the pupils attending school in the Belgian exploitation colony of pre-independent Congo, he received his formal education in French in Catholic primary schools and at the seminary.[2] After independence, the educational policy of providing schooling through the medium of French in Zaïre and more generally in former African exploitation colonies stayed in place – a policy which obviously continued to coerce sub-Saharan populations to access formal education through the language of the former colonial power. This language policy has benefited a small elite that (still today) operates transnationally between the African continent and the Western world, whereas the majority of the African populations functions mainly in their vernacular and vehicular languages (see Mufwene 2008).

At the then *Université Nationale du Zaïre* in Lubumbashi, Mufwene learned a "bookish English" (mostly taught by non-native instructors) when he was preparing his *License en Philosophie et Lettres* (BA in Philosophy and Literature) with a major in English Philology. Thanks to a Fulbright scholarship, he pursued a PhD in formal semantics at the University of Chicago under the supervision and mentorship of the late James D. McCawley. He received his linguistics training while adapting his "bookish" L2 English to real-life interactions and absorbing various aspects of the English varieties he was exposed to. At this point, he must have realised that there was a gap between a constructed standard, neatly presented in books taught by L2 teachers to non-native learners thousands of miles away, and what native speakers do with their languages, sometimes in plain "violation" of the neat and unchallenged rules displayed in the prescriptive grammars from which he had learned. He must also have recognised how individual native speakers' language practices and attitudes to their language varied. Most of all, he must have concluded that speakers, whether speaking a socially valorised language such as English or a stigmatised vernacular such as those spoken in his native Congo, are equally subject to the various pressures of their respective ecologies. Indeed,

---

2.   In colonial sub-Saharan Africa, Catholic schools were believed to be the best for training Africans to become "educated" and suited to work in the colonial administration. Sali, like many other children of caring parents, was sent to the seminary because, his parents believed, this would be the best way to secure his future. Well, Sali was born an iconoclast. An anecdote he often tells is that he was expelled from the seminary for asking too many questions!

by socialising with other Congolese students in the Chicago area, he acquired Lingala, a Congolese vehicular language. Peer pressure was an important ecological factor that drove him to learn the language.

Through his language journey Mufwene has experienced first-hand how S-learners' vehicular languages may change during their lifetime under the pressure of different ecological factors, and how these, in turn, may affect their linguistic behaviour in both comprehension and production. He understands the emotional toll that language attrition may take on speakers, especially when they interact with close relatives left behind. He is attuned to how language ideology shapes people's linguistic behaviour and representations: as a child he was subjected to the psychological and physical violence of being taught in the language of the former coloniser at the expense of his native Kiyansi or Kikongo. He knows how speaking a foreign language (i.e., French or English) "with a unique accent" (hard to be localised by American English or French native speakers, as well as by other L2 speakers of the Diaspora) may lead to disenfranchisement and discrimination, independently of one's actual competence in those languages, one's integration into the community, or one's socioeconomic status and cultural capital. He is aware that language indexicalities do not remain the same across contexts. Lingala, commonly associated with the former Mobutu regime in Zaire and the derided way of living in the capital city, Kinshasa, became for him a new index of his "Congoleseness" in the United States.

Mufwene's life experience has partly motivated his commitment to developing a more comprehensive theory of linguistics that focuses on the interactions between the individual S-learners (more precisely their minds) and their ecologies. The importance of Jamaica as an ecological incubator for his subsequent research on creoles cannot be downplayed here. As a newly appointed assistant professor at the University of the West Indies, he embraced a new line of research thanks to the mentorship of colleagues like the late Mervyn C. Alleyne. In this new ecology characterised by diglossia, he witnessed the ways in which S-learners interacted by using a continuum of linguistic variants. He recognised the challenges that creolistics poses to linguistic theory in general. Since then he has dedicated his academic life to unsettling and debunking deep-seated language theories, including those in creolistics.

2.   Mufwene's language ecology approach

Mufwene has often acknowledged the influence of Robert Chaudenson's historical approach on his ecological perspective, especially regarding the importance of socioeconomic structures of the *settlement colonies* in which creoles emerged

(see Mufwene 2001).[3] In Mufwene's ecological approach, one cannot understand S-learners' language practices if one ignores the specific ecologies in which they are inserted and have evolved. Regarding creoles, this approach entails analyzing their linguistic properties in light of the sociohistorical, political, cultural, and economic factors that fostered their emergence. Mufwene's keen interest in understanding these properties and integrating them into a comprehensive theory of language evolution led him to translate into English and edit Chaudenson's original French book, *Des îles, des hommes, des langues* (Chaudenson 1992). This revised version made Chaudenson's work available to a broader readership, as Chaudenson himself acknowledges in his (2001b) article "Focus on creolist: Salikoko S. Mufwene". Furthermore, Mufwene's reinterpretation of Chaudenson's work made it clear that "Creolisation", a term that linguists now use exclusively to characterize the development of a new language from a pidgin (see the so-called Pidgin-to-Creole Cycle, Hall 1962), involves minimally three factors: space (i.e., a settlement colony), time (seventeenth eighteenth century colonial expansion), and people (i.e., linguistic and cultural agents in contact). These three factors explain why creoles (e.g., Jamaican Creole, Haitian Creole) may appear more distant from their British English and French lexifiers, respectively, than, say, American English and Quebec French. Although the cognitive processes that S-learners of these languages are subject to are the same, their ecologies are not. A fundamental aspect of Mufwene's uniformitarian approach to the evolution of language is that the cognitive processes that led to the emergence of so-called creoles are not different from those at work in other language change processes (e.g., from Latin to French). The modern mind is the same across modern human populations. Therefore, the relevant explanations for observable differences are found in S-learners' respective ecologies (Mufwene 2001). According to this line of thought, the "distance" between French and Haitian Creole is not more striking than that between Latin and French. In both cases, the feature pool (i.e., the inputs to which language learners are exposed and which trigger learning) defines the range of variation (see Aboh 2015; Aboh & DeGraff 2016; Mufwene 2001).

With regard to language change and linguistic typological variation in general, Mufwene has also credited Voegelin, Voegelin & Schutz (1967) and Haugen (1971, 1972) for his ecological approach to language evolution. This model, based on population genetics and macroecology, takes as its premise that a biological approach to evolution is also applicable to languages. The evolution of language is

---

3.   At the time of writing this introduction, we learned of the passing of the French creolist Robert Chaudenson, one of Mufwene's friends and intellectual companions over the past thirty years.

comparable to that of species (see also Croft 2000, and this volume). For instance, processes of POLYPLOIDIZATION and HYBRIDIZATION, which are relevant descriptive concepts for understanding variation within and across individuals, can be analogized to how a speaker may select, recombine, and incorporate different linguistic features into her/his idiolect. Breaking away from the Darwinian view that has been dominant since the nineteenth century, Mufwene argues that languages are better analogized to viral species (rather than organisms). According to him, both languages (as spoken by individual S-learners) and viruses owe their existence to their hosts, with the exception that in the case of language, it is the host who develops a mental grammar that translates into specific linguistic behaviour in communicative settings (Mufwene 2001, 2008, 2018). Although this marks a major difference between language and biological species in general, the analogy allows us to focus on a core aspect of linguistic theory: the ways in which individual S-learners affect each other through communicative acts that eventually lead to the emergence of communal norms (i.e., what Chomsky (1986) refers to as E-language).

However, Mufwene reminds us to use the biological analogy with caution, by underscoring that language phenotypes (i.e., features on which typological classifications are formed) are not equivalent to the genotypes of viruses. Unlike the latter, linguistic phenotypes are acquired in a piecemeal fashion through S-learners' interactional activities.[4] The latters' agency (and sometimes informed choices e.g., in the context of migration or relocation to a new neighbourhood) is crucial to the emergence of mental grammars that are put to use in communicative settings.

For Mufwene, therefore, the ecology includes different, though interrelated, factors that influence how a language evolves locally (i.e., in the mind of the individual S-learner, and at the population level). The mind and the anatomy of S-learners are, according to him, the most critical drivers of language evolution. It is a truism to say that without the development of the hominins' mental and anatomical structures, modern human languages (if they existed) would have evolved very differently from what we know today. However, most linguistic frameworks ignore this aspect. In studies of language contact, for instance, most discussion of contact scenarios is framed as if populations of S-learners meet as armies on a battlefield. Although particular events in history, such as the slave trade or other forms of forced migration, may bring large groups of people into contact with each other within a short period of time, and thus coerce the development of various

---

4.   But see Aboh (2015), who adopts a stronger biological view in which idiolects can be considered to represent to some extent the phenotype of linguistic genotypes as recombined within individual speakers/signers.

L2-acquisition strategies, one should not forget that the mind is where language contact primarily occurs. It is where structural information is processed, competition-and-selection between different linguistic features happens, and recombination of linguistic features leading to new variants occurs. The task of the linguist is to uncover these processes and explain how they may become a "system" at both the individual and population levels (see Aboh 2015, 2020; Mufwene 2001, 2008; Mufwene & Vigouroux 2017).

Socioeconomic, cultural, and historical factors are, in this regard, essential features of the ecological model. In his work on French and English-based creoles, Mufwene draws our attention to POPULATION STRUCTURE and its interactions with economic practices. Likewise, he warns us not to undermine the importance of geographic ecologies, or territories, as they influence settlers' decisions to favour one type of economy over another, for example, sugar cane, tobacco, or cotton (and subsidiary plants such as aloe vera) in the case of island plantation societies in the Caribbean and the Indian Ocean. Although ecology drives language evolution in this framework, one should also pay attention to Chaudenson's (2001a) idea of PERIODIZATION. Regarding French-based creoles, it prompts us to examine how the colonists' economic decisions depended largely on the time of colonization, the availability of capital, and the potential world market interests (Mufwene 2017). Accordingly, the weighting of ecological factors varies in time and space (Mufwene 2018) and must be addressed from a global perspective; hence the discussion of issues of globalization in Mufwene (2002a, 2002b, 2008, 2015).

From a linguistic perspective, the kind of economy developed in a particular locality shapes the types of interactions that take place there (who interacts with whom), their spatialization (i.e., where people interact), their periodicity, and external constraints exerted on these factors. Mufwene's comparison between different plantation economies from the seventeenth to the nineteenth century, however, makes it clear that economic practices alone cannot explain the emergence of creoles in plantation settlement colonies, any more than slavery (or indentured labour) alone can. He takes the example of Brazil, which never produced a Portuguese-based creole despite being engaged in a sugar cane plantation economy and having slaves more than a century before the establishment of the French and English Caribbean settlement colonies. He argues that the population structure in Brazil produced a different linguistic outcome because, unlike the situation in other plantation-based colonies, residential segregation of slaves from the European colonists and indentured servants was not enforced there, despite the presence of clearly racially based discrimination. Further interdisciplinary research should shed light on which relevant ecological factor "rolls the dice" in each setting.

Regarding colonization, Mufwene (2001) argues that different types of colonies led to contrasting linguistic outcomes. Trade colonies that flourished in the

nineteenth century produced pidgins; nineteenth century exploitation colonies led to indigenized varieties of the lexifier (or colonisers' language); and creoles emerged in the settlement colonies of the late seventeenth and eighteenth centuries. This historical approach enables him to debunk the popular and persisting assumption in creolistics that pidgins are the ancestors of creoles, and to challenge some recent claims about creole types (based on pidgin ancestry) (Mufwene 2020). These new insights make it clear that the ecological approach Mufwene advocates has an explanatory power that is often missing in monolithic theories of emergence often mentioned in creolistics or textbooks (e.g., Derek Bickerton's (1984) language bioprogram hypothesis, Claire Lefebvre's (1998) relexification hypothesis, or Ingo Plag's (2008a. 2008b, 2009a, 2009b) interlanguage hypothesis).

The SPEED OF POPULATION GROWTH is another important ecological factor to pay attention to, as the colonies of Cape Verde and the Netherlands Antilles illustrate. In discussing these cases, Mufwene (2008: 39) hypothesizes that the "rapid population replacement in a steadily growing overall population" favoured the emergence of creoles in these territories (see also chapters in Muysken and Smith 2015 and references therein for a similar view on Saramaccan).

In addition to these external ecological factors, Mufwene identifies internal factors as well. INTERNAL ECOLOGY, in his terms, refers to the "dynamics of both intra- and inter-idiolectal variation within a communal system" (2008: 53). From an internal ecology point of view, competition-and-selection is a process inherent in the dynamics of language evolution. Every idiolect (i.e., the externalisation of individual S-learners' mental grammars) draws on a FEATURE POOL (Mufwene 2001) to which individual S-learners contribute variably. Unlike what occurs in the biological gene pool to which the feature pool is analogized, the transmission of a linguistic feature always involves modification and recombination that are particular to each S-learner. Likewise, linguistic features may have multiple sources. For instance, a process of competition arises in the feature pool when different idiolects (e.g., A, B, C) that generate the inputs have variants of the same feature, namely, sounds, morphemes, grammatical structures, lexical items, etc. Although the selection process entails making some variants dominant over some viable competitors, the process does not necessarily lead to the exclusion of all competitors that were selected against. Quite the contrary, many variants remain "latent/recessive" in the minds of S-learners and may become active again given appropriate circumstances (see Aboh 2015, 2017a, 2017b, 2019, 2020). According to Mufwene (2018: 82), therefore, linguistic changes also depend on the composition of the feature pool, which triggers alternative learning hypotheses that S-learners entertain. The feature pool thus defines the range of variation indirectly, even though one should not neglect the creativity of individual S-learners, and

their propensity for generating new variants that are distant from the features of the source languages forming the feature pool.

For Mufwene, external and internal ecologies bear equally on language evolution, and therefore analysts should not privilege one at the expense of the other: we need experts in each subfield to arrive at the level of granularity that this approach requires. For linguists, this view implies that they should step outside their sub-discipline, and engage fully in interdisciplinary collaboration with other fields within the humanities and in biology. In his empirical work, Mufwene uses Jamaican Creole, Gullah, French-based Creoles, and African American Vernacular as a test-bed, but his conclusions reach beyond these varieties as they inform us about language evolution in general.

One cannot embrace an ecological approach without paying particular attention to the social context in which this scientific endeavour takes place. Regarding creolistics, Mufwene has strongly objected to analyses of creole structures that were too influenced by their idiomatic translations into Western European languages. For the past four decades, he has continuously drawn our attention to the fact that the linguistic community, by characterizing the evolution of creoles as unnatural, abnormal, or exceptional, has not really emancipated itself from the late-nineteenth century social prejudices toward non-European populations (DeGraff 2003, 2005; Mufwene 2001). He has repeatedly called out the Western world's hegemonic interpretations and descriptions of language data from the Global South (Mufwene 2020). In so doing, Mufwene has been inviting us *to reopen the books*, as he likes to put it.

Thanks to some of Mufwene's colleagues and *compagnons de route*, we are inviting the reader to open this book by reflecting on some of the numerous issues that he has addressed in his work. The different perspectives the authors adopt here reflect Mufwene's call to be inclusive in our approach to language, while paying particular attention to details about individual S-learners and their ecologies.

3.   The chapters of this book

In Mufwene's (2001) and Croft's (2000) ecological approach to language evolution, language change results from competition-and-selection in which some linguistic features or *linguemes* are selected against others, thus leading to new varieties sometimes relatively distant from the source languages. Although it is conceivable that biological factors like individual differences impact upon the competition-and-selection process, the fundamental questions that arise are what external (social, cultural, economic, historical, geographical) factors are at play, and how the interaction between them eventually determines the structure or social type of the emerging language.

Willam Croft's chapter in this book addresses this question on the assumption that social types constrain the types of contact (and therefore learning) situations in which S-learners find themselves. As a consequence, social and contact types constrain the evolutionary trajectory of languages. To understand this, Croft proposes a sociolinguistic typology of language contact that builds upon Trudgill's (2011) framework to which he adds social contact, social structure, and S-learners' attitude. He submits three socially defined language types: *esoteric*, *exoteric*, and *neogenic* languages. Esoteric languages are only used within a specific speech community. Exoteric languages, on the other hand, serve communication between different speech communities. Finally, neogenic languages emerge in the context of new societies with new identities. As Croft concludes, "each of these three types represent continua of language types that arise under different social circumstances and evolve with different types of linguistic structures and different combinations of linguemes from different speech communities that have been or continue to be in contact." In this view, pidgins are just one extreme type of exoteric language, while creoles represent the type of a neogenic language. Exoteric languages and neogenic languages belong to two different continua, and therefore cannot be seen as resulting from a developmental cycle like the commonly assumed Pidgin-to-Creole Cycle (see also Mufwene 2005, 2008, 2020 for a critique of this developmental scenario). Croft's sociolinguistic typology of language contact not only sheds light on the creole debate and the types of contact languages (e.g., mixed languages, lingua francas) but also provides new analytic tools for understanding the sociohistorical and linguistic complexity of language contact and change.

In this regard, a relevant distinction established by Mufwene (2001, 2002a), and already mentioned above, relates to how different forms of colonization (i.e., trade, exploitation, and settlement) have produced particular socioeconomic and cultural ecologies that eventually impacted upon the structure of the languages in contact, and their evolution. In Mufwene's terms (2001: 204), trade colonization (usually the earliest form of contact between populations) consisted of "sporadic contacts [...] restricted to a specific socioeconomic function, like exchange of commodities. Such contacts restricted regular access to the full lexifier and led to the development of pidgins."

Exploitation colonies (e.g., the mode of colonization of sub-Saharan Africa), were characterized by colonizers controlling the colony administratively and economically for the benefit of European governments and companies. Government representatives and company clerks sent to the various colonies only settled there for the term of their mission. Colonization led to two kinds of new language varieties which diverged structurally from their lexifiers: (1) those lexified by European languages (the indigenized varieties), and (2) those lexified by

languages indigenous to Africa. Settlement colonization was intended to provide "new homes" for the European colonists, who would recreate there parts of the Europe they had left behind. As a matter of fact, many colonial settlements in the Americas were named after regions, villages, or cities in Europe, unlike what occurred in African exploitation colonies. Settlement colonies, which led to the emergence of creoles, usually

> started with intimate interactions between the two parties. [Institutionalised] segregation was subsequent to the increase in the sizes of the European populations and the larger proportions of non-Europeans. Multilingualism led the Africans to adopt the languages of the groups in power as their vernaculars. These were restructured during the appropriation process.   (Mufwene 2001: 171)

Gillian Sankoff's chapter on Tok Pisin and the history of language contact in Papua New Guinea is a meticulous case study of how trade colonization, followed by exploitation colonization, gave rise to different socioeconomic and cultural dynamics that favoured the emergence of the so-called South Pacific Jargon, the ancestor of Tok Pisin. Her contribution focuses on the Buang people in the Snake River Valley of Morobe Province. As Sankoff explains, she "appl[ies] a close-up lens to study how changes in the linguistic landscape brought about by colonization have been assimilated to the local cultural construction of the relationship between language and society."

Sankoff enlists Mufwene's insights to shed light on the history of Tok Pisin. She discusses how South Pacific Jargon resulted from trade in the early part of the nineteenth century, whereas the transition from a trade to an exploitation colony by the late nineteenth century led to the spread of pidgins, then to the emergence of urban vernaculars and nativized pidgins. While painstakingly documenting the history of these transitions, she also examines the intricate roles of socioeconomic and political factors, national identity dynamics, population movements and S-learners' linguistic capital in the spread of language-contact varieties like Tok Pisin. She highlights the fact that such linguistic expansion sometimes occurs at the expense of local indigenous languages – though Tok Pisin itself would eventually be considered "indigenized," as Mufwene has continuously argued in his work. Sankoff's chapter is a superb case study of such "indigenization" processes in parts of Papua New Guinea where she has done extensive fieldwork. Her chapter provides robust empirical support for an ecological approach as developed in Mufwene (2001), while pointing to the necessity of a typology of contact, as Croft advocates in his contribution to this volume.

In Croft's account, pidgins are distinct from mixed languages: the latter (but not the former) are a form of neogeny in which multilingual S-learners (of presumably equal social status) engage in communicative practices, including

code-switching/mixing, which may lead to new speech patterns. The latter may subsequently conventionalize to become a speech form shared by an emerging community with a new identity. It is reasonable to assume that similar communicative patterns contributed to the spread and nativization of Tok Pisin.

Carmel O'Shannessy's chapter on Light Warlpiri, a mixed language spoken in Northern Australia, is another case in point. Building on theories of code-switching/mixing (e.g., Meyers-Scotton 2003), she argues that insertional code-switching (Muysken 2000), whereby a verb phrase from Kriol (an English-based creole) is inserted into a Warlpiri's frame, accounts for the "mixed" structural properties of Light Warlpiri. In a way somehow indicative of the Saussurean systemic approach to language, this pattern triggers the emergence of other creative patterns within the verbal domain and at the nominal and clausal levels. These innovations exhibit aspects of all the source languages, including Warlpiri, Kriol, and English varieties that speakers are exposed to. Though the bulk of its syntax is modelled on Warlpiri, Light Warlpiri displays new grammatical properties not found in its source languages. As O'Shannessy puts it, "the result is a way of speaking that is mixed at every level but draws on Warlpiri structure more than on any of the other sources." Light Warlpiri can therefore be classified as a "mixed language" somehow different from the popular cases discussed in the literature, such as Michif or Media Lengua.

A fundamental aspect of Mufwene's ecological approach to the evolution of language is (individual) variation, and how it plays out in the competition-and-selection process that may ultimately lead to language change at the population level. Such a change can be observed, for instance, in intergenerational differences, with children using more systematically and frequently grammatical patterns than adults (see also DeGraff 2001a, 2001b, 2002, 2003, 2005).

Vivien Dunn, **Felicity Meakins**, and **Cassandra Algy** investigate such an ongoing change in small-scale spatial descriptions of Gurindji Kriol-speaking children. Like Light Warlpiri, Gurindji Kriol is a mixed language which combines the nominal syntax of Gurindji (a Pama-Nyungan language), with the verbal syntax of Kriol (an English-based creole). However, at the clausal level, both Gurindji and Kriol provide for the lexical elements that are computed in the nominal or verbal domains (see Meakins 2013). The two languages differ concerning spatial expressions: speakers of Gurindji generally prefer cardinal direction terms in both small-scale and large-scale space, as opposed to speakers of Gurindji Kriol who avoid this strategy in descriptions of small-scale space relations. Due to contact between these different languages, traditional Gurindji is only spoken by older generations, while younger adults and their children mainly speak Gurindji Kriol. The question that arises is whether this asymmetry may lead to a new system in subsequent generations' small-scale spatial descriptions in Gurindji. Based on the "Man and Tree" task given to children and adolescents, aged between nine and

seventeen, who have acquired the Gurindji cardinal system, the authors contend that this age-group uses a different strategy from their parents. Whereas the latter mainly rely on the cardinal system, the younger generation uses both cardinal- and landmark-based strategies to different degrees. As the authors conclude, however, "the landmark-based strategies are used in a quasi-absolute manner such that the spatial relations system remains conceptually geocentric." This instance of language change (or shift) is in line with Mufwene's ecological approach, in which the emerging system is never isomorphic with the source languages, as it presents new properties that embed aspects of the relevant donor languages.

Another relevant aspect of Mufwene's ecological approach is how the process of competition-and-selection unfolds and (partly) determines essential aspects of individual mental grammars. Their cumulative effect may lead to group linguistic behaviour which, in turn, fosters a change at the population level, i.e., language change. In the words of Mufwene (2001: 162):

> In the competition-and-selection approach […] the language that prevails actually wins a pyrrhic victory, as it adapts itself to its new speakers and contexts of communication, i.e., to part of its changing ecology. This validates again approaching languages as parasitic species and seeing their evolutions in terms of how they adapt to the responses of their new hosts while affecting, or eliminating, other linguistic species that they come in contact with […] The strong version of my approach to language evolution is that the competition-and-selection process has been typical of language change in any community and at any time.

When applied to the context of creole languages, with their alleged special structural type compared to non-creole languages, this approach makes it clear that classical Stammbaum trees used in genetic classifications are partial and imperfect. They are partial because they usually ignore the host, as well as the relevant contact settings concomitant with language change. They are also imperfect, because they treat change from an internal language perspective only. According to Mufwene, creole languages are no more illegitimate descendants of all their "donor" languages than Modern Romance languages (e.g., French, Portuguese, Spanish) would be of Latin. In French-based creoles and Romance families, the prevailing language wins a pyrrhic victory because it is a recombined variant of its donors (Aboh 2015; Mufwene 2002c). Because change is contingent upon recombination, there may be no structural distinction between so-called creoles and non-creole languages. Both come into being through the same process, and the prevailing language bears aspects of its source languages. Therefore, understanding language change requires investigating the genesis of the process in the source languages in contact.

Pieter Muysken[†], like Mufwene, has always stressed that there are no structural traits that set creole languages apart as a type. Creoles, however, can inform us of

particular cognitive processes at play in contact situations. His primary interest in studying substrate influence is to unravel the principles underlying language contact, and show how linguistic phenomena ensuing from such contact can inform us about crucial grammatical aspects of the prevailing language. In his chapter on the formation of Northern Quechua languages, Muysken carefully examines the possibility of substrate influence by taking into account complex socioeconomic, political, and cultural factors that may have led to different situations of language contact or language shift. Because languages are dynamic systems in a state of constant flux, any endeavour to identify substrate influence proves to be very difficult. Muysken provides various guiding principles fully compatible with Mufwene's ecological approach:

1. Avoidance of the "Cafeteria Principle": The sociolinguistic history of the variety in question constrains the choices of source languages for substrate influence;
2. "Mutual Reinforcement": Patterns that exist in more than one substrate are more likely to be adopted by the emerging contact language (see Aboh & Ansaldo 2007);
3. "Conservative Substrate Influence": Properties in the emerging contact language become entrenched if they match properties in (a subset of) the substrate languages;
4. "Multiple Causation": Substrate influence can always converge with some other explanatory principle, with a mutually reinforcing effect between these strategies.

Based on these principles, and keeping in mind the role of acquisition (whether L1 or L2), Muysken surveys a rich array of facts in his comparison of Quechua languages, including fascinating cases of structural changes in Northern Quechua. Among such changes, he analyzes those affecting: case marking; person and number marking on pronouns and verbs; copula constructions; negation markers; markers for evidential, desiderative, intentional, hortative, purposive and potential; diminutive or pejorative suffixes; inalienable kinship; and phonological features, among others. Toward the end of his chapter, he provides a useful table summarizing all possible instances of substrate influence he has analyzed. Muysken's chapter clearly highlights that understanding how languages can influence each other, such as in the case of substrate influence, requires fine-grained analyses of the new varieties and how they compare to their source languages. Such a comparative approach calls for not only a well-informed understanding of the sociohistorical context in which the varieties emerge, but also a detailed knowledge of their grammars.

In creolistics, there have been many studies on various aspects of creole languages to determine which grammatical aspects of their source languages (the so-called substrate languages) contributed to their development (see Muysken and Smith 2015). The aspects of creole languages that figure prominently in such studies involve complementation (Aboh 2006; Winford 1985), serialization (Veenstra 1996), TMA-sequencing (Muysken 1981), noun phrases (Baptista & Guéron 2007, and predication (DeGraff 1992; Winford 1993).

Surprisingly, coordination/conjunction is one of the grammatical phenomena that hasn't received much attention in the field. Although there are some broad characterizations in APiCS (see Michaelis et al. 2013) and some descriptions of coordination/conjunction in individual creoles, not much is known about the syntactic and semantic properties of different coordination types within related creoles, and how these creoles compare to their source languages.

Bettina Migge's contribution aims to fill this gap. Her chapter adopts a three-way comparison between the Eastern Maroon Creoles or Nenge(e) and Western Maroon creoles (Matawai) of Suriname, and their main source languages, namely English and the Gbe languages. Regarding NP coordination, the Suriname Creoles display a relatively common pattern across creole languages. They use a unique form for expressing comitative, instrumental, and NP coordination. However, things are different with VP coordination. Although Haspelmath et al. (2013: 284–287) suggest that creoles do not generally distinguish between NP and VP coordination, the Suriname data show otherwise. Worth noting is how they also involve several VP conjunctions derived from various sources. Migge's chapter sheds light on the undocumented fact in creoles that the different VP coordinators occur in different temporal contexts. Accordingly, there is an interaction between tense/aspect specifications and VP coordination in these creoles. In addition, VP coordinators in these creole languages can be stacked to encode various meanings, including emphasis. Migge's descriptive analysis, based on a very rich data set, shows that the Suriname Creoles involve different coordinator-types, the usage of which suggests different developmental paths. Her chapter provides the first empirical ground for investigating how competition-and-selection operated in the emergence of various coordinate structures in the Suriname Creoles.

Marlyse Baptista's contribution on variation, competition, and change is a direct application of Mufwene's competition-and-selection framework. Based on the Swadesh list (Swadesh 1952), her study seeks to understand variation across five dialects of Cabo Verdean Creole, spoken on different islands in the archipelago, namely, Santiago, Fogo, Brava, Santo Antão, and São Vicente.

In addition to providing evidence in support of Mufwene's framework, Baptista sheds new light on lexical variation in Cabo Verdean varieties. She helps us conceptualize the nuts and bolts of Mufwene's framework, by comparing it

with Darwin's biological evolution theory. She focuses mainly on the impact of idiolect-level variations in Mufwene's competition-and-selection framework, on the one hand, and on the long-term effects of individual-level "infinitesimally small inherited modifications" on natural selection in Darwin's framework, on the other. Baptista's chapter raises two issues that require further investigation under the competition-and-selection model: (1) the granularity of the features subject to competition-and-selection; and (2) the role of semantic nuances in determining the "competing weight" of these features. These questions directly relate to the context of learning, which includes both the inputs the learner is exposed to and the learning biases that s/he is subject to. Baptista's contribution prompts us to address the role of L1 and L2 learners in generating the variation observed across the five Cabo Verdean varieties.

Daniel Véronique's chapter examines the role of naturalistic L2 acquisition by late learners in the development of French-Related Creoles (FRC). This approach may appear counter-intuitive, given the fact that creoles were not the creations of adults only, but emerged from interactions between L2 and L1 acquisition, that is, DeGraff's (2002) L2–L1 cascade (see also Aboh and DeGraff 2014). To understand this cascade, however, one needs to identify which L2 acquisition patterns could have populated the inputs that subsequent L1 acquirers of the creoles could be exposed to (see also Dunn, Meakins and Algy this volume; Muysken this volume). This is the focus of Véronique's chapter, in which he further calls for a more constructive dialogue between the fields of Second Language Acquisition (SLA) and creolistics. Véronique's methodology involves comparing linguistic patterns observed during the acquisition of French as a Second Language (FSL) to similar patterns in FRC. Unlike Plag's (2008a, b, 2009a, 2009b, 2011) interlanguage hypothesis, Véronique seeks to identify which most relevant linguistic features or patterns – arguably the results of acquisition processes – could have generated the inputs that subsequent generations of learners were exposed to. In terms of a uniformitarian approach to learners, Véronique argues that some processes governing the emergence of a mental grammar in the mind of the creole S-learner are the same as those relating to the L2 S-learner. Building on previous work on SLA varieties (e.g., Klein & Perdue 1997; Perdue 1996), Véronique compares aspects of FRC to FSL by examining, for instance, uninflected verb forms, the placement of negation vis-à-vis light verbs, presentationals, and lexical verbs. Based on the striking similarities between aspects of FRC and those of FSL, the author concludes that the restructuring of grammar in creole languages must have followed the same cognitive processes as in SLA. This chapter does not adhere to all aspects of Mufwene's ecological approach. Yet, its results are compatible with this approach in showing that (1) one need not evoke special cognitive processes other than common learning strategies adopted by S-learners to explain the structure of

creole languages; and (2) the speaking-learning ecology can constrain these strate-
gies. These constraints result from language practices and community dynamics
that typically inform us of the socioeconomic, cultural, and political forces at play
within a particular speech community.

Lisa Lim and **Umberto Ansaldo** examine the relevant ecological factors to
explain Peranakans' language practices and identity dynamics in local, transna-
tional, and digital ecologies. They focus particularly on the role of these Malay-
speaking descendants of South Chinese immigrants and Malay-speaking women
in the development of Baba Malay, and, subsequently, in the spread and indi-
genization of English in the region. Thanks to their multilingual repertoire, which
included English, the Peranakans played an important role as power brokers dur-
ing the European colonization of the Straits of Malacca in the nineteenth and
twentieth centuries. As was also the case in other trade and settlement colonies,
these colonial intermediaries managed to convert their linguistic resources into
economic capital. Further inquiry should tell us the extent to which their eco-
nomic success shaped other communities' language ideologies regarding English
in the region.

Peranakans' language and identity dynamics challenge some of the received
ideas about the link between language and identity. Indeed, the Peranakans
embrace their multilingualism (English, Malay, and Chinese) rather than their
vernacular, Baba Malay, for defining their ethnic belonging. Accordingly, they
don't seem to experience the loss of their vernacular, and subsequently, their shift
to English, as a threat to their multifaceted cultural identity. We may wonder
whether their privileged socioeconomic position during the colonial period plays
any role in their strong sense of cultural identity, built independently of their eth-
nic vernacular. Is the fact that their shift to the British colonizers' language was not
coerced through symbolic or physical violence, as was the case in other settlement
colonies, part of the explanation?

What is striking about the Peranakans is less that they no longer speak their
vernacular, but that they don't discursively construct the non-transmission of the
language as an issue. However, Baba Malay is not completely absent from the pub-
lic arena: used in cultural performances such as rap and hip hop, it is resemiotized
as a language of modernity. We can see in this new usage of the language not just
a way of keeping the cultural heritage alive, as Lim and Ansaldo argue, but also
an act of creating new associations, a new cultural history of Baba Malay, for the
younger generation. Also worth noting is how the presence of the Peranakans in
modern Singapore's urban- and mediascape has become more conspicuous, with
their cultural heritage being used as part of the branding of the city-state. The
traditional image they help to construct offers an appealing contrast to the "global
tech hub" aspects of the city.

The Peranakans' case corroborates at least two points made by Mufwene in several of his publications: (1) ethnic languages are not always a defining feature of cultural identity claims; (2) colonization as a factor in language shift and identity dynamics should be evoked in nuanced ways. Besides paying close attention to the type of colonization involved, as already mentioned, linguists should also refine the over-simplistic dichotomy between colonizers and colonized, by examining the population structure of any given territory. Lim and Ansaldo's chapter is in line with the World Englishes scholarship that advocates for the linguistic legitimacy of localized English varieties, in this case Peranakan English. However, we wonder if social stratification based on English variety also exists within the Peranakan community, and whether such stratification correlates with varieties internal to this community.

If ecology and the variation therein roll the dice, the next step for linguists would be to model these factors in order to determine their specific roles in the emergence of a new variety. One possible way to approach this question is to examine loan words as a window into contact dynamics, and investigate how S-learners integrate them into their repertoires.

Liqin Zhang, **Franz Manni**, **Ray Fabri**, and **John Nerbonne**'s chapter takes on this task by applying a computational analysis to loan words in two unrelated language groups: Turkic and Indo-Iranian. Their working hypothesis is that if two words from unrelated languages are similar in meaning and pronunciation, then one of the languages probably borrowed from the other, or even from a third one. Loan words can, therefore, inform us about aspects of the history of a community. Liqin et al.'s analysis rests on Mennecier et al.'s (2016) survey of Central Asian languages. Using the two hundred-word Swadesh list, they collected acoustic and lexical data of eighty-eight informants with a different L1 from the two language families: Kazakh, Kyrgyz, Karakalpak, Uzbek (Turkic), Tajik, and Yaghnobi (Indo-Iranian). They compared three different algorithms used to assess pronunciation similarity: PMI (Frequency)-Based Edit Distance, Spectrogram-Based Edit Distance, and Sound Class Algorithm (SCA). Although all these algorithms identify a large number of loan words, SCA appears to detect loan words most effectively with a superior recall, i.e. "the fraction of the humanly recognized loans that the process detects". Liqin et al.'s chapter offers promising avenues to contact linguists, by providing quantitative tools that will help direct their analytical gaze towards the relevant ecological factors that explain language contact in a given context.

Another aspect that still requires a principled account within the ecological approach is what makes a selected feature spread against others within a speech community: how does the competition-and-selection proceed? Various factors come into play here. As discussed in Aboh (2015), there are purely linguistic factors (like the features related to the interfaces) and also learning factors. Areas of

grammars that are hard to acquire may be more prone to allowing restructuring patterns which may spread across the community.

Nour Efrat-Kowalsky's chapter on grammatical gender in Dutch can be read as a proof-of-concept of this hypothesis. Gender assignment in Dutch has been known to be notoriously difficult for various learner profiles, including L1 learners, 2L1 learners (i.e., bilinguals), and early and late L2 learners. All these learner profiles exhibit overgeneralization of the common gender to some extent. Efrat-Kowalsky's starting point is to investigate how such a non-standard feature could spread within a community.

Through an analysis of tweets, the author shows how language users tend to overgeneralize the common definite determiner *de* with neuter nouns, a feature usually associated with immigrant Dutch varieties in the scholarly literature and in Dutch society. Her analysis concludes that even monolingual L1 speakers use *de* with neuter nouns, although to a lesser extent than bilinguals and L2 learners. The methodology she adopts enables linguists not only to track the life-cycle of a linguistic feature, but also to gather language-in-use data quickly. It also provides access to language users' (mediated) social networks, thanks to their self-declared identification and that of their interactants.

Social Network Analysis has long been a part of the toolkit of linguists studying language variation and change (see for instance Milroy 1987). Although analysing tweets makes language users' communicative behaviour easily measurable, we shouldn't forget, that Twitter is just one of the multiple social networks of an individual, as seemed to be the case of the tweeters examined by Efrat-Kowalsky. Indeed, S-learners belong to different social networks, each of which involves different linguistic practices and therefore competences. The ecologies of S-learners are therefore multiple. Accordingly, Granovetter's (1973) distinction between weak and strong ties is still relevant in the age of increasing online communication. Further analysis should help us to understand whether on- and off-line interactions exert similar sociolinguistic pressure on language users' performance, or complement each other.

The type of analysis offered by Efrat-Kowalsky enables us to formulate new hypotheses about the potential factors that contribute to language variation, and subsequent language change. More importantly, this study highlights the importance of the individual speaker in language variation, as opposed to the status-based social groups emphasized in earlier sociolinguistics work. One of the chapter's main contributions is to hypothesize that *learnability* (i.e., what can be learned or not) is an important ecological parameter that accounts for the spread of linguistic features among a diverse range of speakers. Corroborating Mufwene (1996), Efrat-Kowalsky argues that the less costly or complex it is to acquire, the more likely a feature is to be adopted by individual S-learners and therefore spread

within a community. This hypothesis is also reminiscent of Martinet's (1949, 1955) principle of economy in language, and Zipf's (1949) "principle of least effort". According to both authors, linguistic behaviour is shaped in part by the minimization of effort. Efrat-Kowalsky's chapter is also part of a growing scholarly literature that highlights how online communication provides invaluable linguistic data that can help us better understand how language patterns emerge.

Finally, the book closes with an updated version of a conversation between **Salikoko S. Mufwene** and **Michel DeGraff** (previously published in *Carrier Pidgin* 29, 2001), in which Sali reflects on the intertwinement of his personal and intellectual trajectories, and the genesis of some of his concepts and positions. Sali had no idea that the updated version would be published in this volume, which honours his work and his legacy in creole studies and theoretical linguistics.

References

Aboh, E. O. 2006. Complementation in Saramaccan and Gungbe: The case of C-type modal particles. *Natural Language and Linguistic Theory* 24: 1–55. https://doi.org/10.1007/s11049-005-1815-4

Aboh, E. O. 2015. *The Emergence of Hybrid Grammars: Language Contact and Change*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139024167

Aboh, E. O. 2017a. The emergence of hybrid grammars: A rejoinder to Peter Bakker. *WORD* 63: 207–222. https://doi.org/10.1080/00437956.2017.1347316

Aboh, E. O. 2017b. Population factors, multilingualism and the emergence of grammar. In *Language Contact in Africa and the African Diaspora in the Americas: In Honor of John V. Singler* [Creole Language Library 53], C. Cutler, Z. Vrzić & P. Angermeyer (eds), 23–48. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.53.02abo

Aboh, E. O. 2019. Our creolized tongues. In *Language Contact, Continuity and Change in the Genesis of Modern Hebrew* [Linguistik Aktuell/Linguistics Today 256], E. Doron, M. Rappaport Hovav, Y. Reshef & M. Taube (eds), 287–320. Amsterdam: John Benjamins. https://doi.org/10.1075/la.256.11abo

Aboh, E. O. 2020. Lessons from neuro-(a)-typical brains: Universal multilingualism, code-mixing, recombination, and executive functions. *Frontiers in Psychology* 11: 488. https://doi.org/10.3389/fpsyg.2020.00488

Aboh, E. O. & Ansaldo, U. 2007. The role of typology in language creation: A descriptive take. In *Deconstructing Creole* [Typological Studies in Language 73], U. Ansaldo, S. M. Mathew & L. Lim (eds), 39–66. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.73.05abo

Aboh, E. O. & DeGraff, M. 2014. Some notes on bare noun phrases in Haitian Creole and in Gungbe: A transatlantic Sprachbund perspective. In *The Sociolinguistics of Grammar* [Studies in Language Companion Series 154], Tor A. Åfarli & Brit Mæhlum (eds), 203–236. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.154.11abo

Aboh, E. O. & DeGraff, M. 2016. A null theory of Creole formation based on universal grammar. In *The Oxford Handbook of Universal Grammar*, Ian Roberts (ed.). Oxford: OUP. https://doi.org/10.1093/oxfordhb/9780199573776.013.18

Baptista, M. & Guéron, J. (eds). 2007. *Noun Phrases in Creole Languages: A Multi-Faceted Approach* [Creole Language Library 31]. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.31

Bickerton, D. 1984. The language bioprogram hypothesis. *Behavioral and Brain Sciences* 7: 173–221. https://doi.org/10.1017/S0140525X00044149

Chaudenson, R. 1992. *Des îles, des hommes, des langues*. Paris: L'Harmattan.

Chaudenson, R. 2001a. *Creolization of Language and Culture*, translated and edited by S. S. Mufwene. London: Routledge.

Chaudenson, R. 2001b. Focus on creolist: Salikoko S. Mufwene. *The Carrier Pidgin* 29. <https://mufwene.uchicago.edu/interviewWithRobert.html> (30 August 2020).

Chomsky, N. 1986. *Knowledge of Language: Its Nature, Origin and Use*. Westport CT: Praeger.

Croft, W. 2000. *Explaining Language Change: An Evolutionary Approach*. London: Longman.

DeGraff, M. 1992. Creole Grammars and Acquisition of Syntax: The Case of Haitian Creole. PhD dissertation, University of Pennsylvania.

DeGraff, M. 2001a. On the origin of Creoles: A cartesian critique of neo-Darwinian linguistics. *Linguistic Typology* 5: 213–310.

DeGraff, M. 2001b. Morphology in Creole genesis. In *Ken Hale: A Life in Language*, Michael Kenstowicz (ed.), 53–121. Cambridge MA: The MIT Press.

DeGraff, M. 2002. Relexification: A reevaluation. *Anthropological Linguistics* 44: 321–414.

DeGraff, M. 2003. Against Creole exceptionalism. *Language* 79: 391–410. https://doi.org/10.1353/lan.2003.0114

DeGraff, M. 2005. Linguists' most dangerous myth: The fallacy of Creole exceptionalism. *Language in Society* 34: 533–591. https://doi.org/10.1017/S0047404505050207

Granovetter, M. S. 1973. The strength of weak ties. *American Journal of Sociology* 78: 1360–1380. https://doi.org/10.1086/225469

Hall, R. A. 1962. The life cycle of pidgin languages. *Lingua* 11: 151–156. https://doi.org/10.1016/0024-3841(62)90021-9

Haspelmath, M. & APiCs Consortium. 2013. Nominal and verbal conjunction. In *The Atlas of Pidgin and Creole Language Structures*, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 284–287. Oxford: OUP.

Haugen, E. 1971. The ecology of language. *The Linguistic Reporter*, Supplement 25: 19–26. Reprinted in Haugen, E. 1972. *The Ecology of Language*, A. Dil (ed.), 324–339. Stanford CA: Stanford University Press.

Haugen, E. 1972. *The Ecology of Language*, A. Dil (ed.). Stanford CA: Stanford University Press.

Klein, W. & Perdue, C. 1997. The Basic Variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13: 301–347. https://doi.org/10.1191/026765897666879396

Lefebvre, C. 1998. *Creole Genesis and the Acquisition of Grammar: The Case of Haitian Creole*. Cambridge: CUP.

Martinet, A. 1949. La double articulation linguistique? *Travaux du Cercle Linguistique de Copenhague* 5: 30–7. https://doi.org/10.1080/01050206.1949.10416289

Martinet, A. 1955. *Économie des Changements Phonétiques*. Berne: Francke.

Meakins, F. 2013. Gurindji Kriol. In *The Survey of Pidgin and Creole Languages*, Vol. 3, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 131–139. Oxford: OUP.

Mennecier, P., Nerbonne, J., Heyer, E. & Manni, F. 2016. A Central Asian language survey: Collecting data, measuring relatedness and detecting loans. *Language Dynamics and Change* 6: 57–98. https://doi.org/10.1163/22105832-00601015

Meyers-Scotton, C. 2003. What lies beneath: Split (mixed) languages as contact phenomena. In *The Mixed Language Debate*, P. Bakker & Y. Matras (eds), 73–106. Berlin: Mouton de Gruyter.

Michaelis, S., Maurer, P., Haspelmath, M. & Huber, M. (eds). 2013. *The Survey of Pidgin and Creole Languages*. Oxford: OUP.

Milroy, L. 1987. *Language and Social Networks*. Oxford: Basil Blackwell.

Mufwene, S. S. 1996. The founder principle in creole genesis. *Diachronica* 13: 83–134. https://doi.org/10.1075/dia.13.1.05muf

Mufwene, S. S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511612862

Mufwene, S. S. 2002a. Colonisation, globalisation, and the future of languages in the twenty-first century. *Moss Journal on Multicultural Societies* 4(2): 1–47.

Mufwene, S. S. 2002b. Colonization, globalization and the plight of "weak" languages. *Journal of Linguistics* 38(2): 375–395. https://doi.org/10.1017/S0022226702001391

Mufwene, S. S. 2002c. Competition and selection in language evolution. *Selection* 3(1): 45–56. https://doi.org/10.1556/Select.3.2002.1.5

Mufwene, S. S. 2005. *Créoles, écologie sociale, évolution linguistique*. Paris: L'Harmattan.

Mufwene, S. S. 2008. *Language Evolution: Contact, Competition and Change*. London: Continuum.

Mufwene, S. S. 2015. Colonization, indigenization, and the differential evolution of English: Some ecological perspectives. *World Englishes* 2015: 6–21. https://doi.org/10.1111/weng.12129

Mufwene, S. S. 2017. Language vitality: The weak theoretical underpinnings of what can be an exciting research area. *Language* 93(4): e202–e223. https://doi.org/10.1353/lan.2017.0065

Mufwene, S. S. 2018. Language evolution from an ecological perspective. In *The Routledge Handbook of Ecolinguistics*, A. F. Fill & H. Penz (eds), 73–88. New York NY: Routledge.

Mufwene, S. S. 2020. Creoles and pidgins: Why the latter are not the ancestors of the former. In *The Routledge Handbook of Language Contact*, E. Adamou & Y. Matras (eds), 300–324. London: Routledge. https://doi.org/10.4324/9781351109154-20

Mufwene, S. S. & Vigouroux, C. B. 2017. Individuals, populations, and time space: Perspective on the ecology of language revisited. *Language Ecology* 1(1): 75–102. https://doi.org/10.1075/le.1.1.05muf

Muysken, P. C. 1981. Creole tense/mood/aspect system: The unmarked case? In *Generative Studies on Creole Languages*, P. Muysken (ed.), 181–199. Dordrecht: Foris. https://doi.org/10.1515/9783111392844

Muysken, P. C. 2000. *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: CUP.

Muysken, P. C. & Smith, N. (eds). 2015. *Surviving the Middle Passage: The West Africa-Surinam Sprachbund*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110343977

Perdue, C. 1996. Pre-basic varieties: The first stages of second language acquisition. In *Eurosla 6. A Selection of Papers*, E. Kellerman, B. Weltens & T. Bongaerts (eds). *Toegepaste Taalwetenschap in Artikelen* 55: 135-149.

Plag, I. 2008a. Creoles as interlanguages: Inflectional morphology. *Journal of Pidgin and Creole Languages* 23: 109–130. https://doi.org/10.1075/jpcl.23.1.06pla

Plag, I. 2008b. Creoles as interlanguages: Syntactic structures. *Journal of Pidgin and Creole Languages* 23: 307–328. https://doi.org/10.1075/jpcl.23.2.06pla

Plag, I. 2009a. Creoles as interlanguages: Phonology. *Journal of Pidgin and Creole Languages* 24: 121–140. https://doi.org/10.1075/jpcl.24.1.06pla

Plag, I. 2009b. Creoles as interlanguages: Word-formation. *Journal of Pidgin and Creole Languages* 24: 339–362. https://doi.org/10.1075/jpcl.24.2.05pla

Plag, I. 2011. Creolization and admixture typology, feature pools, and second language acquisition. *Journal of Pidgin and Creole Languages* 26: 89–110.
https://doi.org/10.1075/jpcl.26.1.04pla

Swadesh, M. 1952. Lexico-statistical dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96(4): 452–463.

Trudgill, P. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: OUP.

Veenstra, T. 1996. *Serial Verbs in Saramaccan: Predication and Creole Genesis* [HIL Dissertations 1]. The Hague: Holland Academic Graphics.

Voegelin, C. F., Voegelin, F. M. & Schutz, Jr, N. W. 1967. The language situation in Arizona as part of the Southwest culture area. In *Studies in Southwestern Ethnolinguistics: Meaning and History in the Languages of the American Southwest*, D. Hymes & W. E. Bittle (eds), 403–451. The Hague: Mouton.

Winford, D. 1985. The syntax of Fi complements in Caribbean English Creole. *Language* 61: 588–624. https://doi.org/10.2307/414387

Winford, D. 1993. *Predication in Caribbean English Creoles* [Creole Language Library 10]. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.10

Zipf, G. K. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge MA: Addison-Wesley Press.

CHAPTER 2

A sociolinguistic typology for languages in contact

William Croft

University of New Mexico

Different types of languages evolve in situations of human social contact, depending on the nature of the contact and the attitudes on the part of the speakers towards the societies in contact. Three socially-defined language types are useful for classifying contact languages. One may distinguish between an esoteric language, used for communication within a speech community, and an exoteric language used for communication between different speech communities. The third language type, a neogenic language, results when speech communities merge (e.g., in colonisation). Each language type involves a continuum depending on the degree and asymmetry of the social contact between speech communities. This chapter describes these continua, the "contact" or "mixed" languages that they account for, and their structural linguistic correlates.

Keywords: contact languages, esoteric language, exoteric language, neogenic language, pidgin, creole, mixed language

1. Introduction

Salikoko Mufwene has never been a stranger to controversy. He has argued against the mainstream view that creoles evolved from pidgins through a process of "nativization" (Mufwene 2001), which in turn led him to oppose the received wisdom that pidgins and creoles evolved as languages in a manner qualitatively different from the evolution of languages through "normal transmission" (Mufwene 2005, 2008), which in turn led him to challenge current views about language endangerment, loss and revitalization (Mufwene 2017a, 2017b, and the commentaries in the same electronic issue of *Language*). The overarching intellectual themes of Mufwene's controversial theories, or rather theory, of language ecology, contact and change are his evolutionary approach to language; his focus on language as

practice between individuals in societies; and last but not least, a deep belief that all individuals are fundamentally the same in their linguistic response to their social ecology.

In this chapter, I touch primarily on the first of these themes, namely an evolutionary linguistic perspective on the status of different types of contact languages and their relation to each other and their social ecologies. I will suggest here a sociolinguistic typology that can be thought of as elaborating some of Mufwene's ideas on this topic, which I generally find persuasive (see Croft 2000, 2010). My aim is simply to provide a framework for classifying languages in a way that can serve as an initial step in analyzing the complexity of the social ecology of language that Mufwene has repeatedly demonstrated.

The term "sociolinguistic typology" has been used before, of course, most prominently in Trudgill's monograph of the same name (Trudgill 2011). Like other typological classifications, Trudgill's sociolinguistic typology is intended to allow one to account for patterns of variation in language structure. Trudgill is specifically interested in variation in linguistic complexity or simplicity. He proposes five social determinants of this variation:

1. degree of isolation from other societies (higher isolation ~ greater complexity)
2. degree of social stability (greater stability ~ greater complexity)
3. community size (greater size ~ greater simplicity)
4. social network density (greater density ~ greater complexity)
5. amount of shared knowledge and beliefs, or common ground (Clark 1996: 92–121) (greater common ground ~ greater complexity).

Trudgill argues that small-scale societies that are relatively isolated, stable and (partly due to their small size) made up of a dense network of individuals with much common ground – a characteristic of societies in many but by no means all parts of the world before large-scale European colonization around 1500CE – give rise to high levels of complexity in language structure.

Trudgill's theory is intriguing and his book contains many suggestive examples (see also Andersen 1988 and Thurston 1989 for similar ideas and examples). However, this proposal has not been rigorously examined in detail across a large sample of languages (but see Lupyan & Dale 2010). There also remain challenging issues in defining the structural complexity of languages (see for example Dahl 2004).

1.1   Linguistic features, or linguemes

Here I look at a possibly simpler, though still rather vaguely defined, structural trait of languages, particularly of contact languages: the nature and diachronic

source of their linguistic resources, that is, the particular forms and meanings and form-meaning pairings that make up the inventory of linguistic tools for communication found in a particular language. These linguistic resources or linguistic features, as Mufwene calls them, are called *linguemes* in *Explaining Language Change: An Evolutionary Approach* (Croft 2000), which presents an evolutionary framework for understanding language change that shares much with Mufwene's evolutionary framework (Croft 2010). In Croft (2000), I argue that the core process in language evolution is the replication of linguemes in language use, that is, in communicative interaction between individuals. That is, linguemes are the primary replicators in the evolutionary model of language change that I present – again, not unlike Mufwene's theory, where linguistic features from a variety of sources are replicated with differing frequencies in creole formation in particular and language change in general.

Linguemes can be divided into two broad types: *substance linguemes* and *schematic linguemes* (Croft 2000: 203). Substance linguemes are linguemes in which a pairing of form and meaning is replicated. Substance linguemes may be grammatical morphemes, such as English *-ed* "past" or *will* "future"; basic vocabulary items such as *nose* and *sit*; or other, non-basic vocabulary items such as English *opossum* and *jihad*. (I have deliberately chosen examples of non-basic vocabulary with non-Germanic origins, but the point here is that they are now words of English.)

Schematic linguemes, on the other hand, are linguemes in which only form or only meaning is replicated. An example of a form-only schematic lingueme is the phonetic realization of a phoneme, for example trilled versus uvular /r/ in various continental Western European languages. An example of a meaning-only schematic lingueme is the recruitment of a predicative possession construction, such as the verb "have", combined with a verb denoting an event to express the meaning of the perfect – neither the phonological form of "have" nor the participial verb form is specified – or patterns of co-expression such as a single verb used for both "know a person" and "know a fact". The form-only schematic linguemes are the primary subject of study for phonological typology, and the meaning-only schematic linguemes are the primary subject of study for morphosyntactic and semantic typology.

The distinction between substance and schematic linguemes does not correspond to the distinction between lexical and grammatical elements of language. Lexical items are mostly substance linguemes, as with English *head* or German *Kopf*: Each is a pairing of form and meaning. However, word formation patterns and patterns of semantic extension from one meaning to another are schematic linguemes. For example, English *headhunter* describes a traditional cultural activity with a compound; German uses the same compound pattern but with different forms (*Kopfjäger*; Görlach 2001: 148). English has extended the term

metaphorically to personnel recruitment; and so has German (Görlach 2001: 148). On the other hand, the English grammatical inflection -*ed* is a substance lingueme, contrasting with zero expression of the present: it is a particular form paired with a particular meaning. However, the pattern of contrasting verbal forms for past and non-past tenses is a schematic lingueme: the pattern of the tense contrast, found in many languages (Dahl & Velupillai 2013), does not specify any particular phonological form. Hence, to anticipate later discussion in this chapter, in the context of borrowing, lexical "borrowing" may involve substance linguemes (word forms as well as meanings) or schematic linguemes (calques); and grammatical "borrowing" may involve substance linguemes (grammatical forms as well as meanings) or schematic linguemes (typological traits).

1.2   Evolutionary linguistics and language contact

Linguemes are produced, or more precisely replicated, in utterances between speakers. What makes both Mufwene's and my theories of language change evolutionary theories are three major properties. First, the lingueme pool, or linguistic feature pool, to use Mufwene's term, forms a population, that is, a spatiotemporally bounded entity, namely, the set of linguemes produced in a particular period of time and a particular place. A speech community is also a population of actual persons in a particular time and place, who communicatively interact with each other (obviously, they interact with each other socially in certain ways as well; but we are interested chiefly in their linguistic interactions). Second, variation in the language – the population of linguemes – emerges in the replication process. Finally, and most importantly, the members of the speech community select the linguemes, or variants as they are called by sociohistorical linguists, and the selection process leads to the conventionalization or extinction of the linguemes in question over time (Blythe & Croft 2012).

The popular view of biological evolution is that plants or animals evolve when their population splits up, the populations no longer interbreed – this is called reproductive isolation – and certain biological traits get selected in one population but not another, so that the species divides into two or more new species, each adapted to its biological environment. The linguistic equivalent is the divergence of languages (that is, their linguistic features or linguemes) when a speech community splits, and the new speaker populations no longer communicatively interact, leading to communicative isolation (Croft 2000: 17–19).

The popular view of biological evolution is incomplete, to say the least. Many biological populations are not fully reproductively isolated, and many species (especially plant species) have hybrid origins. The same is even more true of languages. No speech community is totally communicatively isolated. The traditional

tree model in which a speech community splits up, and each new community's language changes due to communicative isolation from the others, is rarely if ever the reality. Human societies are always in some degree of contact with other societies, indeed with many other societies, due to processes of migration, including conquest. In fact, as Mufwene argues, the linguistic feature pool/lingueme pool of all languages has multiple historical sources. This is partly because languages – unlike biological species – may recruit linguemes from any language, not just languages that are genetically closely related to them.

This is the source of Mufwene's interest in, and theories of, language contact, including contact involving pidgin and creole languages; and in contact languages in general. What makes contact languages interesting is that their linguistic feature pool has multiple historical sources, reflecting the social history and ancestry of the speakers who make up the speech community. Hence the value of typologizing languages according to their social function in the ubiquitous case of contact with other societies.

2. Towards a sociolinguistic typology of languages with respect to intersocietal contact

My starting point for the proposed sociolinguistic typology is a distinction proposed by Thurston (1989). Thurston distinguishes between *esoteric* and *exoteric* languages. Esoteric languages are languages used for communication among members of the same speech community. Exoteric languages are languages used for communication between members of different speech communities that remain distinct, that is, where the communication is not a result of the fusion of the speech communities.

The distinction between esoteric and exoteric languages is based on the communicative situation, in particular the speech community membership(s) of the interlocutors. It is not an inherent property, social or otherwise, of the language. From this latter fact follow a number of corollaries. First, a language may be both esoteric and exoteric at the same time; that is, a language may have both esoteric and exoteric functions. Second, a language may change its social type as defined over time. An esoteric language may develop exoteric functions, or an exoteric language may develop esoteric functions. Also, a language that has both esoteric and exoteric functions may come to be restricted to just an exoteric function or just an esoteric function. Of course, in these processes of social change of language use, language structure may also change, leading to a structural divergence of what was once a single language serving multiple functions or changing functions. This

is part of what makes the social processes linguistically interesting (although the social processes are interesting in themselves as well).

To Thurston's esoteric and exoteric types a third type must be added. I qualified the definition of an exoteric language as one in which the interlocutors come from different speech communities that remain distinct. But there is another possible reason why speakers from two different communities are communicating. It could be because the two communities, or at least the relevant subgroups of the communities, are coming together to create a new unitary speech community, for whatever reason. This is one way in which an exoteric language becomes esoteric, but it is a particularly significant way in which this happens. I will call such languages *neogenic languages*.

One might now describe pidgins and creoles in terms of this sketch of a sociolinguistic typology of languages. Pidgins are standardly if simplistically described as exoteric languages, at least at first; one of their commonest uses is in trade relations. Structurally, the lexical substance linguemes of a pidgin are drawn from multiple sources, though one source almost always dominates, the so-called "lexifier" language. The grammatical substance linguemes are generally reduced from those of the "lexifier" language, and many grammatical schematic linguemes and also some grammatical substance linguemes originate in the other source languages, the so-called "substrate" languages. A creole is standardly described as an esoteric language, the language of a speech community where a pidgin has acquired native speakers. In this theory, similarities in the sources of the creole's linguemes to those of pidgins is due to the descent of creoles from pidgins.

One can describe Mufwene's theory of creole formation as a theory in which creole formation is a case of language neogeny: the creation of a new language accompanying the creation of a new society in a colonial context. Mufwene presents a number of arguments in support of his theory. First, pidgins and creoles largely have independent geographical locations and origins (Mufwene 2008: 35, Map 1). Second, creoles emerged largely from plantation colonization in the European expansion post-1500, a social process whereby a new speech community was formed from the fusion of several groups: speakers of non-standard varieties of European languages such as English and French; speakers of minor European languages such as Gaelic and Breton; and enslaved or indentured speakers of either local languages (e.g., Native American languages) or languages of the places from which the enslaved or indentured servants had been taken (e.g., Africa). These groups formed a speech community to the extent that there was communicative interaction between them, and the language of the speech community was the result of the linguistic utterances produced in that interaction. Third, the structures of the creole – both schematic and substance linguemes, in our terms – can be traced to the languages of the speech communities from which the speakers came who fused into the plantation community. Finally, the likelihood of the structures

of a source language surviving to become part of the creole is partly a function of demography (the proportion of speakers of the source language in the new speech community) as well as socioeconomic power structures in the new society.

Of course, the formation of creoles is only one specific case of a neogenic language, just as the formation of pidgins is only one specific case of an exoteric language. They represent only two types of social contact situation leading to two types of contact language. I would like to put this neogenic process of creole formation, and its contrast with the exoteric process of pidgin formation, in a larger context of types of exoteric and neogenic languages. The simple definitions of esoteric, exoteric and neogenic languages hide a continuum of social language types for each of the three broad categories. The continua will be briefly described in the next three sections, along with their apparent structural and social correlates.

3. The exoteric language continuum and its relation to social organization

3.1 Pidgins and the exoteric language continuum

Pidgins represent only one extreme end of the exoteric language continuum; see Figure 2.1. To get there, I will begin from the other end of the continuum and proceed through the intermediate types.

receptive multilingualism ——— symmetric bi-/multi-lingualism ——— asymmetric bi-/multi-lingualism ——— lingua franca ——— pidgin
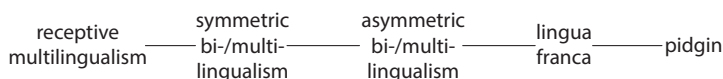
Figure 2.1 The exoteric language continuum

The first type, *receptive multilingualism*, represents the minimum exoteric use of the esoteric languages of the speech communities in contact. In receptive multilingualism, each interlocutor uses their own esoteric language. Exoteric communication succeeds because each interlocutor has at least passive knowledge of the other interlocutor's (otherwise esoteric) language. This type of exoteric communication was described as "semi-communication" by Haugen (1966a/1972) in the Scandinavian context, as "dual-lingualism" by Lincoln (1979/80) in an Austronesian context, and is now referred to as *receptive multilingualism* (see for example Ten Thije & Zeevaert 2007). Receptive multilingualism appears to preserve the esoteric nature of the languages, in that only the speech community member uses their esoteric language in production.

The more widely reported type of bi-/multilingualism occurs where both interlocutors both know and produce one or more of the esoteric languages of their respective speech communities. Such bilingualism may be *symmetric*: either (or both) languages may be used exoterically. This is the next step on the continuum.

Or bilingualism may be *asymmetric*, in that one of the two (or more) languages spoken by the interlocutors is the preferred or exclusive language for exoteric use. Asymmetric bilingualism suggests that one of the esoteric languages in the contact situation is coming to be construed as a specifically exoteric language, in addition to serving as an esoteric language for one of the interlocutors. The asymmetry of bilingualism occurs on a scale, of course: some languages may be more likely than others to be used exoterically, depending on a variety of social factors.

Bi-/multilingualism used for exoteric functions between communities speaking those languages esoterically may have certain structural effects. Stable multilingualism leads to general maintenance of core substance linguemes – grammatical morphemes and basic vocabulary – in the respective speech communities, although non-basic vocabulary linguemes may be borrowed. (These generalizations are probabilistic, of course.) Although some have claimed more extensive borrowing in small-scale (low-population) societies, surveys of borrowing patterns indicate that this is not the case (Alpher & Nash 1999; Bowern et al. 2011); and language phylogenies can be recovered even when there has been up to around 40% of borrowed vocabulary (Bowern et al. 2011; Greenhill et al. 2009).

However, there may be exchange of schematic linguemes, possibly extensively where the societies are in close contact. The most extreme reported case is from the village of Kupwar in India (Gumperz & Wilson 1971). The first example below shows where the Kupwar varieties of Indo-Aryan languages have adopted the morphosyntactic construction for person inflection (contrast the standard Hindi-Urdu construction) from Kannada, a Dravidian language spoken in the village. The second example shows where the Kupwar varieties of Kannada and Marathi have adopted the Indo-Aryan construction for predicate nominal constructions (Gumperz & Wilson 1971: 157, 158):

(1) Kupwar Kannada yəlli   hog   idi   ni
                     where   gone   were(2SG)   you
     Kupwar Marathi kəʈʈə   gel   hotas   twa
     Kupwar Urdu   khã   gəe   te   tu
     Hindi-Urdu   kəhã   gə-ii   th-ii   tu
                     where   gone   were-F   you
                     "Where did you go?"

(2) Kupwar Urdu   ye   tumhar-ə   ghər   həy
                     this   your-SUFF house   is
     Kupwar Marathi he   tumc-ə   ghər   hay
     Kupwar Kannada id   nim-d   məni   eti
     Kannada   i-du   nim   mənə
                     this   your   house
                     "This house is yours."

The sociolinguistic situation in Kupwar is likely to be an extreme case, in that the caste divisions in Indian society lead to a high degree of social separation combined with cohabitation in the same village. There is thus a high degree of exoteric communication combined with a high degree of social segregation. Nevertheless, the same situation is found elsewhere, for example in Oceania (e.g., Thurston 1987), where it has been called "metatypy" (Ross 1996).

The next step is for a language to be more fully separated from its esoteric function, and come to be used as an exoteric language even in a contact situation where none of the interlocutors uses the language esoterically (that is, none are native speakers of that language). A language that has reached this independent degree of exoteric use is generally called a *lingua franca*. A simple example of a lingua franca is the use of English in scientific conferences, say, in Europe, where English may be used by European scientists none of whom are native English speakers.

At this point, exoteric language use has passed beyond the control, so to speak, of the native speakers. This divergence in communicative context may lead to divergence in language structure. For example, there are many lexical items in English as a lingua franca in Europe that are not part of English as an esoteric language in Britain or the United States, as evidenced in Manfred Görlach's *A Dictionary of European Anglicisms* (Görlach 2001).

The last step of separation from any esoteric function is represented by *pidgins*. Pidgins are exclusively used for exoteric functions, possibly highly restricted functions related to certain types of trade or other exchange between distinct societies. Although pidgins may primarily draw their lexicon from a single source, their restricted exoteric function means that both the lexicon and the range of syntactic constructions employed may be limited in comparison to the esoteric language that provides most of the lexicon of the pidgin.

3.2   Exoteric languages and social organization

A number of researchers have observed what they considered a puzzle: the absence of lingua francas or pidgins in areas of high linguistic diversity:

> … the number of reported pidgins is surprisingly small given the linguistic diversity of Papua New Guinea and the number of trading networks in which there was contact between [Austronesian] and other languages.
>
> (Mühlhäusler et al. 1996: 417)

> … basically in the whole interior of Canada, despite the diversity of languages and despite the multiple interethnic contacts … no pidgins are recorded.
>
> (Bakker & Grant 1996: 1152)

> The headwaters of the Xingu river in the Brazilian state of Mato Grosso constitute an area of extraordinary ethnic and linguistic diversity … The tribes of the Upper Xingu entertain close economic and cultural relations … Yet, no lingua franca seems to have developed.   (Adelaar 1996a: 1345)

> Currently, there exists no obvious sociolinguistic or historical explanation for the apparent scarcity of indigenous contact languages in the Southwest; for this area showed as much linguistic diversity, and experienced as extensive intertribal and interethnic contacts, as other regions in North America.
> (Drechsel 1996: 1216; California is also cited in this regard: Drechsel 1996: 1215)

I suggest that the explanation for this apparent anomaly has to do with social organization. Although the analysis of social structure is a complex and controversial area, there are two parameters of variation on which there appears to be some consensus: scale (population size) and stratification. Societies vary considerably in scale. The smallest-scale societies appear to number in the hundreds or low thousands; such societies were found in many parts of the globe before European expansion, and many such societies have survived into the beginning of the twenty-first century.

Societies may also be broadly divided into those that are egalitarian and those that are stratified. Egalitarian societies are those in which "there are few differences between members in wealth, status, and power" (McDowell 2017: 4). Although egalitarian societies do have leaders, they are not hereditary. Stratified societies do have major differences in wealth, status and power (some anthropologists also distinguish an intermediate category of ranked societies, but we will ignore that distinction here).

Egalitarian societies are small-scale while large-scale societies are stratified, presumably since it would be difficult to manage a large-scale society without some degree of stratification. However, societies up to some intermediate size (numbering in the thousands) may also be egalitarian. Thus, we can posit roughly a single scale, with some overlap at intermediate population size, from small-scale egalitarian societies to large-scale stratified societies; see Figure 2.2.

Figure 2.2 A scale of social organization
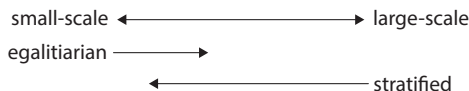
All of the areas described in the quotations above are occupied by small-scale, egalitarian societies. This is of course why there is also great linguistic diversity in these areas: each small-scale society has its own language; without higher-level social integration, the languages will diverge over time; and these areas have been occupied by such communities for a long period of time.

My hypothesis is that *it is only when there is contact with larger-scale, strati-fied societies that lingua francas and pidgins, that is, exclusively exoteric languages, emerge*. In other words, the scale of degrees of exotericity in Figure 2.1 correlates with the scale of social organization in Figure 2.2: if at least one of the languages in contact is farther to the right on the scale in Figure 2.2, then the exoteric language used in contact may be farther to the right on the scale in Figure 2.1. We will explore this hypothesis in the remainder of Section 3, but I first want to suggest a reason for this correlation between social organization and type of exoteric language used.

The question is: why do small-scale egalitarian societies make do with mul-tilingualism in areas of great linguistic diversity (by definition), but larger-scale, stratified societies use more exclusively exoteric languages? My suggestion is that the difference may be due to their engagement (or not) in direct long-distance trade. Small-scale egalitarian societies do not carry on direct long-distance trade, that is, individuals do not regularly travel long distances to trade items. Instead, they trade with neighbouring communities. For this reason, multilingualism involving the local languages will mostly suffice for exoteric function. This is not to deny that trade items may travel long distances through chains of local trade, or that a few individuals may travel long distances beyond neighbouring com-munities. But the trade is not frequent enough or broad enough to motivate the emergence of a lingua franca.

In contrast, large-scale societies have the reach and the resources to engage in long-distance trade with sufficient frequency and breadth to give rise to exclu-sively exoteric languages. This trade may of course involve small-scale egalitarian societies as counterparties, so the latter may also use the exoteric language for trade. The exoteric language is usually, though not always (see below), a descen-dant of the esoteric language of the large-scale society. In the extreme case of social asymmetry between the trading societies and minimal interaction in trade con-tact, a pidgin will arise as the exoteric language.

Pidgins based on European languages, the best-described pidgins, are of this type: they originate with large-scale stratified European societies, and they are used in long-distance colonial trade. The more interesting cases that will test the hypothesis are well-documented non-European lingua francas and pidgins. The question is: are the non-European languages that give rise to lingua francas and pidgins also farther to the right on the scale of social organization in Figure 2.2? If so, then the hypothesis is supported.

The examples of non-European lingua francas and pidgins discussed in the following sections are based primarily on the surveys in *Status and Use of African Lingua Francas* (Heine 1970) and *Atlas of Languages of Intercultural Communica-tion in the Pacific, Asia, and the Americas* (Wurm et al. 1996). These two sources cover most of the world outside the Eurasian continental region where most of the large-scale stratified colonizing societies originated.

3.3   African lingua francas before, or in apparent absence of, European contact

A number of lingua francas in Africa arose before European contact with sub-Saharan Africa. These lingua francas fall into two categories. The first are the languages of states of ultimately Near Eastern origin. Arabic, including a pidginized form, functioned as a major lingua franca for trade between the Arab empire and sub-Saharan societies (Heine 1970: 115–118). The second are languages of sub-Saharan states (called "empires" in the African historical tradition) that arose as a result of contact with Near Eastern-origin states, as a result of long-distance trade with Arab states across the Sahara Desert or the Indian Ocean. Their language was used for political integration but also for trade, and often remained in the latter function after the political collapse of the empire.

The Zenj empire arose as a result of trade along the Indian coast, and Swahili spread with it (Heine 1970: 83). Kanuri developed into a lingua franca along with the Kanem-Bornu empire (Heine 1970: 112). The Hausa "states" developed as a result of trade between the Arab states and the rainforest peoples, leading to Hausa's use as a lingua franca (Heine 1970: 151–153). It appears that Songhai became a lingua franca with the advent of the Songhai empire, and was clearly a lingua franca by the empire's height in the fifteenth-sixteenth centuries (Heine 1970: 159–160). The vicissitudes of the Mali empire starting in the eleventh century led to a series of lingua francas: Mandingo (Heine 1970: 164), which split into Malinke (Heine 1970: 165), Dyula (Heine 1970: 166) and Bambara (Heine 1970: 164), and also apparently gave rise to the pidginized Kangbe (Heine 1970: 170–171). Much later, the Wadai empire was founded in the seventeenth century in what is now eastern Chad by an Arab leader, but Maba as the language of the capital and surrounding area became the lingua franca of the empire (Heine 1970: 115). These lingua francas are summarized in Table 2.1.

Table 2.1 African lingua francas in large-scale societies that predate European expansion

Arabic pidgins	Arab caliphates, states
Swahili	Zenj empire
Kanuri	Kanem-Bornu empire
Hausa	Hausa states
Songhai	Songhai empire (15c–16c)
Mandingo > Malinke, Dyula, Bambara; pidginized Kangbe	Mali empire (11c)
Maba	Wadai empire (17c)

3.4   African lingua francas in the context of European contact

Some societies that controlled trade between Africans and Europeans gave rise to lingua francas. The Duala managed trade between the British and the Cameroon interior from the eighteenth century (Heine 1970: 125). The Wolof benefited from early contact with the Europeans and managed trade on the Senegalese coast (Heine 1970: 147–148). The Ovimbundu managed trade between Europeans and Bantus in south-west Africa from the nineteenth century (Heine 1970: 54–55). Other lingua francas are associated with regional trade networks that arose well after European contact, presumably as a consequence of that contact and the increased trade it stimulated. This appears to be true of Mbum (Adamawa Plateau; Heine 1970: 130), Gbaya (western Central Africa Republic; Heine 1970: 130), Lwena (the Angola-Congo-Zambia border area; Heine 1970: 56) and Jukun (Benue River; Heine 1970: 129).[1]

There is one category of lingua francas that might have arisen before the emergence of states as a result of Near Eastern or European state contact. These lingua francas are found on the Congo River and its tributaries. On the Congo River itself, the early phases of Kituba (Heine 1970: 68) and Lingala (Heine 1970: 73) may antedate European contact, and there may have been a lingua franca preceding Lingala (Heine 1970: 72) there as well. However, Samarin argues against this view (Samarin 1990/1991). Early Sango may have been a trade language on the Ubangi River before European contact (Heine 1970: 131), but Samarin argues persuasively against this view also (Samarin 1982). On the Lualaba River, Tshiluba may have been a pre-European lingua franca (Heine 1970: 65).

The African lingua francas that emerged (certainly or presumably) from long-distance trade with Europeans are summarized in Table 2.2.

3.5   North American lingua francas and trade pidgins

There are a number of trade pidgins and lingua francas reported in North America (Silverstein 1996). Initial European contact was by means of signs and interpreters. In some cases, interpreters knew only local languages (Silverstein 1996: 118); in other cases, it appears that some languages were more widely known, at least among certain individuals, but early European reports were unsophisticated and

---

1.   Yoruba was apparently a lingua franca for trade in the region in and around the kingdom of Dahomey before West Africa was colonized (Heine 1970: 139), possibly between the Gbe and other Oyo communities (Aboh, personal communication). It is not clear if it existed as a lingua franca in this area prior to the establishment of European trading networks.

Table 2.2 African lingua francas that emerged from long-distance trade with Europeans

Lingua franca	Trading partners
Duala	British and the Cameroon interior
Wolof	Europeans on the Senegalese coast
Ovimbundu	European and Bantus in south-west Africa
Mbum	Adamawa Plateau
Gbaya	western Central African Republic
Lwena	Angola-Congo-Zambia border area
Jukun	Benue River
Kituba, Lingala	Congo River
Sango	Ubangi River
Tshiluba	Lualaba River

may describe dialect continua or a bilateral pidgin devised for use with Europeans (Silverstein 1996: 119). However, soon after European contact, extended trade networks and political integration led to some languages becoming lingua francas in the narrow sense, including pidginized forms. Of the latter, the ones that were based largely on indigenous languages and achieved some degree of stability were Montagnais Jargon (early seventeenth century, French-Montagnais contact), Pidgin Delaware (seventeenth century, Delaware-Dutch/Swedish/English; see also Goddard 1997); Apalachee-Spanish Jargon (early eighteenth century, Apalachee/ Spanish); Mobilian Jargon (eighteenth-nineteenth centuries, lower Mississippi societies-French/English); Chinook Jargon (Northwest Coast societies-English); and Pidgin Eskimo, an Inuit-European variety and an Eskimo-Athapaskan variety (Silverstein 1996: 121–131).[2]

  Of the post-contact varieties discussed by Silverstein, two have been argued to be pre-contact in origin, Chinook Jargon and Mobilian Jargon. Hymes (1980) offers some early reports that might be evidence of a pre-contact pidgin antedating Chinook Jargon, but Samarin (1986: 25–26) questions their reliability and the inferences that can be drawn from them. Hymes proposes that Chinook Jargon arose as a result of pre-contact slave trade in the Columbia River area; but a

---

2.   Silverstein suggests the Eskimo-Athapaskan pidgin arose before direct European contact (Silverstein 1996: 120), but Samarin notes that the earliest statement about a precontact Eskimo-Athapaskan pidgin (by Stefánsson) was made 200 years after the first European contact in the Arctic (Samarin 1986: 23).

detailed study of slavery on the Pacific Northwest Coast concludes that extensive slave trading did not occur in the area before contact (Donald 1984, esp. 152–153; see also Samarin 1986: 29). Also, large-scale migrant labour, not small-scale, localized slavery, can give rise to a contact language, but that language is a creole, not a pidgin (Samarin 1986: 29, 30; see Section 4).

These considerations do not rule out the possibility that Chinook Jargon arose as a trade pidgin in pre-contact times. However, linguistic diversity does not require a lingua franca or pidgin for trade, as noted above, and it is likely that trade was restricted to particular members of the society (Samarin 1986: 28–29; see above). Also, the evidence offered by Donald suggests that long-distance, intensive trading networks appear to have arisen as a result of European contact (Samarin 1986; compare Silverstein 1996: 127). Moreover, the European vessels were manned by a large number of indigenous Alaskan and Siberian people, further stimulating the rise of a pidgin (Samarin 1988).

The case for Mobilian Jargon existing before European contact is weaker. The earliest explorers do not mention a lingua franca or pidgin, and used interpreters, as did the Indians among themselves (Crawford 1978: 21–29); there is no positive evidence for a pre-contact origin (Drechsel 1997: 294; Silverstein 1996: 124). After the Europeans established settlements in the Gulf area at the end of the seventeenth century, a lingua franca was reported, and the first examples of the pidgin were recorded (Drechsel 1997: 215–244). Mobilian Jargon was used among Native Americans and between them and European and Africans, peaking in the eighteenth century (Drechsel 1997: 254). Drechsel speculates that a pidgin must have been used among the pre-contact chiefdoms (Drechsel 1997: 285–286), but as we have seen, this is by no means a necessary conclusion. Silverstein argues that Mobilian arose as a pidginized western Muskogean language spoken with the French, with possible contributions from Apalachee-Spanish and from Algonquian languages used in the Mississippi tributaries (Silverstein 1996: 120, 124, 125).

Finally, it has been suggested that Plains Sign Language was a pre-European lingua franca (Taylor 1975, 1981). A Native sign language is reported in the Gulf Coast area by early sixteenth century European explorers. Plains Sign Language spread north and west from the Gulf Coast after European contact (eighteenth to twentieth centuries). Plains Sign Language may have originated among stratified societies in the Gulf Coast area – Native North America included stratified societies in that area – but it spread as a lingua franca only after European contact and the social changes brought about thereby (Samarin 1987).

Table 2.3 summarizes the North American pidgins or lingua francas that appear to have emerged in long-distance trade with Europeans.

Table 2.3 North American pidgins and lingua francas that likely emerged with European contact

North American pidgins and lingua francas	Time period; trading partners
Montagnais Jargon	early 17c, French-Montagnais
Pidgin Delaware	17c, Dutch/Swedish/English-Delaware
Apalachee-Spanish Jargon	early 18c, Spanish-Apalachee
Mobilian Jargon	18c–19c, French/English-lower Mississippi
Chinook Jargon	English-Northwest Coast
Pidgin Eskimo	European-Inuit/Athabaskan
Plains Sign Language	18c–20c, European-Gulf Coast, Great Plains

3.6   Pidgins in Papua New Guinea before European contact

A few pre-contact pidgins have been reported from Papua New Guinea.[3] Although these were used by societies that were at a stage of incipient stratification, they arose in circumstances that favoured their rise even at this stage.

There were a number of bilateral pidgins involving Yimas, a language spoken on the Sepik River. Pidgin Yimas is in fact several bilateral pidgins, each used with a particular linguistic community (Mühlhäusler et al. 1996: 419, based on unpublished work by Jeff Williams; Foley (1988) describes the Yimas-Arafundi pidgin). The Yimas pidgins are used by individual clans, reflecting a common pattern where trade is exclusive to particular families or elites (e.g., Johnson & Earle 2000: 156, 214, 238, 252, 267). Mühlhäusler et al. also cite reports of other bilateral pidgins in the Middle Sepik River area (Mühlhäusler et al. 1996: 420–421). All of these Papuan societies are small-scale and egalitarian, but they are found on a major river and have specialized economic niches. In particular, the Yimas are fisherfolk and incipient stratification has been reported for the Sepik fisherfolk societies (Harrison 1987: 492).

Two bilateral pidgins were used by the Austronesian Motu people for their *hiri* or long-distance trading voyages from the Port Moresby area to the Gulf of Papua (Dutton 1983, 1996). The long-distance trade was required because of the poor agricultural environment of the Motu (Oram 1982: 5) and such voyages were

---

3.   Many other languages used among small-scale societies included in the *Atlas of Languages of Intercultural communication in the Pacific, Asia, and the Americas* (Wurm et al. 1996) are described as "lingua francas", but the term is used in its broad sense and it is likely that they represent one-way multilingualism rather than third-party language use. Descriptions of their use, where given, imply asymmetric bilingualism except where discussed here.

undertaken only when necessary (Oram 1982: 26). The Motu traded with the Eleman and Koriki; they stayed for at least one or two months (Oram 1982: 15) and were largely segregated from the local people (Dutton 1983: 87). The trading languages were predominantly based on Toaripi (an Eleman language) and Koriki respectively; they were the Motu's trading partners. The pidgins were used only in this trade. The Motu were organized as tribes without an overarching political structure (Oram 1982: 3). However, in some villages, headmen were hereditary (Oram 1982: 3), and more significantly, the Motu villages did not make war on one another (Oram 1982: 9), which suggests an emerging regional political identity, a first step towards a larger-scale, stratified society. It appears that the long-distance trade necessitated by food shortages, combined with the long-term absence of the traders from their home society and the segregation of traders from the local people during their stay, were the motivating factors for the emergence of the bilateral pidgin in this small society.

It is very difficult to try to infer social organization, trade and exoteric language use in areas prior to European contact without written records. The changes wrought by European contact were dramatic, and often obscured or erased previous social organization and trading patterns. For example, in Africa, "[s]ince the arrival of the European the contacts between the various ethnic groups had considerably enlarged on account of the stamping out of tribal warfare and the improved means of communication" (Heine 1970: 74). Trade with Europeans along the West African and Congolese coasts led to intensification of long-distance trade with inland tropical forest societies that did not have direct contact with Europeans (Heine 1970: 68, 72). While it is clear that European contact led to social changes that favoured the emergence of exclusively exoteric languages, it is not clear that those social structures existed prior to European contact.

3.7   Social changes and changes in the type of exoteric language

The survey of non-European lingua francas and pidgins in Sections 3.3–3.6 supports the hypothesis that exoteric languages farther to the right of the exoteric language continuum in Figure 2.1 are used when at least one of the societies in contact is larger-scale and more stratified, that is, falls towards the right of the scale of social organization in Figure 2.2.

The survey of synchronic contact situations in those sections can be further supported by evidence of historical changes in the type of exoteric language used in some cases. That is, we can observe the shift from multilingualism for exoteric communication to lingua francas. There is some evidence that lingua francas tend to replace symmetrical multilingualism as societies become larger and more stratified. Tables 2.4 and 2.5 present data from Gapun village, Papua New Guinea

collected in 1987 (Kulick 1993: 118, fn 2), and western and central Kenya in 1968 (Heine 1970: 102).

Table 2.4 Linguistic knowledge in Gapun village, Papua New Guinea in 1987

	Languages spoken; ( ) = passive knowledge		
Sex and age of speakers	Esoteric language	Lingua franca	Neighbouring esoteric languages
Male, > 50yr	Taiap	Tok Pisin	Kopar, Adjora, 1 or 2 others
Male, > 40yr; Female, > 35yr	Taiap	Tok Pisin	Kopar *or* Adjora
Male and Female > 14yr	Taiap	Tok Pisin	(Kopar *or* Adjora)
Male and Female < 14yr	(Taiap)	Tok Pisin	

Table 2.5 Linguistic knowledge in western and central Kenya in 1968

Linguistic knowledge	% of speakers surveyed
Mother tongue only	13.3%
Mother tongue + vernacular(s)	0.6%
Mother tongue + vernacular(s) + lingua franca(s)	17.6%
Mother tongue + lingua franca(s) only	68.5%
Lingua francas: Swahili 85.5%, English 27.8%	

The data from Gapun village document the loss of multilingualism in a small-scale egalitarian society that is now part of Papua New Guinea. Taiap is the esoteric language, spoken only in the village (Kulick 1993: 94). Tok Pisin is the lingua franca, and also the state language of Papua New Guinea; it was introduced into Gapun after World War I (Kulick 1993: 95; see Section 4.1 for further discussion of the evolution of Tok Pisin). The oldest male speakers in 1987 were multilingual in several of the local esoteric languages as well as Tok Pisin, which was at first just another language used for intercultural communication. Younger villagers gradually abandoned use of the local languages in favour of Tok Pisin. The youngest speakers have even abandoned the village language (see Section 4.1).

By 1968 Kenya had been under centralized state control for nearly a century, and contact with Arab and European states had altered trading patterns in the interior for many more centuries. By 1968, when Heine did his survey, virtually none of the multilingual speakers had knowledge of only local vernaculars (esoteric languages), and the vast majority had knowledge of the lingua franca(s) only. Almost all of the multilingual speakers knew Swahili, and about one third knew English, a more recent lingua franca, as well.

Both the Gapun and Kenyan situations involve not just larger-scale societies but also the creation of new societies: the countries of Papua New Guinea and Kenya respectively. I turn to these processes in the next section.

4. The neogenic language continuum

In Section 2, I presented Mufwene's arguments for creoles as neogenic languages, and not as the descendants of pidgins, which are exoteric languages. Creoles, like pidgins, also represent one extreme of a continuum of neogenic languages; see Figure 2.3. Again, I describe the continuum, starting from the opposite end.

language _____ koiné, _____ restructured _____ creole
shift   standard   language

Figure 2.3 The neogenic language continuum

The opposite extreme from a creole, in which there are significant contributions of linguemes from multiple sources, is a neogenic language that arises almost exclusively from a single source. This is *language shift*: the language of the new society eventually shifts to that of the dominant community in the new society. The neogenic language continuum in Figure 2.3 appears to be associated with relatively large-scale societies, although the degree of social asymmetry or stratification (that is, the degree of dominance of one group) appears to play a role, as well as the size of the dominant group relative to that of the other groups brought together in neogeny. Again, the crucial test cases for this hypothesis are non-European contact languages in neogeny, since the European languages are spoken by already large-scale stratified societies.

4.1 Language shift and lingua francas

Language shift is commonly associated with the creation of a large-scale, or at least larger-scale, society by conquest or other means of socioeconomic incorporation. This is most often observed in the creation of modern nation-states, but also took place in the creation of empires in the past, such as the Roman Empire.

Of course, the initial effect of incorporation is the creation of a multilingual society. Language shift does not happen overnight; it frequently takes several generations for complete shift to take place. But incorporation of societies speaking different languages into a new state also leads to larger-scale interactions between the component societies that require an exoteric language. The dominant group's language usually serves as the lingua franca for the longer-distance intersocietal communication that develops within the newly established state. This sort of lingua

franca is therefore the result of incomplete, or not yet complete, language shift. As time goes on, increased economic and political integration of the state leads eventually to complete language shift, unless of course the state or empire breaks up, as happened with the Roman Empire. In the case of the Roman Empire, Latin persisted as a lingua franca among the ecclesiastical and secular elites for many centuries.

Where lingua francas of this type are attested outside the Eurasian area occupied by states since ancient times, they are associated with states, either before or after European contact. Many African lingua francas are primarily languages used for political integration of states, although they may also be used in trade (see Section 3). The Abyssinian empire emerged long before European contact, leading to the use of Amharic as the language among the incorporated societies (Heine 1970: 107). The Mosi 'states' evolved into 'empires', presumably in response to increased trade across the Sahara and with societies in contact with Europeans, and Mosi concomitantly became a lingua franca (Heine 1970: 161–162).

After European contact in Africa, lingua francas arose in newly formed African states. Some African societies responding to European state contact created states as a political defence, incorporating local societies. The Bambara empire arose first after the destruction of the Songhai empire by the Moroccans in the sixteenth century, and survived until the nineteenth century, benefitting from its central location between Arab and European trade routes. Bambara became a lingua franca as a result (Heine 1970: 168). The Ful formed an empire in the nineteenth century at the expense of the Hausa and Bambara states, and the Ful dominated the political administration under French colonization; as a result, Adamawa Ful has become a lingua franca (Heine 1970: 128–129). The rise of the Ashanti empire in the eighteenth century led to Twi being used as a lingua franca (Heine 1970: 141–142).

Table 2.6 summarizes the lingua francas of African states (empires) created before European contact or in response to European contact.

Table 2.6 African lingua francas that emerged in the creation of new states (empires)

Lingua franca	State (empire)
Amharic	Abyssinia
Mosi	Mosi empire
Bambara	Bambara empire (16c–19c)
Adamawa Ful	Ful empire (19c)
Twi	Ashanti empire (18c)

Other African languages became lingua francas in Africa when European states established state colonies and used the local society, or rather one of the local speech communities, as their administrators. Susu had brief prominence around

1790, but is now a contemporary lingua franca in the Sierra Leone estuary (Heine 1970: 146). Ewe became a lingua franca after the Germans took control of the Slave Coast (Heine 1970: 140). Ga became a lingua franca in the nineteenth century after German missionaries chose it for missionary work (Heine 1970: 144). Bulu may have been a lingua franca before its choice by American missionaries in the late nineteenth century, but its clear expansion occurred around that time (Heine 1970: 119–121). After the British took control of Sierra Leone in the nineteenth century, Mende and Temne came to be used as lingua francas as the peoples moved from the interior towards the coast (Heine 1970: 145). Tswana became a lingua franca after the establishment of the Bechuanaland Protectorate (now Botswana; Heine 1970: 52). The language of the Nyanja spread as a lingua franca as the population expanded its territory after the creation of the Nyasaland Protectorate (now Malawi; Heine 1970: 60–61). Ganda became a lingua franca as the British extended their administration through the Kingdom of Buganda (now Uganda; Heine 1970: 105–106). Ewondo (Yaoundé) became a lingua franca as a result of the German colonization of Cameroon (Heine 1970: 122–124), as did Bali to a lesser extent (Heine 1970: 126–127).

Table 2.7 summarizes the African (non-European) lingua francas of newly created European colonies.

Table 2.7 African lingua francas that emerged during the creation of European colonies

Lingua franca	European colonial presence
Susu	Sierra Leone estuary, ca. 1790
Ewe	German Gold Coast
Ga	German missionary work
Bulu	American missionary work
Mende, Temne	British Sierra Leone
Tswana	Bechuanaland Protectorate
Nyanja	Nyasaland Protectorate
Ganda	British Uganda
Ewondo, Bali	German Cameroon

In Central and South America, Nahuatl and Quechua functioned as lingua francas for the Aztec and Inca states respectively, while the Linguas Gerais of Brazil originated in the Portuguese colonial era (Adelaar 1996b; Holm 1989: 605–606).

In Oceania, Tok Pisin began as an exoteric trade language, that is, a pidgin. It is a seeming counter-example to Mufwene's theory because it "creolized" in Papua New Guinea. However, Tok Pisin "creolized" because it became a neogenic language, that is, the language of a colony and later a nation-state, namely Papua New

Guinea. In other words, Tok Pisin began as an exoteric pidgin, then expanded to become a lingua franca in the exoteric sense, and then came to be a lingua franca in the neogenic sense. It expanded in function, and as it became the lingua franca of a newly created state that incorporated a large number of small-scale societies, those latter societies shifted to the lingua franca. In this sense, a pidgin became a "creole", but via processes that are quite general to language neogeny.

It is interesting to compare the development of Tok Pisin in this framework to the development of Latin after the demise of the Roman Empire. Latin served as a lingua franca for the Roman Empire, leading to the shift of many languages under its jurisdiction. After the Roman Empire collapsed, Latin persisted as a lingua franca among the ecclesiastical and secular elites for many centuries. By this point, the Romance languages had diverged and due to its elite prestige, Latin was also used as an elite lingua franca in states with Germanic and Slavic languages. In other words, Latin moved from being a lingua franca in the neogenic sense during the Roman Empire to being a lingua franca in the exoteric sense in Europe, extending long after the fall of the Roman Empire. In other words, Latin and Tok Pisin as lingua francas changed their sociolinguistic functions in opposite directions over their respective histories.

4.2   Neogeny from closely related varieties: Koiné and standard

The next step on the neogenic language continuum is the creation of a *koiné* or a *standard* language out of a set of closely related varieties that have been joined in a newly created state society. A koiné is an exoteric variety used for interdialect communication, in trade or in an incipient state. It differs from complete language shift in that there are multiple sources; however, the sources are closely related varieties. Likewise, a standard language for a state also has multiple sources. Following the classic model of standardization of Haugen (1966b/1972), adapted to the current framework, the steps in the process are: (i) selection of a set of linguemes from different varieties; (ii) codification (in evolutionary biological terms, fixation) of the linguemes for the standard; (iii) elaboration of linguemes for the many functions of the state; and (iv) acceptance by the speech communities in the state, that is, complete language shift to the standard. The standardization process is well documented since it is a concomitant of writing, and there are a number of case studies, for example, French (Lodge 1993), Latin (Clackson & Horrocks 2007) and Greek (Horrocks 2010).

4.3   Neogeny involving speakers of sharply distinct varieties: Restructured varieties and creoles

The next step on the neogenic language continuum is a *restructured variety*. Restructured varieties are language varieties which have undergone some degree

of simplification in comparison with the original language. Restructured varieties stand between simple language shift and creolization in structure and also in demography. Holm (2004) examines five partly restructured varieties, each descended from a different European language (African-American Vernacular English, Afrikaans, Brazilian Vernacular Portuguese, Nonstandard Caribbean Spanish, and vernacular lects of Réunionnais French). Holm concludes that the most significant factor leading to restructuring is the ratio of non-native to native speakers of the European language in the newly created colonial society: the ratio is larger than in simple language shift, but smaller than in creolization.

Restructuring differs from creolization only in degree, not kind. The result of language neogeny is a complex function of the proportion of speakers of the different source languages, the number of different source languages, the social relations between the different linguistic groups of speakers, and so on. In all cases, a new society is created, and a new language with it. The make-up of the new language is the result of a complex but at least partly predictable process, in part due to the major role of demography (see also Tria et al. 2015).

Southern Irish English represents a slightly restructured variety of English due to contact with Irish Gaelic and shift of Gaelic speakers to English. Southern Irish English has a few schematic linguemes drawn from Irish Gaelic, including wider use of the progressive than in British English (3a, b); use of *after* combined with the progressive to express perfect meaning (3d); wider use of the cleft construction (3d, e); and use of elliptical verb phrases instead of *yes/no* in answers to polarity questions (3f, g, h) (Trudgill & Hannah 1994: 106–107):

(3) a. I'm seeing it very well.
    b. This is belonging to me.
    c. I'm after seeing him. (= "I have seen him")
    d. It was very ill that he looked.
    e. Is it stupid you are?
    f. Are you going? – I am.
    g. Is it time? – It is.
    h. Did he come? – He did not.

And phonetic schematic linguemes are frequently drawn from the original language of the shifting speakers; this is, of course, what is known as an accent.

Mufwene also argues that restructuring differs from language shift, and in fact, so-called "normal language transmission", only in degree, not kind (Mufwene 2008). He points out that the history of Indo-European and its descendants is one of migration into South Asia and into Europe in multiple waves, involving conquest, break-up of empires and states, bilingualism of conquered speech communities for generations until those communities eventually shift, and so on.

The amalgamation of speech communities into the societies that spoke the Celtic, Germanic, Romance and Slavic protolanguages, and then their splitting and amalgamation into the societies speaking the modern Indo-European languages, is no less complex or "abnormal" than the amalgamation that led to the creoles in the plantation colonies of the New World. As a consequence, Mufwene argues that creole languages as well as restructured varieties are descendants of the European languages that provide the bulk of the core substance linguemes of these languages (Mufwene 2005, 2008).

5.   "Mixed languages" and the esoteric language continuum

There is also an esoteric language continuum that should be described here in order to complete the sociolinguistic typology presented in this chapter; see Figure 2.4. The esoteric language continuum, along with the neogenic language continuum, will help us to account for one final heterogeneous category of contact languages, namely so-called "mixed languages".
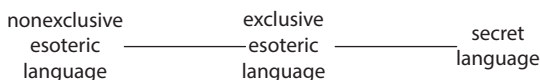
Figure 2.4 The esoteric language continuum

This small continuum starts at one end with *nonexclusive esoteric languages.* A nonexclusive esoteric language is an esoteric language whose exoteric use is not discouraged or avoided. In other words, it is an esoteric language which may more or less freely double as an exoteric language. We tend to think of nonexclusive esoteric languages as the "typical" case, but they actually represent the least esoteric type of esoteric language, since they are open to exoteric function.

An *exclusive esoteric language*, in contrast, is a language in general use in a speech community, but whose exoteric use is discouraged or avoided. The Puebloan languages in the Southwest, particularly those in the Rio Grande Valley in New Mexico, are good examples of exclusive esoteric languages. Most Puebloans do not want outsiders to learn the community language and discourage writing of the language, which of course would make it more available for an outsider to learn.

The most esoteric language of all is a *secret language*. A secret language is a language typically with a limited domain of use that is deliberately designed to exclude overhearers from the communication. A secret language may recruit linguemes from other secret languages, which makes some secret languages a type

of "mixed language" (see below); and/or it may be formed by language games, such as Verlan in French or Cockney Rhyming Slang in English.

Some "mixed languages" appear to be the result of different types of processes of neogeny, while others appear to be the result at least in part of increasing esoteric function; the discussion here is a brief summary of Croft (2003), a revised and expanded version of the analysis in Croft (2000: 213–121).

5.1   Neogeny and "mixed languages": Mixed-marriage languages and semi-shift

What I call mixed-marriage languages are neogenic languages such as Mednyj (Copper Island) Aleut (Thomason & Kaufman 1988: 233–238, drawing on Menovščikov 1969) and Michif (Bakker 1997; Bakker & Papen 1997), in which husbands and wives come from different speech communities but form a new speech community speaking a language with significant portions of linguemes from both source languages. In both cases, European resource exploiters (sealers and trappers) intermarried and settled in Aleut and Cree/Salteaux communities. The linguistic result was distinct grammatical components introduced into the resulting language (Russian verbal inflections in Mednyj Aleut; French words, morphemes and syntax in the noun phrase in Michif). In the case of Mednyj Aleut, there was also displacement of the Aleuts to Copper Island.

There is only one well-documented example of what I call semi-shift, namely Media Lengua (Muysken 1997). In semi-shift, a neogenic language emerges in an apparent process of shifting to the dominant language (Spanish) but with retention of substance and schematic linguemes from the heritage language (Quechua). Media Lengua employs almost entirely Spanish vocabulary (89%; Muysken 1997: 378) but almost entirely Quechua grammatical inflections and constructions, and is spoken by ethnically Quechua speakers. The Media Lengua communities were probably monolingual Quechua around 1900 (Muysken 1997: 374). Men began to work in the Spanish-speaking cities and became fluent in Spanish. Media Lengua arose as a home language in these communities. It is difficult to generalize here since there is only one case.

5.2   Increasing esotericism and "mixed languages": Death by borrowing

The third type of "mixed language" I call death by borrowing (Croft 2000: 217–219, 2003: 52–55). There are a number of well-documented cases in which there is extreme borrowing of increasingly large amounts of vocabulary and even the inflectional morphology of a language. Yet the speakers do not give up their social and linguistic identity.

Well-documented examples of extreme borrowing include the Asia Minor Greek dialects from before World War I (Dawkins 1916; Thomason & Kaufman 1988: 215–222) and the Arabic of Kormakiti in Cyprus (Newton 1964; Thomason & Kaufman 1988: 105–107). The Asia Minor Greek dialects were spoken by Greek communities surrounded by Turkish communities in the Ottoman state. The Greek speakers were commonly bilingual, and some Greek speakers shifted to Turkish (and also to Islam; Thomason & Kaufman 1988: 215). The Kormakiti Arabic speakers are Maronites who left Lebanon sometime after 1191CE, thus originally speakers of Arabic (Newton 1964: 43), and migrated to the medieval state of Cyprus.

Matras argues that the last stage in this process is functional turnover (Matras 2000: 87–91): the basic lexicon and sometimes some grammatical affixes of the original language are restricted to a secret or in-group register of the now-acquired language of the larger society. In other words, a socially increasingly embattled esoteric language ends up becoming even more esoteric, that is, a secret language. Functional turnover occurred in the Romani varieties of Western Europe (including Britain and Scandinavia), called Para-Romani (Matras 2000: 88–90). There is no currently existing Para-Romani language (Matras 2000: 88), but instead an ability to insert Romani vocabulary into utterances in the adoptive society's language (e.g., English or Spanish) to create a secret language.

Matras also suggests that the current status of Ma'a (Mbugu) is the result of functional turnover (Matras 2000: 90–91). Members of the community speak two varieties as first languages. Thomason (1997: 469) describes them as two languages, one being very close to the neighbouring Bantu language Pare. Mous (1994: 175) describes them as two registers of the same language, and calls them Inner Mbugu and Normal Mbugu, the latter being very much like Pare and largely mutually intelligible with Pare (Mous 1994: 176). Inner Mbugu possesses a basic vocabulary, including personal, possessive and demonstrative pronouns, that is largely Southern Cushitic (Thomason & Kaufman 1988: 225), though not from any one existing Southern Cushitic language (Greenberg 1999: 627). But Inner Mbugu also possesses much non-basic vocabulary and almost all grammatical inflections that agree closely with Pare (Mous 1994).

The sociohistorical data suggest that Mbugu is a case of progressively extreme borrowing of Pare/Shambaa by a Southern Cushitic-speaking society (Thomason 1997: 478–482; contra Mous 1994, who argues for a Bantu secret language origin). For example, nineteenth century sources indicate that the Bantu inflections were not yet obligatory at that time (Greenberg 1999: 629). Thomason and Kaufman note that "different (and independent) sources present a picture of the Ma'a people as resisters of total cultural assimilation" (Thomason & Kaufman 1988: 225; see references in Thomason 1997: 472). The present-day Ma'a (Mbugu) community

is scattered among speakers of the Pare and Shambaa societies. The Ma'a/Mbugu people show traces of a pastoralist culture unlike their Bantu neighbours and are physically somewhat different from them as well (Mous 1994: 177). In a more recent period, there has been functional turnover, with Normal Mbugu, essentially a variety of Pare, being used for normal communication.

Two lesser-known societies that have the social history and linguistic effects associated with functional turnover are Ejnu (Lee-Smith 1996) and Wutun (Lee-Smith & Wurm 1996) in north-west China. Both appear to represent social groups that have migrated and been incorporated into a state society. The Ejnu were probably a Shiite group from Persia, the Abdals, who migrated around the eighth century and were later dominated and marginalized by Sunni conquerors. Ejnu retains a substantial core vocabulary of Persian (and Arabic) origin, but much Uighur vocabulary and affixes as well as structural patterns. The Wutun were Chinese from either Sichuan or Nanjing who migrated to Qinghai province by 1585, where they were incorporated into the Tibetan and then the Mongolian state. Wutun possesses a significant amount of Chinese vocabulary but also significant contributions from Tibetan and Bao'an Mongolian, and Tibetan and Bao'an affixes as well as structural patterns.

In sum, we can divide "mixed languages" into two broad categories: those that represent steps on the esoteric language continuum, and those that represent steps on the neogenic language continuum. The steps on the esoteric language continuum represent communities that appear to be retaining the esoteric language by making the language as esoteric as possible. The steps on the neogenic continuum represent an adoption of the neogenic language, but with varying degrees of input from the original esoteric language(s) of the source societies of the neogenic language community.

5.3   Sources of linguemes in "mixed languages"

For each of these categories and subtypes, we can describe the typical sources of the linguemes based on observed examples (Croft 2003: 67). The sources will be labeled E, for the esoteric language of the source community (described as the "heritage language" in Croft 2003) and N, for the neogenic language of the emerging community (described as the "adoptive language" in Croft 2003). Tables 2.8 and 2.9 are adapted from Croft (2003: 67). I should emphasize that the generalizations in Tables 2.8 and 2.9 are probabilistic. They do not imply that the type of lingueme comes exclusively from the language of the relevant speech community. They only imply that the preponderance of linguemes of that type comes from that particular speech community.

Table 2.8 Typical sources of linguemes along the esoteric language continuum

| Esoteric continuum | Substance linguemes | | | |
	Schematic linguemes	Grammatical morphemes	Basic vocabulary	Non-basic vocabulary
Language maintenance	E	E	E	E, other
Death by borrowing	E > N	E > N	E	N, (E?)
Functional turnover	N	N	E	N, (E?)
Secret language	N	N	anything other than N	?

In language maintenance where external linguistic influence is minimal, most of the linguemes are from the esoteric language, although the minimal degree of contact will typically introduce non-basic vocabulary (substance linguemes). In death by borrowing, there remains a strong identity for the community's esoteric language despite incorporation into a larger community with another language. Schematic linguemes and grammatical substance linguemes are introduced from the larger community, as well as non-basic vocabulary; the basic vocabulary is the last bastion of the esoteric language's linguemes. Functional turnover is essentially a transition to a secret language: more and more of the linguemes originate in the larger community, but enough of at least the basic vocabulary remains to prevent outsiders from understanding the utterances of the highly esoteric community. Other secret languages are highly esoteric but use a variety of devices such as word games or linguemes from other secret languages in order to maintain their hyper-esoteric status, as noted above.

Table 2.9 Typical sources of linguemes along the neogenic language continuum

| Neogenic continuum | Substance linguemes | | | |
	Schematic linguemes	Grammatical morphemes	Basic vocabulary	Non-basic vocabulary
Shift (complete)	N, E	N	N	N
Semi-shift (partial)	E	E	N	N
Creolization	N, E+	N	N	N, E
Mixed marriage language	N, E	N, E	N, E	N, E

In complete shift, without a major contribution from the esoteric communities of all but the dominant community, the linguemes of all types are descended from the linguemes of the language of the dominant community. In semi-shift, of which

there is just one well-documented example (Media Lengua), the lexical substance linguemes have their source in the language of the dominant community, but the schematic linguemes and the grammatical substance linguemes have their source in the esoteric language of the original community that is incorporated into the larger community. Creolization is somewhat different in lingueme make-up. Most of the substance linguemes have their source in the language of the dominant community, which is often described as the "lexifier" language, since most of the lexical substance linguemes are from that language. However, a number of schematic linguemes, as well as some substance linguemes, have their source in the esoteric languages of the other communities that formed part of the neogenic society; these are usually described as "substrate" or "adstrate" languages (see for example Muysken & Smith 2015). Finally, mixed marriage languages, of which there are two well-documented cases (Mednyj Aleut and Michif), include linguemes descended from the esoteric languages of both sides of the marriage.

It can be observed that the patterns of lingueme sources in neogenic languages are more varied than the patterns in the esoteric languages. Most of the examples of esoteric languages are perhaps instances of a similar social phenomenon, namely retention of an esoteric language despite incorporation (i.e., a situation that would normally lead to more or less complete language shift).

All of these cases of "mixed languages" share certain social features. They involve contact between a large-scale society and a small-scale society, specifically the incorporation of the latter by the former. There is a very strong but not completely overwhelming social pressure to integrate from the large-scale society, but it is countered by a strong sense of distinct social identity in the incorporated community, or at least some retention of the original distinct social identity even in cases where the communities are more fully integrated.

6.   Conclusion

As noted in Section 1.2, echoing Salikoko Mufwene's long-held view, incomplete communicative isolation is the norm in virtually all speech communities. There is almost always some degree of exoteric language function between distinct speech communities in socioeconomic interaction. And in addition to fission of societies – the classic family tree and normal transmission model of language change – there is constant fusion of societies as well. Mufwene has long argued that linguists must not ignore this complex social story in understanding the evolution of languages and their linguistic structure.

This chapter has attempted to shed a little more light on the types of phenomena that have motivated Mufwene's work, by proposing a sociolinguistic typology

of languages into esoteric, exoteric and neogenic languages. Each of these three types is actually a continuum of language types that arise under different social circumstances, and evolve with different types of linguistic structures and different combinations of linguemes from different speech communities that have been or continue to be in contact.

Although I hope that the sociolinguistic typology presented here sheds some light on the relationship between social contact, social organization and the evolution of languages, it is only a beginning. At best it suggests necessary conditions, not sufficient conditions, for the types of contact languages that emerge. For example, the social origins of Pitkern (Pitcairnese) seem to follow the pattern of the "mixed marriage languages" of Mednyj Aleut and Michif described in Section 5.1. Pitkern arose as a result of the mutiny on the *Bounty* (Holm 1989: 546–551; Mühlhäusler 2013; Ross & Moverley 1964). The nine mutineers had with them nineteen Polynesians (mostly women, mostly Tahitian), and removed them to Pitcairn Island, not unlike the removal of the Aleuts to Mednyj Island. The Polynesian men were murdered early on, as were most of the English speakers, and the second generation were all the product of mixed marriages (Mühlhäusler 2013: 233, 238). Yet the result was not a "mixed language", but an English creole with many Tahitian loanwords and virtually no adoption of grammatical substance linguemes.

I conclude with two observations about sociolinguistic typology that are relevant to the sort of evolutionary linguistics that Mufwene practises (as do I). The first is that some of the social language types and their distinctive structural and diachronic features appear to arise only in large-scale, stratified societies, or in contact with such societies. However, such societies have only emerged quite recently in human history, possibly not before the last Ice Age or the Neolithic Revolution. Hence the language change processes found with these types may not have even existed before the last Ice Age. If so, then we must partly abandon the uniformitarian hypothesis for language change, that language change processes have been the same throughout the history of language. The emergence of certain types of contact languages appears to be dependent on the recent emergence of certain types of social organization, specifically larger-scale stratified societies.

The second observation follows from the first. A side effect of the emergence of new types of social organization, and possibly accompanying new types of language change, is the disappearance of older types of social organization, specifically small-scale egalitarian societies that are in contact almost exclusively with other small-scale egalitarian societies. Social anthropological research on small-scale egalitarian societies returns constantly to the same extant societies, yet these societies exist now surrounded by larger-scale societies and have been incorporated and increasingly integrated into large-scale, highly stratified nation-states. Yet for most of the history of human language, languages were spoken by

small-scale egalitarian societies interacting with similar societies. These recent social, economic and political changes will limit our understanding of sociolinguistic typology and language evolution, at the same time that new forms of social organization and interaction are expanding our understanding in other ways.

References

Adelaar, W. F. H. 1996a. Areas of multilingualism in Northern South America. In Wurm et al. (eds), 1345. https://doi.org/10.1515/9783110819724.3.1345

Adelaar, W. F. H. 1996b. The Tupí-Guaraní languages of Atlantic South America, and Línguas Gerais. In Wurm et al. (eds), 1333–1334. https://doi.org/10.1515/9783110819724.3.1343

Alpher, B. & Nash, D. 1999. Lexical replacement and cognate equilibrium in Australia. *Australian Journal of Linguistics* 19: 5–56. https://doi.org/10.1080/07268609908599573

Andersen, H. 1988. Center and periphery: Adoption, diffusion, and spread. In *Historical Dialectology: Regional and Social*, J. Fisiak (ed.), 39–83. Berlin: Mouton de Gruyter.

Bakker, P. 1997. *A Language of Our Own: The Genesis of Michif, the Mixed Cree-French Language of the Canadian Métis*. Oxford: OUP.

Bakker, P. & Grant, A. P. 1996. Interethnic communication in Canada, Alaska and adjacent areas. In Wurm et al. (eds), 1107–1169. https://doi.org/10.1515/9783110819724.3.1107

Bakker, P. & Papen, R. A. 1997. Michif: A mixed language based on Cree and French. In Thomason (ed.), 295–363. https://doi.org/10.1075/cll.17.12bak

Blythe, R. A. & Croft, W. 2012. S-curves and the mechanisms of propagation in language change. *Language* 88: 269–304. https://doi.org/10.1353/lan.2012.0027

Bowern, C., Epps, P., Gray, R., Hill, J., Hunley, K., McConvell, P. & Zentz, J. 2011. Does lateral transmission obscure inheritance in hunter-gatherer languages? *PLoS One* 6 (9): e25195. https://doi.org/10.1371/journal.pone.0025195

Clackson, J. & Horrocks, G. 2007. *The Blackwell History of the Latin Language*. Chichester: Wiley-Blackwell.

Clark, H. H. 1996. *Using Language*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511620539

Crawford, J. M. 1978. *The Mobilian Trade Language*. Knoxville TN: University of Tennessee Press.

Croft, W. 2000. *Explaining Language Change: An Evolutionary Approach*. Harlow: Longman.

Croft, W. 2003. Mixed languages and acts of identity: An evolutionary approach. In *The Mixed Language Debate*, Y. Matras & P. Bakker (eds), 41–72. Berlin: Mouton de Gruyter.

Croft, W. 2010. Review of Language Evolution: Contact, Competition and Change by Salikoko S. Mufwene. *World Englishes* 29: 306–311.
https://doi.org/10.1111/j.1467-971X.2010.01650.x

Dahl, Ö. 2004. *The Growth and Maintenance of Linguistic Complexity* [Studies in Language Companion Series 71]. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.71

Dahl, Ö. & Velupillai, V. 2013. The past tense. In *The World Atlas of Language Structures Online*, M. S. Dryer & M. Haspelmath (eds). Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/66> (28 January 2019).

Dawkins, R. M. 1916. *Modern Greek in Asia Minor: A Study of the Dialects of Silli, Cappadocia and Pháras with Grammars, Texts, Translations, and Glossary*. Cambridge: CUP.

Donald, L. 1984. The slave trade on the Northwest Coast of North America. *Research in Economic Anthropology* 6: 121–158.

Drechsel, E. J. 1996. Native American contact languages of the contiguous United States. In Wurm, et al. (eds), 1213–1239. https://doi.org/10.1515/9783110819724.3.1213

Drechsel, E. J. 1997. *Mobilian Jargon: Linguistic and Sociohistorical Aspects of a Native American Pidgin*. Oxford: OUP.

Dutton, T. 1983. Birds of a feather: A pair of rare pidgins from the Gulf of Papua. In *The Social Context of Creolization*, E. Woolford & W. Washabaugh (eds), 77–105. Ann Arbor MI: Karoma.

Dutton, T. 1996. Hiri trading languages. In Wurm, et al. (eds), 233–236.

Foley, W. 1988. Language birth: The processes of pidginization and creolization. In *Linguistics: The Cambridge Survey*, Vol. 4, F. J. Newmeyer (ed.), 162–183. Cambridge: CUP. https://doi.org/10.1017/CBO9780511620577.010

Goddard, I. 1997. Pidgin Delaware. In Thomason (ed.), 43–99. https://doi.org/10.1075/cll.17.06god

Görlach, M. (ed.). 2001. *A Dictionary of European Anglicisms*. Oxford: OUP.

Greenberg, J. H. 1999. Are there mixed languages? In *Essays in Poetics, Literary History and Linguistics presented to Viacheslav Vsevolodovich Ivanov on the Occasion of his Seventieth Birthday*, L. Fleishman, M. Gasparov, T. Nikolaeva, A. Ospovat, V. Toporov, A Vigasin, R. Vroon & A. Zaliznjak (eds), 626–633. Moscow: OGI.

Greenhill, S. J., Currie, T. E. & Gray, R. D. 2009. Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society B* 276: 2299–2307. https://doi.org/10.1098/rspb.2008.1944

Gumperz, J. J. & Wilson, R. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border. In *Pidginization and Creolization of Languages*, D. Hymes (ed.), 151–167. Cambridge: CUP.

Harrison, S. 1987. Cultural efflorescence and political evolution on the Sepik River. *American Ethnologist* 14: 491–507. https://doi.org/10.1525/ae.1987.14.3.02a00050

Haugen, E. 1966a/1972. Semicommunication: The language gap in Scandinavia. *Sociological Inquiry* 36: 280–297. Reprinted in *The Ecology of Language: Essays by Einar Haugen*, A. S. Dil (ed.), 215–236. Stanford CA: Stanford University Press. https://doi.org/10.1111/j.1475-682X.1966.tb00630.x

Haugen, E. 1966b/1972. Dialect, language, nation. *American Anthropologist* 68: 922–935. Reprinted in *The Ecology of Language: Essays by Einar Haugen*, A. S. Dil (ed.), 237–254. Stanford CA: Stanford University Press. https://doi.org/10.1525/aa.1966.68.4.02a00040

Heine, B. 1970. *Status and Use of African Lingua Francas*. Munich: Weltforum Verlag/New York NY: Humanities Press.

Holm, J. 1989. *Pidgins and Creoles,* Vol. 2: *Reference Survey*. Cambridge: CUP.

Holm, J. 2004. *Languages in Contact: The Partial Restructuring of Vernaculars*. Cambridge: CUP.

Horrocks, G. 2010. *Greek: A History of the Language and its Speakers*. Chichester: Wiley-Blackwell. https://doi.org/10.1002/9781444318913

Hymes, D. 1980. Commentary. In *Theoretical Orientations in Creole Studies*, A. Valdman & A. Highfield (eds), 389–423. New York NY: Academic Press.

Johnson, A. W. & Earle, T. 2000. *The Evolution of Human Societies: From Foraging Group to Agrarian State*, 2nd edn. Stanford CA: Stanford University Press.

Kulick, D. 1993. Growing up monolingual in a multilingual community: How language socialization patterns are leading to language shift in Gapun (Papua New Guinea). In *Progression and Regression in Language*, K. Hyltenstam & Å. Viberg (eds), 94–121. Cambridge: CUP.

Lee-Smith, M. W. 1996. The Ejnu language. In Wurm, et al. (eds), 851–863.
https://doi.org/10.1515/9783110819724.3.851

Lee-Smith, M. W. & Wurm, S. A. 1996. The Wutun language. In Wurm et al. (eds), 883–897.
https://doi.org/10.1515/9783110819724.3.883

Lincoln, P. C. 1979/80. Dual lingualism: Passive bilingualism in action. *Te Reo* (22/23): 65–72.

Lodge, R. A. 1993. *French: From Dialect to Standard*. London: Routledge.
https://doi.org/10.4324/9780203319994

Lupyan, G. & Dale, R. 2010. Language structure is partly determined by social structure. *PLoS One* 5(1): e8559. https://doi.org/10.1371/journal.pone.0008559

Matras, Y. 2000. Mixed languages: A functional-communicative approach. *Bilingualism, Language and Cognition* 3: 79–99. https://doi.org/10.1017/S1366728900000213

McDowell, P. 2017. Political anthropology: A cross-cultural comparison. In *Perspectives: An Open Invitation to Cultural Anthropology*, N. Brown, L. Tubelle de González & T. McIlwraith (eds). <http://www.perspectives.americananthro.org/Chapters/Political_Anthropology.pdf> (22 August 2018).

Menovščikov, G. A. 1969. O nekotoryx social'nyx aspektax èvoljucii jazyka. In *Voprosy Social'noj Lingvistki*, 110–134. Leningrad: Nauka.

Mous, M. 1994. Ma'a or Mbugu. In *Mixed Languages*, P. Bakker & M. Mous (eds), 175–200. Amsterdam: Institute for Functional Research into Language and Language Use (IFOTT), University of Amsterdam.

Mufwene, S. S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP.
https://doi.org/10.1017/CBO9780511612862

Mufwene, S. S. 2005. *Créole, Écologie Sociale, Évolution Linguistique*. Paris: L'Harmattan.

Mufwene, S. S. 2008. *Language Evolution: Contact, Competition and Change*. London: Continuum.

Mufwene, S. S. 2017a. Language vitality: The weak theoretical underpinnings of what can be an exciting research area. *Language* 93: e202–e223. https://doi.org/10.1353/lan.2017.0065

Mufwene, S. S. 2017b. It's still worth theorizing on LEL, despite the heterogeneity and complexity of the processes (Response to commentators). *Language* 93: e306–e316.
https://doi.org/10.1353/lan.2017.0074

Mühlhäusler, P. 2013. Norf'k. In *The Survey of Pidgin and Creole Languages,* Vol. 1: *English-Based and Dutch-Based Languages*, S. M. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 232–240. Oxford: OUP.

Mühlhäusler, P., Dutton, T., Hovdhaugen, E., Williams, J. & Wurm, S. A. 1996. Precolonial patterns of intercultural communication in the Pacific Islands. In Wurm et al (eds), 401–437.
https://doi.org/10.1515/9783110819724.2.401

Muysken, P. 1997. Media Lengua. In Thomason (ed.), 365–426.
https://doi.org/10.1075/cll.17.13muy

Muysken, P. & Smith, N. (eds). 2015. *Surviving the Middle Passage: The West Africa-Surinam Sprachbund*. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110343977

Newton, B. 1964. An Arabic-Greek dialect. *Papers in Memory of George C. Pappageotes (supplement to) Word* 20: 43–52. https://doi.org/10.1080/00437956.1964.11659849

Oram, N. 1982. Pots for sago: The *hiri* trading network. In *The Hiri in History: Further Aspects of Long Distance Motu Trade in Central Papua*, T. Dutton (ed.), 3–33. Canberra: The Australian National University.

Ross, A. S. C. & Moverley, A. W. 1964. *The Pitcairnese Language*. Oxford: OUP.

Ross, M. D. 1996. Contact-induced change and the comparative method: Cases from Papua New Guinea. In *The Comparative Method Reviewed: Irregularity and Regularity in Language Change*, M. Durie & M. D. Ross (eds), 180–217. Oxford: OUP.

Samarin, W. J. 1982. Colonization and pidginization on the Ubangi River. *Journal of African Languages and Linguistics* 4: 1–42. https://doi.org/10.1515/jall.1982.4.1.1

Samarin, W. J. 1986. Chinook Jargon and pidgin historiography. *Canadian Journal of Anthropology/Revue Canadienne d'Anthropologie* 5: 23–34.

Samarin, W. J. 1987. Demythologizing Plains Sign Language history. *International Journal of American Linguistics* 53: 65–73. https://doi.org/10.1086/466044

Samarin, W. J. 1988. Jargonization before Chinook Jargon. *Northwest Anthropological Research Notes* 22: 219–238.

Samarin, W. J. 1990/1991. The origins of Kituba and Lingala. *Journal of African Languages and Linguistics* 12: 47–77.

Silverstein, M. 1996. Dynamics of linguistic contact. In *Handbook of North American Indians, Vol. 17: Language*, I. Goddard (ed.), 117–136. Washington DC: Smithsonian Institution.

Taylor, A. R. 1975. Nonverbal communications systems in native North America. *Semiotica* 4: 329–374. https://doi.org/10.1515/semi.1975.13.4.329

Taylor, A. R. 1981. Indian lingua francas. In *Language in the USA*, C. A. Ferguson & S. Brice Heath (eds), 175–195. Cambridge: CUP.

Ten Thije, J. D. & Zeevaert, L. (eds). 2007. *Receptive Multilingualism: Linguistic Analyses, Language Policies and Didactic Concepts* [Hamburg Studies on Multilingualism 6]. Amsterdam: John Benjamins. https://doi.org/10.1075/hsm.6

Thomason, S. G. 1997. Ma'a (Mbugu). In Thomason (ed.), 469–487. https://doi.org/10.1075/cll.17.16tho

Thomason, S. G. (ed.). 1997. *Contact Languages: A Wider Perspective* [Creole Language Library 17]. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.17

Thomason, S. G. & Kaufman, T. 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley CA: University of California Press. https://doi.org/10.1525/9780520912793

Thurston, W. R. 1987. *Processes of Change in the Languages of North-Western New Britain* [Pacific Linguistics, Series B. 99]. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Thurston, W. R. 1989. How exoteric languages build a lexicon: Esoterogeny in West New Britain. In *VICAL 1, Oceania Languages: Papers from the Fifth International Conference on Austronesian Linguistics*, R. Harlow & R. Hooper (eds), 555–579. Auckland: Linguistic Society of New Zealand.

Tria, F., Sevedio, V. D. P., Mufwene, S. S. & Loreto, V. 2015. Modeling the emergence of contact languages. *PLoS One* 10(4): e0120771. https://doi.org/10.1371/journal.pone.0120771

Trudgill, P. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: OUP.

Trudgill, P. & Hannah, J. 1994. *International English*, 3rd edn. London: Edward Arnold.

Wurm, S. A., Mühlhäusler, P. & Tryon, D. T. (eds). 1996. *Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110819724

CHAPTER 3

A local history of Tok Pisin

Language contact in Papua New Guinea

Gillian Sankoff

University of Pennsylvania

Research on language contact in Papua New Guinea beginning in the mid 1960s, prior to national independence, revealed linguistic micro-evolution as constrained by society, culture, and people's relationships "on the ground". Among speakers of Buang, most men and women under 40 were already fluent in Tok Pisin. Many also spoke another Austronesian language learned at missionary schools: Yabem. Exploring the micro-evolution of Buang speech varieties along the river valley, this chapter situates 'place' as fundamental in the differentiation of Buang speech varieties. It further attempts to show how the inter-influence of speech varieties in complex repertoires can be culturally framed as natural and unproblematic.

Keywords: trade, exploitation and settlement colonization, language and place, bilingualism, Tok Pisin

1. Introduction

In his review of Nettle and Romaine's *Vanishing Voices*, Salikoko Mufwene remarks that "one of the differences between New Guinea and Australia, where there are very few Aboriginal languages left, lies in the fact that the former underwent exploitation colonization but the latter settlement colonization" (Mufwene 2002: 383). This statement references his tripartite distinction between trade colonization, exploitation colonization and settlement colonization (Mufwene 2001), and their differential effects on the survival of local languages and on the emergence of pidgins and creoles. The first two of these are relevant to the current chapter.

> **Trade colonies** produced pidgin Englishes, owing apparently to the occasional interactions associated with trade between the indigenous people particularly of Africa and the Pacific during the 19th century.
>
> (Mufwene 2015: 13, bold type introduced by G.S.)

In Mufwene's exposition, exploitation colonies constituted the subsequent historical phase.

> [E]**xploitation colonies**…[were a] response to the greater need for raw materials outside Europe created by the Industrial Revolution of the late 18th century. Every European industrial nation wanted some territories to colonize, in order to control the supply of raw materials, in Africa, Asia and the Pacific.
>
> (Mufwene 2015: 14, bold type introduced by G.S.)

In this chapter, I examine the linguistic effects of exploitation colonization in one rural area of Papua New Guinea, using historical records as well as my own field research among Buang villagers in the 1960s and 1970s. In doing so, I apply a close-up lens to study how changes in the linguistic landscape brought about by colonization have been assimilated to the local cultural construction of the relationship between language and society.

The chapter is organized as follows. After briefly laying out the early history of Tok Pisin in the next section, the third section introduces the Buang people, and deals with their earliest involvement in the colonial economy and society. The fourth section describes how Buang proximity to the Bulolo goldfields, discovered in the early 1930s, intensified the social and linguistic changes begun in the previous generation. The fifth section begins with the impact of World War II, following through the subsequent decades. It considers the extent to which twentieth-century Buang history might represent in microcosm the broader national situation in terms of language contact. In the sixth and final section, I raise questions about how linguistic diversity has been and will continue to be experienced and understood in Papua New Guinea.

2.   The early history of Tok Pisin

Pointing to similarities between the historical context of colonization in West Africa and the Pacific islands prior to the early nineteenth century, Mufwene defines the situation of trade colonization as the crucible for the formation of pidgins:

> [P]idgins based on European languages are concentrated on the Pacific islands and on the coast of West Africa … [where] until the early 19th century … the Europeans developed no more than trade colonies in both regions … It is only during the 19th century that the Europeans would transform them into exploitation colonies.   (Mufwene 2008: 212)

The nineteenth-century ancestor of Tok Pisin, pejoratively titled the "South Pacific Jargon" (or often, simply, "broken English"), was indeed a product of trade. In the early decades of the nineteenth century, a maritime version of English was the lingua franca of whaling vessels, crewed by men from all over the world. Ships regularly put in at Pacific island ports to trade goods manufactured in Europe and North America for fish, meat and fresh produce. These visits intensified as of the 1830s, with the beginning of the sandalwood trade (Shineberg 1967). The sandal-wood trade ramped up in the south-western Pacific in the 1840s, giving rise to the term "Sandalwood English", with Pacific islanders joining ships' crews as well as labouring in timber extraction. By the 1840s, when

> the exchange system of the New Hebrides sandalwood trade had built up a de-mand for pearl and tortoiseshell, bêche-de-mer and shell gatherers followed the whalers. They necessarily established closer relations with more islanders … [because] gathering of bêche-de-mer, in particular, was a very painstaking busi-ness.…Pearl and tortoiseshell collecting also required the direct participation of the local inhabitants.   (Corris 1973a: 9)

Bêche-de-mer (the French name for trepang, a sea slug prized in Chinese cuisine) was plentiful off many islands of the south-western Pacific. New England ships in the China trade multiplied their profits by making a stop for trepang along the way. Corris cites a passage from a trader active in the western Solomon Islands in the 1840s: "They can nearly all speak more or less broken English, which they have picked up through their intercourse with whale ships, who often visit them to get supplies of cocoa-nuts and pigs" (Cheyne [1853] 2017: 53).

The transition from trade to exploitation colonization, begun in the first half of the nineteenth century, was fully accomplished in the second half, with the mas-sive labour recruitment that supplied workers for the plantation economy:

> Recruiting for Queensland and Fiji [plantations] had begun in the New Hebrides in the early 1860s and, between 1863 and 1868, 1347 New Hebrideans engaged for Queensland. Between 1864 and 1869, 1647 Melanesians, most of whom were from the New Hebrides, were recruited for Fiji … The Fiji recruiters pushed north to the Solomons in 1870.   (Corris 1973a: 24)

Summarizing the events of the three decades, Corris writes: "In the 1860s many hundreds of New Hebrideans were forced or deceived into entering the recruiters' boats, the same was true in the Solomons ten years later, and yet again around New Guinea and the adjacent islands in the 1880s" (Corris 1973b: xix). Assembling data from a variety of sources, it was possible to construct a partial picture of the linguistic diversity among the islanders involved in this labour system. Table 3.1 accounts for approximately 460,000 of the islanders involved.

Table 3.1 Approximate number of labourers from Vanuatu, the Solomon Islands and Papua New Guinea, taken to Fiji, Samoa, Queensland, New Caledonia and other islands within each area, 1843–1939. (Data sourced from Barclay 1978; Corris 1973a; Denoon 1997a, 1997b; Firth 1997; Mosel & Mühlhäusler 1982; Mühlhäusler 1978; Scarr 1990. Data on Fiji rely entirely on Siegel 1985.)

| | FROM | | | |
TO	Vanuatu	Solomons	Papua New Guinea	TOTAL
Internal to each island group 1843–1939	10,000	c. 67,000	280,000	357,000
Queensland 1863–1904	39,975	13,006	5,797	58,778
Fiji 1864–1911	14,198	8,228	1,618	24,044
New Caledonia 1865–1929	8,000	??	??	8,000
Samoa 1879–1900	c. 1,000	100	6,000	7,100
TOTAL	73,173	88,334	293,415	454,922

In the multilingual plantation context, the trade jargon evolved as a pidgin. Though it is not possible to ascertain whether the varieties spoken in different areas were already diverging, the fact that a proportion of labourers signed on to new contracts for work in more than one location (Corris 1973a) probably contributed to a degree of linguistic unity across the vast area involved.

One data source that illustrates in microcosm the linguistic diversity present on plantations at this period comes from an 1885 publication. In December 1884, the Queensland Legislative Assembly set up a Royal Commission to investigate reported abuses by recruiters (Queensland 1885). The Commission's report, detailing massive abuses and contractual violations discovered from interviewing workers on a number of Queensland sugar plantations, resulted in the repatriation of the islanders involved, and the halting of labour recruiting for Queensland until 1892 (Corris 1968: 105). Of the 481 men interviewed by the commissioners, it was possible to determine with a high degree of certainty the native languages of 310.[1] Spread across 10 different plantations, these 310 men were speakers of 16 different languages.

Thus the transition from trade to exploitation colonization, completed as of the beginning of the plantation period in the 1860s, was responsible for the spread

---

1.   Languages were assigned by checking the names of men's natal villages, as listed in the Village Directory for the Territory of Papua New Guinea, 1960 (Village Directory 1960) against language maps of their home areas. I am grateful to Miriam Meyerhoff for her helpful assistance in carrying out this analysis.

of the pidgin (Crowley 1990; Mosel & Mühlhäusler 1982; Mühlhäusler 1985; Sankoff 1985). And it began to diverge into three separate varieties after plantation labourers returned home to the islands that emerged as three separate countries close to a century later: Vanuatu (Bislama), the Solomon Islands (Pijin) and Papua New Guinea (Tok Pisin).[2] In each of the countries, the former pidgins underwent considerable influence from substrate languages (Keesing 1988; Sankoff 1993) as they developed into urban vernaculars and eventually acquired substantial numbers of first language speakers (Jourdan 1985; Meyerhoff 2000; Romaine 1988, 1989; Sankoff & Laberge 1973; Smith 2002). They all contain features inherited from the original South Pacific Jargon, as well as unique features that are a product of separate development over the 150 years since.[3]

3.   Exploitation colonization in Papua New Guinea: Earliest impacts on Buang villagers

The Snake River Valley is located in one of the rare inland regions of Papua New Guinea where Austronesian languages are spoken. Originating in the high sub-coastal Herzog Mountains south of Lae (Morobe Province), the river runs inland, south-west toward Mumeng, as indicated on Map 3.1. People living in the string of villages along the river, both in the valley itself and on mountainsides overlooking it, have referred to themselves collectively as Buang when dealing with outsiders since at least the first decades of the twentieth century (Waterhouse 2010).

When I began my research on language contact there, new languages had not displaced local vernaculars, which remained the everyday medium of communication for both young and old. Thus it was possible to study linguistic ecology *on the ground*, a way of thinking about the social grounding of language that I learned from people in Mambump Village, where I was welcomed as a resident visitor and began learning the language in August 1966. Mambump villagers referred to their own language as *il ayed* "our language" in Buang, and as their *tok ples* "native language (of a place)" in Tok Pisin. There and in other villages up and down the valley, I found nobody who failed to speak natively the language of their natal village, and nobody who spoke English. Nevertheless, the effects of European colonization,

---

2.   In the 1850s many labourers from the archipelago of the then New Hebrides (Vanuatu) were employed in the trade for trepang. Its French name, *bêche de mer*, "beast of the sea", is the source of the language name Bislama.

3.   In all three countries, they are now the dominant languages in terms of numbers of speakers. Bislama and Tok Pisin are official languages in their respective countries.
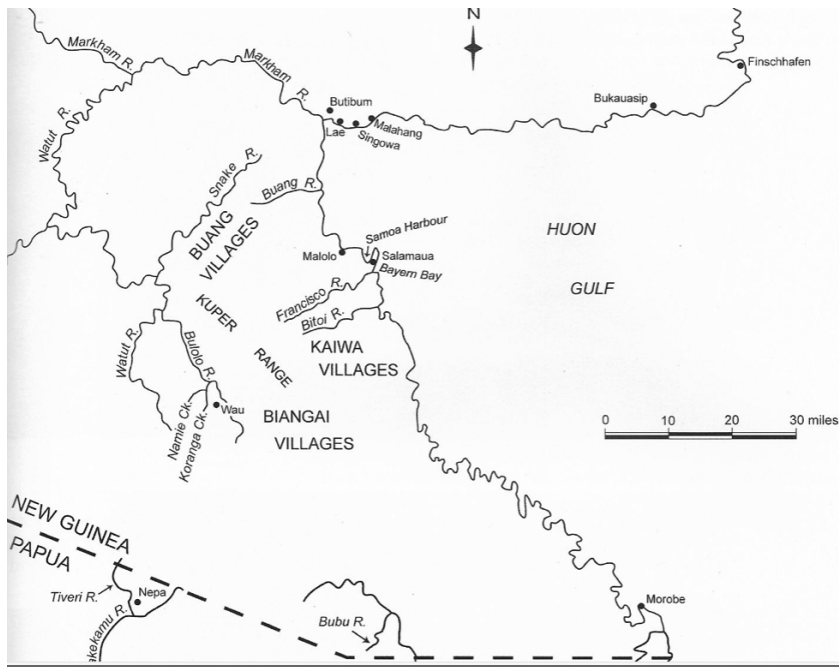
Map 3.1 Location of Buang villages, roughly equidistant between the coastal town of Lae, the Salamaua Peninsula, and the Wau/Bulolo goldfields area. (Source: Waterhouse 2010: 21; reprinted courtesy of cartographer Michael Braund.)

both linguistically and culturally, were ubiquitous. Most people, from teenagers to middle-aged, were also quite fluent in Tok Pisin. Young children were raised in Buang, and when older people occasionally instructed them in Tok Pisin, this occasioned great hilarity on the part of everyone present.

Mambump villagers first experienced exploitation colonization at the beginning of the twentieth century, as a result of the growing demand for coconut oil in Europe. The copra trade in the New Guinea Islands was initially in the commercial hands of Godeffroy's of Hamburg, but in 1884

> the German Empire formally took possession of the northeast quarter of the island and put its administration in the hands of a chartered trading company formed for the purpose, the German New Guinea Company … [In 1899], the German government took control of the colony. [During this time] thousands of local workers were hired as cheap labor on cocoa and copra plantations.[4]

---

4.   <https://en.wikipedia.org/wiki/History_of_Papua_New_Guinea#German_New_Guinea> (last access on 16 December 2020).

Known as Kaiser Wilhelmsland, New Guinea remained a German colony from 1899 until the outbreak of World War I in 1914.[5] However, German barely gained a foothold. Despite an official policy of promoting the German language there,

> [t]he policy was not effectively pursued … Indeed, mission-run schools taught mostly in Tokpisin, which was also the language of commerce and that commonly used by administration officers … By 1909, despite the colonial governor's dislike of a language that many considered brutish, the German administration had established a Tokpisin language school in Alexishafen (Madang) and was encouraging village leaders to learn Tokpisin.[6]   (May 2003: 297)

The grudging acceptance of Tok Pisin by German administrators as the best workable option for communicating with local people did not, however, imply that it was everywhere successful, especially when attempting to communicate with village leaders. The younger generation, men who had been away serving indentures as plantation workers, returned speaking Tok Pisin, but the older men who were deemed suitable to serve as village officials had not had their experience.

> The plan adopted was to confer upon the leading headman an official title, "luluai", the word meaning "village leader" in Rabaul dialect, and hold him responsible for the settlement of minor disputes and the discipline of petty transgressors[7] … The first luluais were usually of advanced years and thus ignorant of pidgin English. A younger official, the "tultul" (another Rabaul word) was therefore chosen to act as interpreter and spokesman. This man had always served a term in European employment … The mark of office was a navy-blue peaked cap with a broad red band for the luluai and two narrow red bands for a tultul.
>
>   (Hogbin 1951: 150–151)

---

5.   In a succinct summary of colonial history, West reports that "the colony was captured by an Australian expeditionary force in 1914, ruled by a military administration for six years, and first retained under civilian control as a Class 'C' mandate under the League of Nations and then as a Trust Territory under the United Nations in 1946" (West 1966: 3). Australia also assumed the administration of Papua in 1946, continuing in this role until Papua New Guinea became an independent country in 1975.

6.   The spelling Tokpisin appears to be unique to May (2003). Hall (1943) proposed the term Neo-Melanesian, also using Melanesian Pidgin, which appeared as well in the title of Mihalic's dictionary and grammar (1971). In recent decades linguists have followed the official Papua New Guinea Government spelling: Tok Pisin.

7.   "Rabaul dialect" refers to the Kuanua language of the Tolai people in whose home area the colonial town of Rabaul was established as the German capital. Kuanua is the source of most of the non-English-origin vocabulary in Tok Pisin and of some of its substrate-origin syntactic features (Mosel 1980; Sankoff 1993, 1996).

Recruited for plantation work in New Britain, Buang men initially joined this throng in the first decade of the twentieth century. Taken there by ship, they served a multi-year indenture before being returned home.[8] This early period, referred to anomalously as *Gut Taim* "good time" in Tok Pisin, was not remembered fondly. One elderly Buang man put it to me succinctly: *Gut Taim i nogut tru* "the *Gut Taim* was really terrible". Working for Europeans involved long hours, capricious and sometimes brutal masters, and difficult living conditions. Nevertheless, those who had been away returned with new and substantial capital, not only in the concrete form of manufactured trade goods bought with their earnings. They also accrued considerable cultural capital gleaned from new contacts with other New Guineans from near and far, and from their exposure to the colonial regime of which Europeans were the masters.

Men arriving on plantations in New Britain as native speakers of a multitude of languages learned the plantation lingua franca, Tok Pisin, which had already gained a foothold among the local Tolai people there (Mosel 1980; Salisbury 1962, 1970). Buang men were no exception. Returning home, ex-plantation workers discovered that fluency in Tok Pisin was linguistic capital with a value in the new colonial society, as described by Hogbin (above) with respect to local *tultuls*. Villagers holding the title of *luluai* or *tultul* were brokers between the village and the secular colonial society.

It is noteworthy that, after World War I, there were Europeans in New Guinea who spoke no English and now found themselves required to communicate not only with the local villagers but also with the Australian administrators. Apparently, "even after 1920 many of the [German, Lutheran] missionaries spoke no English, and in intercourse with administrative officers and other residents they had to employ pidgin" (Hogbin 1951: 234).

Already in the first decades of the twentieth century, Buang people had acquired a reputation for good relations with Europeans. Waterhouse recounts

---

8.   Indentured labour contracts were in force in the south-western Pacific by at least the 1880s (Scarr 1967). Documentation from 1884 reveals that few of the illiterate men who "signed" contracts by "touching pen" or "marking paper" understood the terms of their contracts, in particular with respect to the length of time involved (Sankoff 1985). Reed (1943: 179) notes that "[t]axation of the natives was one of the devices by which the Germans sought to create a stable labour supply, and this expedient has been maintained by the Australian government". In 1931, "recruiters were licensed to go into the villages and sign on men for work as 'general labourers' at six shillings per month, or for 'mining and carrying' (heavy labour) at ten shillings per month, for periods of one to three years. The poll-tax of twenty shillings per year made it necessary for many men, especially in areas remote from white settlement, to go to work" (O'Neill 1979: 2). Deery (2014) documents the end of the practice in 1945.

the saga of Helmuth Baum, a German plantation owner who, when World War I started in 1914,

> disappeared into the jungle rather than surrender to the Australians. He lived among the Buang during the war, out of reach of Australian patrols, prospecting for gold and occasionally visiting missionaries on the coast for supplies.[9] He became very close to the Buang people, helping them when sick and in fights against their enemies. (Waterhouse 2010: 127)

Whether or not Baum learned to speak Buang as a result of the time spent living among them, German would have been an unlikely choice, as Tok Pisin was already known to returned plantation workers, and spreading as a lingua franca.

4.   The Bulolo Gold Rush of the 1930s

Though a few adventurous Buang men made the arduous journey to German plantations prior to World War I, the next commercial development that affected the Buang was much closer to home, and involved many more young men: the Bulolo Gold Rush. In 1921, "Australia officially took over the administration of former German New Guinea (New Guinea). In the same year, it adopted the already well established Tokpisin as its principal (unofficial) language" (May 2003: 298).

Australian prospectors appeared on the scene shortly thereafter, and "the discovery in 1922 of gold at Koranga Creek, a tributary of the Upper Bulolo River" (Sinclair 1981: 23) marked a sea change in the history of Papua New Guinea. Only two days' walk from Headwaters Buang, Sinclair notes that in the Wau/Bulolo area "there were perhaps 50 white miners at work by the end of 1925" (Sinclair 1979: xix). By 1930–31, some 1,900 local workers were employed in mining operations (New Guinea Annual Reports 1941), outnumbering expatriates by at least 20:1. The number of local workers more than tripled in the ensuing decade, to 7,189 in 1940–41. Although by 1939 plantations still accounted for approximately 50% of New Guineans employed in wage labour, mining represented a larger and larger share, rising steadily from 7% of 27,765 employed workers in 1930–31 to 17% of the 41,675 in 1938–39. (New Guinea Annual Reports 1941).

The New Guinean villages closest to the goldfields were populated by speakers of Biangai, a Goilalan language in the Trans New Guinea family of Papuan

---

9.   Waterhouse does not specify the location of the "missionaries on the coast" at that time; however Lutheran historians have documented the establishment of a mission station on the coast south of Lae in 1906 at Bukawa Village, the closest coastal village to the Headwaters Buang (Wagner 1987: 50).

(Non-Austronesian) languages (Pawley & Hammarström 2018: 49). Estimating the Biangai population at somewhere between 1,100 and 1,400 at this time, their language stood no chance of becoming a lingua franca on the goldfields.[10] The local population would have been dwarfed by the tremendous influx of labourers speaking dozens of languages, even if young Biangai men had flocked to the goldfields. To whatever extent Biangai men found employment in the mining operation, I have found no evidence that mineworkers from elsewhere had much to do with the local people, nor of their learning the Biangai language.[11]

To furnish the hundreds of workers required, labour recruiters cast their nets widely. Waterhouse (2010: 88) cites *The Papua and New Guinea Diaries of Sarah Chinnery*, whose writing attests to the numerous *wantok* (Tok Pisin, "same native language") groups present in Bulolo in 1933:

> Large areas of hillside have been cleared to make gardens for the natives. Each *wantok* group is given a garden. Sepiks, Madang, Manus, Aitape … all have their own patch of ground where they grow their own sugar cane, bananas, *kaukau* ["sweet potato"] and paw paws ["papaya"].
>
> (Chinnery cited by Waterhouse 2010: 164, italics in the original.)

Tok Pisin flourished among the workers on the goldfields, and Buang people had a significant presence there. They were noted, among other things, for food they supplied from the sizeable gardens they maintained in Bulolo (Waterhouse 2010). However, it was as carriers hired by both prospectors and *kiaps* "government officers" that they gained most respect. Serving on treks with prospectors across vast swathes of New Guinea territory, they encountered people who spoke dozens of different languages. These expeditions were often dangerous. On at least two occasions, Buang carriers were killed by the people whose territories they were travelling through. Baum, a plantation owner turned prospector who had continued to hire Buang carriers, met his death with them on one of these occasions: "In 1931 the German prospector, Baum, was killed in the Kareeba-Indiwi territory, together with eight of his Buang carriers … In December 1932 the Kapau River Kukukuku

---

10.   A likely village size of approximately 200 would yield a population of only about 1,400, the same number cited by Pawley and Hammarström (2018). The 1973 edition of the *Papua New Guinea Village Directory* lists the same seven villages mentioned by Sinclair, giving a total population figure of 1,074 for the 311 square miles (approx. 805 square kilometres) of the Biangai territory (Papua New Guinea 1973).

11.   The detailed documentation of the situation by O'Neill (1979), Sinclair (1966, 1981) and Waterhouse (2010) contains no evidence that Biangai people were in any way dominant among workers on the goldfields.

murdered the prospectors Clarius and Naylor, together with seven of their Buang carriers" (Sinclair 1966: 8).

Australian prospectors representing both *gavman* "government" and *bisnis* "business"; understood the importance of learning Tok Pisin in order to communicate with New Guineans. Jack O'Neill arrived on the Busama coast on 26 January 1931 at the age of twenty-three, to try his luck searching for gold. His first days were spent being instructed in how to hire the needed New Guinean carriers and labourers for his prospecting ventures – a procedure he found shocking:

> We *bought* the boys, apparently, from Gregor McDonald, who ran a trade store and recruiter's agency.[12] This talk of boys being bought and sold was a shock at first, even if they were not put up on a block for auction. They were not so treated. There was a demand for labour throughout the Territory, and to supply it *recruiters* were licensed to go into the villages and sign on men for work as "general labourers" at six shillings per month, or for "mining and carrying" (heavy labour) at ten shillings per month, for periods of one to three years.
>
> (O'Neill 1979: 2, italics in the original)

Setting out for the trek inland with the twenty men he had hired, only four days later (30 January), his diary mentions the problem of communication: "Most of our boys were bushmen … but fortunately a few of them could speak *pidgin*. We couldn't, but at least their knowledge brought us one step closer to understanding one another, although there was still a wide gap there" (O'Neill 1979: 2, italics in the original).

The only surviving mineworker still alive in Mambump Village in 1966, Gu Danggi (aged approximately sixty-five) spoke eloquently – in fluent Tok Pisin – of the hard work and many dangers of the goldfields.[13] Buang men were not only in great demand as carriers on exploratory foot patrols, but were also employed in the backbreaking work of gold dredging in the mining operations. To get to Bulolo from Mambump Village, the first thirty kilometres or so involved walking down the relatively easy path along the Snake River, where Buang dialects and languages are spoken. The path along the river is labelled "New Buang Track" on Map 3.2, and was one of the routes travelled by miners and their carriers from the coast to the goldfields.

---

12. O'Neill adopts the undeniably derogatory term "boy" in referring to New Guinean men. This practice, universal at the time on the part of all speakers of Tok Pisin, fails to convey the considerable respect O'Neill had for the men who worked for him, revealed in many details of his accounts.

13. Having served in former decades as a *tultul*, Gu Danggi was often referred to by this title, and still had the cap (which he sometimes wore) to prove it.
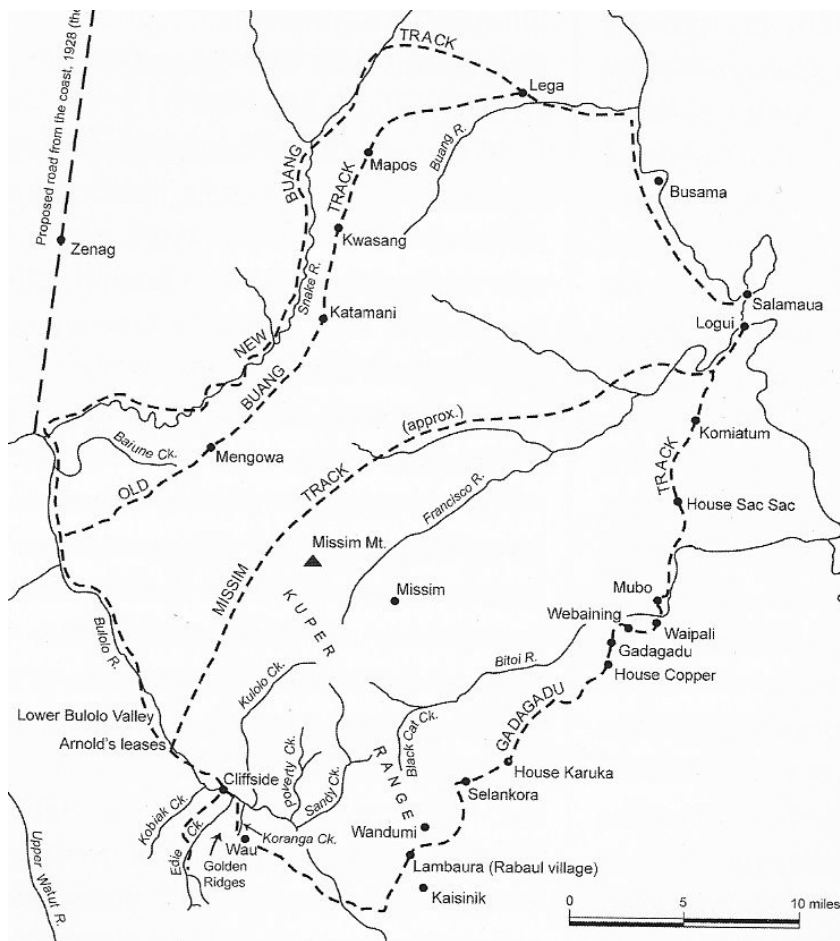
Map 3.2 Tracks from coastal villages to the Wau/Bulolo goldfields. (Source: Waterhouse 2010: 46; reprinted courtesy of cartographer Michael Braund)

In my own experience, native speakers of different Buang varieties, when meeting on this path, would each use their own individual *tok ples* to converse with each other. This was a strategy that combined a modicum of mutual intelligibility with some language-learning from prior linguistic exposure due to relations of trade and intermarriage (Sankoff 1970). To what extent this practice had held four decades earlier I cannot affirm, but it is typical of the passive bilingualism found all over Papua New Guinea. Hogbin cites a specific instance from the Huon Gulf coast south of Lae that he observed in 1944: "[S]ome groups were able to understand but not speak the language of their neighbours. The southern villagers spoke in Gela to the Busama, for example, who replied in Gawa" (Hogbin 1951: 86). Extensive

fieldwork in dozens of languages led Wurm and Laycock to assert that "[p]assive bilingualism, i.e. acquired as contrasted with spontaneous, understanding of a form of speech differing from that of a given community by members of that community, is extremely widespread in New Guinea" (Wurm & Laycock 1961: 136).

From the beginning, Tok Pisin was the lingua franca of the goldfields of Wau and Bulolo, as was the case in the other extractive, commercial and agricultural operations in New Guinea. Some Buang men also joined the Royal Papua New Guinea Constabulary. Extensive travel in these occupations meant that when they retired to their home villages, they were fluent in Tok Pisin as well as in their native Buang *tok ples*. Wives of former policemen who had travelled with their husbands all over the country were the only older Headwaters Buang women in the 1960s to be fluent speakers of Tok Pisin, and they were the ones whose children were also fluent.

5.   World War II and beyond

As indicated on Map 3.1, the Snake River Valley is located high in the mountains about halfway between Bulolo and the coastal port of Lae, which by the end of the 1930s had become the second-largest town in New Guinea. The Lae airport, which had previously serviced the goldfields, was an important target for the Japanese, and it was destroyed by aerial bombing in January 1942. Subsequently, the Japanese sent troops to occupy strategic locations. In March of 1943 they landed at Salamaua (cf. Map 3.1), close to Busama Village on the South Huon coast (Hogbin 1951: 7). At least on this occasion, Tok Pisin was used as a medium of intercommunication between the Japanese and the local villagers: "A small detachment of [Japanese] military police, including an interpreter who spoke pidgin English, went around the countryside with orders that the village officials should attend at the headquarters to receive instructions" (Hogbin 1951: 7).

Landing in Wau eighteen months later, American and Australian forces fought their way back to Salamaua, recapturing it on 14 September 1943 (Hogbin 1951: 10). In later campaigns waged by the Allies, "a large number of the able-bodied men [from the Salamaua area] were … conscripted to act as carriers" (Hogbin 1951: 12). However, the closeness of the Headwaters Snake Valley to Lae (only about thirty kilometres, as the crow flies) gives a deceptive picture of its accessibility from there, since the high valley is ringed by mountains. Nowhere in the literature have I found any evidence that Buang people were involved in World War II. Indeed, I heard no World War II stories from villagers I spoke with in the 1960s and 1970s. Perhaps the relative seclusion provided by their mountainous homeland protected them from being conscripted by either side.

Beginning in 1945, the immediate post-war period saw the reconstruction of Lae, which had been destroyed in the heavy bombing. By 1948, Sinclair was able to travel from Lae by jeep to Wau, where he began his career as a *kiap* "patrol officer". The journey started with the loading of the vehicle onto a barge, and crossing of the wide mouth of the Markham River to the beginning of a scarcely passable road to the goldfields.

Buang men continued to be in demand over the next decades, and many signed on to work as carriers and cooks (Sinclair 1966, 1981).[14] In 1948, Sinclair reported being puzzled at the readiness of one hundred young Buang men to return on a patrol with him to the Upper Watut area where relatives and friends had been killed in the previous decade. He made inquiries to Tol, the Paramount *Luluai* at Mapos Village, as to why this was the case. Tol's reply is cited in an English that clearly reflects its Tok Pisin original. "Oh *kiap*," he quotes Tol as saying, "they are sorry for the Buang men who died in there in the good time" (Sinclair 1981: 75). Tol explained that since the spirits of the dead men would be lingering in that faraway place, a visit would be a way to provide comfort to them, and he would authorize the young men to "go with the government".[15]

My own inquiries in 1966–67 yielded an estimate that about half of the men then living in Mambump Village had spent some time away, usually for a year or two in their late teens or early twenties, working as carriers, as domestic servants, as policemen, or even (in more recent times) as truck drivers or mechanics. All of these men were bilingual in Tok Pisin. After a stint away, most had returned to take up subsistence farming on their own land. But the late 1960s also saw the expansion of the cash economy to rural areas. Under the auspices of the *didiman* "agricultural officer" with headquarters in Mumeng, more than a few villagers had begun to grow coffee as a cash crop ideally suited to their hillside climate. Thus in addition to their own food crops, Buang farmers were planting, processing and

---

14.   As a former *kiap* himself, Sinclair's own words make the point convincingly: "The Buang people were famous carriers. Slim and wiry, they could out-perform any of the Morobe District tribes and for many years had been sought after by prospectors and patrol officers" (Sinclair 1981: 74). Indeed, my own research among the Buang was a result of their being in good repute with Australian administrators. When my initial choice of a research site was refused by the *kiap* in charge of that area, and the second proved inaccessible, I was directed by another *kiap* to visit the Buang because of their reputation as welcoming of foreigners.

15.   Note the use by Tol of the term "good time" (*Gut Taim*) to refer to the pre-World War II era, and of "the government" to mean "the government officers/representatives". Translating Sinclair's citation back into Tok Pisin, Tol probably said something close to the following: " *O kiap, ol i sori long ol man bilong Buang ol idai long hap long Gut Taim. … ol yangpela man bilong mi bai ol i go wantaim gavman.*"

selling coffee to two cooperative societies: the Sake Society fostered by the *didiman* and the Lutheran-sponsored Namasu (Sankoff 1969).[16] At that time coffee was sold at the airstrip located near the Headwaters village of Wagau, and transported out via a single-engine airplane. Returning in 1971, I discovered that a road capable of handling vehicular traffic had linked the Headwaters area to the national road network, and that the Sake Society had purchased a truck. Prior to the completion of the road, Buang women in 1966 had begun to travel the considerable distance to sell produce at the large weekly market in Lae. Carrying up to about twenty kilograms of oranges in string bags (Tok Pisin, *bilum*) anchored on their foreheads, these women walked one of several routes before meeting a road where they could pay a fee to catch a ride for the rest of the journey in the back of a truck. Like the men walking to Bulolo some four decades earlier, they could follow the riverside track to Mumeng, where a truck ride north-east to Lae took them another eighty kilometres. Passengers on any given truck always hailed from a variety of linguistic backgrounds, and Tok Pisin was the normal language of intercommunication on these trips, as it was in the market itself. Alternatively, the women could follow the shorter, but very steep mountain track from the Snake River Headwaters descending to Gurakor Village, from which the distance to Lae by truck was only forty kilometres. In an age before rural electrification, many young village women were willing to undertake a long trek carrying heavy loads, and indeed looked forward to the bright lights of the town and the money they would earn.

Another post-war change brought a new language into the linguistic repertoire of Mambump Village. This was the installation in the Snake River Headwaters of the first resident missionary, a Lutheran pastor named Barnabas from Lababia, a coastal village south of Salamaua. Lutheran missionaries had arrived on the coast in the first decade of German colonization, and over the ensuing decades most coastal villagers became converts.[17] Outreach to Headwaters Buang people, however, did not take place until after World War II. And whereas elsewhere in New Guinea many other missions used Tok Pisin, the Lutherans in Morobe District (now Morobe Province) had adopted Yabem, an Austronesian language

---

16.   The Sake Society's name derives from the first person singular pronouns in Central and Headwaters Buang (*sa* in the former and *ke* in the latter). The s/k correspondence in the two dialects was well understood by most adults I discussed language with in Headwaters villages. In addition to *sa* and *ke* as 'I', they would tell me that "We say *ku* 'cry', they [the Central Buang people they referred to as *Vring*] say *su*; we say *vɔkev* 'yesterday', they say *vɔsevɔŋ*; we say *kes* 'hit', they say *sis*", etc.

17.   German Neuendettelsau missionaries "established their first station at Simbang [a village on the north Huon coast] in 1886 … and at Malolo, a Bukawa-speaking village on the south coast of the Gulf in 1907" (Bradshaw 2016: 62).

distantly related to Buang, as a lingua franca. When I met with Barnabas shortly after my arrival in 1966, we spoke in Tok Pisin, but Yabem was the language he used with people who had attended the primary school he had set up. It was also the language of services he conducted on Sunday mornings at the church built on land where local people had invited him to establish the church, the school and his own residence. By this time Barnabas had been a resident in the community since probably the early 1950s. In 1966, many grown men in their forties and fifties had become church leaders and were fluent in Yabem. This was the language that I observed being used extensively at a regional conference of the Evangelical Lutheran Church of New Guinea, hosted by Headwaters Buang people in 1967.[18] In the late 1960s, I also observed Yabem being used to a limited extent by Morobe area Lutherans as a lingua franca in non-religious contexts. This was usually treated more as a jovial way to exchange pleasantries and affirm social ties with co-religionists, with participants soon reverting to Tok Pisin when getting to the meat of a conversation. During the 1970s, with the advent of independence in Papua New Guinea, the renamed Evangelical Lutheran Church of Papua New Guinea reversed the language policy in favour of Tok Pisin, as can be observed on its website where Tok Pisin is used along with English.[19]

While conducting research on Tok Pisin in Lae in 1971, I met many Buang people who had taken up residence there. There was also a large Buang contingent residing in the capital, Port Moresby. For many urban children growing up at that time, Tok Pisin was the default choice for communication, not only outside the family but also with their own relatives. Most of them could also speak Buang, and many were learning English in school. Although most native Tok Pisin-speaking children recorded by Suzanne Laberge and myself at that time had parents who did not share a *tok ples* (Sankoff & Laberge 1973), rapid urbanization was in itself an important factor in producing Tok Pisin native speakers (Romaine 1989; Smith 2002). In fact urbanization is cited as the major impetus for the nativization of Tok Pisin's sister language, Solomon Islands Pijin (Jourdan & Keesing 1997).

For the Buang people I knew in Mambump Village, linguistic diversity was a given. Learning other speech varieties, whether dialects somewhat similar to their

---

18.   During the 1960s, missionary linguists were in residence farther down the Snake River Valley (Bruce and Joyce Hooley in Mapos Village, Central Buang; and Joan Healey and Roma Hardwick in Manga Village). Their presence did not alter the linguistic landscape since their goal was to use the Buang and Manga languages in evangelization. Following the Summer Institute of Linguistics policy of Bible translation in indigenous languages, both teams also provided literacy education in these languages.

19.   <http://www.elcpng.org/> (last access on 16 December 2020).

own, or quite different languages, was taken as natural. Language was closely tied to the notion of place, and when a person had dealings with people from another place, she naturally found herself learning their language. Though village exogamy was not the rule, there were in every village a few wives from other villages who natively spoke another language or dialect (and a few such husbands in some). No-one cited a language difference as a reason for not marrying into a different language or dialect.[20] Taking Mambump Village as a central compass point, people could supply names for every village within an approximately sixty-kilometre radius, and were able to tell me how long it would take, if you went there to live, to understand the language. Speaking it would naturally come later. I made a point of interviewing in-married women whenever I could, and they generally explained that someone (for example, a new sister-in-law) had sat them down and instructed them in some basic vocabulary. By and large, however, it was a case of language learning by immersion in daily life.

Including Manga, Central and Headwaters Buang (the latter two grouped together as one language with two dialects), Buang speakers have been numbered between 10,000 and 12,000 in recent decades. By this measure, the two Buang languages would be mid-sized languages in Papua New Guinea. Based on sources available in 1976, my investigation into the size of 629 languages (of perhaps 700 enumerated by that time) revealed only 9 with more than 30,000 speakers and 219 with fewer than 500 speakers (Sankoff 1977 [1981]). All reports I consulted on such very small languages mentioned bilingualism.

Discussing the two smallest members of the Ndu language family in the Sepik area, Laycock reports that "Ngala is spoken by the inhabitants of a single village, Swagup … [and] it is with the Wogamusin speaking inhabitants of Washkuk that the Swagup natives have most of their trading contacts … a large number of adult males in Swagup speak Wogumusin" (Laycock 1965: 131). Laycock further states that Wogamusin itself has only 336 speakers. The second very small language he mentions is Yelogu, which again "is spoken by the inhabitants of a single village. All 63 inhabitants appear to be bilingual, speaking the unrelated Kwoma language as well as Yelogu" (Laycock 1965: 139).

Some evidence points to speakers of very small languages choosing to shift to another language. Reporting on the languages spoken in the Finisterre Range, Claassen and McElhanon note that "the Mamaa language is spoken in the village of Mamaa (pop. 200) … the people are being assimilated by the Finugwan people

---

20.   Stanford (2009) describes the common practice of inter-linguistic marriage among Sui people in China, and Sorenson (1967) and Jackson (1974) report on compulsory linguistic exogamy in the Vaupes region of the Upper Amazon.

and most of them are bilingual" (Claassen & McElhanon 1970: 56). They give a population figure of 400 for Finugwan.

Though very small languages accounted for only 2.4% of the population (of a total of 1,904,954) that I included at that time (Sankoff 1980: 97, Table 5-1), it is safe to conclude that bilingualism occurs, virtually unremarked by the speakers, across many of the language boundaries between New Guinea's very numerous languages.

6.   Discussion

The history of language contact among the Buang prompts me to further consider two questions: (1) Is the assimilation of languages introduced during exploitation colonization the result of a world view that takes language diversity and multilingualism to be normal? and (2) What is the future of linguistic diversity in Papua New Guinea?

First, we may ask, how did the idea of a natural association between language and place accommodate the new languages of the colonial era? Yabem, which has receded in importance in recent decades, was always seen as legitimately belonging to Lutheran space, and to those who claimed it.[21] For its part, Tok Pisin was a language belonging to those spaces created by exploitation colonization, but the Buang people I talked with did not view it as *bubum ayez* "white people's language". Nor did other New Guineans refer to it as such. Spaces like plantations, goldfields and towns were indeed a product of colonization. However, New Guineans far outnumbered the colonists in every settlement, and they soon came to claim Tok Pisin as their own, disparaging the poor quality of the Tok Pisin of most expatriates (Mühlhäusler 1981). Indigenization (Mufwene 2015) began early in the history of Tok Pisin, and its embrace as the language of urban life has only enhanced its success.

Language and place are still tied together. After national independence in 1975, the people of Papua New Guinea have come to reclaim the entire country as their own, and in that country, Tok Pisin occupies a central place. With a total population exceeding 7 million, Papua New Guinea is listed as having 4,122,000 speakers of Tok Pisin, 4 million of whom are cited as being bilingual.[22] As the language

---

21.   In Headwaters Buang, the concept of Lutheran space was literal geographically, as it referenced land that had been provided by traditional landowners in several villages for the Lutheran church, school and pastor's home and food gardens.

22.   According to the National Census Office of Papua New Guinea, the most recent census (2011) established the population at 7,275,324. <https://nso.gov.pg> (dead link on 16 December 2020).

most widely spoken in the country, it is the "most frequently used language in Parliament and commerce" and the "de facto language of national identity".[23]

As in so many other countries around the world, English remains the language of the educated elite, spoken by only 1–2% of the population.[24] Is it a language of somebody's place? Yes, but despite its status as one of four official languages in Papua New Guinea, that place is far away.[25] Its role in Papua New Guinea may be somewhat larger than it is in countries lacking a colonial history of English, yet its situation closely resembles Mufwene's characterization:

> English is typically considered a foreign language in the relevant territories, such as Japan and South Korea, despite huge investments by an elite class of the latter country in the English education of their children. Its function as a lingua franca is primarily for communication with people of other nationalities, regardless of whether it is used inside or outside the country.   (Mufwene 2015: 10)

What about language loss in Papua New Guinea, often characterized as the most linguistically diverse country in the world? Including several references to Papua New Guinea, Mufwene appropriately refocuses the discussion of language loss and language shift on speakers and their adaptive choices. He reminds us that "languages do not kill languages; speakers do, precisely those who are expected to use them rather than those speaking the alternatives to which they shift" (Mufwene 2002: 376). In a later work he questions the assumption that "speakers will lose much of their identities if their languages and cultures either change or are replaced" (Mufwene 2015: 377), observing that speakers themselves may actually prefer new identities, and that the languages they shift to may incorporate important elements – both linguistic and cultural -- from the substrate influences resulting from the process. As described above, very small languages in pre-colonial Papua New Guinea appear to have been lost because their speakers adopted a language of their neighbours. Additionally, a present-day language may register a significant break from the ancestral version of that language, when language contact has radically transformed it (Bradshaw 2017; Thurston 1987). Such processes will no doubt continue, as speakers of these languages make practical choices according to their particular language ecology. Current trends show some diminution in linguistic diversity, but as documented in this chapter, traditional acceptance of

---

23. *Tok Pisin*. <https://www.ethnologue.com> (last access on 16 December 2020).

24. <https://en.wikipedia.org/wiki/Languages_of_Papua_New_Guinea> (last access on 16 December 2020).

25. In addition to Tok Pisin and English, the official languages of Papua New Guinea include Hiri Motu and Papua New Guinean Sign Language <https://en.wikipedia.org/wiki/Languages_of_Papua_New_Guinea> (last access on 16 December 2020).

bilingualism means that language choice in Papua New Guinea is not a zero-sum game. The steadily increasing number of Tok Pisin speakers does not automatically result in steadily decreasing numbers of speakers of other languages.

As discussed above, the vast majority of speakers of Tok Pisin, itself the majority language in the country, are bilingual. In rural areas, bilinguals probably speak at least one of the languages of their ancestors. In urban contexts, however, bilingualism for many may instead involve Tok Pisin and English.[26] Of the country's official languages, Hiri Motu runs a distinct third, with the number of its second language speakers listed at 120,000 for 1989. Moreover, it is noted as being in decline in favour of Tok Pisin and English.[27] Passive bilingualism, characterized as traditionally "extremely widespread" (Wurm & Laycock 1961: 136), is nowhere quantified, but as we have seen, Papua New Guineans have consistently treated multilingualism as unproblematic. It seems entirely likely that their choices going forward will result in linguistic diversity in Papua New Guinea for some time to come.

Acknowledgement

---

26.   With 13.1% of the population now living in urban areas, a percentage reported for 2017 as having been quite stable over the past decade (https://knoema.com/atlas/Papua-New-Guinea/Urban-population, (last access on 16 December 2020), it is likely that the percentage of English speakers, both bilingual and monolingual, will increase with every generation.

27.   Hiri Motu <https://en.wikipedia.org/wiki/Hiri_Motu> (last access on 16 December 2020).

References

Barclay, G. 1978. *A History of the Pacific from the Stone Age to the Present Day*. London: Sidgwick and Jackson.

Bradshaw, J. 2016. Papuan Kâte and Austronesian Jabêm: Long contact, little convergence. *Language & Linguistics in Melanesia* 34: 59–76.

Bradshaw, J. 2017. Evidence of contact between Binanderean and Oceanic languages. *Oceanic Linguistics* 56: 395–414.   https://doi.org/10.1353/ol.2017.0019

Cheyne, A. [1853]2017. *A Description of Islands in the Western Pacific Ocean*. Miami FL: Hard-Press. (Kindle edn).

Claassen, O. R. & McElhanon, K. A. 1970. Languages of the Finisterre Range, New Guinea. *Pacific Linguistics A*-23: 45–78.

Corris, P. 1968. "Blackbirding" in New Guinea waters, 1883-84: An episode in the Queensland labour trade. *Journal of Pacific History* 3: 85–105.
https://doi.org/10.1080/00223346808572126

Corris, P. 1973a. *Passage, Port and Plantation: A History of Solomon Islands Labour Migration 1870–1914*. Melbourne: Melbourne University Press.

Corris, P. 1973b. *Editor's Introduction to The South Sea Islanders and the Queensland Labour Trade. 1870–1914*, P. Corris (ed.), xix–xlii. Canberra: Australian National University Press.

Crowley, T. 1990. *Beach-La-Mar to Bislama: The Emergence of a National Language in Vanuatu*. Oxford: OUP.

Deery, C. 2014. A New Deal for the Territories? The Abolition of the Indentured Labour System in Colonial Papua and New Guinea. Honours dissertation, University of Melbourne.

Denoon, D. 1997a. Land, labour and independent development. In *The Cambridge History of the Pacific Islanders*, D. Denoon, S. Firth, Jocelyn Linnekin, M. Leleisea & K. Nero (eds), 152–184. Cambridge: CUP.   https://doi.org/10.1017/CHOL9780521441957

Denoon, D. 1997b. New economic orders: Land, labour and dependency. In *The Cambridge History of the Pacific Islanders*, D. Denoon, S. Firth, J. Linnekin, M. Leleisea & K. Nero (eds), 218–252. Cambridge: CUP.   https://doi.org/10.1017/CHOL9780521441957

Firth, S. 1997. Colonial intervention and the invention of the Native. In *The Cambridge History of the Pacific Islanders*, D. Denoon, S. Firth, J. Linnekin, M. Leleisea & K. Nero (eds), 253–288. Cambridge: CUP.   https://doi.org/10.1017/CHOL9780521441957.009

Hall, R. A., Jr. 1943. *Melanesian Pidgin English: Grammar, Texts, Vocabulary*. Baltimore MD: Linguistic Society of America.

Hogbin, I. 1951. *Transformation Scene: The Changing Culture of a New Guinea Village*. London: Routledge & Kegan Paul.

Jackson, J. 1974. Language identity of the Colombian Vaupés Indians. In *Explorations in the Ethnography of Speaking*, R. Bauman & J. Sherzer (eds), 50–64. Cambridge: CUP.

Jourdan, C. 1985. Sapos Iumi Mitim Iumi: The Social Context of Creolization in the Solomon Islands. PhD dissertation, Australian National University.

Jourdan, C., & Keesing, R. 1997. From Pisin to Pijin: Creolization in process in the Solomon Islands. *Language in Society* 26: 401–420.   https://doi.org/10.1017/S0047404500019527

Keesing, R. 1988. *Melanesian Pidgin and the Oceanic Substrate*. Stanford CA: Stanford University Press.

Laycock, D. C. 1965. *The Ndu Language Family. Pacific Linguistics, Series C, No.1*. Canberra: Research School of Pacific Studies.

May, R. J. 2003. Harmonizing linguistic diversity in Papua New Guinea. In *Fighting Words: Language Policy and Ethnic Relations in Asia*, M. E. Brown & S. Gamguly (eds), 293–318. Cambridge MA: The MIT Press.

Meyerhoff, M. 2000. *Constraints on Null Subjects in Bislama (Vanuatu): Social and Linguistic Factors*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.

Mihalic, F. 1971. *The Jacaranda Dictionary and Grammar of Melanesian Pidgin*. Madang, Papua New Guinea: Jacaranda Press.

Mosel, U. 1980. *Tolai and Tok Pisin: The influence of the substratum on the development of New Guinea Tok Pisin. Pacific Linguistics Series B 73: Complete issue*. Canberra: Australian National University.

Mosel, U. & Mühlhäusler, P. 1982. New evidence of a Samoan origin of New Guinea Tok Pisin (New Guinea Pidgin English). *Journal of Pacific History* 17: 166–175.
https://doi.org/10.1080/00223348208572445

Mufwene, S. S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP.
https://doi.org/10.1017/CBO9780511612862

Mufwene, S. S. 2002. Colonization, globalization and the plight of 'weak' languages. *Journal of Linguistics* 38: 375–395. (Review article, Nettle D. & Romaine, S. *Vanishing Voices: The Extinction of the World's Languages*. Oxford: OUP). https://doi.org/10.1017/S0022226702001391

Mufwene, S. S. 2008. *Language Evolution: Contact, Competition, and Change*. London: Continuum.

Mufwene, S. S. 2015. Colonization, indigenization, and the differential evolution of English: Some ecological perspectives. *World Englishes* 34: 6–21. https://doi.org/10.1111/weng.12129

Mühlhäusler, P. 1978. Samoan Plantation Pidgin English and the origin of New Guinea Pidgin. *Pacific Linguistics Series A* 54: 67–119.

Mühlhäusler, P. 1981. Foreigner talk: Tok Masta in New Guinea. *International Journal of the Sociology of Language* 28: 93–113. https://doi.org/10.1515/ijsl.1981.28.93

Mühlhäusler, P. 1985. The number of Pidgin Englishes in the Pacific. *Pacific Linguistics Papers in Pidgin and Creole Linguistics Series A* 72: 25–51.

New Guinea Annual Reports. 1941. *Permanent Mandates Commission Minutes Relating to Examination of Reports on New Guinea*. Report of a Commission to Inquire into the Matter of Native Labour in the Territory. <http://www.notapoormansfield.com/index.php/new-guinean-experience/nge-stats> (dead link on 16 December 2020).

O'Neill, J. 1979. *Up from South*. Melbourne: Oxford: OUP.

Papua New Guinea. 1973. *Papua New Guinea Village Directory*. Konedobu, Papua New Guinea: Department of the Chief Minister and Development Administration.

Pawley, A. & Hammarström, H. 2018. The Trans New Guinea family. In *The Languages and Linguistics of New Guinea: A Comprehensive Guide*, B. Palmer (ed.), 21–175. Berlin: Mouton de Gruyter.

Queensland. 1885. *Report of the Royal Commission on Recruiting Polynesian Labourers in New Guinea and Adjacent Islands*. Notes and Proceedings of the Legislative Assembly, Queensland, Vol. 2. 1885.

Reed, S. W. 1943. *The Making of Modern New Guinea*. Philadelphia PA: The American Philosophical Society.

Romaine, S. 1988. Some differences between spoken and written Tok Pisin in Papua New Guinea. *English World Wide* 9: 243–269. https://doi.org/10.1075/eww.9.2.07rom

Romaine, S. 1989. Tok Pisin and English in Papua New Guinea. In *World Englishes*, K. Watson-Gegeo (ed.). Special issue of the *South Pacific* 64: 1–19.

Salisbury, R. F. 1962. Notes on bilingualism and linguistic change in New Guinea. *Anthropological Linguistics* 4: 1–13.

Salisbury, R. F. 1970. *Vunamami: Economic Transformation in a Traditional Society*. Berkeley CA: University of California Press.

Sankoff, G. 1969. *Wok bisnis* and Namasu: A perspective from the village. *New Guinea Research Bulletin* 28: 61–81.

Sankoff, G. 1970. Mutual intelligibility, bilingualism and linguistic boundaries. *International Days of Sociolinguistics*, 839–848. Rome: Istituto Luigi Sturzo. (Republished in *The Social Life of Language*, G. Sankoff, 1980, 133–141. Philadelphia PA: University of Pennsylvania Press).

Sankoff, G. 1977. Multilingualism in Papua New Guinea. In *New Guinea Area Languages and Language Study, Vol. 3: Language, Culture, Society and the Modern World. Pacific Linguistics*, S. A. Wurm (ed.), C 40: 267–307. (Revised including updated statistics in *The Social Life of Language*, G. Sankoff, 1980, 95–132. Philadelphia PA: University of Pennsylvania Press).

Sankoff, G. 1980. Multilingualism in Papua New Guinea. In *The Social Life of Language*, G. Sankoff, 95–132. Philadelphia PA: University of Pennsylvania Press. https://doi.org/10.9783/9781512809589

Sankoff, G. 1985. Contracts in Queensland: Touching pen and marking paper. In *History and Ethnohistory in Papua New Guinea. Oceania Monograph 28*, D. Gewertz & E. Schieffelin (eds), 100–126. Sydney: University of Sydney.

Sankoff, G. 1993. Focus in Tok Pisin. In *Focus and Grammatical Relations in Creole Languages* [Creole Language Library 12], F. Byrne & D. Winford (eds), 117–140. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.12.10san

Sankoff, G. 1996. The Oceanic substrate in Melanesian Pidgin/Creole revisited: A tribute to Roger Keesing. *Pacific Linguistics Series C* 133: 421–250.

Sankoff, G. & Laberge, S. 1973. On the acquisition of native speakers by a language. *Kivung* 6: 32–47.

Scarr, D. 1967. Recruits and recruiters: A portrait of the Pacific Islands labour trade. *Journal of Pacific History* 2: 5–24. https://doi.org/10.1080/00223346708572099

Scarr, D. 1990. *The History of the Pacific Islands: Kingdoms of the Reefs*. Melbourne: Macmillan Company of Australia. https://doi.org/10.1007/978-1-349-15138-7

Shineberg, D. 1967. *They Came for Sandalwood: A Study of the Sandalwood Trade in the South-West Pacific, 1830-1865*. Carlton: Melbourne University Press.

Siegel, J. 1985. Origins of Pacific Island labourers in Fiji. *Journal of Pacific History* 20: 42–54. https://doi.org/10.1080/00223348508572504

Sinclair, J. B. 1966. *Behind the Ranges Patrolling in Papua New Guinea*. Melbourne: Pacific Publications.

Sinclair, J. B. 1979. *Foreword*. In *Up from South*, J. O'Neill, vii–xxiii. Melbourne: OUP.

Sinclair, J. B. 1981. *Kiap: Australia's Patrol Officers in Papua New Guinea*. Sydney: Melbourne University Press.

Smith, G. 2002. *Growing up with Tok Pisin*. London: Battlebridge.

Sorensen, A. 1967. Multilingualism in the North West Amazon. *American Anthropologist* 69: 670–684. https://doi.org/10.1525/aa.1967.69.6.02a00030

Stanford, J. 2009. "Eating the food of our place": Sociolinguistic loyalties in multidialectal Sui villages. *Language in Society* 38: 287–310. https://doi.org/10.1017/S0047404509090502

Thurston, W. R. 1987. *Processes of Change in the Languages of North-Western New Britain*. Canberra: Research School of Pacific Studies, Australian National University. *Pacific Linguistics*, Series B. 99: Complete issue.

Village Directory, 1960. *Territory of Papua New Guinea: Department of Native Affairs*. Port Moresby: W.S. Nicholas, Government Printer.

Wagner, H. 1987. Beginnings at Finschhafen: The Neuendettelsau Mission jointly with the Australian Lutheran Church. In *The Lutheran Church in Papua New Guinea: The First Hundred Years, 1886–1986*, H. Wagner & H. Reiner (eds), 31–98. Adelaide: Lutheran Publishing House.

Waterhouse, M. 2010. *Not a Poor Man's Field: The New Guinea Goldfields to 1942: An Australian Colonial History*. Canberra: Halstead Press.

West, R. J. 1966. The historical background. In *New Guinea on the Threshold*, E. K. Fisk (ed.), 3–19. Canberra: Australian National University Press.

Wurm, S. A. & Laycock, D. C. 1961. The question of language and dialect in New Guinea. *Oceania* 32: 128–143. https://doi.org/10.1002/j.1834-4461.1961.tb01747.x

CHAPTER 4

Conventionalized creativity in the emergence of a mixed language – A case study of Light Warlpiri

Carmel O'Shannessy

ANU College of Arts and Social Sciences

Questions of how contact languages are best categorized and how their paths of development are most accurately described remain contentious. Mixed languages or split languages raise many unanswered questions about mechanisms of change in multilingual contexts, and how they lead to specific structural outcomes. An examination of the structure of Light Warlpiri, a mixed language spoken in a Warlpiri community in northern Australia, shows that the main mechanism of change is insertional code-switching where a verbal phrase from one language, Kriol (an English-lexified Creole) is inserted into a string of another language, Warlpiri. The result is a way of speaking that is mixed at every level, and can be categorized as underlyingly Warlpiri with innovative insertions and reanalyses from Kriol.

Keywords: Warlpiri, Light Warlpiri, mixed language, Creole, code-switching,

1. Introduction

Light Warlpiri is a newly emerged way of speaking in one Warlpiri community in Australia's Northern Territory, that has developed in approximately the last forty years (O'Shannessy 2005). The name represents the term older Warlpiri use to refer to the way younger people speak, *Warlpiri rampaku* "Warlpiri light", meaning Warlpiri with some changes (O'Shannessy 2012). It emerged in a two-stage process from code-switching practices in a specific pattern when adults were talking to young children in a baby talk register, that is, a register that adults use when talking to babies and very young children (O'Shannessy 2012). As that cohort of children grew up, they internalized the code-switching as a single linguistic system and added innovations in the verbal structure. They retained this way of speaking

into adulthood and transmitted it to their children. The new way of speaking is now the primary way of interacting among adults under about the age of forty and everyone younger than that in the community. Light Warlpiri speakers also learn and speak Warlpiri, and varieties of English and Kriol, and have some knowledge of other languages and varieties spoken in the area.

The terms "Warlpiri" and "classic Warlpiri" in this chapter refer to Warlpiri, a Ngumpin-Yapa language, as described in the literature (Hale et al. 1995; Nash 1986; Simpson 1991). The term Kriol refers to an English-lexified Creole spoken in the north of Australia (Sandefur 1979; Schultze-Berndt et al. 2013), and the term "varieties of English" refers to Standard Australian English and Aboriginal varieties of English (Butcher 2008; Eades 1993; Harkins 1994; Malcolm & Kaldor 1991).

Most of the nominal structure of Light Warlpiri is drawn from Warlpiri, in a very consistent and systematic manner, as illustrated in Section 3.2 below (O'Shannessy 2005, 2013). Much of the verbal structure is drawn from Kriol, and from varieties of English. But importantly, there are also innovations in the verbal structure, drawing on all of the sources of Light Warlpiri (O'Shannessy 2013). The term creativity in this chapter refers to processes of re-analysis and innovation on the part of speakers, as described throughout.

In this chapter, I argue that the structure of Light Warlpiri is thoroughly mixed, according to the criteria of Myers-Scotton (2003), yet its structure is nevertheless underlyingly Warlpiri. I also argue that perhaps all that is needed to form a mixed language is conventionalized insertional code-switching, where a verbal phrase from one language, Kriol (an English-lexified Creole) is inserted into a string of another language, Warlpiri (cf. Auer 1999, 2014; Myers-Scotton 2003).

The structure of the chapter is as follows. I first briefly outline two theories of how mixed languages develop from code-switching practices, those of Myers-Scotton (2003) and Auer (1999, 2014). In Section 3, relevant properties of Light Warlpiri are presented. In Section 4, I show how Light Warlpiri aligns with the theories of Myers-Scotton (2003) and Auer (1999), and how the data illustrate that Light Warlpiri is both thoroughly mixed and underlyingly Warlpiri. The conclusion is in Section 5.

2.   Theories about mechanisms of a mixed language emerging from code-switching practices

Two prominent theories of how mixed languages emerge from code-switching practices are those of Myers-Scotton (2003), and Auer (1999, 2014). These two theories are presented here because they see the possibilities of new languages arising from code-switching practices, yet in different ways. For some time such

a transition was hypothesized by several researchers without clear empirical evidence of it occurring (Auer 1999: 324). But more recently empirical evidence for such a sequence has been obtained (McConvell & Meakins 2005; O'Shannessy 2012, 2013).

Auer's (1999) theory sees a mixed language emerging from code-switching practices. Auer illustrates a process of structural sedimentation where a community of speakers moves from code-switching to a fused lect. In this process speakers first code-switch between languages, and the code-switching is socially meaningful in that the choice of one language or another indexes, for instance, discourse or social information. The speakers gradually conventionalize the code-switching patterns until the language contrasts are less socially meaningful within a conversation, but are meaningful in a wider sense, for example of signalling group identity (see Croft, this volume, on neogeny and mixed languages). The juxtaposition of the languages may involve insertions of elements of one language into another, or alternations of phrases or clauses in each language, and at this point is termed language mixing. Over time the way of speaking conventionalizes further, and the choice of elements from each language becomes less fluid and variable, as functional specialization develops. This means that a specific grammatical function is always performed by elements from a specific language. For example, in Light Warlpiri both core and peripheral grammatical functions (e.g., ergative, dative, locative, allative, possessive [to some extent]) employ Warlpiri case-markers. At this point the way of speaking that began as code-switching has become a fused lect.

Corroborating this, in discussing the emergence of Mednij Aleut, a combination of Aleut and Russian spoken in the Bering Strait, Golovko (1994) suggests that "codemixing is a necessary but not sufficient condition for the emergence of a mixed language" (Auer 1999; Golovko 1994: 118). Golovko points to social factors being essential conditions. In the case of Mednij Aleut these are adult aspirations of a new social identity, with code-switching being an ingroup code.

Myers-Scotton (1988, 1993, 2000, 2002, 2003, 2004) developed a structural model of code-switching that lays a foundation for how the combinations of elements from two languages play out in language contact phenomena, including code-switching practices and mixed languages, which Myers-Scotton prefers to call split languages (Myers-Scotton 2003: 73). I will explain the essential concepts in the theory through the structure of classic code-switching, and then turn to composite code-switching, which is more relevant to Light Warlpiri. In the Matrix Language Frame and 4-M model (referring to four types of morpheme) of classic code-switching (Myers-Scotton 2003, 2004), there is a consistent asymmetry in the structure provided by each of the two languages involved. Although surface morphemes are drawn from both languages, only one language provides the

underlying morphosyntactic frame of a clause, and this language is the matrix language (Myers-Scotton 2004: 106). The other language is the embedded language and elements from it are inserted into the matrix language, or can occur as embedded islands with full morphology, but embedded into a larger matrix language frame. In this model, the matrix language supplies the order of morphemes (with surface forms sourced from either language), and also supplies specific types of grammatical morpheme, called late outsider system morphemes. Morphemes are categorized as content morphemes, which receive or assign thematic roles, e.g., nouns and verbs, and system morphemes, which do not. System morphemes are categorized as early system morphemes, which provide conceptual information to their head, such as English determiners that provide information on specificity, or late system morphemes. Late system morphemes signal grammatical information and are of two types, late bridge and late outsider morphemes. Late bridge morphemes, such as English *of*, expressing possession or association, "connect content morphemes with each other or integrate larger structures" (Myers-Scotton 2004: 111). Late outsider morphemes co-index larger constituents, and their form depends on information from outside their immediate maximal projection, for example, subject agreement markers on verbs, or nominal case-markers.

Myers-Scotton (2003: 88) argues that in bilingual contexts there can be a matrix language turnover, in which the former embedded language gradually becomes the matrix language, as "the morphosyntactic frame changes from that derived from one language to that derived from the other" (Myers-Scotton 2003: 88). Importantly, in this theory a turnover can be arrested midway, in which case the matrix language only partly changes, so the way of speaking then has abstract grammatical material from both languages, or a composite matrix language frame. In Myers-Scotton's terms this is the scenario for the emergence of a split language. A context for a matrix language turnover not being completed would be when the socially dominant language in a community shifts from one language to the other. If there has been an arrested matrix language turnover, the result is that in at least one component of the morphosyntactic frame, structure is sourced from both contributing languages, and this is Myers-Scotton's definition of a split language. Myers-Scotton (2003) shows how the abstract material from different languages can be present in different components of a word or phrase. By abstract material she means lexical-conceptual structure, predicate-argument structure, and morphological realization (Myers-Scotton 2003: 85). Different combinations of abstract material at the three levels result in different surface forms.

The matrix language turnover model of Myers-Scotton (2003) is invoked as an explanation of another mixed language emerging from code-switching practices, Gurindji Kriol, in an Aboriginal community only 107 kilometres from the site of the study reported on in this chapter (McConvell & Meakins 2005). Gurindji

Kriol is the result of conventionalized code-switching between Gurindji and Kriol by adults. The next generations speak the mixed language as their primary language, and have passive knowledge of the traditional language, Gurindji (McConvell & Meakins 2005: 11). Some changes in grammatical structure have also taken place. The ergative marker has become a nominative marker (Meakins 2015), a serial verb construction has developed (Meakins 2010), and there have been other changes to the nominal case-marking system (Meakins 2011).

Both Auer's (1999, 2014) and Myers-Scotton's (2003) theories involve code-switching practices preceding the emergence of a split or mixed language and offer structural mechanisms for the process. Myers-Scotton's relies on the notion of arrested matrix language turnover, whereas Auer's relies on conventionalization of code-switching practices, which is at least partly socially motivated. In keeping with Golovko's (1994) notion, the structural sedimentation occurs for social reasons of signalling group identity and ingroup practices.

3.   Structure of Light Warlpiri

The structure of Light Warlpiri is drawn from all of its sources, yet is underlyingly Warlpiri. In this section, I illustrate the relevant structures of Light Warlpiri, and show where they are and are not retained from Warlpiri. In the examples, elements drawn from Warlpiri are in italics and elements drawn from English and Kriol are in roman font.

3.1   Verbal structure of Light Warlpiri

The structure of Light Warlpiri sentences is illustrated in Examples (1), (2) and (3).

(1)   *Ngalipa   jalang* wi-m   kam   ka-*kurl   nyampu-kurra*
       1PL.INCL today 1PL-NFUT come car-COM here-ALL

       ngurra-kurra.
       home-ALL

       "Today we came in the car here, to your place."   (ElicitLA21_AC58_2014)[1]

(2)   Yu-m   pud-im-on   *mayi nyampu ngaju-nyang?*
       yu-NFUT   put-TR-LOC   Q   DET   1SG-POSS
       "Did you put this one, mine, on?"   (C06_14)

---

(3) Botul-*ing*   i-m   *panturn*-um taya.
     bottle-ERG 3SG-NFUT pierce-TR   tyre
     "A bottle pierced the tyre."   (ERGstoryLC39_2010)

Examples (1) to (3) illustrate functional specialization in Light Warlpiri in terms of which source language supplies the grammatical and surface form material for which functions. Most of the nominal structure is from Warlpiri, in the form of case-marking, seen in (1) in the case-markers *kurl* "COMITATIVE", and *-kurra* "ALLATIVE", and in (3) *-ing* "ERGATIVE" (from Warlpiri *-ngki* "ERGATIVE"). Much of the verbal structure is from Kriol, seen in (1) with *kam* "come", and in (2) with *pud-im-on* "put-TRANSITIVE-LOCATIVE". In Kriol most transitive verbs take a transitive suffix, especially those that are higher in transitivity (Batchelor 2017; Sandefur 1979; Schultze-Berndt et al. 2013). In (1), *Ngalipa* "1 PL.INCLUSIVE" is a free pronoun retained from Warlpiri. It is cross-referenced by *wi* "1PL" in an innovative auxiliary complex, explained below (and see also example 42).

The verbal structure, however, is complex, both in the lexical verbs and the auxiliary verb structure. First, although verbs from Kriol are by far the most common type of verb in Light Warlpiri, verbs derived from Warlpiri also occur, as in (3). Importantly, the Warlpiri-derived verb has been re-analyzed so that it takes a Kriol transitive suffix, *-im* (with at least one allomorph, *-um*) "TRANSITIVE". In Warlpiri the verb structure is *pantu-rnu* "pierce-PAST". But in Light Warlpiri the final vowel is omitted and the stem re-analyzed to include part of the past tense inflection, and a Kriol transitive suffix is attached to the re-analyzed stem (Meakins & O'Shannessy 2012). Warlpiri verb stems in Light Warlpiri can include complex verbs, as in examples (4) and (5). It is so far unclear whether the coverbal category is productive. In Warlpiri, complex verbs consist of a coverb and an inflecting verb, and most coverbs cannot occur independently of an inflecting verb and do not take inflections themselves (Nash 1982). The entire complex verb forms a single phonological phrase (Meakins & O'Shannessy 2012; Pentland & Laughren 2005). In Light Warlpiri the complex verbs in (4) and (5) are re-analyzed to fit Light Warlpiri verbal structure, for instance the Kriol transitive suffix occurs on the Warlpiri re-analyzed verb stem in (4).

(4) Maja-wan   *inya* i-m   *kurnta-ngarrirn*-im.
     mother-NMLZ DEM 3SG-NFUT shame-tell-TR
     "That mother there she told him off."   (ERGstoryLA92_2015)

All Warlpiri-derived verbs that occur in Light Warlpiri are re-analyzed to fit a Light Warlpiri verbal frame, but not all transitive verbs take the Kriol transitive marker, as in (5).

(5) Ah a-m   *kari-nyan*   yu *wiyarrpa*.
     DIS 1SG-NFUT not.recognize 2SG AFFECT
     "Ah, I didn't recognize you, dear one."   (ElicitLA21_AC58_2015)

In example (4) the verb *kurnta-ngarrirn-im* "shame-tell-TRANSITIVE" takes a Kriol-derived transitive suffix, but the verb *kari-nyan* "not.recognize" in (5) does not. In Warlpiri the verb *kari-nyanyi* "not.recognize" selects an ergative-absolutive case array, meaning that it is a transitive verb, but in this Light Warlpiri example it does not take a Kriol-derived transitive suffix. The verbal transitive suffix in Light Warlpiri is variably applied, and in the absence of a thorough investigation of the variation I hypothesize that verbs of lower transitivity are marked less often. Whether there are restrictions on Warlpiri-derived verbs that can occur in Light Warlpiri is not clear, with more forms being documented as more data are collected.

The constructions *wi-m* "1PL- NONFUTURE" and *yu-m* "2SG-NONFUTURE" in (1) and (2) are an auxiliary verb structure, formed from Kriol and English word shapes but with altered grammatical structure and semantics. The auxiliary structure is an innovation in Light Warlpiri that draws on English, Kriol and Warlpiri, but differs from all of them (O'Shannessy 2013). Kriol and English contribute the word shape of the pronoun form, *wi/yu* "1PL/2SG", and in English an auxiliary verb could be attached (we're, we'll, you're, you'll). The form *-m* occurs in English as a contracted auxiliary form, as in *I'm*. The semantics of *-m* in English *I'm* is 1SG present tense. A bilabial nasal occurs word-finally in English pronouns *him* and *them*, and in Kriol pronouns *im* "3SG" (< him) and *dem* "3PL" (< them), but is simply part of the pronoun, without any independent meaning. In hearing English and Kriol, speakers heard forms such as those given in Table 4.1.

Table 4.1 Pronominal forms in English and Kriol that influenced the verbal auxiliary in Light Warlpiri

| Row number | Person and number | Auxiliary form | Derived from | | Sample verb form |
			English	Kriol	
1	1SG	I'm	√		going "go-PROG"
2	3SG	im		√	faindim "find-TRANS"
3	3SG	im		√	lafing "laugh-PROG"
4	3PL	dem		√	gedim "get-TRANS"

The pronoun forms in rows 2–4 of Table 4.1 were re-analyzed to take on the structure in row 1, that of a pronoun-plus-TMA element, rather than a lexical pronoun. The result was the new structure, seen in Table 4.2. Once there was re-analysis of the bilabial nasal, *-m*, as a separate meaningful morpheme, it was available to be part of a regularized paradigm and to attach to other pronouns, specifically *yu-* and *wi-*, as shown in rows 5 and 6, creating the innovative structure.

Table 4.2 Pronominal forms and structure in Light Warlpiri

Row number	Person and number	Form in English or Kriol	New structure in Light Warlpiri		Sample verb form
			Pronoun	TMA element	
1	1SG	I'm	a	m	going "go-PROG"
2	3SG	im	i	m	faind-im "find-TRANS"
3	3SG	im	i	m	laf-ing "laugh-PROG"
4	3PL	dem	de	m	ged-im "get-TRANS"
5	1PL	wi	wi	m	ged-im "get-TRANS"
6	2SG	yu	yu	m	ged-im "get-TRANS"

Part of the motivation for the re-analysis of Kriol pronouns to a pronoun-plus-TMA element would have been that Warlpiri verbal structure has the form TMA-plus-pronominal clitic, as in example (6).

(6)   *Ngajulu-rlu   ka-rna-ngku   nyuntu   nya-nyi.*
       1SG-ERG   PRES-1SGS-2SG.NS   2SG   see-NPST
       "I see you."   (Hale et al. 1995: 1430, adapted)

The TMA-plus-pronominal clitic structure, *ka-rna-ngku* "PRES-1SGS-2SG.Non-Subject" is entirely consistent in Warlpiri. In (6), *ngajulu* "1SG" is a free pronoun, cross-referenced by the pronominal element in the auxiliary, *-rna* "1SG". Note that in Warlpiri the auxiliary carries information for both subject and object but in Light Warlpiri it only carries information for subject. In past perfective clauses in Warlpiri the TMA element is null and the pronominal element attaches to the first constituent of the clause, as in example (7).

(7)   *Wati-ø-li   ya-nu.*
       man-PERF-3PL.S   go-PST
       "The men left."   (Hale et al. 1995: 1435, adapted)

Warlpiri also motivates the reasons that the new Light Warlpiri TMA element *-m* would take on the semantics of non-future (present or past) rather than only present, as in English and Kriol. In Kriol there is a dedicated past tense morpheme, *bin* "PAST", an independent word that occurs pre-verbally (Sandefur 1979; Schultze-Berndt et al. 2013). But this morpheme only occasionally occurs in Light Warlpiri. Instead, the semantics of *bin* "PAST" has been combined with the present tense meaning of English *m*, resulting in the meaning of present and/or past, or, NON-FUTURE. Additionally, the semantics of the new auxiliary system in Light Warlpiri map onto the semantics of Warlpiri combinations of auxiliary and verb, as shown in Table 4.3 (Laughren 2013), although the structures differ.

Table 4.3 Equivalence between Warlpiri auxiliary+verbal inflection and Light Warlpiri auxiliary, adapted from Laughren (2013: 8)

Auxiliary base	Verb form	Meaning	Factivity	Light Warlpiri auxiliary	
lpa	PAST	past imperfective	actualized (factive)	*-m*	"non-future"
ø	PAST	past perfective			
ka	NONPAST	non-past			
ø	NONPAST	permissive/directive; future prediction under negative operator	unactualized (future/ potential)	*-rra*	"future"
lpa	IRREALIS	unrealized in present	unactualized (irrealis)	*-na*	"desiderative"
ø	IRREALIS	unrealized in past			

More details of the Light Warlpiri auxiliary system are given in Table 4.4.

Table 4.4 Detail of the Light Warlpiri verbal auxiliary system, using 1SG as the pronoun form

Row number		1SG	1PL	2SG	3SG	3DU/PL	2DUAL
1	future	a-l	wi-l	yu-l	i-l	de-l	yutu/yumob garra
2	future/should	a-rra	wi-rra	yu-rra	i-rra	de-rra	yutu/yumob garra
3	might/should	a-rra	wi-rra	yu-rra yu mada	i-rra	de-rra	yutu/yumob beta / mada
4	might/should much less often	a gada	wi gada	yu gada	i gada	dei gada	yutu/yumob gada
5	want to	a-na	wi-na	yu-na	i-na	de-na	yutu/yumob wana
6	non-future (i.e., present or past)	a-m	wi-m	yu-m	i-m	de-m	yutu/yumob / yutu/yumob bin
7	past completed much less often	a bin	wi bin	yu bin	i bin	dei bin	yutu/yumob bin
8	past progressive	a was	wi was	yu was	i was	dei was	?
9	might	a mait	wi mait	yu mait	i mait	dei mait	yutu/yumob mait
10	past negative	a neban	wi neban	yu neban	i neban	dei neban	?
11	present negative	a neba	wi neba	yu neba	i neba	dei neba	yutu/yumob neba
12	future negative	a won	wi won	yu won	i won	dei won	yutu/yumob won

In Table 4.4 it is clear that the auxiliary pronoun-plus-TMA element pattern is very consistent in rows 1–3 and 5–6, but does not occur for all aspectual constructions, several of which involve a separate modal word, drawn from both English

and Kriol. It is also clear that the cliticized forms do not occur on second person dual and plural pronouns, seen in the final column. The forms in rows 4 and 7 are Kriol forms that occur occasionally in Light Warlpiri.

The form *neban* "NEGATIVE.PAST" is another innovation in the Light Warlpiri verbal auxiliary system. The word form is a combination of Kriol *neba* "NEGATIVE", and *bin* "PAST". When "*neba bin*" is pronounced quickly, syllable elision occurs and the bilabial stop consonant is only pronounced once, resulting in *neban*, the new conventionalized form, as in example (8).

(8) Yu   neban   go *mayi* Satadei   yu-m   jeinj-im   *mayi.*
     2SG NEG.PST go Q   Saturday 2SG-NFUT change-TR   Q
     "You didn't go on Saturday, did you? You changed it, did you?"
                                                          (ElicitLA21_AC58_ 2014)

Another innovation in the Light Warlpiri verbal auxiliary system is in row 3 of Table 4.4, *mada* "SHOULD". This is presumably a contraction of *yumob garra* "2PL SHOULD" from Kriol, conventionalized as *yu mada*. An example of *mada* "SHOULD" is given in (9).

(9) Yu   mada   sidan   prog-progi-*kijak.*
     2SG should sit.down   frog-REDUP-EVI
     "You should sit down, in case the frog gets you."

The forms in row 3 of Table 4.4 are attested in Kriol, and those in row 4 are attested in the neighbouring mixed language, Gurindji Kriol (Meakins 2011). But those in row 6, critical to the analysis of the auxiliary, are only attested in the region in Light Warlpiri, as are the *neban* (row 10) and *mada* forms (row 3).

In Light Warlpiri, free pronouns from Warlpiri occur for emphasis, as in (1), or clarification, as in (10), and are cross-referenced with the subject pronoun of the auxiliary.[2]

(10) Wi-m   go *ngalijarra*   wi-m   teik-im *kurdu-kurdu.*
      1PL-NFUT go 1DUAL.INCL   1PL-NFUT   take-TR child-REDUP
      "We two went, we took the children."   (ElicitLA21_AC58_ 2014)

In (10) the Warlpiri free pronoun *ngalijarra* "1DU.INCL" indicates the scope of the auxiliary pronoun *wi* "1PL".

---

2.   However, a similar re-analysis has taken place independently in the English-lexified Yarrie Lingo (Angelo et al. 2019), spoken in Queensland, but with slightly different semantics. In Yarri Lingo the *-m* element indicates past tense only. Speakers of the two varieties live about 3,000 kilometres apart and would rarely interact, so there is unlikely to be any influence of speaker interaction.

A further similarity to Warlpiri is that in both Warlpiri and Light Warlpiri, tense and aspect readings are realized through combinations of verb and auxiliary morphology. For example, the auxiliary future *-rra* combines with the absence of transitive marking on a verb to indicate an irrealis potential threat or warning, as in (11).

(11) I-rra   bait   yu   dat   *jurlpu.*
      3SG-FUT   bite-Ø you that bird
      "That bird might bite you."   (O'Shannessy 2005: 41)

3.2   Nominal structure of Light Warlpiri

The nominal structure of Light Warlpiri is almost entirely from Warlpiri. Core arguments of verbs are mostly indicated by Warlpiri case-marking in ergative, absolutive (null marking) and dative functions. Dixon's (1979) distinctions of A (transitive subject), S (intransitive subject) and O (transitive object) are used to discuss these.

Ergative case-marking

A arguments are optionally ergatively marked, as ergative marking occurs in 85% of them. In (12) the A argument takes an ergative suffix, but in (13) it does not.

(12) *Watiya-ng* i-m   *katirn*-im.
      tree-ERG   3SG-NFUT squash-TR
      "The tree squashed it."   (ERGstoryLC17_2010)

(13) *Nyampu-ju* laitning   na   i-m   straik-im dat lil   boi.
      DET-TOP   lightning FOC 3SG-NFUT strike-TR that little boy
      "Here lightning struck the little boy."   (ERGstoryLA59_2015)

The forms of the ergative marker in Light Warlpiri are reductions of allomorphs that occur in classic Warlpiri. The reduced forms occur in contemporary Warlpiri alongside other allomorphs, but in Light Warlpiri regularization has taken place so that there are now two allomorphs almost exclusively, *-ng* "ERG" on vowel-final stems and *-ing* "ERG" on consonant-final stems (O'Shannessy 2016b). As Light Warlpiri has emerged, the amount and conditioning of ergative case marking has evolved. In earlier data (O'Shannessy 2005) ergative case-marking occurred on approximately 59% of A arguments, but in later data (O'Shannessy 2016a) it occurs more often, on approximately 85% of A arguments. When the word order is not SVO, ergative marking occurs on 95% of A arguments (O'Shannessy 2016b), as in example (14), so it is largely conditioned by word order, but to some extent also by the prominence and agentivity of the A argument (Meakins & O'Shannessy 2010).

(14)   ... an *jinta-kari wirliya-nga* i-m   puk-um *jilkarla-ng.*
        CONJ one-other leg-LOC   3SG-NFUT poke-TR thorn-ERG
        "And the thorn pierced the other one on the leg."   (ERGstoryA57_2008)

As in Warlpiri, a multi-word A argument can take ergative marking on the final
word of the phrase, as in (15), or on each word of the phrase, as in (16).

(15)   *Karnta-jarra inya-jarra-ng* de-m   jeis-im   bihain.
        woman-DUAL   DEM-DUAL-ERG 3PL-NFUT chase-TR behind
        "The two women there chase it, they're behind it."   (ERGstoryLA92_2015)

(16)   *Nyampu* dey shoot-ing-it   uuju *manu* puluk *karnta-ng*
        DEM   3PL shoot-PROG-TR horse and   cow   woman-ERG

        wati-ng   an *kurdu-ng.*
        man-ERG and child-ERG

        "Here the woman, man and child are shooting the horse and cow."

                                                        (ERGstoryLAC58_2015)

Ergative marking in Light Warlpiri also occurs on non-core arguments in transitive
clauses, as it does in Warlpiri. In examples (17) and (18) it occurs on instrument NPs.

(17)   I-m   *pantirn*-im naif-*kurlu-ng.*
        3SG-NFUT pierce-TR   knife-COM-ERG
        "He pierced it with a knife."   (C03_C04_3_3)

(18)   De-m   fix-im-bat   kurupa-*kurlu-ng.*
        3PL-NFUT fix-TR-ITER crow.bar-COM-ERG
        "They fixed it using a crowbar."   (ERGstoryLC09_2015)

In addition, as in Warlpiri, instruments may be marked with only the comitative
suffix, as in (19).

(19)   Raiful-*kurl* i-m   get-im dat *nganayi*   krakadail.
        rifle-COM   3SG-NFUT get-TR DEM you.know crocodile
        "He got the you know, crocodile, using a rifle."   (ERGstoryLA62_2008)

Again, as in Warlpiri, NPs indicating the time that a transitive event takes place
may be ergative-marked as in (20).

(20)   *Ngajarra*   wi-rra   do-im *jukurra-ng*   rekoding *kuja mayi*?
        1DUAL.EXCL 1PL-FUT do-TR tomorrow-ERG recording thus Q
        "We'll (not you) do the recording tomorrow, like this?"   (ElicitLAC58 _2014)

Also, as in Warlpiri, NPs that indicate manner can be ergative-marked, as in (21).

(21)   I-m   puk-um *kuja-ng* *kilji-ng*   *nganayi-ng*   jap-wan-*ing.*
        3SG-NFUT poke-TR thus-ERG hard-ERG you.know-ERG sharp-NMLZ-ERG
        "It poked it like this, hard, the you know, the sharp thing did."

                                                        (ERGstoryLC09_2005)

Example (22) is provided to show that it is not the case that the suffix on these NPs has lexicalized in Light Warlpiri, as the same elements occur in intransitive clauses without ergative marking.

(22) *Ngapa* i-m   fal-dan   *kilji-nyayirni.*
      rain   3SG-NFUT fall-down hard-INTENS
      "It rained really heavily."   (ERGstoryLAC58_2015)

Ergative marking in Light Warlpiri differs from that in Warlpiri in not marking adjuncts of location. Example (23) shows a Warlpiri sentence in which the locative NP *ngapa-ngka* "water-LOC" is ergatively marked, indicating that the transitive event took place in that location. But there is no construction like this in the Light Warlpiri data.

(23) *Ngarrka-ngku ka-ø-ø*   *yankirri luwa-rni*   *ngapa-ngka-rlu.*
      man-ERG   PRES-3SGS-3SG.NS emu   shoot-NPST water-LOC-ERG
      "The man is shooting the emu at the waterhole."   (Hale 1982: 268)

There are two types of construction in Light Warlpiri in which ergative case-marking occurs in a clause containing an intransitive verb (25, 26) and (27, 28). In both of them the ergative indicates a transitive or causative function, even though it does not attach to an A argument. One of these functions perhaps draws on the juxtaposition of intransitive and transitive clauses in Warlpiri as spoken in Laja-manu Community, as in (24), where the ergative construction indicates an ablative function referring to an event expressed in a prior intransitive clause.

(24) *Yakarra! Wanti-ja-ø   mayi*? *Watiya-ngu mayi*?
      DIS   fall-PST-3SGS Q   tree-ERG   Q
      "Oh no! Did he fall? Because of the tree?"
                              (C03_C04_5_3, Light Warlpiri speaker, speaking Warlpiri)

In (24) the use of the ergative suffix on *watiya* "tree" and *mayi* "Q" questions whether the event of falling was "because of the tree" or "caused by the tree". In Warlpiri typically an ablative would be used for this function. It is not uncommon in languages for there to be syncretism in ablative and ergative forms (e.g., Bickel & Nichols 2009). In Light Warlpiri a similar type of ergative-as-ablative construction occurs, as in (25) and (26).

(25) An *jinta* i-m   fal-dan   *watiya-ng.*
      and one 3SG-NFUT fall-down tree-ERG
      "And one [of them] fell because of the tree."   (ERGstoryLC05_2010)

(26) *Watiya-ng mayi* i-m   fal-dan?
      tree-ERG   Q   3SG-NFUT fall-down
      "Is it because of the tree that he fell?"   (ERGstoryLC33_2010)

The second type of construction is in clauses where the goal of a transitive action is not achieved or is in progress, and the verb includes *trai* 'try' as in (27) and (28).

(27) *Kuuku-ng* i-m   trai-in   to get-im na *jarntu-pawu.*
      monster-ERG 3SGS-NFUT try-PROG to get-TR now dog-DIM
      "Now the monster is trying to get the dog."   (ERGstoryLC157_2015)

(28) An *boy-pawu-ng* an *jarntu-ng* dei trai get-im *kanta*   *wana*
      and boy-DIM-ERG and dog-ERG 3PLS try get-TR bush.coconut DIS

      kankarl watiya-nga.
      high   tree-LOC

      "And the boy and the dog try to get the bush coconut, you know, up high in
      the tree."   (ERGstoryLA93_2015)

In (27) the monster does succeed in getting the dog in the next sequence of the story, but in (28) the boy and the dog do not succeed in getting the fruit from the tree. This appears to be the only context in which an S argument is marked with an ergative suffix.

Dative case-marking

As in Warlpiri, the dative in Light Warlpiri occurs in ergative-dative (29), ergative-absolutive-dative (30), and absolutive-dative case arrays (31) (O'Shannessy 2016b: 232). The form of the dative in Light Warlpiri is a reduced form from classic Warlpiri that occurs in contemporary Warlpiri. But in Light Warlpiri the reduced form has conventionalized so that two new allomorphs occur almost exclusively, *-k* "DAT" on vowel-final stems and *-ik* "DAT" on consonant-final stems (O'Shannessy 2016b: 232). This regularization is parallel to that of ergative allomorphy.

(29) *Pakarra-ng* i-m   luk-raun   futbal-*ik.*
      name-ERG 3SGS-NFUT look-around football-DAT
      "Pakarra looked around for the football."   (O'Shannessy 2016b: 232)

(30) *Ngaju-ng* a-m   gib-im shet *kurdu-pawu-k.*
      1SGS-ERG 1SGS-NFUT give-TR shirt child-DIM-DAT
      "I gave a shirt to the child."   (O'Shannessy 2016b: 232)

(31) *Kurdu-kurdu wita* i-m   shaut-ing   *jarntu-k.*
      child-REDUP small 3SGS-NFUT shout-PROG dog-DAT
      "The small children are shouting to the dog."   (O'Shannessy 2016b: 232)

Some dative functions are indicated using forms from English/Kriol but with dative meanings. In Warlpiri the dative case is selected by the verb "to speak". In Light Warlpiri the dative selected by the verb "talk" is realized through a form from English/Kriol. This construction is not attested in English or Kriol.

(32)   A-rra   tok   fo   *Nungarrayi   junga.*
       1SGS-FUT   talk   DAT   skin.name[3]   true
       "I'll tell Nungarrayi, really!"   (C06_17)

A difference in dative marking between the two languages is that in Warlpiri the dative is registered in the auxiliary but in Light Warlpiri it is not. Recall that in Warlpiri the auxiliary has information for both subject and non-subject pronouns but in Light Warlpiri there is only information for subjects.

There are morphological differences between the two languages in possessive case-marking, where English/Kriol forms occur in some types of construction and Warlpiri forms in others (Meakins & O'Shannessy 2005), as in (33) and (34).

(33)   Get-im   kap   *Nungarrayi-kirlang   kurnta-nga*!
       get-TR   cup   skin.name-POSS   shame-LOC
       "Get the cup, its Nungarrayi's, shame!"   (Meakins & O'Shannessy 2005: 59)

(34)   Mai   aus   i-m   *yinya-wana   kankarl.*
       1SG.POSS   house   3SG-NFUT   DEM-PERL   high
       "My house is along there up high."   (Meakins & O'Shannessy 2005: 59)

This section has shown that nominal morphology in Light Warlpiri is largely retained from Warlpiri, with some exceptions and some re-analyses.

3.3   Sentence-level syntax and verb-argument structure in Light Warlpiri

Light Warlpiri exhibits several features of sentence-level syntax retained from Warlpiri, specifically, variable word order, pronoun drop and discontinuous constituents. These are the features that contribute to Warlpiri being categorized as non-configurational (Hale 1983). As Table 4.5 illustrates, core arguments need not be overt in verbal clauses. In clauses with overt arguments word order is predominantly SVO, but also varies. In stories told from picture stimuli (O'Shannessy 2004) by 8 Light Warlpiri-speaking adults in 2015, there is a total of 684 verbal clauses. Of those, 41% are transitive, and only 37% of the transitive clauses have an overt A argument. Eighty-five per cent of those are marked ergative, but when the order is VA, 95% are marked ergative. This indicates that a complementary relationship between word order and case-marking is emerging. The information is summarized in Table 4.5.

---

3.   Warlpiri has a complex system of naming kinship relations, widespread in Australian languages, and retained in Light Warlpiri.

Table 4.5 Word order and non-overt A arguments in 8 Light Warlpiri narratives

Category	Raw numbers	%	Explanatory note
Total clauses	684		
Transitive clauses	287	41	41% of total clauses are transitive
Transitive clauses with overt A	102	37	37% of transitive clauses have overt A
Transitive clauses, overt A, with ERG marking	87	85	85% of transitive clauses with overt A have ergative marking
Transitive clauses, VA word order	21	21	21% of transitive clauses with overt A have VA order
Transitive clauses, VA word order, with ERG marking	20	95	95% of A arguments in VA order have ergative marking

Variable word order in Warlpiri is constrained in that the auxiliary element must occur in second position in a clause, as in (35).

(35)   *Ngajulu-rlu ka-rna-ngku   nyuntu nya-nyi.*
       1SG-ERG   PRES-1SGS-2SG.NS 2SG   see-NPST
       "I see you."   (Hale et al. 1995: 1430)

In (35) we see that the auxiliary and verb elements in Warlpiri need not be adjacent. Additionally, the first element in a sentence is usually the most informative, or the element in focus (Hale 1992; Simpson & Mushin 2008). In Light Warlpiri the constraint on word order differs. Word order in Light Warlpiri is variable, but the auxiliary and verb are always adjacent, in auxiliary-verb order, and this unit can occur in any position in the clause, as in (36), (37) and (38).

(36)   Yu-m   *winjirn*-im haf-wan   *kuja-ng.*
       2SG-NFUT pour-TRANS   half-NMLZ   thus-ERG
       "You poured it halfway, like this."   (C03_17)

(37)   *Jalang* wi-m   go krik-*kirra kurdu-kurdu-kurl.*
       today   1PL-NFUT   go creek-ALL   child-REDUP-COM
       "Today we went to the creek with the children."   (Elicit_LA21_AC58_2014)

(38)   *Kala nyarrpara-rla nyuntu-ju* yu-m   bugi?
       DISJ where-LOC   2SG-TOP   2SG-NFUT   swim
       "But where did you swim?"   (O'Shannessy 2015: 290)

The adjacent position of the Light Warlpiri auxiliary and verb is probably motivated by both English and Kriol, as both languages have pre-verbal subject pronoun and auxiliary placement. The information structure of Light Warlpiri has not been examined in detail other than in terms of its relationship to ergative

case-marking, but I hypothesize that focus or prominent information occurs in initial position, regardless of the syntactic status of the initial element.

Consistent with the constraint that the Warlpiri auxiliary must be in second position, as in (35), is that only a constituent can precede it. It is this rule that shows that the verb does not form a constituent with its internal argument in surface structure, since a verb and an object NP cannot occur together in initial position, preceding the auxiliary. Therefore, there is no constituent consisting of a verb-plus-object complement. Light Warlpiri auxiliary and verb placement does not suggest a comparable constraint in constituent content.

Another feature of Warlpiri syntax is discontinuous constituents (Hale et al. 1995: 1434).

(39) *Jarntu-ng* i-m   jeis-im   pujikat-*pawu wita-pawu wiri-jarlu-ng.*
      dog-ERG 3SG-NFUT chase-TR cat-DIM   small-DIM big-INTENS-ERG
      "A big dog chased a small cat."   (ElicitLA31_2015)

Example (39) shows a Light Warlpiri transitive sentence with a discontinuous A argument, *jarntu-ng … wiri-jarlu-ng* "dog-ERG … big-INTENS-ERG". The elements marked ergatively are components of the A argument, but other elements occur between them. The example is from directly elicited data that probed whether a construction like this can occur in Light Warlpiri. Not all speakers produced this structure, and some preferred a modifying element to occur next to a head NP. Some arguably discontinuous phrases have been found in spontaneous Light Warlpiri speech, but their interpretations are not unambiguously those of discontinuous constituents. The argument here is that this type of constituent is permitted, but its occurrence is uncommon in spontaneous speech, and there is no consensus about its acceptability.

In Warlpiri, adverbial and adjectival functions are achieved through the semantics of elements which function syntactically as noun phrases, demonstrated by their ability to host nominal case-markers and auxiliary clitics. In Light Warlpiri, Warlpiri-derived adverbial and adjectival elements also have nominal structure. Examples are given in (40), and in (39) and (21) above.

(40) Sii yu   kaan   draiv-im *pulya-ng*
      See 2SGS can.NEG drive-TR careful-ERG
      "See, you're not driving carefully."   (ERGstoryLAC58_2015)

4. Light Warlpiri and theories of ML emergence

Light Warlpiri shows features that support Myers-Scotton's (2003) theory in two ways. It shows abstract information from all sources in at least one component of its morphosyntactic frame, and therefore is thoroughly mixed, in keeping with

Myers-Scotton's definition of a split language (Myers-Scotton 2003). And it has late outsider morphemes from both Warlpiri (ergative case marking) and English/ Kriol (auxiliary pronouns). Example (41) illustrates how abstract information in the Light Warlpiri morphosyntactic frame is drawn from all sources.

(41)   *Karnta-pawu-ng* i-m   get-im *jarntu* *nyan-nyang*.[4]
       woman-DIM-ERG 3SG-NFUT get-TR dog   3SG-POSS
       "The woman got her dog."   (ERGstoryLC10_2010)

Table 4.6 shows the sources of each of Myers-Scotton's (2003) three types of abstract material in Light Warlpiri.

Table 4.6 Elements of a Light Warlpiri sentence in terms of Myers-Scotton's (2003) three types of abstract material[5]

	Source language	
Type of material	**Warlpiri**	**English/Kriol**
Lexical / conceptual	*karnta* "woman"	i "3SG" get
	-pawu "DIM"	
	nyan "3SG"	
	-nyang "POSS"	
Predicate argument structure	ERG-ABS case	
	NOM-ACC case	NOM-ACC case
Morphosyntactic realization	*-ng* "ERG"	i "3SG AUX PRO"
	-m (draws on all sources)	-im "TRANS"

As seen in Table 4.6, at each level of Myers-Scotton's abstract material, material is drawn from Warlpiri and English/Kriol. In terms of Myers-Scotton's 4M model, example (41) is presented again in Table 4.7 to illustrate the types of morphemes present.

Table 4.7 A Light Warlpiri sentence in terms of Myer-Scotton's (2004) 4M model

karnta	*-pawu*	*-ng*	i	-m	get	-im	*jarntu*	*nyan*	*-nyang*
child	-DIM	-ERG	3SG AGR	NFUT	get	TRANS	dog	3SG	POSS
content	content	late outsider	late outsider	early system	content	early system	content	content	content

---

4. *nyan-nyang* is a contracted form of classic Warlpiri *nyanungu-nyangu* "3SG-POSS".

5. Many thanks to Carol Myers-Scotton for discussion of the data.

Late outsider morphemes are drawn from both Warlpiri (ergative case-marking) and English/Kriol (the auxiliary pronoun, labelled as an agreement element to make the connection to Myers-Scotton's theory clear); both depend on information from outside their maximal projection for their form. The early system morphemes *-m* "NFUT" and *-im* "TRANS" provide conceptual information of temporality and transitivity respectively. The Light Warlpiri morphosyntactic frame, then, is a composite frame, involving abstract material from all sources.

The structure of Light Warlpiri supports Myers-Scotton's (2003) theory, but the notion of an arrested matrix language turnover cannot be tested empirically with the current data. In the Light Warlpiri situation, there has not been a shift in terms of which language is socially dominant in the community, one of the motivations for an arrested matrix language turnover.

Although Light Warlpiri is clearly a mixed or split language, it is also underlyingly Warlpiri. First, sentence level structure is retained from Warlpiri. Core arguments may be elided if they are recoverable from the discourse, word order is variable, and discontinuous constituents are possible. Second, nominal morphology is largely from Warlpiri, with some exceptions. Further, within nominal morphology, it is clear from Section 3.2 that, for instance, ergative marking occurs on nominals as it does in Warlpiri, indicating instruments and adjuncts of manner and time. Third, the verbal auxiliary structure draws on all of the sources, including Warlpiri, as explained above. Additionally, a relevant part of the verbal structure retained from Warlpiri is that transitive verbs select ergative-marked A arguments. While ergative marking is frequently presented as belonging to the domain of nominal structure, its occurrence on A arguments of transitive verbs makes it part of the verbal structure also. Even if the theory that nominals external to the verbal complex are adjuncts, not core arguments (Jelinek 1984) is adhered to, whether or not the NPs take ergative marking is still dependent on the transitivity of the verb. Similarly, while a well-formed sentence does not require overt A, S or O arguments, these arguments are part of the basic clausal structure.

The scenario posed by Auer (1999, 2014) as structural sedimentation of code-switching patterns, is a plausible one for the development of Light Warlpiri. I have argued elsewhere for a specific mechanism whereby the pre-Light Warlpiri code-switching patterns were conventionalized (O'Shannessy 2012). The argument is that a specific code-switching pattern that later became the blueprint for Light Warlpiri structure was the main speech style directed to young children in a Baby Talk register, and was internalized by the children as a single system (O'Shannessy 2012). The pattern is that of code-switching, where a unit from one language – more than a single word – is inserted into a string of the other. This code-switching pattern is a combination of what have been described as two types of code-switching patterning by Muysken (2000). It is insertional in that a unit from one language is inserted into another. Yet in the case of Light Warlpiri the inserted element is

a grammatical phrase, consisting of a verb and a verbal auxiliary, so is also like Muysken's (2000) alternational switching, where the language changes at a phrasal or clausal boundary. In Myers-Scotton's (2003, 2004) terms the inserted element is an embedded island, as the grammar of one language (the Kriol transitive marker) is attached to the inserted lexical items.

When the practice was one of adult code-switching, before Light Warlpiri emerged, an English/Kriol pronoun and verb were inserted into a Warlpiri string, as in (42).

(42) *yakarra nyanya* wi   hab-im *nyanya wana ngalipa   nyanya*
     DIS   food   1PLS have-TR food   DIS   1PL.INCL food
     "Gosh, we have food, food, you know, us, food"   (O'Shannessy 2012: 325)

The combinations of structures can be stylized as follows:

Warlpiri string + <u>English/Kriol pronoun & verb</u> + Warlpiri string

When the young children to whom this speech style was directed conventionalized it as they grew up, they added innovations, so that now, the verbal structure is not simply from English and Kriol, but contains structures from all of the source languages, as argued above in Section 3.1 and in O'Shannessy (2012, 2013).

The mechanism whereby the innovations are brought in by children is the creativity commonly seen in first language acquisition, where child language acquirers search for patterns in the speech they receive as primary input, and form hypotheses accordingly. Some of their hypotheses might produce surface forms that do not align with the way of speaking in the community, for instance, by involving over-generalizations (e.g., *bringed* for *brought* in English). Usually, when there is no language change involved, the child later changes the hypothesis and constructs a new one that conforms to adult speech styles. But in some contexts the children might retain their non-conforming hypotheses as they grow up and create a slightly different way of speaking (O'Shannessy 2020). The contexts include those where there is multilingualism, variable input and change in progress, and the speakers form a tightly-knit, bounded community with some motivation to create a speech style that signals their identity (O'Shannessy 2020). Similar scenarios have been presented for the development of innovative structures in Creoles by, for example, Le Page & Tabouret-Keller (1985), Mufwene (2001), Aboh and Ansaldo (2006), Aboh (2009, 2015) and Croft (this volume).

Auer's (1999) theory of structural sedimentation of code-switching practices, if taken up and internalized by children, may be enough to explain the mechanisms of the emergence of a mixed language. It highlights social motivations for the structural sedimentation or conventionalization to occur. I argue that as the way of speaking conventionalizes, innovations may be brought in, but they are not

required for a mixed language to emerge, given a specific constellation of social factors interacting with the code-switching practices.

5.   Conclusion

In this chapter I have argued that the structure of Light Warlpiri is thoroughly mixed, yet underlyingly draws on Warlpiri more than on any other source. Myers-Scotton's 4M model of code-switching and language contact phenomena (Myers-Scotton 2004), and her notion of three levels of abstract material in a phrase (Myers-Scotton 2003) have been used to show that the abstract material in at least two components of Light Warlpiri, the verbal and nominal systems, are drawn from all of its sources. A comparison of sentence-level syntax of Light Warlpiri with Warlpiri (Hale et al. 1995) has shown that the two languages share the properties of pronoun drop, variable word order and discontinuous constituents. In terms of the mechanisms of emergence, I have argued that it may be that conventionalized code-switching practices alone are sufficient for a mixed language to emerge.

References

Aboh, E. O. 2009. Competition and selection. That's all! In *Complex Processes in New Languages* [Creole Language Library 35], E. O. Aboh & N. Smith (eds), 317–344. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.35

Aboh, E. O. 2015. *The Emergence of Hybrid Grammars: Language Contact and Change*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139024167

Aboh, E. O. & Ansaldo, U. 2006. The role of typology in language creation. In *Deconstructing Creole* [Typological Studies in Language 73], U. Ansaldo, S. Matthews & L. Lim (eds), 39–66. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.73.05abo

Angelo, D., Fraser, H. & Yeatman, B. 2019. The art of recognition: Visualising contact languages with community vernacular language posters. *Babel* 54(1–2): 34–36.

Auer, P. 1999. From code-switching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *The International Journal of Bilingualism* 3: 309–332. https://doi.org/10.1177/13670069990030040101

Auer, P. 2014. Language mixing and language fusion: When bilingual talk becomes monolingual. In *Congruence in Contact-Induced Language Change: Language Families, Typological Resemblance and Perceived Similarity*, J. Besters-Dilger, C. Dermarkar, S. Pfänder & A. Rabus (eds), 294–335. Berlin: De Gruyter. https://doi.org/10.1515/9783110338454.294

Batchelor, T. 2017. Making Sense of Synchronic Variation: Semantic and Phonological Factors in Transitivity Marking in Barunga Kriol. BA (Honours) dissertation, University of Sydney.

Bickel, B. & Nichols, J. 2009. Case marking and alignment. In *The Oxford Handbook of Case*, A. Malchukov & A. Spencer (eds), 304–321. Oxford: OUP.

Butcher, A. 2008. Linguistic aspects of Australian Aboriginal English. *Clinical Linguistics and Phonetics* 2: 625–642. https://doi.org/10.1080/02699200802223535

Dixon, R. M. W. 1979. Ergativity. *Language* 55(1): 59–138. https://doi.org/10.2307/412519

Eades, D. 1993. *Aboriginal English*. Melbourne: Primary English Teaching Association.

Golovko, E. V. 1994. Mednij Aleut or Copper Island Aleut: An Aleut-Russian mixed language. In *Mixed Languages: 15 Case Studies in Language Intertwining*, P. Bakker & M. Maarten (eds), 113–121. Amsterdam: Institute for Functional Research into Language and Language Use.

Hale, K. 1982. Some essential features of Warlpiri verbal clauses. In *Papers in Warlpiri Grammar: In Memory of Lother Jagst*, S. Swartz (ed.), 217–314. Darwin: Summer Institute of Linguistics-Australian Aborigines Branch.

Hale, K. 1983. Warlpiri and the grammar of non-configurational languages. *Natural Language and Linguistic Theory* 1: 5–47. https://doi.org/10.1007/BF00210374

Hale, K. 1992. Basic word order in two "free word order" languages. In *Pragmatics of Word Order Flexibility* [Typological Studies in Language 22], D. Payne (ed.), 63–82. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.22.03hal

Hale, K., Laughren, M. & Simpson, J. 1995. Warlpiri. In *An International Handbook of Contemporary Research*, Vol. 2, J. Jacobs, A. von Stechow, W. Sternefeld & T. Vennemann (eds), 1430–1449. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110142631.2.21.1430

Harkins, J. 1994. *Bridging Two Worlds: Aboriginal English and Crosscultural Understanding*. Brisbane: University of Queensland Press.

Jelinek, E. 1984. Empty categories, case and configurationality. *Natural Language and Linguistic Theory* 2: 39–76. https://doi.org/10.1007/BF00233713

Laughren, M. 2013. A revised analysis of Warlpiri verb inflections plus auxiliary combinations: Their makeover in "Light" Warlpiri. In *Proceedings of the 2012 Conference of the Australian Linguistic Society*, J. Henderson, M.-E. Ritz & C. Rodríguez Louro (eds), 1–14. Perth: Australian Linguistic Society.

Le Page, R. & Tabouret-Keller, A. 1985. *Acts of Identity: Creole-Based Approaches to Language and Ethnicity*. Cambridge: CUP.

Malcolm, I. & Kaldor, S. 1991. Aboriginal English: An overview. In *Language in Australia*, S. Romaine (ed.), 67–83. Cambridge: CUP. https://doi.org/10.1017/CBO9780511620881.005

McConvell, P. & Meakins, F. 2005. Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics* 25: 9–30. https://doi.org/10.1080/07268600500110456

Meakins, F. 2010. The development of asymmetrical serial verb constructions in an Australian mixed language. *Linguistic Typology* 14: 1–38. https://doi.org/10.1515/lity.2010.001

Meakins, F. 2011. *Case-Marking in Contact: The Development and Function of Case Morphology in Gurindji Kriol* [Creole Language Library 39]. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.39

Meakins, F. 2015. From absolutely optional to only nominally ergative: The life cycle of the Gurindji ergative suffix. In *Borrowed Morphology*, F. Gardani, P. Arkadiev & N. Amiridze (eds), 190–218. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9781614513209.189

Meakins, F. & O'Shannessy, C. 2005. Possessing variation: Age and inalienability related variables in the possessive constructions of two Australian mixed languages. *Monash University Linguistics Papers* 4: 43–64.

Meakins, F. & O'Shannessy, C. 2010. Ordering arguments about: Word order and discourse motivations in the development and use of the ergative marker in two Australian mixed languages. *Lingua* 20: 1693–1713. https://doi.org/10.1016/j.lingua.2009.05.013

Meakins, F. & O'Shannessy, C. 2012. Typological constraints on verb integration in two Australian mixed languages. *Journal of Language Contact* 5: 216–246.
https://doi.org/10.1163/19552629-006001001

Mufwene, S. S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP.
https://doi.org/10.1017/CBO9780511612862

Muysken, P. 2000. *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: CUP.

Myers-Scotton, C. 1988. Code-switching as indexical of social negotiations. In *Code-Switching*, M. Heller (ed.), 151–186. Berlin: Mouton de Gruyter.

Myers-Scotton, C. 1993. *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: OUP.

Myers-Scotton, C. 2000. Four types of morpheme: Evidence from aphasia, code switching, and second-language acquisition. *Linguistics* 38: 1053–1100.
https://doi.org/10.1515/ling.2000.021

Myers-Scotton, C. 2002. *Contact Linguistics*. Oxford: OUP.
https://doi.org/10.1093/acprof:oso/9780198299530.001.0001

Myers-Scotton, C. 2003. What lies beneath: Split (mixed) languages as contact phenomena. In *The Mixed Language Debate*, P. Bakker & Y. Matras (eds), 73–106. Berlin: Mouton de Gruyter.

Myers-Scotton, C. 2004. Precision tuning of the Matrix Language Frame (MLF) model of codeswitching. *Sociolinguistica: International Yearbook of European Sociolinguistics* 18: 106–117. https://doi.org/10.1515/9783484604650.106

Nash, D. 1982. Warlpiri verb roots and preverbs. *Work Papers of SIL-AAB* 6: 165–216.

Nash, D. 1986. *Topics in Warlpiri Grammar*. New York NY: Garland.

O'Shannessy, C. 2004. *The Monster Stories: A Set of Picture Books to Elicit Overt Transitive Subjects in Oral Texts*. Unpublished series. Nijmegen: Max Planck Institute for Psycholinguistics.

O'Shannessy, C. 2005. Light Warlpiri: A new language. *Australian Journal of Linguistics* 25(1): 31–57. https://doi.org/10.1080/07268600500110472

O'Shannessy, C. 2012. The role of code-switched input to children in the origin of a new mixed language. *Linguistics* 50: 328–353.

O'Shannessy, C. 2013. The role of multiple sources in the formation of an innovative auxiliary category in Light Warlpiri, a new Australian mixed language. *Language* 89: 328–353. https://doi.org/10.1353/lan.2013.0025

O'Shannessy, C. 2015. Multilingual children increase language differentiation by indexing communities of practice. *First Language* 35: 305–326. https://doi.org/10.1177/0142723715609227

O'Shannessy, C. 2016a. Distributions of case allomorphy by multilingual children. *Linguistic Variation* 16: 68–102. https://doi.org/10.1075/lv.16.1.04osh

O'Shannessy, C. 2016b. Entrenchment of Light Warlpiri morphology. In *Loss and Renewal: Australian Languages Since Contact*, F. Meakins & C. O'Shannessy (eds), 217–252. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9781614518792-014

O'Shannessy, C. 2020. How ordinary child language acquisition processes can lead to the unusual outcome of a mixed language. *International Journal of Bilingualism*.
https://doi.org/10.1177/1367006920924957

Pentland, C. & Laughren, M. 2005. Distinguishing prosodic word and phonological word in Warlpiri: Prosodic constituency in morphologically complex words. In *Proceedings of the 2004 Conference of the Australian Linguistic Society*, I. Mushin (ed.), 1–17. Sydney.

Sandefur, J. 1979. *An Australian Creole in the Northern Terrritory: A Description of Ngukurr-Bamyili Dialects*, (Part 1) Series B. Vol. 3. Darwin: Summer Institute of Linguistics, Australian Aborigines and Islanders Branch.

Schultze-Berndt, E., Meakins, F. & Angelo, D. 2013. Kriol. In *English-Based and Dutch-Based Languages*, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 241–251. Oxford: OUP.

Simpson, J. 1991. *Warlpiri Morpho-Syntax: A Lexicalist Approach*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-3204-6

Simpson, J. & Mushin, I. 2008. Clause-initial position in four Australian languages. In *Discourse and Grammar in Australian Languages* [Studies in Language Companion Series 104], I. Mushin & B. Baker (eds.), 25–58. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.104.05sim

CHAPTER 5

Acquisition or shift?

Interpreting variation in Gurindji children's expression of spatial relations

Vivien Dunn, Felicity Meakins and Cassandra Algy
ARC Centre of Excellence for the Dynamics of Language, School of Languages and Cultures, University of Queensland

This chapter examines the spatial description system employed by Gurindji children in Kalkaringi (Northern Territory, Australia) to describe ternary relations in small-scale space. While Gurindji is the traditional language of Kalkaringi, a new variety, Gurindji Kriol, has developed as a result of language contact, and is now the first language of young adults and children. Speakers of Gurindji use cardinal directions in descriptions of both small-scale and large-scale space, whereas cardinal terms are almost never used to describe small-scale in naturally-occurring Gurindji Kriol speech. We show that the strategies preferred by children differ from those used by their parents, who produce predominantly cardinal descriptions. Instead, Gurindji Kriol-speaking children show variable use between cardinals and landmark-based strategies.

Keywords: language contact, mixed languages, spatial relations, cardinal directions, language change

1. Introduction

This chapter investigates the frames of reference (FoR) employed by Gurindji Kriol-speaking children from Kalkaringi in northern Australia to describe the angular relation between objects in small-scale space. A FoR is a coordinate system used to specify the position of one entity with respect to another when the two entities are separated in space (Levinson & Wilkins 2006: 3–4). There are three major categories of FoR found in the world's languages: intrinsic, relative and absolute (Levinson 2003). In an intrinsic FoR, the coordinate system is based on the inherent facets of one of the entities (e.g., the front and back of the object);

in a relative FoR, the coordinates come from the observer's own body (e.g., the observer's left and right side) and are projected onto the scene; and in an absolute FoR, the coordinates used are abstract fixed bearings (e.g., cardinal directions) (Levinson & Wilkins 2006: 20–21) (Section 2). Of interest for this chapter are absolute systems and the intergenerational effects of language contact on the use of cardinal directions.

The traditional language of Kalkaringi is Gurindji, a Pama-Nyungan language which has an absolute frame of reference based on cardinal directions for descriptions of small-scale space. For example, Gurindji speakers will talk about flies landing on their north shoulder or direct someone to put the sugar on the east side of the flour on a kitchen shelf. However, Gurindji has undergone significant changes since colonization, including in the domain of spatial language (Meakins 2011: 44). Traditional Gurindji is now only spoken fluently by older members of the community, with young adults and children speaking a mixed language known as Gurindji Kriol. FoR studies have been conducted with numerous speech communities across the world, but the effects of language contact and shift on these systems have not been fully explored. The few studies which exist show that in colonial situations where the introduced language is a European language with a relative FoR system, speakers of languages with absolute systems tend to shift towards a relative system either by borrowing terms for "left" and "right" from the colonial language or by repurposing body part terms for "left-hand" and "right-hand" from their own languages (Adamou 2017; Dasen & Mishra 2010; Polian & Bohnemeyer 2011) (Section 2).

In the case of Kalkaringi, Gurindji-derived cardinal terms are now only used by Gurindji Kriol-speaking adults in natural discourse to describe large-scale space, with deictic terms and absolutely-orientated gestures favoured instead in descriptions of small-scale space. Nonetheless, adult speakers of Gurindji Kriol, like speakers of Gurindji, rely on the Gurindji-derived cardinal system for describing small-scale space when they are obscured from each other's view, rendering deictics and gesture ineffectual (Meakins 2011). Interestingly, adult speakers have not adopted English left/right terms despite Western schooling. Further intergenerational change has been observed in Gurindji Kriol-speaking children. Meakins & Algy (2016) tested children's knowledge of the Gurindji cardinal terms, which they found to be weaker than that of their parents' generation. However, they did not examine usage, i.e., if not cardinals, then what linguistic resources do children bring to the task of describing spatial relations (Section 3)?

This study builds on Meakins & Algy (2016) by examining what FoR strategies Gurindji Kriol-speaking children use in descriptions of small-scale space. We report on the results of the "Man and Tree" task administered to thirty Gurindji Kriol-speaking children and a control group of ten Gurindji Kriol-speaking adults

(Section 3). This director–matcher task is designed to reveal a speaker's FoR strategies for descriptions of small-scale space when they are obscured from their interlocutor's view (Levinson et al. 1992). We find that the children's strategies are much more variable than those of the adults. The children use cardinals, and occasionally left/right terms, but landmarks make up the majority of their descriptive repertoire. Here we argue that the use of landmarks is not indicative of a shift away from the abstract geocentric system, but may represent the development of a quasi-absolute system (Section 4). Of course, interpreting the results of a child language study in a community where language shift is taking place is fraught with complexities. Nevertheless, we make the case that intergenerational differences are representative of language change rather than acquisition (Section 4).

2.   The absolute FoR, bilingualism and language change

Descriptions of spatial relations involve locating an entity, often referred to as "figure", in space with respect to a reference entity, or "ground" (Talmy 1983: 232). When the figure and ground coincide in space, the relationship between the two entities can be specified using a non-angular topological description (often involving case-markers or prepositions);[1] when the figure and ground are separated in space, however, the spatial relation becomes more conceptually complex, and a coordinate system is needed to specify the angular relation between the entities (Levinson & Wilkins 2006: 3–4). These coordinate systems are known as FoRs.

   According to the framework used by Levinson (2003), there are three major types of linguistic FoR found in the world's languages: intrinsic, absolute and relative.[2] Of particular interest for this chapter are the absolute and relative FoRs, which involve mapping external coordinate systems to locate the figure with respect to the ground (Levinson 2003: 20). In the absolute FoR, the coordinates come from a fixed bearing system such as cardinal directions, for example "the dog is *west* of the house". In the relative FoR, the viewer's own body axes are projected

---

1.   An example of a topological description in English is "the cat is in the box". The figure (cat) and ground (box) are not separated in space and their spatial relationship to each other is indicated by a preposition.

2.   The intrinsic FoR involves dividing the ground (or the figure) into facets which are then projected onto the surrounding space and used as the basis for the coordinate system (Levinson 2003: 42). The phrase "the dog is in front of the house" employs this FoR. Here, the location of the figure (the dog) is described in relation to the ground (house), with the ground-centred coordinate "front" providing a projective search domain within which the figure can be found.

onto the ground, for example "the dog is *right* of the tree" (Levinson & Wilkins 2006: 21). Also of interest for this chapter is the use of landmarks to express the relative position of a figure and ground, but there is significant disagreement in the literature about how to treat landmarks (see Section 3 and Section 4).

Cross-linguistic differences in the use of FoRs have been demonstrated systematically through the use of the "Man and Tree" task, which was created specifically to elicit language data that reveal a speaker's FoR preference(s) for descriptions of small-scale spatial relations (described in Section 3). This task has been used in languages which have a relative system such as Dutch and English, but also in languages which use different kinds of absolute systems based on compass points (N-S-E-W), e.g., Hai//om (Namibia), Totonac (Mexico), Arrernte (Australia) and Juchitán Zapotec (Otomanguean, Mexico); river drainage (upstream/downstream), e.g., Jaminjung (Australia); slope (uphill/downhill), e.g., Tzeltal (Mexico); and seaward/inland systems, e.g., Longgu (Solomons) (Brown 2006; Levinson 2003; Majid et al. 2004; Pederson et al. 1998; Schultze-Berndt 2006). These studies demonstrate that, although speakers of languages such as English have both the relative and the absolute FoRs in their spatial "toolbox", they rely on only one of them in small-scale spatial descriptions. For example, English and Dutch speakers consistently use the relative FoR while speakers of Arrernte rely on the absolute FoR in tasks such as the "Man and Tree" game.

Much of the literature on FoRs has focused on monolingual speakers, and has only recently begun to address bilingualism and language shift in speech communities (see e.g., Adamou 2017: 176; Dasen & Mishra 2010; De León 1995; Edmonds-Wathen 2014; Marghetis et al. 2014; Meakins 2011; Meakins & Algy 2016; Meakins et al. 2016; Polian & Bohnemeyer 2011). Studies of language shift and spatial relations in colonial settings typically find a shift towards a relative system, which is often the main FoR used by speakers of the colonial language. For example, Polian & Bohnemeyer's (2011: 887) study of the Tenejapan community found an increase in the linguistic use of the relative FoR, particularly among bilingual Spanish-Tzeltal speakers. Similarly, Adamou (2017) notes that contact with Spanish influenced the development of the relative FoR for Ixcatec-Spanish bilinguals, which is reflected in their use of borrowed Spanish terms, e.g., *izquierda* "left", and also the innovated projective use of Ixcatec body-part terms, e.g., *ku1tʃʔe1* "left hand". Dasen and Mishra (2010: 83) found that left/right terms in Balinese, which were previously only used intrinsically to denote sides of the body, are increasingly being used projectively by Balinese children and adults who are bilingual in Indonesian, a language whose speakers use terms for left and right in a relative FoR. One question for this chapter is whether Gurindji Kriol-speaking children have begun adopting the English left/right system as a result of Western education.

A complicating factor for studies of language change is participant age, and whether differences between adults and children can be attributed to language change or acquisition. Studies of language change assume that age-stratified variation reflects change in progress rather than acquisition (cf. the apparent-time hypothesis, which assumes that an individual's vernacular does not change once they are past a certain age [Labov 1963]). For this reason, it is necessary to know when the acquisition of absolute systems takes place.

One of the first studies to investigate absolute FoR acquisition was De León's (1994) work with Tzotzil children in Mexico, whose language has an absolute linguistic system based on an "upland"/"downland" axis. De León (1994: 857) reports that the Tzotzil children acquire this system between four and five years of age; however, she goes on to state that the system does not become "terrain-independent" until the age of nine (De León 1994: 878). According to Levinson et al. (2002: 172–173), a coordinate system is not truly absolute unless it is abstracted from environmental features (i.e., "terrain-independent"). Similarly, Brown and Levinson (2000: 168) report Tzeltal children using the absolute uphill/downhill system by the "unexpectedly early" age of six to seven years, but acknowledge that their conceptualization of the system may not yet be "fully adult" at this age (Brown & Levinson 2000: 191) (for other acquisition studies see also Cablitz 2002; Dasen & Mishra 2010).

These studies highlight a problem for researchers investigating the age of acquisition of a "true" absolute FoR system: given that many absolute coordinate systems have been abstracted from local environmental features (Levinson 2003: 49), it is difficult to determine when children are using the linguistic terms in an abstract way. Although Brown and Levinson (2000: 178) note that these kinds of absolute systems can be used away from the speech community's territory, and are therefore terrain-independent, acquisition studies by Brown and Levinson (2000, 2009) and by De León (1994) were conducted on the communities' home territory, where the absolute and landmark-based uses of the terms are conflated (Brown & Levinson 2000: 181; Polian & Bohnemeyer 2011: 874).

In addition to investigating *when* children acquire absolute systems, researchers are interested in *how* children acquire these systems. One area of particular interest is whether child learners of absolute languages go through a "landmark stage" before an abstract conceptualisation and use of the coordinate system is acquired. De León (1994: 878–880, 1995: 26–32) has proposed a developmental path for the acquisition of absolute language based on her research with Tzotzil children in Mexico and Guugu Yimithirr children in Australia. She claims that children in both communities pass through a stage in which the spatial terms are used in a landmark-esque, "locally anchored" manner before the system becomes "abstractly coordinated" and "terrain-independent".

As the next section will outline, changes in Gurindji have occurred rap-idly, making it necessary to include data from children in this intergenerational snapshot of spatial language use (see Meakins & Algy 2016; Meakins et al. 2016; Meakins & Wilmoth 2020; Van den Bos et al. 2017 for other apparent time studies including Gurindji Kriol-speaking children). Nonetheless, in order to avoid con-flating acquisition and language shift in examining changes in the spatial relations system, we have only included children above the age of nine, in line with previous studies of the acquisition of absolute systems.

3.   Changes in the use of the absolute frame of reference

Gurindji is a Pama-Nyungan language spoken at Kalkaringi, in the southern Victoria River District of the Northern Territory, Australia (see Figure 5.1). Language shift has taken place in the community due to contact between the Gurindji people and *kartiya* (non-Indigenous people), which began following the colonization of the area in the mid to late nineteenth century. Nowadays, Gurindji in its traditional form is spoken only by older generations, with younger adults and children speaking a new language known as Gurindji Kriol. Gurindji Kriol is a mixed language which blends the noun phrase grammar of Gurindji with the verb phrase grammar of Kriol, an English-based creole language, and combines nouns and verbs from both languages in a single linguistic system (Meakins 2013).

Gurindji is noted for its absolute FoR which is used in small-scale spatial description, as is the case in many Australian languages (e.g., Haviland 1998; Laughren 1978; McGregor 2004: 211–221; Schultze-Berndt 2006; Wilkins 1989, 2006). Gurindji has an absolute system consisting of four cardinal terms, each with twenty-eight inflected forms, as well as a river drainage system (Meakins 2011: 45; Meakins et al. 2013: 33–35). Gurindji also has terms for "left hand" (*jampukarra*) and "right hand" (*jutumparra*), but neither is used projectively (Meakins et al. 2013).

The river drainage system is not found in Gurindji Kriol, and the cardinal system paradigm, which has been significantly reduced, does not appear often in descriptions of small-scale space in natural discourse (Meakins 2011: 45). Unlike some of the case studies from other colonial contexts (Section 2), speakers have not supplemented this apparent linguistic "gap" with English left/right terms (despite exposure to standard Australian English and Western schooling), nor have they developed a projective usage of the Gurindji right-hand and left-hand terms. Instead they have opted for deictic references and co-speech gesture for describing small-scale space in everyday contexts Meakins 2011: 43). Nonetheless, Meakins (2011) found that when Gurindji Kriol-speaking adults are obscured
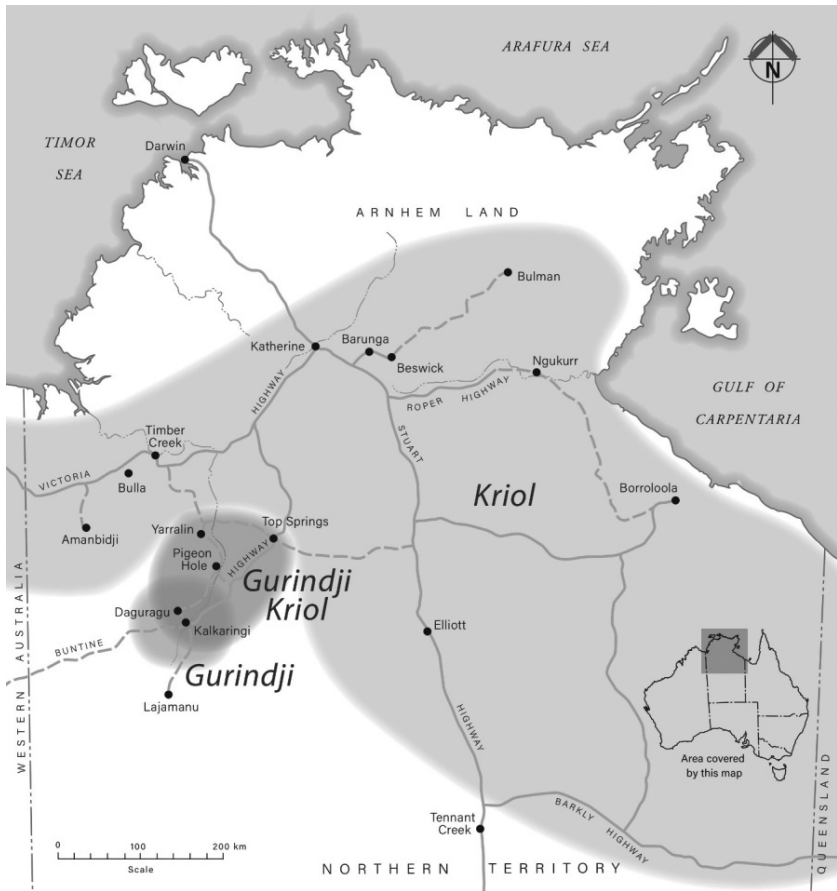
Figure 5.1 Areas in northern Australia where Gurindji, Kriol and Gurindji Kriol are spoken. Gurindji Kriol is now spoken at Kalkaringi, Daguragu, Pigeon Hole and Yarralin (Meakins et al. 2019: 299).

from their interlocutor's view (which means gesture and deictic terms cannot be used) such as in "Man and Tree" tasks, they use cardinal terms to describe small-scale space, thus "revealing" their geocentric mental map. This geocentric orientation was further confirmed using the "Animals in a Row" rotation task, which is a non-linguistic task used to show cognitive orientation (Meakins et al. 2016).

Further intergenerational change in the spatial relations system was explored using the "Which-way-is" task. In this task, Gurindji Kriol-speaking children and a control group of adults were tested on their active and passive knowledge of the cardinal system through a series of questions such as "Which way is [researcher points east, cardinal response given by participant]?" and "Which way is [researcher says "east", participant points towards corresponding direction]?". This task revealed some intergenerational differences; for example, Gurindji Kriol-speaking children

had a lower productive knowledge of cardinal terms than adults (Meakins & Algy 2016: 496), but were found to have a good passive knowledge of the terms, and gave significantly more correct east/west responses than north/south responses. Meakins and Algy (2016: 479) claimed these results were indicative of the children's "continuing attention to geocentric cues" related to the trajectory of the sun. Indeed, in the "Animals in a Row" task, Gurindji Kriol-speaking children also showed similar results to adults, providing further evidence that their mental map of the world is also based on fixed bearings (Meakins et al. 2016).

This chapter extends our understanding of Gurindji children's descriptive repertoire in the domain of space. The "Which-way-is" task was designed to test children and adults' active and passive knowledge of the Gurindji cardinal system, but it did not reveal what linguistic systems children actually utilise in spatial descriptions when they cannot see each other, and therefore cannot use deictic terms. For example, how do children express the north-south axis if the Gurindji terms are not in their spatial inventory?

4. Experimental task: "Man and Tree"

4.1 Procedure

The "Man and Tree" task was developed by Levinson et al. (1992) at the Cognitive Anthropology Research Group (later known as the Language and Cognition Group) at the Max Planck Institute for Psycholinguistics (MPI).[3] The "Man and Tree" task is

---

3. Note that the image number in our study corresponds to a different number in the MPI Field Kit:

Image number in our study	Corresponding image in Levinson (1992)
1.1	3.3
1.2	3.5
1.3	3.6
1.4	3.7
1.5	3.2
1.6	3.8
1.7	3.4
1.8	3.1
2.3	2.3
2.4	2.4
2.5	2.5

a director–matcher task, in which two people sit side-by-side, facing the same direction, with a small wooden board placed vertically between them so they cannot see each other's pictures. The game consists of four sets of minimally different images: Set 1 depicts two men in various standing configurations facing the same direction; Set 2 contains various standing and facing configurations of a man and a tree (see Figure 5.2); Set 3 has the two men facing in different directions; and Set 4 contains images of a man, two trees and a group of pigs in various formations.
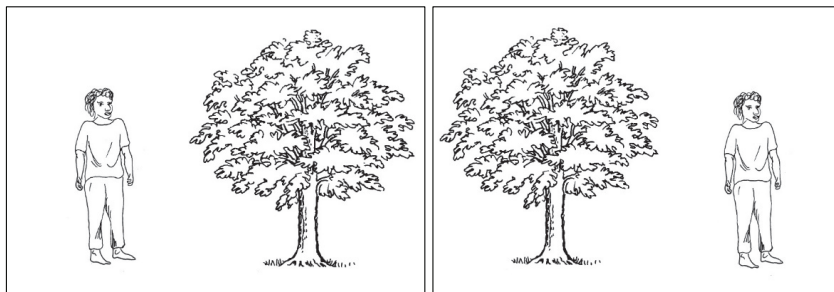
Figure 5.2 Sketches of Set 2 "Man and Tree" images[4]

Image number in our study	Corresponding image in Levinson (1992)
2.6	2.6
2.7	2.7
2.8	2.8
3.1	4.9
3.2	4.10
3.3	4.12
3.4	4.7
3.5	4.11
3.6	4.8
3.7	4.5
3.8	4.6
4.1	4.3
4.2	4.4
4.3	2.1
4.4	4.2
4.5	4.1

4. Note that these are just sketches of the actual MPI pictures, which were photos of a toy man and tree. See Pederson et al. (1998: 564) for reproductions of the complete set of "Man and Tree" pictures.

The director of each session describes each of the images in turn to the matcher. The matcher can ask the director for further information if needed, then attempts to identify the corresponding image from their set and hand it to the director. If the matcher guesses correctly, the director moves on to describe the next image; if the matcher guesses incorrectly, the director hands back the picture, and the pair continue negotiating until the correct image is identified.

4.2   Participants

The "Man and Tree" tasks were conducted with thirty Gurindji Kriol-speaking children and ten Gurindji Kriol-speaking adults. The children's tasks were conducted by Meakins and Algy on two occasions within a two-year period: twenty-one children participated in 2015, and an additional fourteen children participated in 2017. Five sessions were eliminated from the analysis due to recording problems; thus the present study is based on the remaining thirty sessions. Twenty-two of the child directors were female, and eight were male.[5] The participants ranged from nine to seventeen years of age. All but one of the "Man and Tree" sessions were run in outdoor locations.[6] There were always other community members present aside from Algy. Meakins was also present. Caregivers were paid for the children's participation. All sessions were audio-recorded only.

The adult data were collected by Meakins (2009, 2015) under the same task conditions. Thirteen adults ranging from twenty-one to thirty-six years of age participated, but only ten are included in the present study. Two of the thirteen data sets were excluded due to technical problems with data collection and recording, and one of the sessions was performed by Algy and was therefore discounted for methodological reasons. All adult participants were female, and all adult sessions took place outdoors. Following the data collection sessions, the audio-recordings were transcribed, annotated, and translated using CLAN (MacWhinney 2000) with each utterance sound-linked.

4.3   Coding

For this study, each spatial description produced by participants was coded as one of five strategies, based on Brown & Levinson (2009): cardinal, landmark,

---

5. Gender is not thought to play a significant role in linguistic and non-linguistic FoR preferences (e.g., Levinson 2003: 190–197; Van Leeuwen 1978), so the high proportion of female participants should not have an effect on the results (cf. Bohnemeyer & Stoltz 2006: 308–309).

6. See Li & Gleitman (2002) and the subsequent response from Levinson et al. (2002) for a debate on the possible influence of indoor versus outdoor location in similar experimental tasks.

intrinsic, relative or deictic. The facing direction of the speakers had been noted for each session, and thus each cardinal token was further coded as either "correctly oriented" or "incorrectly oriented". Any utterance prompted by an onlooker present during the recording session was discarded, thus only spontaneous exemplars of the strategies are included in the analysis. The strategies, as well as some English-translated examples of these, are given in Table 5.1.

Table 5.1 Examples of spatial description strategies used in the "Man and Tree" task (English translations – except relative terms "left" and "right" which are code-switched from English)

Cardinal	Landmark	Intrinsic	Relative	Deictic
tree on north side	man looking at [NAME]	men facing each other	tree on left side	man facing this side
men looking west	men looking at television	man facing tree	man on the right	men facing this way and that way
one looking north, one looking south	tree towards shop	tree behind man	men facing forwards and back	tree on that side
	men looking at Yarralin (100 km away)			

The main interests for our study are the cardinals and landmarks (shown in the first two columns of Table 5.1). "Cardinal" refers to tokens using the Gurindji-derived terms *kayirra* "north", *kurlarra* "south", *kaarnirra* "east" and *karlarra* "west". This system is based on fixed bearings and thus considered to be an absolute FoR system. These were further analyzed and coded as "correctly oriented" or "incorrectly oriented" (as discussed in Section 4.4.1).

"Landmark" is a less straightforward category to code, because there is still no real consensus in the literature on how to define and treat FoR descriptions involving landmarks (see e.g., Levinson 2003: 47–50; Li & Gleitman 2002; O'Meara & Pérez Báez 2011: 841–843; Pederson 1993: 295; Terrill & Burenhult 2008: 122–125).[7] In this study, we take our lead from Brown & Levinson (2009) and

---

7. Levinson et al. (2002: 172) argue that landmarks are basically a type of (large-scale) intrinsic system, and classify them as such (see also Levinson 2003: 81). Dasen and Mishra (2010: 135) consider landmarks to be an intermediate between intrinsic and absolute FoRs, while Li and Gleitman (2002) classify them as "allocentric" (as opposed to "egocentric"). Brown and Levinson (2009: 456), who include a separate landmark category in their FoR coding schema, maintain that landmark systems are conceptually very different from abstract

use the term "landmark" to refer to any concrete picture-external entity, including speech act participants and onlookers. We distinguish between two broad sub-categories: "visible" and "non-visible", with the "visible" category further divided into "community" (specific fixed landmarks known to all community members) and "local" (terms with moveable referents or referents which do not have a single fixed referent at the community level). See Table 5.2 for exemplars of each subcat-egory of landmarks. This categorisation becomes relevant in Section 4.5.2.

Table 5.2 Categories and exemplars of landmarks used in the "Man and Tree" task

Visible: local	Visible: community	Non-visible
car, wall, television, kitchen, speech act participants, third person (onlookers), house, gate, fence	shop, school, church, clinic, office, club, police station, hill, river, arts centre, Lawi[*]	Halls Creek, Nitjpurru, Top Springs, Yarralin[**]

Notes:
*Lawi refers to a river crossing located west of the Daguragu community.
**Halls Creek refers to a place 500 km west of Kalkaringi, Nitjpurru is 100 km north, Top Springs is 170 km north and Yarralin is 200 km north.

4.4   Summary of major strategies used

Table 5.3 provides a summary of the Gurindji Kriol-speaking children and adults' spatial description strategies used in the "Man and Tree" sessions. The results for each strategy are also given as a percentage of the total spatial descriptions pro-duced by each age group.

Table 5.3 Spatial description strategies used in the "Man and Tree" task

	Cardinal	Landmark	Intrinsic	Deictic	Relative	Total
Children	404 (34.5%)	556 (47.5%)	111 (9.5%)	53 (4.5%)	44 (4%)	1168
Adults	329 (77.5%)	19 (4.5%)	75 (17.5%)	2 (0.5%)	0 (0%)	425

---

absolute systems. Even amongst studies that do treat landmarks as a separate category, the cri-teria used for classifying a ground as a "landmark" are not always specified or clear, and vary from study to study. In particular, instances where the speaker or other persons present are used as the ground in a spatial relations description have eluded consistent FoR classification in the literature. For example, Danziger (2010) proposes that descriptions with a speech act participant as ground be classified under a fourth FoR, "Direct", whereas Terrill and Burenhult (2008: 120–121) believe descriptions such as "the man is facing me/you/her" (which they call "person deixis") can be classified as intrinsic. However, many other authors classify speech act participants and third person references as landmarks (e.g., Bohnemeyer & Stoltz 2006: 305; Brown & Levinson 2009: 457–459; Polian & Bohnemeyer 2011: 878; Senft 2006: 226).

Table 5.3 shows that there is variation in the strategies chosen by the Gurindji child participants to express spatial relations. Cardinals and landmarks are the most frequently-used strategies. This use of cardinals and landmarks is exemplified in Example (1), where a child director establishes the man's facing direction using an out-of-sight landmark cue, *Daguragu*, and describes the position of the tree relative to the man using a cardinal, *kaarnirra-said* "east-side". Note that in all examples, Gurindji-derived words are italicised, Kriol-derived words are given in Roman font and spatial terms are given in bold.[8]

(1) Dat *karnti* **kaarnirra**-**said** dat man luk-in-at-*karra*   **Daguragu**
     the tree   east-SIDE   the man look-CONT-at-CONT Daguragu
     "The tree is on the east side and the man is looking at Daguragu."
                       (JB, 15yrs: FM15_a281: 2.25min: Image 2.3, facing north)[9]

Table 5.3 also shows that cardinals were the preferred strategy for Gurindji Kriol-speaking adults during the "Man and Tree" task, as illustrated in (2). The adult data show very little variation, with the use of landmarks (and other strategies) rare.

(2) ***Kurlarra-k*** *ngumpit=ma* im feis-ing   *karnti=ma* im
     south-ALL   man=TOP   3SG face-CONT and tree=TOP   3SG

     top **kaarnirra**-**said**-*ta.*
     be   east- SIDE-LOC
     "The man is facing southwards and the tree is on the east side."
                       (AR, 27yrs: FM09_a113: 5.22min: Image 2.6, facing south)

All other strategies – intrinsic, relative, deictic – are present, but marginal. Interestingly, there is only a small uptake of relative spatial terms in the child data, in this case, "left" and "right", which are borrowed from English, for example in (3).

(3) Wan man feis-ing   Halls Creek roud an dat *karnti* im **left**-**said**.
     a   man face-CONT Halls Creek road and the tree 3SG left-SIDE
     "One man is facing Halls Creek road and the tree is on the left-hand side."
                       (EM, 13yrs: FM15_a279: 3.52min: Image 2.8, facing north)

The Gurindji terms *jampukarra* "left-hand" and *jutumparra* "right-hand" have not developed beyond body-part terms, i.e. they are not used projectively by younger

---

8.   Abbreviations used in the examples: 3 = third person, ALL = allative, CONT = continuative, DU = dual, LOC = locative, NMLZ = nominaliser, S = subject, SG = singular, TOP = topic.

9.   The metadata in each example contain the speaker's initials, e.g., "JB" (identifiable to the speech community, anonymous to outsiders, as per consent conditions); the recording number, e.g., "FM15_a281" (FM = Felicity Meakins, 15 = 2015, a = audio, 281 = recording session number); the start time of utterance in the recording, e.g., "2.25min"; the "Man and Tree" image number, e.g., "Image 2.3"; and the speaker's facing direction, e.g., "facing north".

generations. The marginal uptake of the English relative terms stands in contrast to studies of similar language contact situations (outlined in Section 3, where a shift to a relative FoR is evident.

A small number of deictic terms were also used but this probably relates to the design of the task. The adult participants had little difficulty adapting to the non-naturalistic constraints of the task, which prevented them from using deixis and gesture, but the child participants generally found the task more difficult. It is there-fore not surprising that some children were inclined, at least initially, to describe the images using deixis, as they would in day-to-day naturalistic conversation. This is illustrated in (4), where HE instinctively attempts to use both deixis and gesture to communicate spatial relation information to her matcher. The use of a deictic term by a child director was typically followed by a question of clarification from the researchers, and the director was usually able to rephrase their description using a non-deictic strategy, as illustrated by CW and Cassandra Algy's exchange in (5).

(4) Dat man im feis-ing   streit-ap   bat dat tri   im   dat-said
     the man 3SG face-CONT straight-up but the tree 3SG that-SIDE

     luk   bro.
     look   bro

     "The man is facing straight up but the tree is on that side, look mate
     [addressing matcher]."
                    (HE, 11yrs: FM15_a277: 2.48min: Image 2.6, facing south-west)

(5) CW: Dij-said.
          this-side
          "This side."

     CA: Wat   dijei?
          what   this.way
          "What [do you mean by] 'this way'?"

     CW: *Kaarnirra.*
          East
          "East."
          (CW, 10yrs; CA, 27yrs: FM15_a270: 2.49min: Image 2.8, facing south)

Intrinsic, relative, deictic strategies will not be discussed further in this chapter. The next two sections consider the use of cardinals and landmarks in more detail.

4.4.1   *Children's use of cardinals*

Cardinals were one of the major spatial relations strategies used by Gurindji Kriol-speaking children during the "Man and Tree" task. A total of 404 cardinal descrip-tions occurred in the data, representing 34.5% of the 1,168 child descriptions. This is proportionally less than in the adult group, where cardinals represented 77.5%

of all spatial descriptions. Table 5.4 shows the relative use of the different cardinals in both the child and adult data.

Table 5.4 Cardinal tokens produced by Gurindji Kriol-speaking children and adults

	East (*kaarnirra*)	West (*karlarra*)	South (*kurlarra*)	North (*kayirra*)	Total
Children	150 (37%)	134 (33%)	71 (18%)	49 (12%)	404
Adults	75 (23%)	97 (29%)	78 (24%)	79 (24%)	329

Table 5.4 shows that the cardinal most commonly used by the children was *kaarnirra* "east" (n = 150, 37%), followed closely by *karlarra* "west" (n = 134, 33%), with *kurlarra* "south" (n = 71, 18%) and *kayirra* "north" (n = 49, 12%) appearing much less frequently in the data. In contrast, the four cardinals occurred with fairly equal frequency in the adult data.

In terms of individual performances, twenty-seven of the thirty children attempted to use at least one cardinal term. Twenty-one children produced the term *kaarnirra* "east" on at least one occasion during the task. Twenty children produced *karlarra* "west", fifteen produced *kurlarra* "south", and eight produced *kayirra* "north". Seven of the thirty children produced at least one token of each cardinal, but no child under the age of thirteen produced all four cardinal terms. The utterance in (6) shows the use of the cardinal term *kaarnirra* "east" to provide information about the facing orientation of the man, and the cardinal term *karlarra* "west" to describe the location of the tree.

(6) Dat boi luk-in-at   ***kaarnirra*** an dat tri im ***karlarra*-said**.
     the boy look-CONT-at east   and the tree 3SG west-SIDE
     "The boy is looking east and the tree is on the west side."
                                (XR, 11 yrs: FM15_a271: 2.52min: Image 2.4, facing south-west)

Although cardinals were a major strategy for the children, they were not always used correctly. Table 5.5 shows the number of cardinal tokens used by children with the correct orientation in comparison to adults. This table also shows the correct uses of each cardinal as a proportion of the total uses, which are given in Table 5.5.

Table 5.5 Cardinal tokens used with correct orientation by children and adults. (Percentages represent correct cardinal tokens as a proportion of the total uses of that cardinal in the data.)

	East (*kaarnirra*)	West (*karlarra*)	South (*kurlarra*)	North (*kayirra*)	Total
Children	125 (83%)	85 (63%)	39 (55%)	37 (76%)	286 (71%)
Adults	70 (93%)	87 (90%)	74 (95%)	65 (82%)	296 (90%)

Table 5.5 shows that Gurindji adults produced cardinals correctly (90% of the time) more often than children (71% of the time). With respect to the individual cardinals, children produced *kaarnirra* "east" (83%) and *kayirra* "north" (76%) correctly more often than *karlarra* "west" (63%) and *karlarra* "south" (55%). In terms of their individual performances, twenty-one of the twenty-seven children who produced cardinal terms used at least one cardinal term consistently with the correct orientation. Nine participants used one term correctly, seven used two, three used three, and two children used all four terms correctly. Many of the errors were instances of the children using *kurlarra* "south" to designate west, and *karlarra* "west" to designate south. This confusion is likely due to the phonological similarity of *kurlarra* and *karlarra*, and to the fact that some children may have been aware of the researchers' interest in the Gurindji cardinal system and thus attempted to use cardinal terms when they were not confident to do so.

Children's knowledge of cardinals can also be shown to follow an implicational hierarchy: *kaarnirra* > *karlarra* > *kurlarra, kayirra*; i.e. consistent correct usage of *karlarra* "west" by an individual implies correct usage of *kaarnirra* "east", but not necessarily correct usage of the other cardinals. If a child uses *kurlarra* "south" or *kayirra* "north", or both, with the correct orientation, they also used both *kaarnirra* "east" and *karlarra* "west" correctly.

4.4.2   *Children's use of landmarks*

Table 5.6 shows that landmarks represent 47.5% of the total 1,168 spatial descriptions given by children, making landmarks their dominant strategy. Almost all the children made use of landmark cues to some extent, and seventeen of the thirty children utilised landmarks for over 50% of their spatial descriptions. The highest proportion of landmark use by an individual child was 95%. In contrast, landmark descriptions in the adult data made up just 4.5% of the total strategies. Only three of the ten adults utilised any landmark cues, and the highest percentage of landmark uses by an individual adult was 41.5%.

Table 5.6 summaries the different types of landmarks children used to orientate the entities pictured in the "Man and Tree" images. Also provided is the percentage of the overall strategies (given in Table 5.6) that each landmark substrategy use represents.

Table 5.6 Landmark strategies in Gurindji Kriol child and adult data including percentage of overall strategies

	Visible: Local	Visible: Community	Non-visible	Total
Children	118 (10%)	341 (29%)	97 (8.5%)	556 (47.5%)
Adults	14 (3.5%)	5 (1%)	0 (0%)	19 (4.5%)

Visible community-level landmarks were the most common strategy used, in particular buildings of central importance to the community, as well as natural features of the landscape. These landmarks represent 29% of all spatial description strategies in the child data. An example is given in (7), where "hill" is used to describe the man's facing orientation as well as the position of the tree. Similarly, in (8) the community church is used to describe the location of the tree. Visible local landmarks were also not uncommon, constituting 10% of all spatial description strategies used. The use of *ngayu* "me" in (8) is one such example, where the speaker uses herself as a local-level "landmark" to give information about which way the man is facing. Other common local landmarks were non-fixed cues such as another human in the recording context, a car, or generic cues such as houses, gates and fences. Finally, non-visible landmark cues typically involved the names of distant towns or communities, such as Nitjpurru, a community fifty kilometres north of Daguragu, as shown in (9) (see also Figure 5.1 for the location of other places, and (1) and (3) for more examples of landmarks).

(7)   Dat   man   im   luk-in-at-*karra*   **hil-said**   an   dat   tri
      the   man   3SG   look-CONT-at-CONT   hill-SIDE   and   the   tree

      im   **hil-said**.
      3SG   hill-SIDE

      "The man is looking towards the hill, and the tree on the hill side."
                  (KH, 11 yrs: FM15_a272: 2.55min: Image 2.5, facing south-west)

(8)   Dat   man   luk-in-at   *ngayu*   an   dat   tri   im
      the   man   look-CONT-at   1SG   and   the   tree   3SG

      dat-said   **church-said**.
      that-SIDE   church-SIDE

      "The man is looking at me and the tree is on that side, the church side."
                  (XR, 11 yrs: FM15_a271: 4.17min: Image 2.7, facing south-west)

(9)   Dat   boi   im   luk-in-at-*karra*   **Nitjpurru-area-*ngka***.
      the   boy   3SG   look-CONT-at-CONT   Pigeon.Hole-AREA-LOC
      "The boy is looking in the direction of Pigeon Hole."
                  (GI, 14 yrs: FM15_a280: 3.23min: Image 2.4, facing south-west)

Non-visible landmarks were the least frequently used, but this may be related to the task location, as most of the sessions were performed in communities where visible landmarks were present.

4.5   Discussion

It is evident from the "Man and Tree" results that the children's usage of cardinals differs from that of the adults. Almost all of the adults in the study used all

four cardinal terms as their dominant strategy, and used them consistently and accurately. In contrast, the cardinal system was not the children's dominant spatial relation strategy, with the use of landmark terms more frequent. In this section, we argue that variation in the use of cardinal and landmark terms is not simply a case of incomplete or continuing acquisition, but is representative of language change (Section 4.5.1). Furthermore, we argue that children are not just using landmarks to supplement their fragmented knowledge of cardinal terms, but rather that the variation represents the emergence of a combined repertoire of terms and strategies, and a principled reorganisation of the spatial relation system that is still largely grounded in geocentric cues (§ 4.5.2).

4.5.1   *Acquisition or language change?*

In this section, we make the argument that the generational differences are indicative of language change (or at least competition between variants), rather than acquisition. As outlined in §2, based on De León's (1994) and Brown & Levinson's (2000) studies, it appears that children begin to use absolute terms in an abstract way at around eight to nine years of age. The Gurindji Kriol-speaking children who participated in this study were aged nine years and older, and had therefore probably already reached their ultimate attainment of cardinal terms. Figure 5.3 shows the number of correctly-oriented cardinals used by each participant (as a percentage of their total strategies) plotted against their age; there is a visible difference between the children's and adults' cardinal usage in the "Man and Tree" task, but within the child age group there is no correlation between a child's age and their use of (correctly-oriented) cardinal tokens.
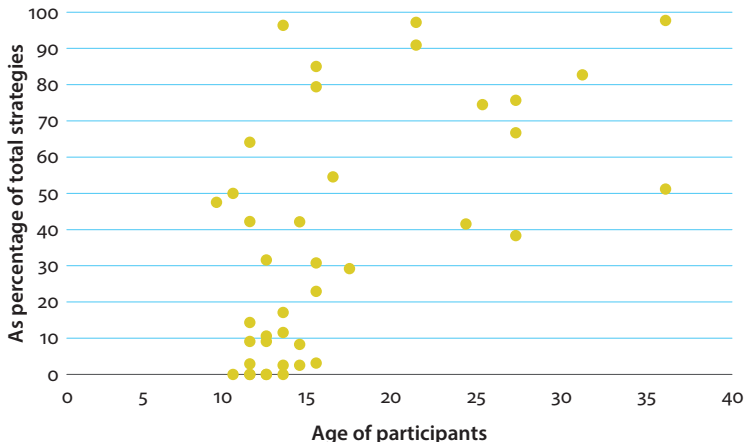

Figure 5.3 Individual use of cardinals with the correct orientation (as a percentage of total strategies). Each dot represents an individual child (9–17 years of age) or adult (21–36 years of age).

Figure 5.3 also shows that children from the same age group can have vastly different knowledge and usage of the cardinal system. For example, 15-year-old DB produced only three cardinal tokens during the task (only one of which was used correctly). In contrast, 15-year-old RA used 38 cardinals (90% of which were correctly oriented), which represents 85% of her total spatial description strategies. Furthermore, some of the youngest children, such as nine-year-old AS and ten-year-old CW, produced cardinal terms far more frequently and accurately than many of the older children.[10]

This generational difference, coupled with large individual differences among child participants of similar ages, is also reflected in the landmark data. Figure 5.4 shows the use of landmarks (as a percentage of total strategies) of each participant plotted against their age.
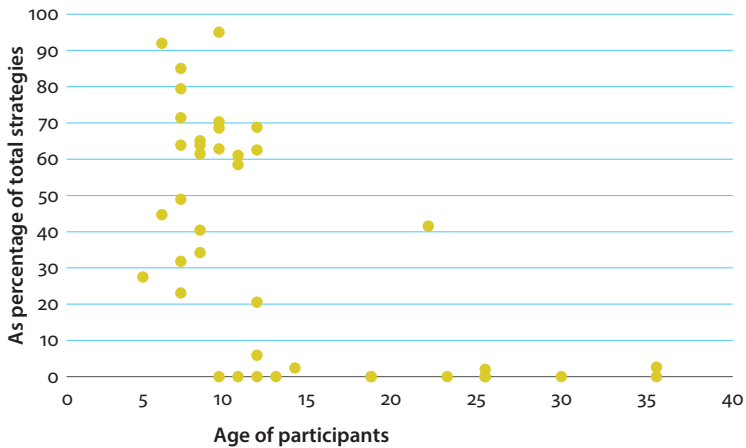

Figure 5.4 Individual use of landmarks (as a percentage of total strategies). Each dot represents an individual child or adult participant.

Figure 5.4 also shows there is a clear difference between the two generations, but no correlation between age and landmark usage within the child group. Similar to the cardinal data, children of the same age often demonstrate different degrees of landmark usage; for example, landmarks make up 85% of 11-year-old KH's spatial description strategies, while they represent just 23% of strategies for 11-year-old JJ. Similarly, landmarks represent 69% of strategies for 15-year-old JB, in comparison

---

10.   Nine-year-old AS produced 25 cardinals, 76% of which were correct (47.5% of her total strategies), and 10-year-old CW produced 20 (95% correct, 50% of total strategies). In contrast, 15-year-old DB used just three cardinals (33% correct), and although 17-year-old NS used cardinals frequently only 31% were correct.

to just 21% for 15-year-old MI. This pattern provides extra evidence for our claim that children's use of spatial description strategies can be attributed to intergenerational language change in the Gurindji community, rather than to individual children's acquisitional stage or cognitive development.

Although we make the claim for intergenerational language change in the Gurindji community's spatial relation system, there are probably some acquisitional factors at play in the use of visible local landmarks. Figure 5.5 shows that there is a decreasing reliance on this subcategory of landmarks with age.
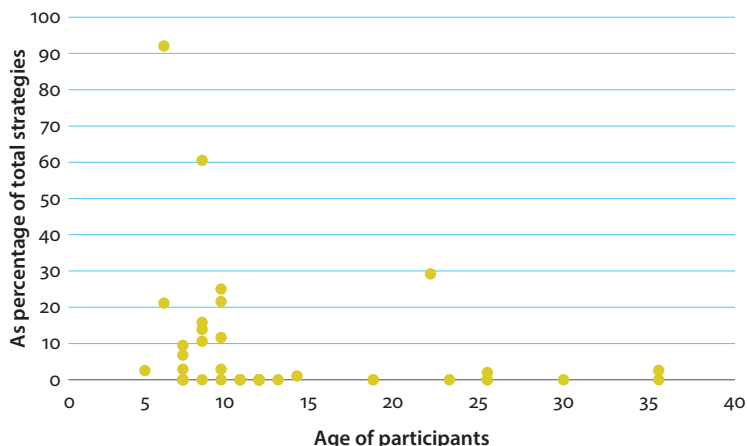

Figure 5.5   Local landmarks as a percentage of total strategies

In terms of individual performances, one of the youngest children, 10-year-old SY, used visible local landmarks in almost all her descriptions, and aside from one adult (who is an outlier within her age group), no person over the age of 13 used local landmarks in more than 3% of their spatial descriptions. It therefore appears that the use of local landmarks tapers off with age and may be related to language acquisition. It is possible that the use of local landmarks is a gateway to the use of community landmarks. However, no children below the age of nine were tested, and previous studies have shown that much of a child's linguistic FoR acquisition happens at or before this age; we would thus need to test younger children for stronger evidence of this pattern.

We could hypothesise further that these community landmarks are ultimately a stepping stone to non-visible landmarks as the children's mental maps of their world expand and develop. Note that relatedly, De León (1994: 878–880, 1995: 26–32) claims that children acquiring absolute systems pass through a "landmark-stage" first (in our data, however, we do not find direct evidence of this particular acquisitional pattern).

Although acquisition provides an explanation for some aspects of the children's performance, the lack of a correlation between age and cardinal and landmark usage suggests that the children have fully acquired the Gurindji cardinal system to the extent that they will at or before nine years of age (the age of the youngest participant in the study). That is, although they are not using the full Gurindji cardinal system, they have fully acquired the spatial system which they will use into adulthood. This finding is in line with Meakins & Algy's (2016: 490–491) study, which also shows that differences between the child and adult data in the "Which-way-is" task can be attributed to intergenerational language change (see § 3). Note, though, that re-testing the children in a decade may reveal different usage patterns. For example, children may have reached their ultimate attainment in acquiring a spatial relations system by the age of nine, but they may also increase the usage of the cardinal system over the landmark system as they enter adulthood and are exposed more to travel. The repertoire will likely remain combined, but the relative frequencies may change. The question then is: what does the shift from a cardinal system to a combined repertoire of cardinal and landmark terms represent?

4.5.2   *A landmark-based or quasi-absolute system?*

The case for a shift from an absolute system to a landmark system is compelling, given that landmarks are used extensively by children, but rarely by adults. Indeed, even the change in the relative frequency of cardinal use is suggestive of such a shift. Recall that the children's use of cardinals follows an implicational hierarchy, i.e., they are more likely to use *kaarnirra* "east" and *karlarra* "west" consistently with the correct orientation than the other two terms. Moreover, *kaarnirra* emerges as the term most likely to be used and to be used correctly by each individual child. Thus, their orientation to the east-west axis remains strong, which is in line with Meakins & Algy's (2016: 479) findings showing that Gurindji Kriol-speaking children are also more likely to give correct responses for the east-west axis in the "Which-way-is" task. They claim that this result is explained by the children's "continuing attention to geocentric cues" (2016: 479): namely, the location of the rising and setting sun on the horizon.

Nonetheless, it is not entirely clear whether cardinal points remain the abstract basis of an absolute FoR system, or have become orientating objects within a landmark system. One interpretation of the continued use of the east-west axis is that the sun (more specifically, the location of sunrise and sunset) has become a landmark, and that the cardinal system is no longer abstracted to the other quadrants of the world, i.e., "north" and "south". Indeed many of the child participants mixed cardinals with landmarks during the "Man and Tree" task, for example ED's use of *shop* [west] and *kaarnirra* [east] in (10).

(10)   Wan man im feis-ing **shop** an   wan-bala   im
      one   man 3sg face-CONT shop   and one-NMLZ 3sg

      feis-ing   *kaarnirra.*
      face-CONT   east

"One man is facing the shop [west] and one is facing east."
(ED, 11 yrs: FM15_a278: 5.24min: Image 3.2, facing south-west)

Other researchers have called attention to the fact that determining whether a seemingly absolute linguistic system is actually based on abstract fixed bearings is more difficult than has commonly been claimed (e.g., Bohnemeyer & Stoltz 2006: 305; Senft 2006: 226–227; Terrill & Burenhult 2008: 122–125). Many of the absolute coordinate systems reported in the literature are linguistically abstracted from visible landmark features such as slope or river drainage (Levinson 2003: 45). Even quite abstract systems have been argued to be landmark-based. For example, Terrill & Burenhult (2008: 124–125) report that it is likely that the cardinal terms used by Lavukaleve speakers, which seem to be absolute at first glance, "do not really represent abstract bearings, but operate in a way similar to landmarks". In cases where "north" and "south" are not encoded, the visible basis of the system is less concrete and therefore it is harder to make the argument for abstraction.

Note too that there are clear cases where cardinal terms have become more concrete in nature and fixed to a location. For example, many Guugu Yimithirr children use *naga* "east" to refer to a specific beach located east of their community (De León 1995: 26). Even in the Gurindji community, the cardinal term *kayirra* "north" is often used in a landmark-like way to refer specifically to Daguragu when people are travelling there from Kalkaringi (which can be many times per day).

One way to determine whether an absolute system is really abstract is to see if it is used by speakers in areas that are unfamiliar to them. When "Man and Tree" participants are tested in their local community, as has been the case with this and most FoR studies, it is not possible to distinguish landmark-like uses of seemingly absolute terms from truly abstracted uses, because the two are conflated (Polian & Bohnemeyer 2011: 875–876). This method is a potential future path for exploring absolute systems and abstraction.

Finally, spatial descriptions involving landmarks have been notoriously difficult to define and categorise. Brown and Levinson (2009: 456) give conceptual reasons for treating landmarks differently from cardinals, stating that a language with an absolute FoR requires its speakers to have an "internalised mental compass", with the implication being that language with a landmark-based system does not. Although this may be true for systems only involving visible local landmarks, Gurindji child participants also referred to landmarks that were not visible and were many kilometres away. For example, in (11), MI describes the two men's

facing direction by accurately referring to a place more than fifty kilometres away from the task location. His ability to invoke a faraway, non-visible landmark with minimal obvious reflection and effort is surely an indication that landmark-based FoR systems, much like absolute systems, *can* require speakers to have an "internalised mental compass".

(11)   Man dat-tu *kutij-karra* long-wei an dei feis-ing **Wave.Hill**.
       Man the-DU stand-CONT long-WAY and 3PL.S face-CONT Wave.Hill
       "The two men are standing far apart and they're facing Wave Hill."
                                      (MI, 15 yrs: FM17_a481: 4.24min: Image 1.3, facing north)

This example is of particular interest because MI's "Man and Tree" session was performed outside of the community, away from visible community landmarks such as buildings. MI had no trouble coopting non-visible landmarks to express the relative position of the man and tree. It could be that this strategy is a part of the children's spatial strategy repertoire, but only relied upon when other strategies such as visible landmarks are not possible. Indeed, Brown & Levinson (2009: 460) found that adult participants in the Tzeltal community produced landmark terms quite commonly in "Man and Tree" tasks despite having an absolute system. They attributed this use of landmarks to participants opting for position descriptions with more angular accuracy than was permitted by absolute slope terms. Meakins and Algy (2016: 15) also found that Gurindji Kriol-speaking children's ability to correctly point to non-visible locations is high (68%). Moreover, Meakins et al.'s (2016) study using the non-linguistic rotation task "Animals in a Row" found that the majority of Gurindji Kriol-speaking children gave viewpoint-independent responses, also suggestive of children's awareness of geocentric coordinate systems.[11] Thus the shift to a landmark system does not seem to be at odds with having a mental compass, and it could be that landmarks are used as quasi-absolute terms. Further exploration of the children's use of cardinals outside of their local community is required, however, before we can discount the possibility that they are using cardinal terms in a "landmark-esque" manner.

Thus, although much variation is present, some emergent dominant patterns belie an interpretation of random variation or simply the fragmentation of a geocentric system. Landmarks are coopted to fill in gaps in the paradigm left by a decreasing linguistic knowledge of the north-south axis. Although the system is

---

11.   One caveat is that Levinson et al. (2002: 159) note that "such an experiment … distinguishes between egocentric and allocentric reference frames, but it does not precisely distinguish what kind of allocentric reference frame is involved. Allocentric frames of reference include both absolute and intrinsic ones."

reconstituted linguistically, it has the same geocentric basis which is grounded in the children's intimate knowledge of their individual experienced worlds.

5.   Conclusion

This chapter has shown that intergenerational changes in the linguistic expression of spatial relations is under way in the Gurindji speech community, but the conceptual system remains largely geocentric. Results from the "Man and Tree" task demonstrate that the Gurindji community's preferred strategies for describing small-scale space are shifting, with landmarks emerging as the dominant strategy for the child generation. We argue that the intergenerational change in the spatial relation system involves a linguistic shift in terminology (cardinal → landmark), but that the new terms are perhaps being integrated into a quasi-absolute system such that the conceptualisation of space remains largely geocentric; further testing of children outside of their local communities is required to confirm this. This mental map takes time to develop in children. They first go through a stage of coopting local landmarks in their immediate communicative context to describe the relative position of objects. This later expands to community landmarks, as children develop a mental map of their immediate experienced surrounds. We hypothesise further that more distant landmarks, including places not visible at the time of spatial description, become a part of the children's linguistic repertoire used to position objects in space.

Throughout this chapter we have juxtaposed the "Man and Tree" results from children with those of adults, and have generally claimed that the differences observed can be attributed to intergenerational language change, rather than acquisitional factors. Nevertheless, there are gaps in the literature on how and when children acquire the absolute FoR systems. Even less is known about the relationship between absolute and landmark FoR systems. The complex language ecology and ongoing language shift at Kalkaringi further complicate matters. Longitudinal spatial relations data from the current generation of Gurindji children will clarify the analysis presented in this chapter. Ultimately, retesting child participants from this study in ten years' time will tease out processes of language acquisition from those of language change.

Acknowledgements

References

Adamou, E. 2017. Spatial language and cognition among the last Ixcatec-Spanish bilinguals (Mexico). In *Multidisciplinary Approaches to Bilingualism in the Hispanic and Lusophone World* [Issues in Hispanic and Lusophone Linguistics 13], K. Bellamy, M. W. Child, P. González, A. Muntendam & M. C. Parafita Couto (eds), 175–207. Amsterdam: John Benjamins. https://doi.org/10.1075/ihll.13.08ada

Bohnemeyer, J. & Stoltz, C. 2006. Spatial reference in Yukatek Maya: A survey. In *Grammars of Space: Explorations in Cognitive Diversity*, S. C. Levinson & D. Wilkins (eds), 273–310. Cambridge: CUP. https://doi.org/10.1017/CBO9780511486753.009

Brown, P. 2006. A sketch of the grammar of space in Tzeltal. In *Grammars of Space: Explorations in Cognitive Diversity*, S. C. Levinson & D. Wilkins (eds), 230–272. Cambridge: CUP. https://doi.org/10.1017/CBO9780511486753.008

Brown, P. & Levinson, S. C. 2000. Frames of spatial reference and their acquisition in Tenejapan Tzeltal. In *Culture, Thought, and Development*, L. Nucci, G. Saxe & E. Turiel (eds), 167–197. Mahwah NJ: Lawrence Erlbaum Associates.

Brown, P. & Levinson, S. C. 2009. Language as mind tools: Learning how to think through speaking. In *Crosslinguistic Approaches to the Psychology of Language: Research in the Traditions of Dan Slobin*, J. Guo, E. V. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura & S. Ozcaliskan (eds), 451–464. New York NY: Psychology Press.

Cablitz, G. 2002. The acquisition of an absolute system: Learning to talk about space in Marquesan (Oceanic, French Polynesia). In *Space in Language: Location, Motion, Path, and Manner*, E. V. Clark (ed.), 40–49. Stanford CA: CSLI.

Danziger, E. 2010. Deixis, gesture, and cognition in spatial frame of reference typology. *Studies in Language* 34: 167–185. https://doi.org/10.1075/sl.34.1.16dan

Dasen, P. & Mishra, R. 2010. *Development of Geocentric Spatial Language and Cognition: An Ecocultural Perspective*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511761058

De León, L. 1994. Exploration in the acquisition of geocentric location by Tzotzil children. *Linguistics* 32: 857–884. https://doi.org/10.1515/ling.1994.32.4-5.857

De León, L. 1995. *The Development of Geocentric Location in Young Speakers of Guugu Yimithirr* [Working Paper 32]. Nijmegen, Netherlands: Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics.

Edmonds-Wathen, C. 2014. Influences of indigenous language on spatial frames of reference in Aboriginal English. *Mathematics Education Research Journal* 26: 169–192. https://doi.org/10.1007/s13394-013-0085-4

Haviland, J. 1998. Guugu Yimithirr cardinal directions. *Ethos* 26: 25–47. https://doi.org/10.1525/eth.1998.26.1.25

Labov, W. 1963. The social motivation of a sound change. *Word* 19: 273–309. https://doi.org/10.1080/00437956.1963.11659799

Laughren, M. 1978. Directional terminology in Warlpiri. *Working Papers in Languages and Linguistics* 8: 1–16.

Levinson, S. C. (ed.) 1992. *Space Stimuli Kit 1.2: November 1992*. <http://fieldmanuals.mpi.nl/volumes/1992/man-tree-space-games/> (24 May 2020).

Levinson, S. C. 2003. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511613609

Levinson, S. C., Brown, P., Danziger, E., De León, L., Haviland, J. B., Pederson, E. & Senft, G. 1992. Man and Tree & Space Games. In *Space Stimuli Kit 1.2: November 1992*, S. C. Levinson (ed.) 7–14. Nijmegen: Max Planck Institute for Psycholinguistics. <http://fieldmanuals.mpi.nl/volumes/1992/man-tree-space-games/> (20 April 2020).

Levinson, S. C., Kita, S., Haun, D. B. M. & Rasch, B. H. 2002. Returning the tables: Language affects spatial reasoning. *Cognition* 84: 155–188. https://doi.org/10.1016/S0010-0277(02)00045-8

Levinson, S. C. & Wilkins, D. 2006. The background to the study of the language of space. In *Grammars of Space: Explorations in Cognitive Diversity*, S. C. Levinson & D. Wilkins (eds), 1–23. Cambridge: CUP. https://doi.org/10.1017/CBO9780511486753

Li, P. & Gleitman, L. 2002. Turning the tables: Language and spatial reasoning. *Cognition* 83: 265–294. https://doi.org/10.1016/S0010-0277(02)00009-4

MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk*. Mahwah NJ: Lawrence Erlbaum Associates.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M. & Levinson, S. C. 2004. Can language restructure cognition? The case for space. *Trends in Cognitive Sciences* 8: 108–114. https://doi.org/10.1016/j.tics.2004.01.003

Marghetis, T., McComsey, M. & Cooperrider, K. 2014. Spatial reasoning in bilingual Mexico: Delimiting the influence of language. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, P. Bello, M. Guarini, M. McShane & B. Scassellati (eds), 940–945. Austin TX: Cognitive Science Society.

McGregor, W. B. 2004. *The Languages of the Kimberley, Western Australia*. London: Routledge.

Meakins, F. 2009. The case of the shifty ergative marker: A pragmatic shift in the ergative marker in one Australian mixed language. In *The Role of Semantics and Pragmatics in the Development of Case* [*Studies in Language Companion Series 108*], J. Barðdal & S. Chelliah (eds), 59–91. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.108.06mea

Meakins, F. 2011. Spaced out: Intergenerational changes in the expression of spatial relations by Gurindji people. *Australian Journal of Linguistics* 31: 43–77. https://doi.org/10.1080/07268602.2011.532857

Meakins, F. 2013. Gurindji Kriol. In *The Survey of Pidgin and Creole Languages*, Vol. 3, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 131–139. Oxford: OUP.

Meakins, F. 2015. From absolutely optional to only nominally ergative: The life cycle of the Gurindji Kriol ergative suffix. In *Borrowed Morphology*, F. Gardani, P. Arkadiev & N. Amiridze (eds), 189–218. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9781614513209.189

Meakins, F. & Algy, C. 2016. Deadly reckoning: Changes in Gurindji children's knowledge of cardinals. *Australian Journal of Linguistics* 36: 479–501. https://doi.org/10.1080/07268602.2016.1169973

Meakins, F., Hua, X., Algy, C. & Bromham, L. 2019. The birth of a new language does not favour simplification. *Language* 95(2): 294–332. https://doi.org/10.1353/lan.2019.0032

Meakins, F., Jones, C. & Algy, C. 2016. Bilingualism, language shift and the corresponding expansion of spatial cognitive systems. *Language Sciences* 54: 1–13. https://doi.org/10.1016/j.langsci.2015.06.002

Meakins, F., McConvell, P., Charola, E., McNair, N., McNair, H. & Campbell, L. 2013. *Gurindji to English Dictionary*. Batchelor, Northern Territory: Batchelor Press.

Meakins, F. & Wilmoth, S. 2020. Overabundance resulting from language contact: Complex cell-mates in Gurindji Kriol. In *The Complexities of Morphology*, P. Arkadiev & F. Gardani (eds), 81–104. Oxford: OUP. https://doi.org/10.1093/oso/9780198861287.003.0004

O'Meara, C. & Pérez Báez, G. 2011. Spatial frames of reference in Mesoamerican languages. *Language Sciences* 33: 837–852. https://doi.org/10.1016/j.langsci.2011.06.013

Pederson, E. 1993. Geographic and manipulable space in two Tamil linguistic systems. In *Spatial Information Theory*, A. U. Frank & I. Campari (eds), 294–311. Berlin: Springer.

Pederson, E., Danziger, E., Wilkins, D., Levinson, S. C., Kita, S. & Senft, G. 1998. Semantic typology and spatial conceptualization. *Language* 74: 557–589. https://doi.org/10.1353/lan.1998.0074

Polian, G. & Bohnemeyer, J. 2011. Uniformity and variation in Tseltal reference frame use. *Language Sciences* 33: 868–891. https://doi.org/10.1016/j.langsci.2011.06.010

Schultze-Berndt, E. 2006. Sketch of a Jaminjung grammar of space. In *Grammars of Space: Explorations in Cognitive Diversity*, S. C. Levinson & D. Wilkins (eds), 63–114. Cambridge: CUP. https://doi.org/10.1017/CBO9780511486753.004

Senft, G. 2006. Prolegomena to a Kilivila grammar of space. In *Grammars of Space: Explorations in Cognitive Diversity*, S. C. Levinson & D. Wilkins (eds), 206–229. Cambridge: CUP. https://doi.org/10.1017/CBO9780511486753.007

Talmy, L. 1983. How language structures space. In *Spatial Orientation*, H. L. Pick & L. P. Acredolo (eds), 225–320. New York NY: Plenum Press. https://doi.org/10.1007/978-1-4615-9325-6_11

Terrill, A. & Burenhult N. 2008. Orientation as a strategy of spatial reference. *Studies in Language* 32: 93–136. https://doi.org/10.1075/sl.32.1.05ter

Van den Bos, J., Meakins, F. & Algy, C. 2017. Searching for "Agent Zero": The origins of a relative case. *Language Ecology* 1: 4–24. https://doi.org/10.1075/le.1.1.02van

Van Leeuwen, M. S. 1978. A cross-cultural examination of psychological differentiation in males and females. *International Journal of Psychology* 13: 87–122. https://doi.org/10.1080/00207597808246615

Wilkins, D. 1989. Mparntwe Arrernte (Aranda): Studies in the Structures and Semantics of Grammar. PhD dissertation, Australian National University, Canberra.

Wilkins, D. 2006. Towards an Arrernte grammar of space. In *Grammars of Space: Explorations in Cognitive Diversity*, S. C. Levinson & D. Wilkins (eds), 24–62. Cambridge: CUP. https://doi.org/10.1017/CBO9780511486753.003

CHAPTER 6

Substrate influence in Northern Quechua languages

Pieter Muysken[†]

Radboud University, Nijmegen

Quechua language varieties spread northward into parts of Ecuador, Colombia, and Northern Peru, and were adopted as a native language by speakers of earlier Pacific, Highland, and Amazonian languages in a process of language shift. This process started in the fifteenth century with the Inca state, and is still going on in some regions in the Piedemonte, where speakers of smaller languages are acquiring Quechua as a second, and their ultimately primary language. These Quechua varieties underwent numerous changes which this chapter investigates from the perspective of substrate influence, building on knowledge gathered in creole studies. The chapter further discusses the extent to which substrate influence is relevant to all Northern Quechua varieties or only a subset thereof.

Keywords: Norther Quechua varieties, Northern expansion, substrate influence, cafeteria principle, mutual reinforcement

1. Introduction

Quechua language varieties spread northward into parts of Ecuador, Colombia and Amazonian northern Peru, and were adopted as a native language by speakers of the languages already present there from the Pacific, the Highland and the Amazonian regions, in a process of language shift.[1] This process started in the

---

[†] Pieter unfortunately left this world, before we received the proofs of his chapter. From wherever he is now, we hope he will like the final version of this chapter he wrote to honor a friend.

1. This chapter is respectfully dedicated to Salikoko Mufwene, with whom I have discussed the issue of substrates ever since his attendance at a conference that Norval Smith and I organized in Amsterdam in April 1985 (whose proceedings were published as Muysken & Smith

fifteenth century with the Inca state and is still going on in some regions in the Andean foothills, the Pie de Monte, where speakers of smaller languages such as Wao Tededo and Shuar have been acquiring Quechua varieties as a second and sometimes ultimately their primary language. Thus, Quechuan varieties have played and still play a key role in the multilingual ecologies of the Pie de Monte.

The Quechua varieties that spread northward, which I will label Northern Quechua here, have undergone numerous changes, as is typical in such a situation of mostly non-demic expansion and ethno-genesis. These changes fall under a number of headings:

– Koineization and disappearance in the new language of original distinctions present in the contributing Quechua varieties from different parts of Peru;
– Morphological regularization and paradigm simplification, presumably as the result of second language learning;
– Autonomous developments and compensatory strategies to recuperate morpho-syntactic and semantic distinctions lost in the simplification process;
– Influence of the substrate languages, the original languages of the speaker communities, on the new language.

While all four developments are of great interest and crucial to our understanding of the diversification in the Quechua language family, this chapter focuses on the role of influence by the substrate languages, influence which may have interacted with the other developments mentioned. It critically reviews the possible influence of these languages on Quechua varieties. I begin with a brief outline of the socio-linguistic history and the position of Northern Quechua in the Quechuan family (Section 2). Then some theoretical and methodological issues will be discussed (Section 3), while Sections 4–8 discuss potential substrate-related changes in all Northern Quechua varieties: Ecuadorian Highland Kichwa, Ecuadorian Lowland Kichwa, Colombian Inga and Southern Pastaza Quechua. Section 9 summarizes and concludes the overview.

In the context of Creole studies, substrates have been a topic of sustained debate, since substrate influence is notoriously difficult to actually prove. Likelihood is often the only attainable goal. The same problem holds in the cases

discussed below, which can serve almost as textbook examples of how complex the phenomenon of substrate influence is.

2. The sociolinguistic history and the position of Northern Quechua in the Quechuan family

The history and internal diversification of the Quechuan language family has been described in great detail elsewhere (e.g., Adelaar 2004) and is still subject to much debate. I cannot summarize this debate here, but will present the findings of my own group.[2] I will only present the bare outlines here, as there is not much in the argument of the chapter that hinges on the classification of the family and narrative of its early history.

The family can be dated back to the early centuries of our era in Central Peru, and is generally assumed to have two main branches, Quechua I and Quechua II. Quechua I varieties did not undergo significant subsequent expansion, and their distribution reflects their emergence. The Quechua II varieties, in contrast, did spread considerably. Around 1000 CE I assume that varieties of Quechua II were taken to northern Peru as part of the Huari Empire, where they underwent some influence from neighbouring Quechua I varieties (the early northern expansion). Varieties were diffused from Cajamarca and the neighbouring regions into the lowlands, San Martin, as late as the colonial period.[3] Around 1200 Quechua II varieties spread into the Aymara territory of Cuzco, where they replaced Aymara and Puquina and came to be associated with the Inca civilization. In the Inca period, after 1300, Cuzco-type Quechua II varieties extended their territory to Bolivia and in colonial times they spread to Argentina. After 1480, Quechua II varieties were taken to Ecuador, including Cuzco Quechua varieties, and after 1540 Quechua spread into Colombia (the late northern expansion). From 1600 onwards it started diffusing into the Amazonian foothills, and to some extent it is still spreading to this day. The family tree for the Quechuan language is given in Figure 6.1 (there is much discussion about details in the branching, but this tree will serve for the purposes of the discussion in this chapter).

---

2. For a set of other interdisciplinary perspectives see for example the volumes that came out of a series of meetings organized by Heggarty and Beresford-Jones (e.g., Heggarty & Beresford-Jones 2012).

3. Note that the languages that date back to the early northern expansion are not part of what I call "Northern Quechua" in this chapter. I am using the term "Quechua" for members of the Quechua language family in general and "Kichwa" for varieties spoken in Ecuador.

The languages in the early northern expansions (Cajamarca, Ferreñafe, Chachapoyas, San Martín) and the late ones (what I call Northern Quechua here: Ecuador, Colombia and Amazonian northern Peru) are thus assumed to be only distantly related. They share a number of features, and I assume these to be either the result of similar substrate influence or in some cases direct contact between them.
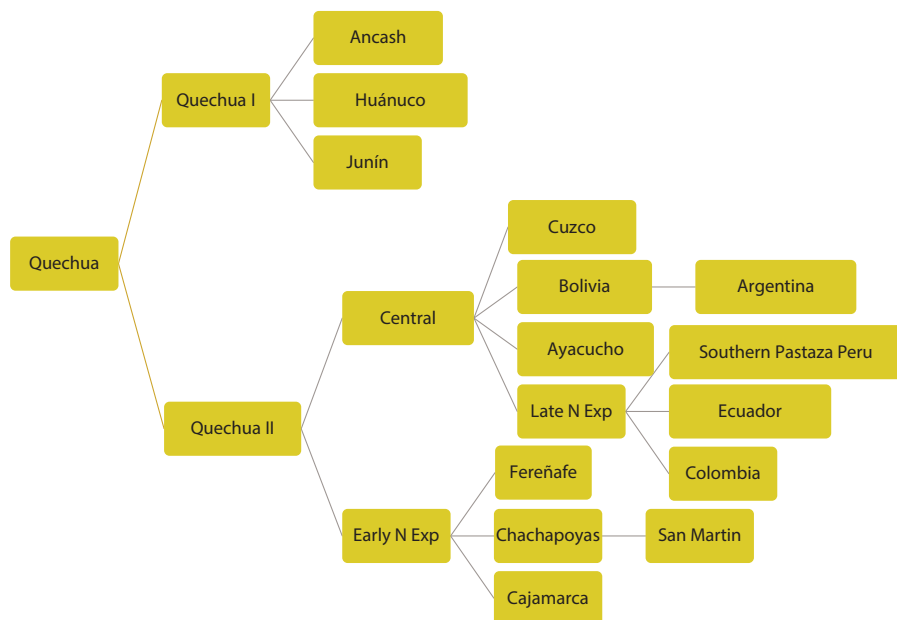
Figure 6.1. Schematic overview of main branches of the Quechuan language family (based on Adelaar 2012; Van de Kerke & Muysken 2014).* Note: * N Exp = northern expansion

While most expansions had been gradual and took place in the era long before the Spanish invasions, the late northern expansions into Ecuador, Colombia and Southern Pastaza Peru at least partially followed the Spanish invasion and were much more abrupt. In fact they involved a great deal of second language learning. In Muysken (1977, 2000) I have argued that this led to a considerable amount of restructuring in the language, particularly in the nominal and verbal reference paradigms. This has then produced a number of grammatical changes, which clearly set the northern varieties apart from those of Peru out of which they ultimately originated.

Northern Quechua emerged primarily through second language acquisition, not through population movement. First, the Incas may have come into contact with Chicham (Jivaroan) speakers on their way north, and subsequently Barbacoan speakers, who eventually became the largest groups of "new Quechua" or Kichwa speakers in the highlands. As Quechua spread into Colombia, Kamen'tsa or Kamsá speakers interacted with it. In the lowlands, Quechua varieties came into

contact with Waorani and Záparo. Finally, the variety that spread to the Southern Pastaza region in Peru came into contact with Candoshi. This is schematically represented in Figure 6.2.[4]

substrates*

15thC	16th C	17th C	18th C	19th C	20th C

-->

KAMSÁ

↕

COLOMBIAN INGA

▲

CUZCO Q   ⬂

CHINCHAY Q ► KICHWA of the HIGHLANDS

↑   ↑   ▼   WAO TEDEDO   ZAPARO

CHICHAM   BARBACOAN   ↕   ↓

⬂   KICHWA of the LOWLANDS

▼

SOUTH PASTAZA QUECHUA

↕

CANDOSHI
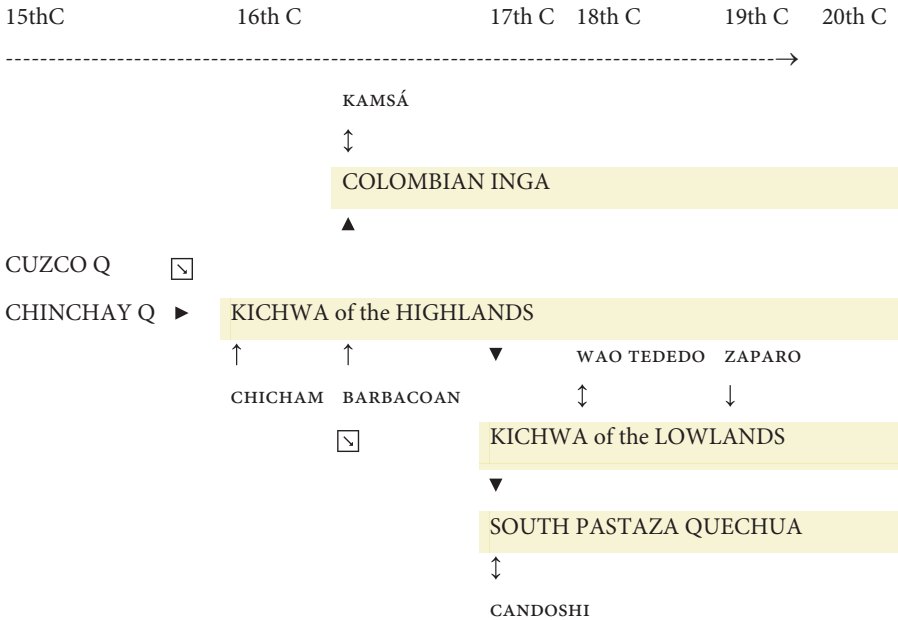
Note: * ►= descent; ⬂ = influence through contact; ↕ = mutual influence

Figure 6.2. The genesis and spread of the Northern Quechua varieties and their potential substrates*

In this chapter I will highlight possible influences that Chicham, Barbacoan, Kamsá, Waorani, Záparo and Candoshi may have had on the different Northern Quechua varieties at different stages of their development. Kamsá, Candoshi and Waorani continue to be spoken alongside varieties of Quechua. Barbacoan has receded to a few communities and no longer influences Kichwa. Chicham languages influence Kichwa in specific zones (for which I do not have data), but not generally. I think we can safely conclude that there is strong evidence for substrate influence overall, but that for each individual case, more careful analysis and argumentation are needed.

---

4. The precise nature of the original Quechua varieties that spread into Ecuador, Cuzco and Chinchay is not relevant to the discussion here, since the features discussed characterize all Peruvian varieties. Possibly as many as four distinct varieties were involved.

3.   Theoretical and methodological considerations

In Pidgin and Creole studies the issue of substrates concerns the influence that the original languages of the speaker communities (in the midst of which the Creole languages developed) have had on the new languages. This has been an issue for a long time, and probably will never be fully resolved. However, a number of theoretical principles and methodological considerations have been developed in this domain of research which can be useful in the study of substrate influences anywhere. These principles include the following.

The *Cafeteria Principle* was coined by Dillard (1970) and brought into the substrate debate by Bickerton (1981). It basically says that when you try to establish substrate contributions you cannot simply shop in a wide variety of possible contributing substrate languages spoken in the wider region as if it were a cafeteria. The most convincing cases of substrate influence involve very precise analyses in which the geographical location of the varieties in contact, and the exact period of the multilingual contact are exactly specified, as well as the contact feature and its contributing source features are precisely circumscribed. In the context of this chapter this means that the sociolinguistic history sketched in Figure 6.2 should be our guide in trying to establish possible substrates. This means we should also consider the question of whether these substrate influences have affected all of Northern Quechua or just a subset.

The second principle is that of *Mutual Reinforcement* of different substrates, as argued by Singler (1988): when different substrates share a pattern, it is more likely that this will end up in the new language. Thus, if Chicham and Barbacoan share a pattern, the chance that it will end up in the eventual new language, Northern Quechua, will increase.

A third principle to be considered is that of *Conservative Substrate Influence* (see Keesing 1988): many potential substrates, e.g., Chicham and Barbacoan, share typological features with Quechua: nominalizations, switch reference, SOV, case marking, verbal suffixes (Adelaar 2008; Birchall 2014a, 2014b; Krasnoukhova 2012). Thus, the conservative substrate influence that these languages may have exerted could play a role in explaining why Quechua preserved many of its features. Languages do not always change in an expansion context, and in this case second language learners of Quechua varieties often encountered categories that were also present in their own languages (this is what is called positive transfer in the second language acquisition literature; cf. e.g., Jarvis & Pavlenko 2008). This principle has so far not attracted sufficient attention, but I think it is an important one in contact situations across the globe.

We should also envisage the phenomenon of possible *Multiple Causation*, when both substrate and autonomous explanations are available, and these influences may have reinforced each other. For instance, substrate influence may

reinforce a compensatory strategy, when the loss of an original morphological distinction leads to the creation of a new distinction, where a pattern from a substrate language is adopted as a model.

There is an important temporal dimension that we need to take into account. Not all changes in Kichwa happened at once. Can we still attribute slower changes, which were only completed after the disappearance of the original community languages, to substrate influence?

Substrate influence can have several linguistic consequences, as we will see in Sections 4–6. Often it is when categories are absent in their own first languages that we may expect changes in communities where speakers are involved in language shift, often leading to *Loss* or simplification due to what is termed *Negative Influence*, where a particular category is not present in the substrate languages. Of course, loss due to substrate influence is a thorny issue. When Quechua learners of Andean Spanish show problems with the Spanish gender distinctions (lacking in Quechua itself), this may be viewed as negative substrate influence, but also as the result of universal principles of simplification in a second language learning context.

Things are different in the case of *Calquing*, when specific patterns from the substrate languages are copied. Calquing can apply to morphology (a suffix changes meaning), to syntax (a pattern is reinterpreted) or to the lexicon (a word acquires a specific meaning).

In the case of *Morphological Borrowing*, where actual morphemes are copied, the other language continues to be spoken. In this case we speak of *Adstrate* influence, as is the case with Aymaran suffixes used in Puno Quechua (Adelaar 1987).

Finally, we often find *Pragmatic Shift*, where in a contact setting a grammaticalized form acquires new grammatical meanings. In some varieties of Ecuadorian Spanish, the separate word *también* "also" has been reduced to *-tan* and may have an indefinite meaning similar to local Quechua *-pish*, as in *Onde-tan han ido?* "Where oh where did they go?"

Sometimes, it may be impossible to come to a final verdict and we need to envisage the possibility that some innovative features of Kichwa were not the result of direct influence from specific substrate languages, but arose from general processes of structural reduction or simplification due to second language learning, and at the same time a case of multiple substrates reinforcing each other.

4. General features of Northern Quechua varieties possibly due to substrate influence

I will begin by outlining the structural changes that have affected all Northern Quechua varieties together that I have been able to potentially link to substrate

influences, starting with case marking. I will not discuss lexical changes here. This is a promising area of research for individual varieties (see e.g., Urban 2018), but this work so far has not considered all Northern Quechua varieties together.

4.1 Collapse of the genitive and benefactive cases

A first example is the collapse of the genitive and benefactive cases in all Northern Quechua varieties (1):[5]

(1) **PeruvianQ (PeQ)** > **Northern Quechua**
-pa -paq   *-pa(k)*
GEN BEN   GEN, BEN

The Quechua languages, including Peruvian prestige varieties such as Cuzco Quechua, distinguish between two case markers: benefactive *-paq* and genitive *-pa*/*-q* (< *p*)/*-q-pa*. In Cuzco Quechua the form *-paq* is a benefactive as in example (2) or purposive as in (3):

(2) *Pi-<u>paq</u>-taq   chay punchu-ta-ri   awa-sha-nki*
who-BEN-EMP that poncho-ACC-CNTR weave-PR-2SG
"For whom are you weaving that poncho?"

<div align="right">Cuzco Quechua (Cusihuamán 1976: 135)</div>

(3) *allin   runa   ka-na-y-<u>paq</u>-mi   eduka-ku-sha-ni*
good person be-NMLZ-1SG-BEN-EVI educate-RFL-PR-1SG
"I am educating myself to be a good person."

<div align="right">Cuzco Quechua (Cusihuamán 1976: 135)</div>

The genitive form *-pa* occurs primarily in attributive nominal constructions, and co-occurs in the DP with the nominal possessive suffix *-n* on the possessed element:

(4) *[añas-<u>pa</u> t'oqo-<u>n</u>-mi] chahay-qa*
fox-GEN   hole-3-EVI yonder-TOP
"That over there is the hole of a fox."

<div align="right">Cuzco Quechua (Cusihuamán 1976: 136)</div>

It is clear that the possessor and possessed element form a constituent here. The possessor reference marker and the genitive case marker co-occur.

In Ecuador, in most varieties, the same form is used for both genitive as in (5) and benefactive as in (6):[6]

---

5. This section is partially based on Muysken (2017).

6. In fact, Muysken (2017) suggests that underlyingly the categories may be distinguishable in some varieties, but the basic observation holds.

(5) *Maria-pak   kusa*
    Maria-GEN   husband
    "the husband of Maria"   Chimborazo Kichwa

(6) *Maria-pak   randi-ni*
    Maria-BEN   buy-1SG
     "I buy (that) for Maria."   Chimborazo Kichwa

In Tsáfiki, the same form *-chi* is also used for both genitive and oblique, suggesting that the collapse of the two forms in Kichwa could be a case of substrate. In (7) *-chi* marks possession, while in (8) it marks dative and in (9) it marks direction:

(7) *Maria J. Cristo-chi   aya*
    Maria J. Cristo-OBL   mother
    "Maria is the mother of J. Cristo."   Tsáfiki (Seler 1902)

(8) *Pedro-chi   kuwa-de*
    Pedro-OBL   give-IM
    "Give it to Pedro."   Tsáfiki (Moore 1966: 96)

(9) *jun-chi   ji-na-yo-e*
    there-OBL   go-PR-CNG-DCL
    "I am going over there."   Tsáfiki (Moore 1966: 96)

However, the collapse of genitive and benefactive in Northern Quechua varieties may well be a case of multiple causation because *-pa* and *-paq* were already similar.[7] There are three other reasons to assume possible independent development co-occurring with substrate influence, and hence multiple causation:

a. The Peruvian Cuzco Quechua target for the genitive is unclear, alternating between [-pa] ~[-p] ~[-q] in different phonological contexts;
b. In Ecuador itself there is often a -CVC ~-CV alternation in suffixes, which could have been extended to the *-pa~-paq* alternation;
c. The canonical [possessor-GEN N-person] configuration present in Peru (see 4) had disappeared in Ecuador, and this domain was the principal context for use of the genitive. This makes the link between a noun marked genitive and the possessed noun looser, possibly facilitating the reinterpretation of the genitive as an oblique case.

---

7.   A reviewer notes that perhaps there is a historical link between the suffixes. Other dialects of Quechua in Peru have different patterns to those in Cuzco, and even in Cuzco the *-pa* genitive becomes *-q* after a vowel. Some have hypothesized that the origin of *-paq* may be original *-pa* reduplicated, with final vowel loss and change to [q], leading to *-pa-q*.

Altogether, the collapse of genitive and benefactive may have been substrate-influenced, but most likely was at least reinforced by other changes.

4.2   The loss of the inclusive/exclusive distinction

While all of Peruvian Quechua makes an inclusive/exclusive distinction for first person plural (possibly originally the result of Aymaran influence), this has disappeared in Northern Quechua (10):

(10)   **PeQ**   >   **Northern Quechua**
       Ñuqanchis   ñuqayku   ñukanchis
       1PL.IN.PRO   1PL.EX.PRO   1PL.PRO

This has repercussions in the pronouns, but also in verbal personal reference markers.

At the same time, Chicham and Barbacoan languages have no inclusive/exclusive distinction (in contrast to several other families in the general region; see Crevels & Muysken 2005). Thus, the loss in Kichwa could be a case of negative substrate influence: distinctions not made in substrates disappear in the new language. The alternative, of course, is to consider it a case of simplification. However, this does not do justice to the fact that many distinctions made in Peruvian Quechua did survive in Kichwa. Why simplify one thing but not another?

4.3   Phonological features

The topic of phonology merits a much more extended treatment than is possible here. I will just mention some of the changes in the Northern Quechua varieties that may possibly be linked to substrates, as presented in Table 6.1.

Altogether, Northern Quechua varieties started sounding much more like the substrate languages in the area (voicing, higher CV ratio, absence of uvular sounds), but it is not easy to point to specific cases of substrate. The phonological influences may have been cases of mutual reinforcement.

4.4   Nominal reference marking

One of the defining features of Kichwa and Colombian Inga is the loss of nominal reference marking (11).

(11)   **PeQ**   >   **Kichwa and Inga**
       mama-yki   kan-pak   mama
       mother-2SG   2SG.PRO-GEN   mother
       "your mother"

Table 6.1 Phonological features in Kichwa that may be linked to substrates

Change	In the substrate	Alternatives	Evaluation
Loss of the ejective consonants /p'/, /t'/, /č'/, /k'/. Replacement by either plain stops or aspirated stops.	No language in Ecuador has ejectives.	Simplification	Possible
Partial loss of aspirated consonants; when lost, replaced by plain stops.	Only Cofán has aspirated consonants.	Simplification	Possible
Conflation of the velar and the glottal stop: /k/ and /q/ > /k/.	Chicham, Cofán and Siona have glottal /ʔ/. Absence of /q/.	Simplification	Possible
Tendency towards sonorization, lenition and loss of consonants: /wasi-pi/ 'house-LOC/ > [wasi-bi], [wasi-i]. Also in some roots.	Cha'palaachi, Tsáfiki, Waorani, Siona, Kamen'tsa and Esmeraldeño have voiced consonants.	Natural tendency	Above all in the center and north, hence possibly linked to substrate as Quechua spread northward.
Fricativization of palatal laterals.	There are palatal fricatives in some languages, but very different ones.		Not convincing, needs more study.
Change of vowels: /-man/ 'DIR' > [-mun] /-tak/ 'EMP' > [-tik].	There are mid vowels in Cha'palaachi (only /e/), Tsáfiki, Waorani, Siona, Cofán, Kamen'tsa and Mochica.		Especially the center of Ecuador. Needs further exploration.
Fricative Φ in Imbabura Kichwa: /phuyu/ 'cloud' > [Φuyu].	Present in Tsáfiki	Natural tendency	Possible
Insertion of vowels in Loja Kichwa consonant clusters: /ačka/ 'mucho' > [ačika].	Many restrictions on clusters in original languages, but odd that it would be only in Loja; occurs optionally in many Quechua varieties.	Natural tendency	Not very likely

In Chicham and Barbacoan, only the possessor is marked, not the possessed element, and hence this could be a case of substrate influence. An alternative explanation for this change is simplification due to L2 learning, which is indeed a possibility.

Several things speak against this alternative explanation, however. The change probably started early (it is already referred to in the eighteenth century sources), but one of the branches of Northern Quechua, South Pastaza Quechua (SPastazaQ), does have nominal inflection, and the relevant substrate language Candoshi has head-marking for possession. This suggests that ongoing contact did play a role in the other varieties, after SPastazaQ split off. This reasoning is also an example of applying the Cafeteria Principle: only and exactly for SPastazaQ is Candoshi the relevant substrate.

The loss of nominal inflection for person had important consequences for the grammar of Kichwa and Inga. In (12) it is shown to have affected not only possession (12a), but also nominalization (12b), and adverbial subordination (12c).

		PeQ		**Kichwa**
(12)	a.	mama-yki	a.	*kan-pak mama*
		mother-2sg		2sg.pro-ben mother
		"your mother"		"your mother"
	b.	*riku-sqa-yki*	b.	*riku-sqa-yki*
		see-nmlz-2sg		see-nmlz-2sg
		"that you have seen"		"that you have seen"
	c.	*riku-pti-yki*	c.	*(kan) riku-kpi*
		see-ds-2sg		(2sg.pro) see-ds
		"if you see"		"if you see"

The nominalizations in (12b) can occur in object position (typically), but also in other argument positions. The adverbial subordinate clauses in (12c) typically occur either at the beginning or at the end of the sentence.

Thus, a single morphological change has had important grammatical consequences, leading to a broad range of grammatical differences between the Ecuadorian and the Peruvian varieties of Quechua (Muysken 1977).

4.5   The reinterpretation and reduction of the Quechua copula *ka-* as a clitic under the possible influence of Chicham

A very complex issue concerns the reinterpretation and reduction of the Quechua copula *ka-* as a clitic in some Northern Quechua varieties (Muysken 2010). The issue is complex for four reasons: (a) the chain of argumentation for the clitic status of the copula is not simple in itself; (b) the crucial facts have not yet been studied for all varieties involved; (c) the reinterpretation of the copula as a clitic

is part of a more general grammaticalization process involving the copula in the Quechua language family; (d) the copula and non-verbal predicate constructions in the relevant substrate languages require more analysis.

Originally, I started thinking about a possible substrate relation because in both Arajuno (Lowland Ecuador) Kichwa, and Shuar, the southern neighbour, we find superficially similar constructions. In this Kichwa variety, the copula is *a-* and it can either occur with a suffix as a separate word as in (13a) and (13b), or be attached to the affirmative marker *-mi,* which is then reduced and functions as a proclitic on the copula, as in (13c).

(13) a. *ñuka*   *Pedro-mi*   *a-ni*
        1SG.PRO Pedro-AF   be-1SG

   b. *ñuka*   *Pedro*   *a-ni*
       1SG.PRO Pedro be-1SG

   c. *ñuka*   *Pedro*   *m-a-ni*
       1SG.PRO Pedro AF-be-1SG
        "I am Pedro."   Arajuno Kichwa

The predicate construction in Shuar does not involve a separate copular verb, but rather an enclitic of suffixal copula *-yt-* preceded by a euphonic linking vowel *-a-*:

(14) *peŋker-a-yt-I*
       good-EU-COP-3SG
       "It is good."   Shuar (Pellizzarro 1969: 9)

Thus the Tena Kichwa (a variety spoken in the lowlands) copula *a-* superficially resembles the Shuar euphonic marker *-a-* which precedes the phonologically dependent copula in Shuar. Furthermore, the Quechua copula is no longer a separate word.

In some highland Kichwa varieties (and in some other lowland varieties) the copula is not *-a-*, but a clitic nonetheless. The central varieties of Salcedo and Salasaca have a copula *ga-* which is not a root but also a clitic, as argued in Muysken (2010). It cannot occur by itself (15a), has a fixed position with respect to the predicate (15b), and is devoiced when the preceding segment is voiceless (15c):

(15) a. **ga-ni*
         be-1SG
        "I am, I exist."

   b. *Martina ga-ni / *ga-ni Martina*
       Martina be-1SG   be-1SG Martina
       "I am Martina."

   c. *shamu-k ka-ni*
       come-AG be-1SG
       "I used to come."   Salasaca Kichwa

However, in other varieties, such as Imbabura (16), the copula does not have clitic status in the same way:

(16)   a.   *Xwan-ga*   *Pedro-ka*
            Juan-TOP   Pedro-TOP

       b.   *Xwan ka-ni.*   *Pedro   ka-ni.*
            Juan   be-1SG   Pedro   be-1SG
            "I am Juan."   "I am Pedro."   Imbabura Kichwa

Differently from Salcedo, in Imbabura Kichwa the copula in (16b) is not sensitive to the (nasal) preceding environment to become voiced (hence *Xwan ka-ni*, not *Xwan ga-ni*), in contrast with the topic marker in (16a), which does adapt to the preceding nasal (hence *Xwan-ga*, not *Xwan-ka*).

Thus the phonological argument I brought in for Salcedo does not work for Imbabura Kichwa. It needs to be seen whether the other arguments do hold for different Northern Quechua varieties. In addition, the reinterpretation of *ka-* as a clitic needs to be seen as part of a more general process of specialization of meanings of *ka-*. In Peruvian varieties, *ka-* has the full range of meanings that could be attributed to a copula, but in Ecuador it is limited to predication and identification, as can be seen in (17). Bolivian varieties, studied in Muysken (2010), are somewhere in between.

(17)

Function	**Peru**	**Ecuador**
possession	ka-	chari-
existence	ka-	tiya-
location	ka-	tiya-, ka-
auxiliary	ka-	ka-, (chari-)
predication	ka-	ka-
identification	ka-	ka-

This specialization is part of a general grammaticalization process. Clitic status is a logical possibility, but not a necessary outcome.

In the Barbacoan languages, predication is part of a complex series of positional verbs, and it needs to be seen whether these can be related to the developments in Northern Quechua. Altogether, it is not clear whether the developments in the copula have to do with possible substrate influence. This requires more systematic investigation.

4.6   The spread in use of a negative existential verb *illa-*, possibly modelled on Chicham or Barbacoan

The Northern Quechua varieties have grammaticalized the meaning of the verb *illa-* as a negative existential. In Peruvian varieties the verb means something like

"to lack". Consider (18a), whereas many Peruvian varieties would have something like (18b):

(18) a. *kay-bi   illa-n*
          this-LOC NEG.EXIS-3
          "Here there is no one/nothing."   Salasaca Kichwa

      b. *kay-pi   mana   ka-n-chu*
          this-LOC not   COP-3-NEG
          "Here there is no one/nothing."   Cuzco Quechua

The form *illa-* is present everywhere in Ecuador, and also in SPastazaQ and Inga. It may have had a wide use in different varieties of Peruvian Quechua, but it is not present any more in Cuzco Quechua. In Ayacucho Quechua, there is a verb *illa-* meaning "being off on a trip, and consequently not being there". In Cajamarca Quechua in northern Peru the verb is absent, but it does occur in the Peruvian variety of San Martin, with the same meaning as in Ecuador.

Both the Chicham and the Barbacoan languages have a negative existential, as the examples (19), (20) and (21) from Shuar and Tsáfiki show.

(19) *acá-way*
      NEG.EXIS-3SG
      "There is not."   Shuar (Juank 1982: 6)

(20) *yaunchu-ka   colegio-nama-ka,   atsu-ia-yi*
      past-FOC   high school-LOC-FOC EXIS.NEG-PAS2-3SG.DCL

      shuar   chicham
      Shuar   language

      "In the past, there was no Chicham language in the high school."
                                                             Shuar (Saad 2018)

(21) *osi-chun   i'to-yo-e*
      sell-DS   NEG.EXIS-CNG-DCL
      "There is nothing to sell."   Tsafiki (Moore 1966: 79)

Consider also the change mentioned in the previous section in which the Peruvian verb *ka-* "be" gradually lost its existential meaning and was replaced by *tiya-* in that meaning. Perhaps in tandem with this process, the more grammatical use of *illa-* could be a case of mutual support of different substrates, leading to the further grammaticalization of the Peruvian form as a negative existential.

4.7   Change in the status of the evidential

In Northern Quechua varieties, the enclitic *-mi* (22) is often glossed as affirmative (cited from a folk story [IIAD 1975] about the *Tiempos Urdimales,* a mythical past when everything was different):

(22)   *ima-lla-bish*   *pasa-k-lla* *tuku-k-lla-mi*   *ka-shka*
       what-DEL-IND   happen-AG-DEL-AF COP-PAS2
       "Anything used to happen (in the old days)."   Salasaca Kichwa (urd 1)

In Peruvian Quechua, the same suffix *-mi* "denotes that the speaker has seen or participated personally in the realization of the event that he describes, or that he knows in some direct way that the event referred to is taking place or will take place in the near future" (Cusihuamán 1976: 240).[8] As Faller (2002: 168) phrases it, "[it] encodes that the speaker has the best possible or strongest evidence in relation to the type of information conveyed".

It is clear that in the example just cited, there is no strong evidence for the events recounted. The statement is simply an affirmation. In contrast, in Tsáfiki the marker *-e* indicates simple affirmative. Possibly, in bilingual settings the Quechua direct experience evidential is reinterpreted as a simple affirmative marker (which is also the pragmatic implicature of many cases of the direct experience evidential as used in Peruvian Quechua), under the influence of the Tsáfiki usage. I should mention that no careful study has been done so far on the use of *-mi* in Ecuador, but it seems to involve information structure and affirmation much more than evidentiality. The reportative marker *-si/-shi* is also absent in many Northern Quechua varieties; instead, *ni-n* [say-3] "they say" is used.

5.   Highland Ecuadorian Kichwa

I have found a few features that only occur in Highland Ecuadorian Kichwa that are possibly linked to substrate influence.

5.1   The development of a desiderative involving the verb "say" in the southern highlands

There is an interesting construction in Highland Kichwa, which has been argued to have a possible source in Chicham. In such a case, it would be an example of the Founder Principle proposed by Mufwene (1996), which predicts that early populations play a key role in substrate formation. Since Chicham languages are further south than Barbacoan, probably their speakers were in contact with Quechua speakers before those of Barbacoan further north. However, a

---

8.   "*denota que el hablante ha visto o participado personalmente en la realización del evento que él describe, o que conoce en forma directa que el referido evento está en plena ejecución o que va ocurrir en un futuro cercano*". (Cusihuamán 1976: 240)

reviewer notes that the Inca advance was so fast here that really there would not have been any significant difference in the antiquity and duration of contact with Quechua-speakers. Also, it is not so clear how far south Barbacoan languages were spoken.

The relevant development involves the replacement of a want + infinitive construction by a say + first person future construction with the same meaning (23):

(23) **PeQ** > **Kichwa**
miku-y-ta   *muna-ni*   *miku-<u>sha</u>*   <u>*ni*</u>*-ni*
eat-INF-ACC want-1SG   eat-1FU   say-1SG
"I want to eat."

Some examples of a parallel construction in Shuar are (24) and (25). In both cases, a form of the verb "say" has the meaning of intention or desire, and is combined with a verb marked for first person.

(24) *Išičik*   *úm-iny-á-yt-hey*   *nampék-ay-h*   *tu-sa-n*
modest drink-AG-EU-COP-1SG get.drunk-NEG-1SG say-SS-1SG
"I am a modest drinker, not wanting to get drunk."   Shuar (Juank 1982: 12)

(25) *tsuama-ra-tah*   *ta-ku-r*
heal-PFV:PL-SG.HORT   say.IMPFV-SIM-1PL:SS
"when we want to heal him (lit. when we say let me heal him)"
Shuar (Saad 2018)

This construction does not occur in the lowlands, and neither do we find it in SPastazaQ or Inga. In the highlands, it is not found in parts of Imbabura. However, there is evidence that it was already a grammaticalized construction in the eighteenth century, as evidenced in the grammatical sketch by Nieto Polo (1753/1964).

Notice that it is also found in Tsáfiki (26):

(26) *Ji-chi*   *ti-na-yo-e*
go-INCP say-PR-CNG-DEL
"I want to go."   Tsafiki (Connie Dickinson p.c.)

We must assume this pattern to be a somewhat later development, starting somewhere in the south or centre of the highlands.

5.2   Dual hortative

In highland Kichwa, it is claimed that there is a distinction between two related forms (27a) and (27b), one used with two people, and one with more people:

(27) a. *Haku.*
go.HORT
"Let's go!"

b.   *Haku-chi.*
go.HORT-PL
"Let's all go."   Kichwa

It could be that this related to a similar distinction in Shuar (28a) and (28b):

(28)   a.   *Wɨtai.*
wɨ-tai
go.PFV-HORT
"Let's go!" (frequently the speaker and one other person)

b.   *Wɨ-ara-tai.*
go.PFV-PL-HORT
"Let's all go."   Shuar (Saad 2018)

However, the possibility is not to be excluded that the distinction in Kichwa is more the result of pragmatic implicature (since hortative requires at least two people, plural hortative may be interpreted as involving more than two). In this case it could be the result of parallel development rather than copying a direct grammatical contrast, or possibly a case of multiple causation. It is difficult to decide this just on the basis of the Ecuadorian data; possibly a comparative typological survey of how frequent this pattern is may help decide.

5.3   Switch reference in purposive nominalization

The possible introduction of switch reference in the purposive nominalizations of northern highland Ecuadorian Kichwa as a compensatory strategy may be based on a Barbacoan model, as argued for by Bruil (2008). Is it possible Tsafiki influence (Adelaar 2004; Bruil 2008)? While in Peruvian Quechua varieties, morphological marking distinguishes who is doing the eating, as in (29), the loss of nominal person marking in Kichwa that can be seen in (30a) would obscure this distinction. The suggestion is that the extension of the use of the original hortative marker *-chun* in Kichwa compensates for the loss, as in (30b).

(29)   a.   *miku-na-y-paq*
eat-NMLZ-1SG-BEN
"for me to eat"

b.   *miku-na-n-paq*
eat-NMLZ-3-BEN
"for X to eat"   Peruvian Quechua

(30)   a.   *miku-nga-pak   randi-rka-ni*
eat-NMLZ-BEN   buy-PAS-1SG
"I bought it to eat."

b. *miku-chun randi-rka-ni*
  eat-DS.BEN buy-PAS-1SG
  "I bought [it] for X to eat."   Highland Kichwa

The contrast between (31) and (32) in Tsafiki is argued to be the model for this development.

(31) *fi-chuu   cáyo-e*
  eat-SS:BEN I.bought
  "I bought it to eat."   Tsafiki (Moore 1979: 48, 49)

(32) *sona   mera-sa   tayo-e*
  woman listen-DS.BEN I.have
  "I have it for my wife to listen."   Tsafiki (Moore 1979: 48, 49)

An alternative explanation to the copying from a Barbacoan substrate would be that the construction emerged autonomously from the reduction of a quotative (33), as argued by Ross (1963).

(33) *Pay   ñuka-ta   kacha-rka willa-chun ni-shpa*
  3.PRO 1SG.PRO-ACC send-PAS   tell-HORT   say-ss
   "She sent me, saying: let him tell."   Highland Kichwa

This would require assuming that the quotative gerund *ni-shpa* "saying" was deleted as the construction was grammaticalized. However, this may amount to just suggesting a path through which the substrate influence from Tsafiki was realized over time.

6.   Lowland Kichwa: Intentional or future with "do"

The early history of lowland Kichwa is not known. It was probably brought into the lowlands from the highlands in the early colonial period, and soon spread as a lingua franca in trade relations and through active diffusion by the missionaries. Groups of speakers of Zaparo, Waorani and Chicham languages became bilingual in Kichwa, and many ultimately shifted to Kichwa. Currently, it is still an important second language in other communities that have resisted quechuization so far. I have found one specifically lowland feature that could be linked to a substrate.

A very distinctive feature of lowland Kichwa is the use of the verb "to do" in a reduced form as an intentional or future marker, as in (34) and (35):

(34) *shamu-nga   ra-u-ni*
  come-NMLZ do-PR-1SG
  "I am going to come."   Lowland Kichwa

(35)   *wañu-nga   ra-u-nchi*
  die-NMLZ   do-PR-1PL
  "We are going to die."   Lowland Kichwa

Jaeger (2006), in a typological survey of periphrastic "do"-constructions, shows that "do" as a future or intentional marker is rare typologically. However, it is found in some languages not far from where lowland Kichwa is spoken. The neighbouring language of several lowland Kichwa varieties, Waorani, shows the following pair. In (36a) the suffix *-kæ* marks intention, while in (36b) the same form occurs as a verb marking activity, "do":

(36) a. *bo-tō taa-miæ̃~-**kæ**-bo-ī-pa*
    1-PRO   cut-CL.tail-INC-1SG-IFR-AF
    "I am going to cut off his tail." Waorani (Holman 1988: 63)

  b. *æ-bā-nō biwii **kæ**-kā-ta-wo*
    how bird do-3-PAS-DUB
    "I wonder how the little bird is doing?" Waorani (Holman 1988: 65)

Thus the development of an intentional construction involving the verb "do" in lowland Ecuadorian could quite possibly reflect the influence of Waorani.

7.   Colombian Inga

Little is known about the early history of Inga, except that it is likely a very early colonial off-shoot of northern highland Kichwa. It is spoken in the same area as the Colombian linguistic isolate Kamen'tsa or Kamsá. Inga has been documented by the SIL linguist Levinsohn (1976) as well as by the anthropologist McDowell (1989).

7.1   The introduction of diminutive or pejorative suffixes

Levinsohn (1976: 95) lists the following "deprecative" suffixes (37) in Inga (see also Levinsohn 1974):

(37) *-ajim* [axim] "sympathetic, compassionate"
  -ashim [aʃim] "teasing in a deprecative manner"
  -uj [ux] "demanding, pleading, unbelieving"
  -ash [aʃ] "slightly deprecative, referring to subject"
  -aj [ax] "moderately deprecative, referring to subject"
  -ans [ans] "strongly deprecative, referring to subject"   Inga

The most frequent suffix is *-ag* [ax], as shown in example (38), from Levinsohn (1976: 24, 27):

(38) *ñamby-ag* "path"
      indy-ag "sun"
      tambw-ag "ranch"   Inga

These are not found in other Quechua varieties. Can they be traced back to Kamsá? The language has not been well documented so far, unfortunately, and there is no direct mention of the same class of suffixes in the language. Fabre (2002) notes however, in his preliminary grammatical sketch, that the language has three suffixes which resemble a rich system of classifiers: *-tem(a)* "diminutive", *-jem(a)* /*-kwem(a)* "augmentative", and *-xem(a)* "caritative". These elements can be attached to different parts of speech, including verbs. The following example (39) shows *-xema:*

(39) *sə-n-x-i-ʃatʃe*   *sapo-xema*
      1SG-EV.SEEN-PRED-TM-grab frog-CAR
      "I grabbed the poor frog."   Kamsá (Fabre 2002: 186)

Obviously, much more work is needed to really understand the facts in Kamsá, but it is likely, in my view, that there is a link between Kamsá *-xem(a)* and Inga *-ajim* [axim]. There is both a meaning and a form correspondence. It remains to be seen whether the other suffixes in the Inga class can likewise be linked to Kamsá sources, but in situations of adstrate (bilingual coexistence) it is not unusual for affective pragmatic markers from one language to be used in the other one. It should be noted also that it would be ideal if we had earlier sources for Kamsá, since the contact between Inga and this language must go back several centuries.

7.2   The potential marker *-ntra*

Inga has a modal or future marker *-ntra* [ntra], which does not come from Quechua and is not present in other Quechuan varieties (40).

(40) *ojalá tukuy familia chasa-lla wiya-**ntra**-kuna nuka   palabr-ita*
      may every family thus-DEL hear-INT-PL   1SG.PRO word-DIM
      "May every family thus hear my words."   Inga (Levinsohn 1976: 42)

Quite possibly, this marker comes from Kamsá (Juajibioy Chindoy & Wheeler 1974: 71) *-cha(n)* glossed as "intention" (unfortunately full glosses are lacking), as in (41a). Fabre (2002: 172) glosses *-cha* as a future marker (41b):

(41) a. *ye-**chan**-jokeda ko-**cha**-yán   **cha**-waise bokna*
          "she will stay"   "you will say"   "s/he should go"

      b. *ku-bu-**cha**-n-x-ua-nác*
          2SG-1SG-FU-EV.SEEN-PRED-PRED-TM-take
          "I will take you"   Kamsá

It could be that the Kamsá marker *-cha-n* is a frozen form containing both the future *-cha-* and the evidential *-n-*. When a more complete analysis of Kamsá has been made, this matter can be cleared up, but the correspondence is not unlikely. Just as in the case of the deprecative suffixes, the match is not exact but suggestive.

The sequence /tra/ in the Inga suffix *-ntra* does not correspond to other Quechua forms and suggests that this is a borrowing. I do not have enough information about Kamsá to see whether it matches the *cha* of the Kamsá suffix. The other problem is the initial /n/ of the Inga suffix. Possibly it is an analogy to the initial /n/ in two Inga future suffix forms, of which only the first and third correspond to a Peruvian source and the second is an innovation in Inga, possibly modelled on a periphrastic construction in Northern Quechua:

(42)   *-sa*   1SG.FU   < Peruvian *-saq* "1SG.FU"
　　　 -nka   2SG.FU   <? Northern Quechua *-na ka-*-"-NMLZ COP-"
　　　 -nga   3SG.FU   <Peruvian *-nqa* "3.FU"

Altogether, I think the Inga potential marker *-ntra* may be copied from Kamsá, but the match is not a direct one, and involved reinterpretation.

7.3   Object marking

Levinsohn reports that the portmanteau suffix *-yki* that in Peru marks the combination of a first person subject with a second person object (glossed as 1 > 2) in Inga can also refer to a third person object (43):

(43)   *roga-ku-g   ri-rka-yki*
　　　 beg-PR-AG   go-PAS-1>23
　　　 "I went to beg him."   Inga (Levinsohn 1976: 48)

It is possible that this change, which is not seen anywhere else in the large number of Quechua varieties, is modeled on Kamsá, where third person object can be marked:

(44)   *i-bo-xawiyana*
　　　 3OB.SG-3SU.DU-say
　　　 "he said it to him"   Kamsá (Howard 1977: 59)

8.   Peruvian South Pastaza Quechua

Peruvian South Pastaza Quechua (SPastazaQ) is an offshoot of Ecuadorian lowland Kichwa, as can be seen in a number of diagnostic features (see also Muysken 2020). A number of its properties suggest influence from neighbouring languages.

8.1 Nominal person marking

A striking property is that it has retained some archaic features such as nominal person marking (Tödter et al. 2002: 27), as in (45).

(45) *wawa-yni* [child-1SG.PO] "my child"
      mama-yki [mother-2SG.PO] "your mother"
      yaya-n   [father-3.PO]   "her/his father"

It is odd to think of this as a substratal feature since it is part of Peruvian Quechua. Since this variety is spoken in Peru, it is tempting to link it directly to other northern Peruvian varieties but this is highly unlikely, since it has undergone the diagnostic changes characteristic of Ecuadorian varieties. In Ecuador, nominal person marker was gradually lost in all varieties (including the Ecuadorian lowland varieties from which SPastazaQ directly descends); it was variable already in the eighteenth century, and still mentioned with some words like *mama-y* "my mother" at the end of the nineteenth century. The most likely scenario is that nominal person marking was variable in all varieties at least until the eighteenth century, was part of the varieties that were taken to the Amazonian region in the seventeenth century, and then became obligatory again only in the region where the local language, Candoshi, had nominal person marking. This must be viewed as a case of conservative substrate influence, since Candoshi has suffixal nominal person marking (46):

(46) *shiip-i* [nose-1SG.PO] "my nose"   Candoshi (Tuggy 1966: 234)

In the other areas, where the local languages had no nominal person marking, it disappeared from local Quechua varieties as well.

8.2 Inalienable kinship

One striking feature of SPastazaQ is that kinship terms may not occur by themselves, without some marker. Thus **mama* "mother" is not well-formed as a bare citation form.

(47) *mama-yki*   *mamá-w*   **mama*
      mother-2SG   mother-VOC
      "your mother"   "mother!"   SPastazaQ (Tödter et al. 2002: 27)

This is not the case in all other known varieties of Quechua, but Candoshi has the same alternation between either a possessed kinship term or a vocative (48):

(48) *ataata* "my mother"
      atáa   "mother (VOC)"   Candoshi (Tuggy 1966: 3)

It is not clear from the sources available to me how the vocative is formed, but it is clearly a separate category.

In Záparo it is also necessary to mark nuclear kinship terms with a possessive pronoun (49):

(49)   *kwa anu*   *anu
       1SG   mother
       "my mother"   Záparo (Beier et al. 2014: 197)

8.3   Vocative

Tödter et al. (2002: 26) show that SPastazaQ has developed a vocative. With nouns ending in -*i* and -*u* this vocative is marked with stress shift; in nouns ending in -*a* it is marked with -*w* combined with a stress shift (50):

(50)   *Yayá-w* [father-VOC]   "father!"
       Sará-w [Sara-VOC]   "Sara!"

While there is no vocative case in Quechua, many languages in the area of SPastazaQ do have this category. Saad (2018) gives the form -*a* for Shuar, and Taylor and Chau (1983) give the following example (51) for the related Chicham language Achuar:

(51)   *apa-wá-chi-ru-na-k-a*
       father-VOC-DIM-1.PO-DEIC-TOP-VOC
       "oh my own little father"   Achuar

Candoshi has a vocative as well (Tuggy 1966: 238), but I could not determine the form it takes from Tuggy's description. A reviewer notes that, in Peruvian Quechua, often the first person marker -*y* is used as a vocative (*mama-y* "my mother"), but this is different from the way vocative occurs in SPastazaQ.

8.4   Indirect object marking

Tuggy (1966) notes that indirect objects are marked with -*ta* "accusative" in SPastazaQ (52):

(52)   *Tasa-ta*   *ku-shka-ni*   *mama-yni-ta.*
       basket-ACC   give-PAS2-1SG   mother-1SG.PO-ACC
       "I gave a basket to my mother."   SPastazaQ (Tödter et al. 2002: 61)

The form *mama-yni-man [mother-1SG.PO-DIR], which would be general in all other varieties of Quechua, is impossible. This is reminiscent of the situation in Candoshi, where -*átsi* or -*á* occurs with all complements (Tuggy 1966: 243). However, the information available on this aspect of Candoshi is limited.

9. Conclusions and discussion

We can conclude that on the whole there is substantial evidence for substrate influence in Northern Quechua varieties, most of the time involving sounds, structural patterns or distinctions, but sometimes actual morphemes. The caveat in all the previous sections is, however, that most of the languages have not been adequately described so far, making definite claims impossible. Notice also that the chain of argumentation is often rather complex. Thus this chapter can best be seen as a proposal for a research programme, pending adequate reference grammars for all the varieties involved, rather than as a definitive statement. It also nicely illustrates the complexity of working on substrates in general. Table 6.2 summarizes my main findings so far.

Table 6.2 Overview of the potential substrate features discussed in the chapter

§	Feature	HIK	LOK	ING	PAS	*Source*	*Type of contact*
	Profile	x	x	x	x	All	Conservation
4.1	No distinction benefactive/genitive	x	x	x	x	Barbacoan	Loss, multiple causation
4.2	First person plural collapse	x	x	x	x	Chicham, Barbacoan	Loss
4.3	Phonology						Various, through mutual reinforcement
4.4	Absence of N-person	x	x	x		Chicham, Barbacoan	Loss
4.5	Reinterpretation of the copula as a clitic	(x)	(x)			Chicham	Phonological similarity, grammatical reinterpretation
4.6	*illa-*	x	x	x	x	?Chicham	Lexical calque
4.7	Change in status evidential	x	x			Chicham, Barbacoan	Pragmatic shift
5.1	*-sak ni-*	(x)	(x)			Chicham	Syntactic calque
5.2	Dual hortative	x				Chicham	Morphological calque
5.3	Switch referencer in purposives	x	(x)			Barbacoan	Syntactic reinterpretation
6.1	*ra-* inchoative		x			Waorani	Morphological calque

(*Continued*)

Table 6.2 (*Continued*)

§	Feature	HIK	LOK	ING	PAS	Source	Type of contact
7.1	Affective suffixes			x		Kament'sa	Morphological borrowing
7.2	*-ntra-* modal future			x		Kament'sa	Morphological borrowing
7.3	Third person object marking			x		Kament'sa	Syntactic extension
8.1	Nominal person reference marking				x	Candoshi	Conservation
8.2	Inalienable kinship				x	Candoshi, Záparo	Semantic calque
8.3	Vocative				x	?Candoshi	Morphological borrowing
8.4	Indirect object				x	Candoshi	Syntactic reduction

It seems that the Barbacoan languages played the most important role, possibly with secondary influence from the Chicham languages. There are also local effects of other languages such as Kament'sa or Kamsá, Candoshi and Waorani. The role of Záparo needs to be explored further. There may also be cases of negative influence, as in the case of nominal person marking, where its retention in SPastazaQ possibly due to the influence of Candoshi suggests that substrate does play a role.

References

Adelaar, W. F. H. 1987. Aymarismos en el quechua de Puno. *Indiana* 11: 223–231.

Adelaar, W. F. H. 2008. Towards a typological profile of the Andean languages. In *Evidence and Counter-evidence: Essays in Honour of Frederik Kortlandt,* Vol. 2: *General Linguistics*, A. Lubotsky, J. Schaeken & J. Wiedenhof (eds), 23–33. Amsterdam: Rodopi. https://doi.org/10.1163/9789401206365_004

Adelaar, W. F. H. 2012. Cajamarca Quechua and the expansion of the Huari state. In *Archeology and Language in the Andes: A Cross-Disciplinary Exploration of Pre-History* [Proceedings of the British Academy 173], P. Heggarty & D. G. Beresford-Jones (eds), 197–217. Oxford: OUP. https://doi.org/10.5871/bacad/9780197265031.003.0008

Adelaar, W. F. H. (in collaboration with P. Muysken). 2004. *The Languages of the Andes*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511486852

Bickerton, D. 1981. *Roots of Language*. Ann Arbor MI: Karoma.

Beier, C., Bowser, B., Michael, L. & Wauters, V. 2014. *Diccionario Záparo Trilingüe: Sápara-Castellano-Kichwa / Castellano-Sápara y Kichwa-Sápara*. Quito: Abya-Yala.

Birchall, J. 2014a. Argument Marking Patterns in South American Languages. PhD dissertation, Radboud University.

Birchall, J. 2014b. Verbal argument marking patterns in South American languages. In *The Native Languages of South America: Origins, Development, Typology*, L. O'Connor & P. Muysken (eds), 223–249. Cambridge: CUP.

Bruil, M. 2008. Innovations in the Ecuadorian Converb Systems: Grammatical Change in Language Contact Situations. Master's dissertation, Leiden University.

Crevels, M. & Muysken, P. 2005. Inclusive/exclusive distinctions in the languages of South-Western South America. In *Clusivity, Typology and Case Studies of the Inclusive–Exclusive Distinction* [Typological Studies in Language 63], E. Filimónova (ed.), 311–338. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.63.15cre

Cusihuamán, A. G. 1976. *Gramática Quechua Cuzco-Collao*. Lima: Instituto de Estudios Peruanos.

Dillard, J. L. 1970. Principles in the history of American English: Paradox, virginity, and cafeteria. *Florida Foreign Language Reporter* 8: 32–33.

Fabre, A. 2002. Algunos rasgos tipológicos del Kamsá (valle de Sibundoy, Alto Putumayo, Sudoeste de Colombia) vistos desde una perspectiva areal. In *Current Studies on South American Languages*, M. Crevels, S. van de Kerke, S. Meira & H. van der Voort (eds), 169–198. Leiden: CNWS.

Faller, M. 2002. Semantics and Pragmatics of Evidentials in Cuzco Quechua. PhD dissertation, Stanford University.

Heggarty, P. & Beresford-Jones, D. G. (eds). 2012. *Archeology and Language in the Andes: A Cross-Disciplinary Exploration of Pre-History* [Proceedings of the British Academy 173]. Oxford: OUP. https://doi.org/10.5871/bacad/9780197265031.001.0001

Holman, T.W. 1988. Waorani verb affixes. In *Workpapers Concerning Waorani Discourse Features,* Vol. 10: *Language Data* [Amerindian Series], E. G. Pike & R. Saint (eds), 57–69. Dallas TX: SIL.

Howard, L. 1977. Esquema de los tipos de párrafo en camsá. In *Estudios en camsá y catío*, L. Howard & M. Schöttelndreyer (eds), 1–67. Lomalinda, Colombia: Instituto Lingüístico de Verano, Editorial Townsend.

IIAD (Instituto Interandino de Desarrollo). 1975. *Urdimal Tiempomunda*. Ambato: Instituto Interandino de Desarrollo.

Jaeger, A. 2006. *Typology of Periphrastic 'Do'-Constructions*. Bochum: Brockmeyer.

Jarvis, S. & Pavlenko, A. 2008. *Crosslinguistic Influence in Language and Cognition*. Abingdon: Routledge. https://doi.org/10.4324/9780203935927

Juajibioy Chindoy, A. & Wheeler, A. 1974. *Bosquejo Etnolingüístico del Grupo Kamsá de Sibundoy, Putumayo, Colombia*. Bogotá: Instituto Lingüístico de Verano and Ministerio de Gobierno.

Juank, A. 1982. *Aujmatstai Yatsuchi: Manuel de Aprendizaje de la Lengua Shuar*, Vol. 1, 2nd edn. Quito: Mundo Shuar.

Keesing, R. M. 1988. *Melanesian Pidgin and the Oceanic Substrate*. Stanford CA: Stanford University Press.

Krasnoukhova, O. 2012. The Noun Phrase in the Languages of South America. PhD dissertation, Radboud University.

Levinsohn, S. H. 1974. Sufijos despectivos en inga. *Revista Colombiana de Antropología* 16: 357–361. https://doi.org/10.22380/2539472X.1532

Levinsohn, S. H. 1976. *The Inga Language*. The Hague: Mouton.
https://doi.org/10.1515/9783110819120

McDowell, J. H. 1989. *Sayings of the Ancestors: The Spiritual Life of the Sibundoy Indians*. Lexington KY: University Press of Kentucky.

Moore, B. R. 1966. Diccionario Castellano-Colorado, Colorado-Castellano. *Llacta* 22: 95–221.

Moore, B. R. 1979. *Método Para Aprender el Idioma Colorado (Gramática Pedagógica)*, Vol. 1, Lecciones 1 a 6. Quito: ILV.

Mufwene, S. S. 1996. The founder principle in creole genesis. *Diachronica* 13: 83–134. https://doi.org/10.1075/dia.13.1.05muf

Muysken, P. 1977. *Syntactic Developments in the Verb Phrase of Ecuadorian Quechua*. Lisse: Peter de Ridder Press. https://doi.org/10.1515/9783112420201

Muysken, P. 2000. Semantic transparency in Lowland Ecuadorian Quechua morphosyntax. *Linguistics* 38: 973–988. https://doi.org/10.1515/ling.2000.018

Muysken, P. 2010. The copula in Ecuadorian Quechua. In *Linguistics and Archeology in the Americas: The Historization of Language and Society*, E. B. Carlin & S. C. van de Kerke (eds), 191–206. Leiden: Brill. https://doi.org/10.1163/9789047427087_011

Muysken, P. 2017. Collapse of genitive and benefactive case in Ecuadorian Quechua? *Stellenbosch Papers in Linguistics* 48: 255–260.

Muysken, P. 2019. *El kichwa ecuatoriano: Orígenes, riqueza, contactos*. Quito: Abya Yala.

Muysken, P. Forthcoming. The early stages of Ecuadorian Quechua. In *Contact, Structure, and Change: A Festschrift for Sarah G. Thomason*, A. Babel & M. Sicoli (eds). Ann Arbor MI: University of Michigan Press.

Muysken, P. & Smith, N. 1986. *Substrata versus Universals in Creole Genesis* [Creole Language Library 1]. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.1

Nieto Polo de Aguila, S. J. T. 1753/1964. Breve Instrucción o Arte para entender la Iengua común de los indios, según se habla en la provincia de Quito. *Llacta* 20: Complete issue.

Pellizzaro, S. M. 1969. *Shuar: Apuntes de Gramática*. Quito: Offsetec.

Ross, E. 1963. *Introduction to Ecuador Highland Quechua or Quichua in Ten Easy Lessons*. Quito: Instituto Lingüístico de Verano.

Saad, G. M. 2018. A Grammar Sketch of Shuar: Focus on the Verb Phrase. Master's dissertation: Radboud University, Nijmegen.

Seler, E. 1902. Notizen über die Sprache der Colorados von Ecuador. In *Gesammelte Abhandlungen zur Amerikanischen Sprach- und Alterthumskunde*, 3–17. Berlin: A. Asher.

Singler, J. V. 1988. The homogeneity of the substrate as a factor in pidgin/creole genesis. *Language* 64: 27–51. https://doi.org/10.2307/414784

Taylor, A. C. & Chau, E. 1983. Jivaroan magical songs: Achuar anent of connubial love. *Amerindia* 8: 1–27.

Tödter, C., Zahn, C., Waters, C. W. & Wise, M. R. 2002. *Shimikunata Asirtachik Killka Inka-Kastellanu (Diccionario Inga-Castellano)* [Serie lingüística Peruana 52]. Lima: Instituto Lingüístico de Verano.

Tuggy, J. C. l966. *Vocabulario Candoshi de Loreto* [Serie Linguistica Peruana 2]. Pucallpa: Instituto Lingüístico de Verano.

Urban, M. 2018. The lexical legacy of substrate languages: A test case from the southern Ecuadorian highlands. *Transactions of the Philological Society* 116: 435–459. https://doi.org/10.1111/1467-968X.12129

van de Kerke, S. & Muysken, P. 2014. The Andean matrix. In *The Native Languages of South America: Origins, Development, Typology*, L. O'Connor & P. Muysken (eds), 126–151. Cambridge: CUP.

CHAPTER 7

Coordination in the Suriname Creoles

Comparing Nenge(e) and Matawai

Bettina Migge
University College Dublin

Although there is a fair amount of research on the morphosyntax of the Suriname Creoles, coordination has received surprisingly little attention. Coordination can be seen as a fundamental aspect of grammar that intersects with a number of seemingly different constructions. Using quantitative and qualitative methods, this chapter provides a comparative overview of coordinate structures in the Eastern Maroon Creole or Nengee, and in one of its Western Maroon sister languages, Matawai. It then focuses on coordinate constructions involving the juxtaposition of sentential constituents. The comparison of the distributional, semantic, and pragmatic properties of coordinate constructions within and across the two sets of Suriname Maroon Creoles reveals striking differences reported here for the first time.

Keywords: Coordination, Creoles, Matawai

1. Introduction

Cross-clausal phenomena, such as serial verb constructions (e.g., Lefebvre 1991; McWhorter 1992; Migge 1998; Sebba 1987; Veenstra 1996) and complementation (Aboh 2006; Byrne 1987; Migge & Winford 2013; Plag 1993), have received a fair amount of attention in research on Creole languages, in part because they display unique features from the perspective of European languages. Both types of construction have played a role in discussions of Creole genesis, being seen simultaneously as evidence for innovation and for substrate influence. There is, however, a conspicuous relative absence of systematic research on coordinate structures; descriptive grammars and grammatical sketches mention coordination but rarely examine it in much detail. This is surprising given that coordination also diverges from equivalent constructions in European languages and is thus also a possible

candidate for substrate (see Holm 2000; Lefebvre 2015), universal or contact-based restructuring influence (e.g., Mufwene 2001). The absence of research on Creoles is probably also due to the lack of research on coordination in general.

The aim of this chapter is to attempt to give a holistic overview of coordination constructions in two of the Creoles of Suriname belonging to the two related language clusters, the Eastern Maroon Creoles or Nenge(e) (Pamaka and Ndyuka varieties) and one little studied Western Maroon Creole, Matawai. They developed from the language varieties that had emerged on the Surinamese plantations (Early Sranan Tongo) and formed into separate sociolinguistic entities as the result of their creators' flight from the Surinamese sugar plantations.[1] The Matawai formed as a separate community in the 1730s (Migge 2017) while the three Eastern Maroon communities emerged in the eighteenth century, with Ndyuka and Aluku forming in the early part and Pamaka in the mid to later part of the century (Migge 2013: 40). After their initial establishment, the Maroon languages underwent partially separate developments from the plantation varieties due to their relative isolation from mainstream society. Maroons settled in relatively inaccessible locations and their movement to and from urban areas was highly policed by the colonial government. Since the beginning of the twentieth century and particularly from the 1960s onwards, Maroons have migrated first temporarily – to take up temporary cash labour opportunities – and since the 1980s increasingly permanently to Suriname's and French Guiana's urban areas (Van Stipriaan 2015). In both countries they nowadays make up about one third of the population and are predominantly urban populations. Maroons have also expanded their linguistic repertoires, speaking Sranan Tongo and having various types of competence in either or both of the two countries' official languages, Dutch and French. While Eastern Maroon varieties have been rising in importance – in line with their speakers' social upward trajectory and demographic expansion (Migge & Léglise 2015) – Matawai is endangered as it is hardly transmitted across generations in the urban areas where most of its speakers currently reside (Migge 2017). In the multilingual contexts of Suriname and French Guiana, Maroon languages are influenced by the official languages, French and Dutch, and by the widely spoken language Sranan Tongo, while dominant or widely represented Maroon languages such as Ndyuka also have an impact on other Maroon languages (Migge & Léglise 2015; Yakpo et al. 2015).

The data for this paper come from naturalistic recordings of interactions among speakers of Eastern Maroon or Nenge(e) varieties (Pamaka and Ndyuka)

---

1.   The direct descendant of the planation varieties is Suriname's national language, Sranan Tongo.

recorded in Suriname and French Guiana (both rural and urban contexts) between 1995 and 2010, and among speakers of Matawai recorded in 2013 in the traditional Matawai villages, and in Paramaribo, Suriname. Some elicitation work on Nenge(e) was also carried out in 2017 and 2018 – the elicited data is flagged with an *E*. All data were collected and processed by the author with the help of speakers of the language. Miriam Steerman also helped with interpreting some of the Matawai data. The paper gives an overview of the kinds of structures that exist, including the functions, meanings and distribution of the different coordinators and their interrelationships in these two languages as well as assesses to what extent their properties resemble those found in other Pidgin and Creole languages (P/Cs).

The paper is structured as follows. Section 2 briefly defines coordination from a typological perspective and summarizes the literature on coordination in P/Cs. Section 3 briefly compares coordination in the two related languages and their predecessor, Early Sranan. Section 4 discusses conjunctive coordination in Nenge(e) and Matawai and Section 5 deals with disjunctive and adversative coordination.

2.   Coordination

2.1   Defining coordination

Coordination can be broadly defined as "the juxtaposition of two or more *conjuncts* often linked by a *conjunction* such as *and* or *or*" (Goodall 2017). Haspelmath (2005: 3) further clarifies that "the two parts A and B [must] have the same status (in some sense that needs to be specified further)" because if one part is more prominent than or subordinated to the other phrase then it does not constitute coordination. When the two units are combined via coordination, they generally "still have the same semantic relations with other surrounding elements" (Haspelmath 2005: 34).

Structurally, coordination may theoretically involve conjuncts belonging to any kind of lexical category and conjuncts may be single words or phrases. However, in practice languages differ with respect to the types of word classes that may be coordinated. Languages generally also have several kinds of coordinate structures and coordinators whose distribution differs with respect to their meanings and the kinds of lexical elements and phrases that they can conjoin. Thus in some languages, such as German, the same coordinator (*und* "and") may conjoin verbal and nominal conjuncts, while in other languages, such as Fongbe, different coordinators are required (Lefebvre 2015). According to Haspelmath (2005: 10), differentiation between nominal and phrasal coordinators is rather common in the languages of the world; syndetic coordination is rather common,

but coordination may also be achieved through juxtaposition or asyndetic coordination. In syndetic coordination researchers also distinguish between monosyndetic coordination, "which involves only a single coordinator … and bisyndetic coordination, which involves two coordinators" when not more than two coordinands are present (Haspelmath 2005: 4). Coordinators are usually part of one of the two coordinands, and then are either found to head or to be placed at the end of that coordinand. In relation to movement, it is generally held that extraction must simultaneously take place out of both coordinands (Haspelmath 2005: 28). Other topics dealt with in the literature are the constraints on ellipsis and gapping (see Haspelmath 2005: 31ff on gapping).

Semantically, Haspelmath (2005: 5) suggests that languages differentiate between conjunctive ("and"), disjunctive ("or") and adversative ("but") coordination. In the case of conjunctive coordination, clause coordinators often mean "and then" rather than "and". However, it is difficult to judge "to what extent this sequential meaning is part of the coordinators' meaning and to what extent it simply derives from contexts in which sequences of events are reported" (Haspelmath 2005: 13). In some Oceanic languages, for example, types of coordinators are distinguished based on animacy, while in others they are differentiated based on whether coordinands are common or proper nouns or involve tightly or loosely related concepts (Haspelmath 2005: 13). Finally, types of coordinate constructions may also be distinguished based on the relationships that exist between subjects and objects in the coordinands. Lefebvre (2015), for instance, argues that identity between the subjects of clausal coordinands determines the choice of coordinators in Fongbe.

The literature provides a good overview of the structural aspects of coordination such as the syntactic and semantic properties of the coordinands and the coordinators. However, we find out very little about the discursive or pragmatic properties of different coordinate constructions.

2.2   Coordination in Creoles

Although coordination often receives little systematic treatment in descriptive grammars and in research papers on individual languages and overview chapters on P/Cs (e.g., Winford 2008), typological research has revealed a number of patterns. Holm (2000: 227) points out that in the case of conjunctive coordination, NP coordination, including coordination of single verbs, and clausal (VP or sentence) coordination are distinguished in a number of Atlantic varieties and in their African substrate languages. The former employs a form that is homophonous with a preposition meaning "with" and is also used to express instrumental and comitative meanings. Instrumentals are nouns that express a means for accomplishing an action (e.g., cut <u>with a knife</u>) and comitatives express accompaniment

(travel <u>with my friend</u>); coordinate constructions simply conjoin two NPs which can have inanimate or animate referents. Comitatives usually only have animate referents, while instrumentals usually have inanimate ones. The examples in (1) come from Haitian Creole, which uses *ak* (< French *avec* "with") for instrumentals (1a), comitatives and NP conjunction (1b).

(1) Haitian Creole

    a. Bouki te marye ak yon bèl wosiyòl.
       B ANT marry with IND pretty nightingale
       "Bouki was married to a beautiful nightingale." (DeGraff 2007: 117)

    b. Annou goumen pou lapè ak/e lajistis.
       let-1PL fight for peace and justice
       "Let us fight for peace and justice." (DeGraff 2007: 122)

In the case of phrasal or sentence coordination, the coordinator sometimes derives from the pronoun "he", as in Principe Creole and in a number of Atlantic languages. Sentence coordination may also involve juxtaposition of the two conjuncts. Holm (2000: 228) also points out that the distributions of the nominal and sentence level coordinators may overlap (see also Hassamal 2016 for Mauritian); he suggests that in the case of French Antilles and French Guianese Creole this "blurring of the distinction" might have emerged later, possibly due to greater contact with French. In some cases, there is a semantic difference. In Papiamentu *i*, typically used for verbal conjunction, encodes a looser association or consecutive meaning, while *ku* is typically used to link single verbs or nouns, indicating a non-propositional, joint or simultaneous meaning (Hassamal 2016 for Mauritian).

(2) Papiamentu

    a. e muhé ta kanta **ku** baila.
       DET woman IMP sing and dance
       "The woman is dancing and singing (at the same time)."

    b. e muhé ta kanta **i** baila.
       DET woman IMP sing and dance
       "The woman is singing on some occasions and dancing on others."
                                (Richardson 1977 cited in Holm 2000: 228–229)

This distinction does not seem to be universal, however, since Tok Pisin and Nubi, for instance, do not currently distinguish nominal and verbal coordination, possibly due to processes of levelling or substrate influence. On the basis of a larger database, the *Atlas of Pidgin & Creole Language Structures* (APiCs, Michaelis et al. 2013b, 2013c), and comparison with the *World Atlas of Language Structures* (WALS, Haspelmath et al. 2005), Maurer et al. (2013a: 276–277) show that 72% of APiCs languages compared with only 24% of WALS languages use the same element for comitative and instrumental constructions, possibly due

to input from European languages where this pattern is also found. If there are two coordinators, their distributions only (partially) overlap in 16% of APiCs languages (10% in WALS) : either one coordinator is only used for comitatives, as *ansanm* in Réunion Creole, and the other (*ek*) is used for both, or, as in Bislama, one is found in instrumental constructions (*long*) and the other (*wetem*) may occur in both. Finally, differention of use is crosslinguistically the most common pattern (WALS 66%). However, it is found in only 12% of APiCs languages such as Sri Lankan Malay, where *samma* is employed in comitative and *-ring* in instrumental constructions.

Maurer et al. (2013b) also found three kinds of relationships between coordination of NPs and comitatives. In 41% of APiCs languages (44% of WALS), NP coordination is differentiated from comitatives as in Krio, where *ɛn* conjoins two NPs whereas *wit* appears in comitatives. In the second case, APiCs languages (34%, compared with 56% of WALS) such as Cap Verdean Creole of Brava use the same form (*ku*) for both NP coordination and comitatives. In the third case, coordinators in APiCs languages (25% and none in WALS) show overlap. In one pattern exemplified by Virgin Island Dutch Creole, *mi* "with" is used in comitatives and for NP coordination but *en* only conjoins NPs. In another pattern, a construction like "together with", e.g., Casamances *juntu ku*, is used to differentiate comitatives from NP coordination expressed with *ku*. Maurer et al. (2013b) suggest that these patterns are, in part, due to the input languages, since identity in form (second case) is common among African languages and differentiation (first case) is found among European, South and Southeast Asian languages.

Discussing the relationship between NP and verbal conjunction, Haspelmath et al. (2013: 284–287) found that lack of differentiation is common in APiCs languages, particularly in those influenced by English. A word derived from "and" in European languages is often employed for that purpose. A number of APiCs languages differentiate the two types of coordination, where "with" is used for nominal coordination and the form derived from "and" in the European languages is employed for verbal conjunction. In some APiCs languages one coordinator, usually derived from "and", is used in both contexts and "with" only in nominal contexts such as in Réunion Creole, or they have different meanings (Hassamal 2016). Finally, Pichi has three conjunctive coordinators: *àn* for both, *wit* for NPs and *we* for clausal conjuncts.

Haspelmath et al. (2013: 285) suggest that juxtaposition generally varies with overt coordinators – with the exception of Pidgin Hindustani. Quint (2008: 35–37) shows that juxtaposition is the dominant strategy for verbal coordination in the Cap Verdean of Santiago. It has two semantic values, successiveness (3a) and simultaneity (3b). It may also be used for disjunctive and adversative conjunction.

(3) Cape Verdean of S.
    a. E lába kel frida, e kába, e xugâ -i -ei.
        3 wash DEM wound 3 finish 3 rince 3 3
        "She washes the wounds, she finishes [to wash] [and then] rinses them."

    b. (…) ta tinha batáta, [H], ta tinha kána, ta
        it there-have sweet-potato it there-have sugar-cane it

        tinha kusa, ki ta pagába mudjeris (…)
        there-have thing REL it pay woman

    "There was sweet potato, [at the same time] there was sugar cane,
    [at the same time] there was the thing that paid the women."

In the case of co-referential subjects where only the first one is expressed, juxtaposition has a consequence or result meaning (Quint 2008: 39) (4).

(4) Kópu kaprí -l di mó, kebra.
    Glass slip 3 of hand break
    "The glass slipped [from] his hand [and] broke."
                           (Mendes & al. 2002: 229 cited in Quint 2008: 39)

Several of the descriptive articles in Michaelis et al. (2013a) also mention that disjunctive and adversative conjunction are distinguished from conjunctive conjunction by separate coordinators but do not discuss them in detail.

    Coordinate constructions in the Creoles of Suriname have certainly been discussed in grammatical descriptions (e.g., Goury & Migge 2003/2017; Huttar 2007; Huttar & Huttar 1994; McWhorter & Good 2012; Migge 2013; Van den Berg & Smith 2013; Winford & Plag 2013) but to my knowledge there has not been a systematic investigation of coordination within one of these six languages or a comparative study. This chapter attempts to contribute towards filling this gap by comparing coordination in the Eastern Maroon Creoles or Nenge(e) with Matawai, a representative of the Western Maroon cluster. The remainder of this chapter examines to what extent coordination in the two Suriname Creoles aligns with and differs from the patterns established in the literature, and investigates the factors that condition the distribution of the different types of coordinators.

3. Coordination in Nenge(e) and Matawai: An overview

Table 7.1 gives an overview of coordinate constructions and markers in Nenge(e), Matawai and their predecessor, Early Sranan Tongo (Van den Berg & Smith 2013: 11) based on the literature.

Table 7.1 Overview of types of coordinators

Coordinators					Coordinators	Function		
Nenge(e)	Matawai	Early Sranan	gloss			Conjunct.	Disjunct.	Adversat.
anga		nanga, langa	with, and	NP, comitative		X		
	ku		with	NP, comitative		X		
da	da	dan	and then	finite S		X		
ne(en)			and then	finite S		X		
	hen		and then	finite S		X		
	nɔɔ		and then	finite S		X		
Ø	Ø		and then	finite S non-finite S		X		
en	en	en	and	finite S		??X		
		kaba	and (then); but	finite S		X		X
soseefi	soseepi	so srefi	in the same way, also	finite S		X		
		(da) so	in the same way; thus	finite S		X		
efu	ee	efu, ofu	or	finite S			X	
ma	ma	ma(ra)	but	finite S				X
		noso	but					X
toku (so)	(ma) toku	tog, toku	still; yet	finite S				X

Most generally, Table 7.1 shows that the Suriname Creoles present the less common APiCs pattern in that they distinguish nominal and verbal coordination. While they have several different verbal conjunctive constructions (finite S) each, only one coordinator derives from "and" in a European language (*en*) and there is only partial overlap between verbal conjunctive coordinators in the three varieties.[2] Early Sranan *kaba* and (*da*) *so* are absent in the Maroon Creoles, and the Maroon coordinators *neen*, *hen*, *nɔɔ* and Ø are not found in Early Sranan. While the two Maroon varieties share *da*, Ø and *en*, they differ with respect to several others (*ne*(*en*) versus *hen*, *nɔɔ*). They also employ different NP coordinators. Moreover, adversative *noso* is not found in the Maroon Creoles. This crude comparison suggests that this area of grammar has undergone change over time and

---

2.   *En* may be traced to Dutch *en* "and" or English *and*.

involves a degree of heterogeneity. Given the variety of conjunctive constructions and their salience in the literature on coordination in P/Cs, the chapter will focus primarily on these constructions; non-conjunctive coordination constructions will be discussed in Section 4.

4. Conjunctive coordination

4.1 Nominal coordination

The Suriname Creoles are English-lexified but they do not employ coordinators derived from English *and* or *with* to express comitative or instrumental notions, or nominal coordination. However, like many APiCs languages, examples (5–7) use the same form, *anga* (< English *along*) in Nenge(e) and *ku* (< Portuguese *com* or < Fongbe *ku*) in Matawai to express these three meanings.[3]

(5) Ndyuka
    a. Bowta á be mu kon a toobi anga businengee.
       name NEG PST MOD come LOC trouble with Maroon
       "Bowta should not have gotten into trouble with Maroons."

    Matawai
    b. Ma a an dɛ ku i.
       but 3 NEG be with 2
       "But he is not with you."

(6) Pamaka
    a. A be o dongo anga boto.
       3 PST FUT go-downriver with boat
       "He would have gone downriver with the boat."

    Matawai
    b. I ta kɔti ɛn ku ofangi.
       2 IMP cut 3 with machete
       "You can cut it with a machete."

(7) Ndyuka
    a. Eyee, tu ondoo dunsu anga tu balin kasolini a mu
       Yes two hundred thousand with two ton gas 3 MOD
       gi mi.
       give 1
       "Yes, two hundred thousand SRD and two tons of gasoline he has to give me."

---

3. There are also a few instances of *anga* in the Matawai data.

Matawai

b. A   bui   en   mau so, seebi dei, en   futu ku
    DET bracelet 3POSS arm so seven day 3POSS foot with

    en   mau.
    3POSS hand

    "He put a ritual bracelet on his arm like this, seven days, his leg
    and arm."

Interestingly, the overwhelming majority of the comitative constructions consist of a subject NP involving coordination of two NPs joined by *anga* and *ku* (8).

(8) Ndyuka
     a. En anga a   baya feti   de.
         3   with DET guy fight there
         "He and the guy fought there."

     Matawai
     b. U   ku   ɛn de a   tela   aki.
         1PL with 3 be LOC shore here
         "We and him are on the shore."

There are also semantic distinctions in that in (5) the subject is the agent and the person referred to in the phrase headed by *anga/ku* is inadvertently involved in the action. By contrast, in (8) the two are equal agents and have the same degree of involvement and agency in the action (p.c. 2018 M. Moese for 8a). The notion of jointness is thus encoded by varying the positioning of the *anga/ku* phrase.

Comitative and instrumental phrases headed by *anga/ku* can be extraposed (9) and constitute a separate intonational phrase. This is not the case in (7) and (10) though, as they are subject to the coordination constraint according to which extraction from a coordinate phrase is prohibited. The two coordinands are part of the same intonational phrase.

(9) Pamaka
     a. Anga sawtu mi boli   en.
         with salt 1 cook   3
         "I cooked it with salt."

     Matawai
     b. Ku   wan mii   de pali.
         with DET child 3PL give.birth
         "With a child they gave birth to."

(10) Pamaka
     a. Sama   anga u   be   e   gi   toli?
         person with 2PL PST   IMP give story
         "Who did you (excl.) discuss with?"

    b.   *__   anga u be e gi toli, sama? E
    c.   *sama anga __ be e gi toli, u?   E

While there is agreement among speakers that *anga* and *ku* are essentially nominal coordinators (and perform instrumental and comitative functions), "[i]n modern written Ndyuka, *anga* is [also] occasionally used by some writers for coordination of independent sentences" (Huttar 2007: 234–235; Huttar & Huttar 1994: 229). The examples provided by Huttar and Huttar (11) come from Bible translation work.

(11)   Ndyuka   (Huttar & Huttar 1994: 229)
        Melkisedek be   de   a   konu fu Salemu kondee. Anga a be
        Name   PST COP DET king for name   country and   a PST
        de   wan dominee fu Masaa Gadu.
        COP DET priest   for lord   god
        "M. was the king of Salem. And he was a priest of the Lord God."

Similar instances were also produced during French-Nenge(e) translation work and more formal academic written work in French Guiana. Example (12) comes from a bilingual pedagogy document produced in French Guiana during a teaching workshop. It resembles (11) in that it also involves coordination of two independent intonational clauses and expresses a sense of concurrentness. Note, however, that in (12) the second phrase is non-finite, as the subject and the aspectual marker are not repeated as in (11).[4] Although this usage of *anga* is not common, this extension is also found in French Creoles in the region, where it has been traced to contact with a European language (Holm 2000).

(12)   Pamaka
        A   ilm e   gi   a   toli   fu   a   buku anga yeepi fu balibali.[5]
        DET Ilm IMP give DET story for DET book with help for shoutshout
        "The mother tongue teacher tells the story of the book and helps the
        children respond."

4.2   Phrasal conjunctive coordination

Nenge(e) and Matawai have two main verbal coordinators, *da* (< Dutch *dan*) (9a) and *ne(en)* (9b) in Ndyuka/Pamaka, and *hen* and *nɔɔ* (Matawai).[6] In the Matawai data, we also find *da* and rarely also *ne(en)*. Occurrence of *ne(en)* is most likely due

---

4.   Example (12) might also be interpreted as coordination of two complex NPs, where *yeepi* would be a noun. This ambiguity might be at the heart of *anga*'s extension to verbal coordination, in addition to the fact that there is no other non-consecutive "and" coordinator in the language.

5.   ILM stands for *Intervenant en Langue Maternelle*, currently the official term for mother tongue teachers.

6.   I am not sure what the origins of (*ne*)*en*, *hen* and *nɔɔ* are, but it would be plausible that the first two derive from emphatic third person singular pronouns (*neen > na en*; *hen > hɛn*) (cf. Holm 2000) while *nɔɔ* is probably linked to the discourse marker "now" in English.

to recent influence from varieties of Nenge(e).[7] However, *da*'s growing presence in (urban) Matawai speech suggests that it is starting to compete with *nɔɔ*; further quantitative investigation is needed. Typically, all these coordination markers are part of and head the second coordinand, though *da* may appear with both coordinands at the same time (Huttar 2007: 224). The coordinands always involve finite phrases and belong to different intonational phrases; omission of subjects was judged ungrammatical by Nenge(e) speakers.

(13)   Ndyuka

    a.   Te   i si en ya,   da   *(i) koti lontu.
        when 2 see 3   here then   2   cut   around
        "When you see him here, (then) you avoid him."

    E-Pamaka

    b.   Di   mi be gi   en, ne   *(a) fika   ala nati poli, neen
        when 1   PST give 3   then   3   leave all wet   spoil then

        *(mi) go, neen *(mi) go tapu a   baafu.
        1   go then 1   go cover DET meat

        "When I gave him [grilled cassava], (and then) he left it wet and to spoil, (and then) I went (and then) I went to cover the sauce [and the meat]."

Functionally, these phrasal coordinators express a sense of succession (of events) that can be translated as "and then" (13–14; see 18 for *da*).

(14)   Matawai

    a.   Di   de   opo a   kuutu   paya, i go a   i   pisi,
        when they open DET meeting IDEO 2 go LOC 2POSS part

        nɔɔ i go a   i pisi, **nɔɔ** i kon   drai piisii   taa "we,
        then 2 go LOC 2 part then 2 come turn happiness say well

        wi   e   diingi nyan tee."
        1PL IMP drink   eat   a.lot

        nɔɔ di   i opo   **nɔɔ** i an   ta   fii   bunu, **nɔɔ** i toona
        then when 2 get.up then 2 NEG IMP feel good then 2 return

        go   a   data,   toona go a   data,   **nɔɔ** i ta   dede kai.
        go LOC doctor return go LOC doctor then 2 IMP die   fall

        i   dɛdɛ kaa!
        2 die   finish

        "When they started the meeting, you go to your quarter, **and then** you celebrate saying 'we'll drink and eat', **and then** you get up, **and then** you do not feel well, **and then** you go to the doctor, **and then** you fall down dying. You are already dead!"

---

7.   This form is closely associated with varieties of Nenge(e) and not generally found in Matawai or its sister language, Saamaka (McWhorter & Good 2012). Matawai speakers' contact with varieties of Nenge(e) has been on the rise in Suriname since the 1980s (Migge 2017).

b.　Den　naki　i　aiti　　bongi,　fo　　bata　　daan　fu di　i　lɛgɛdɛ,　**hen**
　　　they　hit　　2　eight　bundle　four　bottle　rum　since　2　lie　　　　then

den　taki　en　yei　　ye,　**hen**　a　piki　　　den　man　taki　"nei,　mi
3PL　say　3　hear　ASS　then　3　respond　DET　man　say　no　　1

an　　kɛ　　mɛki　toobi　da　　B."
NEG　want　make　trouble　give　B

"They fined you eight bundles and four bottles of rum for lying, **and then** they said he understood **and then** he replied to them saying 'no, I do not want to make trouble for B.'"

The meaning of succession is also found if the first phrase does not involve an action verb but a focus construction (15a) or another kind of non-verbal clause (15b).

(15)　Pamaka
　　　a.　Na　　a　　sani　fu　mi　ya,　**neen**　fosi　i　e　　yee　　den
　　　　　FOC　DET　thing　for　me　here　then　first　2　IMP　hear　DET

sani　di　　a　du.
thing　REL　3　do

"It's [due to] my plight, **and then** you first hear about the things that she did."

Matawai
b.　Fu　di　　se　　aki,　**nɔɔ**　mi　　　owma　disi,　hen　　pali
　　for　DET　side　here　then　1POSS　granny　DEM　3EMP　give.birth

mi　　　mama.
1POSS　mother

"For this side here, **thus/then** it's my granny, she gave birth to my mother."

However, *da* and *nɔɔ* and to a lesser extent *ne*(*en*) and *hen* also perform functions that do not express succession in the strict sense. For example in (16a–d) the first *da*, *neen*, *nɔɔ* and *hen* function to introduce a summary statement and the second *da* and *neen* introduce an aside remark.

(16)　Ndyuka
　　　a.　**Da**　ala　fa　　i　yee,　en　anga　a　　baya　feti　de　　**da**　a
　　　　　thus　all　how　2　hear　3　and　DET　guy　fight　there　and　DET

baya　fu　　wi　　ogii　　tuu.
guy　POSS　1PL　mean　true

"Thus all you hear, he and the guy fought there, and our guy is mean too."

b.　**Neen**　now　**neen**　den　man,　den　man　seefi　a
　　thus　now　and　3PL　man　3PL　man　self　PRE

koniman　　　　　　　neen　de　　taki　"A　bun!"
intelligent-person　then　3PL　say　It　good

"Thus now, and the men are sly persons, then they said 'OK!'"

Matawai

c. Ne a taki "A.K. dede o kii mi **nɔɔ** mi da di
   then 3 say name death FUT kill me thus I give DET

   basia da i."
   assistant give you

   "Then he said, 'A.K.'s death will kill me thus I give you the post of
   assistant to the leader.'"

d. **Hen** u dɛ anga di basia tee tata a kon dede.
   thus we COP with DET assistant until elder he come die
   "Thus we were with the assistant to the leader position until the elder
   died."

The kind of summative or result-introducing function of *da* occurs relatively fre-
quently in the data. The aside remark marking *da* (16) is rarer, but may also occur
sentence-initially as part of a second turn in a discussion (17). In Matawai the
summative or result-introducing function of *nɔɔ* is also present (see 15b).

(17) Pamaka

A: A dati, J. di abi en moy uwii.
   3 that name REL have her nice hair
   "That's right, J. who has nice hair."

B: Da a uwii e lebi gbey.
   then DET hair IMP red IDEO
   "and the hair is even turning red."

There are also differences in the distribution of the two phrasal conjunctive coordi-
nators in each language. *Ne(en)* in Nenge(e) and *hen* in Matawai coordinate phrases
that have a past temporal reference where predicates are either unmarked or are pre-
ceded by the past temporal marker (*be*, *bi*), such as in a narrative of past events (18).

(18) Ndyuka

a. A gi wan kabiten f'en anga a taa kabiten
   3 give a village.head POSS-3 and DET other village.head

   ne a taa wan kon konkuu gi mi taki, a man á
   then DET other one come lie give 1 say DET man NEG

   gi en fi en kaasinaoli ete. **Ne** mi taki, "pe u
   give 3 POSS 3POSS fuel yet then 1 say where 2PL

   fende den?" **Ne** a taki a Da P. baya gi en den.
   find them then 3 say FOC elder P guy give 3 3PL

"He gave one village head and the other one his share. **Then** the other came
and lied to me, saying that the man did not give him his fuel yet. **Then** I said
'where did you find them?'. **Then** he said it's Elder P.'s son who gave it to
him."

Matawai

b. Fa   i   si   dee   sembε de   aa, ala dee   wosu u   de
    how 2   see DET.PL people COP there all DET.PL house POSS 3PL

    bi   dε,   **hen** i   si   den flosi   komutu, **hen** den ko
    PST there then 2   see 3PL relocate leave   then 3PL come

    aki, **hen** u   ko   de   aki.
    here then 1PL come COP here

    "When all the people were there, all their houses were there, **then** you
    saw them leaving and relocating, **then** they came here, **then** we came to
    be here."

In contrast to *ne*(*en*) and *hen*, *da* in Nenge(e) (14a, 20) and *nɔɔ* (and *da*) in Matawai
(19) may link phrases that have different kinds of temporal or aspectual orienta-
tions. In (19a), for instance, the speaker is talking about an event that recurred sev-
eral times in the past. In (19b), the person is not talking about a specific event but
about something generic that is likely to happen. In (19c) the *nɔɔ*-headed phrase
introduces a future proposition.

(19) Matawai

a. Te   mi kumutu aa,   mi ko   a   Makayapingo. Teeeee,
    When 1   leave   there 1   come LOC name   a.lot

    en   na   yai buka, **nɔɔ** u   ta   ko   aki, u   pasa yai
    and FOC year mouth then 1PL IMP come here 1PL spend year

    ku   dee   sembε aki, **nɔɔ** u   gowe.
    with DET.PL people here then 1PL leave

    "When I left from there, I went to Makayapingo. Only after a long time,
    at the end of the year, **then** we were coming here, spending the new
    year with people here, **then** we went back."

b. Di   i   si   buulu sipan tee   a   sipan **nɔɔ** a   ta   bos.
    when 2   see blood tense a.lot 3   tense then 3   IMP burst

    "When you see that the blood pressure is very high (in the vein), it is
    going to burst open."

c. Mi ku   den bi   ko,   ma di   sikoo tapa, **nɔɔ** de   o   go
    1   and 3PL PST come but DET school close then 3PL FUT go

    a   dee   mama baka.
    LOC 3POSS mother back

    "I and them came together, but the school is closed, and then/thus
    they'll go back to their mother."

(20) Pamaka

a. U   begi paadon, **da** u   e   begi i, **da** ala sani mu
    We ask pardon then 1PL IMP ask 2 then all thing MOD

        waka bun.
        walk   well

        "We asked for forgiveness, beseeching you that all must work out."

  b. B.,   mi o   tya   nyanyan kon da   i o   puu   poti
      name 1   FUT carry food   come then 2 FUT remove put

      gi   den.
      give 3PL

      "B, I will bring the food out for you to distribute it to them."

The different temporal orientations of the two coordinators is also visible in discourses where they co-occur. In (21), a past event (following another) is introduced with *ne*(*en*) in line 1. In line 5, however, *da* is used to introduce an event that was on the brink of happening. When K then recounts what he actually said to the BO, that is again introduced by *ne*(*en*) (line 5) in the same way that the non-arrival of the paramount chief is introduced by *ne*(*en*) (line 6). In contrast, the assumption of what possibly happened – lack of BO's understanding in lines 7 and 11 – is again introduced by *da*.

(21)   Variation between *da* and *ne*(*en*) in Nenge(e): Ndyuka

    1 K: **Ne**   a   bosikopu kon taki gaaman ná o   kon.
        then DET message   come say chief   NEG FUT come
        "**And then** the message arrived saying that the paramount chief won't come."

    2 F: A ná o   kon, a á   kaba anga en   wooko.
        3 NEG FUT come 3 NEG finish with 3POSS work
        "He is not finished with his work."

    3 K: Ehee! A á   kaba a   wooko.
        No   3 NEG finish DET work
        "Yes! He has not finished his work."

    4 F: M!

    5 K: **Da** a   BO wani piki   taki a be taygi en so. **Ne**
        then DET BO want respond say 3 PST tell   3 so then

        mi taki,
        1   say

    6   "No! Efu a   man be taygi i taki a o   kon,
         NEG if   DET man PST tell   2 say 3 FUT come

    7   **ne** a á   kon, **da**   kande yu á   be fusutan   a
         then 3 NEG come then maybe 2   NEG PST understand DET

        man bun."
        man well

"**Then** the BO wants to reply saying he told him so (that he'd come).
Then I said, 'No! If the man had told you that he'll come, **and then** he
did not come, **then** maybe YOU did not understand him well.'"

8 F: Weeno "Exactly"

9 K: Bika   mi   á   wani a taki taki a   koli a
      Because   I   NEG want he say say he   lie he
      koli en.
      lie   him
      "Because I do not want him to say that he lied to him."

10 F: Nono!
      "No."

11 K: **Da** kande i no be fusutan   a   man bun.
      then maybe 2 NEG PST understand DET man well
      "Maybe you didn't understand the man well."

12 F: Weeno!
      "Exactly"

In (22) the speaker is narrating events that took place in the past, introducing
them with *hen*. However, when she introduces something that is still current, the
glass, she uses *nɔɔ* or *da*.

  (22)   Variation between *da* and *hen* in Matawai

        1 K: Ma fu famii, mi   hati mu koto.
            but for family 1POSS heart MOD cold
            "But for the family, I have to be calm."

        2 P: Weeno! "Exactly!"

        3 K: **Da** mi o   du wan sondi da   den.
            then 1   FUT do a   thing give them
            "**Then** I'll help them with something."

        4 P: Ya.
            "Yes."

        5 K: **Hen** mi manda wan u   di basia   taki go lai   wata,
            then 1   send   a   POSS DET assistant say go charge wáter

        6   **da** mi abi wan gasi so, so a de   ku   wata,
            and 1   have a   glass so so 3 COP with water
            "**Then** I send one of the assistants to pour water, **and** I have a glass like
            that, it is full up to here, **then** I went to stand up right in the middle of
            the meeting house."

        7 P: **hen** mi go taanpu leti mindi fu   di   gaangasa.   Ya.
            then 1   go stand right middle POSS DET meeting.house "Yes."

        8 K: **Hen** mi taki "kapten B, te   sembɛ ná   kon a   di
            then 1   say leader B when people NEG come LOC DET

9   dede fi   i   aki   taa de   misi mau a   i,
    death POSS 2 here say 3PL miss hand LOC 2

10 wi libisembɛ aki, **nɔɔ** u   ta   taki
    we humans   here then we   IMP say

11 we   sende en du ku   i…”
    well send 3   do with 2

"**Then** I said 'village head B, when people do not come to your burial here saying that they did something bad to you, we humans here, **then** we can send him to do the necessary with you…'"

As in (21), in (22) the hypothetical action and the current state are introduced by *da* (lines 3, 5) or *nɔɔ* (line 11) while the events that occurred in the past are introduced by *hen* (lines 5, 7, 8).

The data also include an instance where the two coordinators are juxtaposed (23). They do not appear to be false starts based on intonational data. It seems that the first element functions as a summative focusing device (cf. 16–17). Interestingly *da/nɔɔ* generally precede *ne(en)* and *hen* in such a construction.

(23)   Ndyuka
a.   Soo! **Da ne** u   sidon   na   a   taki de.   Da   a e
     right then then 1PL sit.down LOC DET talk there then 3 IMP

     kon   enke gaaman be   piki a   BO taki …
     come like chief PST tell DET BO say

     "Exactly! **Thus then** we held the meeting. Then it seemed to be that the paramount chief had told the BO that…"

Matawai
b.   Hen a balaki kon, da   di   a balaki ko,   **nɔɔ hen**
     then 3 vomit come then when 3 vomit come then then

     den   famii tei faya.
     DET.PL family tie fire

     "Then he poured it all out, then when he poured it all out, **thus then** the family members stopped the public discussion."

4.3   Coordination with *en*

En derives from Dutch where it is used as an all-purpose coordinator like *and* in English. However, in Nenge(e) and Matawai, apart from the occasional code-switched structures (Migge 2017) it occurs rarely. It is used to coordinate two finite phrases, but the second phrase essentially emphasizes the truth value or content of the previous phrase (19; see 24 for Ndyuka) and also marks the second phrase as an aside remark; it resembles one of the usages of *da* (16). In (24), the phrase headed by *en* emphasizes that the speaker knows that the listener knows this too (p.c. 2018 M. Moese).

(24) E-Ndyuka

    a.   A e   taki ingiisi, **en** a e   taki en bun.
         3 IMP talk English and 3 IMP talk 3   good
         "She speaks English and she speaks it well."

    b.   Na Polu fufuu a   moni, en i sabi.
         FOC Paul steal DET money and 2 know
         "It's Paul who took the money and you know that."

4.4 Zero coordination

Both data sets also include instances where two or more phrases that each have falling intonation at the end of the phrase are juxtaposed but are not conjoined by a coordinator. There are different types of such constructions. In one type different finite phrases denoting successive events are juxtaposed, and the last one may present the result (25a) of these actions. In (25) the phrases could have also been conjoined by *da* (25a), *da* and then *ne(en)* (25b) and *hen* (25c). The presence/absence of *da/ne(en)* does not appear to create a difference in meaning as such, in that both options (zero versus coordinator) express successiveness and results. However, presence of a coordinator puts focus on the successiveness nature of the events.[8]

(25) Pamaka

    (a)   Den bay sani anga ala nyan te   den kaba kelle. A
         they buy thing with all food until they finish IDEO DET

         m'manten ya   den go bay wan pisi kasi Ø   den go
         morning here they go buy a   piece cheese then 3PL go

         bay dii   beele oli na   ana Ø   den kon Ø   den
         buy three bread hold LOC hand then 3PL come then 3PL

         diingi te   Ø   wan wiki Ø   tu wiki Ø   moni á
         drink until then one week then two week then money NEG

         de   moo.
         COP more

         "They buy things and food until they are finished. In the morning, they buy a piece of cheese Ø they buy three breads to hold in the hand Ø they come Ø they drink tea Ø one week Ø two weeks Ø there is no more money."

    Ndyuka

    b.   Ne   den go lay den   boto te   a   lay. Di   den e
         then they go load DET.PL boat until 3SG load when 3PL IMP

         pasa a   wan fu   den   sula Ø   mi yee den sunguu
         pass LOC one POSS DET.PL rapid then 1   hear 3PL sink

---

8.   A speaker of Nenge(e) explained that adding *da/neen* in (21) makes the narrative very "heavy" and is "too much". I interpret this to mean that is it "formal" and "redundant".

huu Ø a boto koti a mindii a boto koti a
IDEO then DET boat cut LOC middle DET boat cut LOC

mindii kelen Ø ala a masini gwe.
middle IDEO then all DET machine leave

"Then they overcharged the boat. When they passed through one of the rapids there Ø I heard they sank to the ground Ø the boat broke fully in the middle Ø the outboard motor floated away."

Matawai

c.   Mi teki wan busje de Ø mi lei teeee Ø u ko
1 take a bus there then 1 drive a.lot then 1PL come

a wan kamia dɛ Ø a tya mi go a Zorg en Hoop
LOC a path there then 3 carry 1 go LOC name

Ø mi taa "we na ala a plein, na ala u saka
then 1 say well FOC there LOC plane FOC there 1PL land

ku plein" Ø u go a wan hoogeneete di sembɛ
with plane then 1PL go LOC a important.person DET person

di u go nɛ en, hen an hopo yete Ø u buta
REL 1PL go LOC 3 3EMP NEG get.up yet then 1PL put

kobo de…
bag there

"I took a bus there Ø I stayed in it for a long time until we came to a path there Ø it brought me to the Airport of Zorg en Hoop Ø I said 'well there is the plane, there we landed with the plane' Ø we went to an important woman – the person to whom we went, she had not gotten up yet Ø we left the luggage there…"

Juxtaposition also encodes simultaneity. In (26a) the speaker describes several events/actions that are taking place concurrently; only the last action (*den á wani yee…*) is dependent on/ succeeds the previous one (*moni sani teke a liba*). Conjoining them with *ne(en)* would create a sense of successiveness. In (26b) the suggested actions are like a list of things that should happen but not in that order; insertion of *nɔɔ* would create this sense of successiveness. Note also the structural parallelism that also creates a linkage between them.

(26)   Ndyuka

a.   Man de anga boketi e go a soo Ø man de anga
man COP with bucket IMP go LOC shore and man COP with

gaan langa angina saka e seli (unclear) Ø den e seli
big long rice bag IMP sell and 3PL IMP sell

na a kondee ya te sama di á poy moo. Di
LOC DET village DEM until person REL NEG MOD more the.one

e teke sawtu Ø a teke sawtu Ø di e teke sopi.

IMP take salt   and 3 take salt   and the.one IMP take rum

We   di   a   moni sani teke a   liba Ø   den ná   e
well when DET money thing take DET river then 3PL NEG IMP

wani yee   den   sani de   moo.
want hear DET.PL thing DEM anymore

"There are men with buckets [of fish] at the waterfront Ø men are with long rice bags selling (unclear) Ø they are selling in the village until there are no more people (to sell to?). The one who takes salt Ø s/he takes salt Ø the one who takes rum. Well, when the money took over on the river Ø they do not want to hear the things any more."

Matawai

b. De ta   yaika [d]i soni man aki ta   kɛɛ: mɛ   a an   ta
    3PL IMP listen DET thing man DEM IMP cry   make 3 NEG IMP

kɛɛ moo anga en   Ø   mɛ   a fika da   mi wawan Ø
cry more with 3   and make 3 leave give 1   alone   and

mɛ   a fika da   mi wawan Ø   mɛ   a taki ko.   I si
make 3 leave give 1   alone   then make 3 say come 2 see

di   u   go sindo   aki nɔɔ a ta   kɛɛ te   anga now a
when 1PL go sit.down here then 3 IMP cry when with now 3

an   man taki di   soni.
NEG MOD say DET thing

"They heard how that man was crying: make him not cry anymore Ø make him stay alone with me Ø make him stay alone with me Ø make him speak up. You see, when we came to sit down here then he was crying until now/then he could not talk about the matter."

Juxtaposition may also involve non-finite sentences encoding simultaneity (27a) or successiveness (27b, 27c) if the subjects are coreferential. Note that only the imperfective marker (*e/ta*) may be repeated in both phrases (27a and 27e). In the case of transitive verbs coreferential objects/patients are also absent (27d). Some of these constructions resemble resultative-type serial verb constructions (e.g., 27c).

(27) Pamaka

a. a   di   mi de ape, ne   mi si a e   dongo   Ø
    FOC when 1 COP there then 1   see 3 IMP go.down.river and

e   pasa anga a   siki pikin.
IMP pass with DET sick child

"When I was there, I saw him going downriver **and** passing with the sick child."

b. Da   a fika a   boto fu en anga en   uman boli
    then 3 leave DET boat POSS 3 with 3POSS wife cook

    Ø   nyan.
    then eat
    "Then he left the boat for him and his wife to cook **and** eat."

c.   Na en anga en libi Ø   meke a   pikin ya.
    FOC 3   with 3   live then make DET child DEM
    "Then they lived together **and** had this child."

    Matawai

d.   I ná a   fu go begi eijs Ø   bebe moo.
    2 NEG have for go ask ice then drink more
    "You do not have to ask for ice **and** drink [ice] anymore."

e.   A ta   kendi Ø   ta   naki vip.
    3 IMP hot   and IMP beat IDEO
    "It was hot **and** beating."

Finally, property items in attributive (28a, 27b) and predicative (28c) position are also conjoined using juxtaposition – note that property items in predicative position are verbs (Huttar & Huttar 1994; McWhorter & Good 2012 for Saamaka; Migge 2013 for Nenge(e)). In this case juxtaposition expresses a sense of simultaneity. In (28a) and (28b), the two property items are part of the same intonational curve and it is not possible to insert *anga/ku* or any of the verbal coordinators. In the case of (28c), the two property items are part of two separate phrases as there is falling intonation on both. Juxtaposition may be replaced with the coordinator *da* (28d, 28e) but in the case of attributive contexts (cf. 28b and 28e), this would entail a change from a single nominal phrase to two phrases, and the second is a finite clause where the property item now functions as a verb (28e) rather than a nominal modifier (28b), because *da* can only coordinate verbal phrases. The recordings did not include such examples.

(28)   E-Ndyuka

a.   A   uman de   abi   wan langa Ø   fini neki.
    DET woman DEM have a   long   and thin neck
    "This woman has a long and thin neck."

b.   A abi   wan langa Ø   baala boto.
    3 have a   long   and wide boat
    "He has a long and wide boat."   (p.c. 2018 M. Moese)

c.   En   boto langa Ø   baala.
    3POSS boat long   and wide
    "His boat is long and wide."   (p.c. 2018 M. Moese)

d.   En   boto langa **da**   a baala.
    3POSS boat long then 3 wide
    "His boat is long and it is wide."   (p.c. 2018 M. Moese)

e. A abi   wan langa boto **da**   a baala.
    3 have a   long boat then 3 wide
    "He has a long boat and it is wide."   (p.c. 2018 M. Moese)

The Matawai data did not include sentences with two property items. This needs
to be further investigated. The data also did not include instances where juxtaposi-
tion was used to express disjunctive or adversative meanings.

4.5   *Soseefi*

The coordinator *soseefi* in Nenge(e) and *soseepi* in Matawai (<English *so self)* is
used to link two propositions and expresses similarity between them. Syntacti-
cally, the two phrases are either finite clauses (29a) or the one headed by *soseefi*
may also be a reduced phrase where the verb of the second phrase may be implied
if it is coreferential with that of the first phrase (29b). These constructions are not
very frequent in Nenge(e) and Matawai. Note that both elements can also be com-
bined with *da* (Nenge(e)) and *da/nɔɔ*.

(29) Ndyuka
      a. Den á   mu   de   sama di   abi   bigi fasi
          3PL NEG MOD COP person REL have big manner

          soseefi den á   mu   de   sama di   e   waka
          so-self 3PL NEG MOD COP person REL IMP walk

          diingi te   enke den e   duungu.
          drink until like 3PL IMP drink

          "They mustn't be arrogant people; also they mustn't be people who go
          around drinking till they're practically drunk."

                                          (Huttar & Huttar 1994: 228)
      b. Fa   i si du fu   den, du fu   u   e   taanga gi   den,
          How   2 see POSS 3PL   POSS 1PL IMP difficult give 3PL

          a   soseefi du fu   den.
          FOC so-self POSS 3PL

          "See theirs [language], ours is difficult for them, in the same way that
          theirs [is difficult for us]."

      Matawai
      c. Na   di   soni f'en an   ta   kon   a   mi soseepi yu tu
          FOC DET thing POSS3 NEG IMP come LOC 1   so-self 2   also

          i go a   someni wosu…
          2 go LOC so-many house

          "Her matter was not brought up to me, in the same way you also went
          to so many houses (but did not tell me)."w

5.   Disjunctive and adversative constructions

Disjunctive coordination involves the presentation of options. It is most typically expressed by the conditional marker *efu* (or *ofu*) in Nenge(e) (< English *if* and Dutch *of*) and by its counterpart *ee* in Matawai (30). They both coordinate finite sentences only, which may be direct or indirect statements or questions (Huttar & Huttar 1994: 233–235). Coreferential subjects must be repeated, as in (30b). This function of *efu* and *ee* does not occur frequently in the data.

(30)   Pamaka

a.
Efu	den	abi	ati	fu	komoto	a	BT.	N.	an	sabi
if	3PL	have	heart	for	leave	LOC	name	name	NEG	know

san	a	goontapu.	**Efu**	a	S.	e	kali	en	**efu**	na	en
what	COP	world	or	FOC	name	IMP	call	3	or	FOC	3

seefi	fustan	e	kali	en.
self	mind	IMP	call	her

"If they are couragerous enough to leave BT., then N. does not know anything about how the world works or it's S. that's making her do it or her own mind is making her do it."

Matawai

b.
Nɔɔ	i	o	yei	fa	di	libi	fu	u	dɛ,	wo	go	**ee**
thus	2	FUT	hear	how	DET	life	POSS	1PL	COP	1PL.FUT	go	or

u	ná	o	go?
1PL	NEG	FUT	go

"Thus you will hear about our life, will we go or will we not go?"

Adversative coordination involving expression of counter-expectation is typically conveyed by *ma* (< Dutch *maar*; <Portuguese *ma*) in both Nenge(e) and Matawai (31).

(31)   Pamaka

a.
Mi	boli	gi	en	ma	a	an	nyan	moo.
I	cook	give	3	but	3	NEG	eat	more

"I cooked for her but she did not eat anymore."

Matawai

b.
Den	pali	wan	mii	a	Foto	aki	ma	da	M.,	den	tye
3PL	birth	one	child	LOC	town	here	but	then	M.	PL	carry

en	ko	ala.
it	come	there

"They gave birth here in Paramaribo, but then M., they brought him there."

Additionally, there is also the adversative coordinator *toku* (*so*) (< Dutch *toch*). From a functional perspective, it seems to be synonymous with *ma*. It is possible

that *ma* and *toku* (*so*) are stylistic variants, as *toku* (*so*) seems to occur more typically in formal speech such as formal meetings. It is often also combined with *ma* (32b and 32c). Huttar & Huttar (1994: 323) also suggest that there is "an ascending order of strength of the counter-expectation" between *ma*, *toku* and *ma toku*; I have not been able to confirm this so far.

(32)   Pamaka
     a. Efu   na   wan   taa   sama   toku a   gaaman abi   fu   go
        if   FOC   a   other person yet   DET   chief   have for   go

        a   ini.
        LOC   in

        "If it was another person [to whom it happened] but the paramount chief has to intervene."

     b. A   sani de   u   be   kuutu en   kaba   […] ma toku da
        DET   thing DEM   1PL   PST   discuss 3   COMPL   but yet   then

        u   be   yee, wan   pikin foo pasa a   u   yesi taki da   a
        1PL   PST   hear a   little bird pass LOC 1PL   ear say COOR 3

        wani tya   kuutu   baka.
        want carry discussion back

        "This matter was arbitrated before but yet we heard, a little bird passed by our ears, that the matter needs to be arbitrated again."

    Matawai
     c. U   bi   taaki mɛ   toobi   an   pasa   ma tɔku, fa   u   taki
        we   PST   say   make trouble NEG   happen but yet   how we   say

        seepi, a   to   pasa.
        self   it   FUT   happen.

        "We had said let trouble not happen but yet as we speak, it will happen."

6.   Summary and conclusion

The investigation of coordination in two Suriname Creoles, the Eastern Maroon Creoles and one of the Western Maroon Creoles, Matawai, showed that they resemble a great number of P/Cs in that there is identity between the comitative, instrumental and nominal coordinator. However, Nenge(e) differs from many other P/Cs in that this element, *anga*, did not derive from English *and*. *Ku* in Matawai may derive from Portuguese *with*, though. *Anga* also resembles some other P/Cs in that it has been extended to clausal conjunction, albeit in a very limited and specialized context. In both Nenge(e) and Matawai nominal conjunction is distinguished from verbal conjunction. However, unlike most other P/Cs they

Table 7.2 Summary of syntactic and functional properties of coordinators in Nenge(e) and Matawai

Property	Nenge(e)										Matawai								
	anga	da	ne	en	Ø	soseefi	efu	ma	toku	ku	da	nɔɔ	hen	en	Ø	soseepi	ee	ma	toku
Syntactic properties																			
NP/N conjunction	X									X									
comitative	X									X									
instrumental	X									X									
finite clause conjunction	(X)	X	X	X	X	X	X	X	X	?	X	X	X	X	X	X	X	X	X
non-finite conjunction (if coref)	(X)				X	X									X	X			
Linking of adjs					X										(X)				
combination with *da / nɔɔ*			X			X							X	X					
combination with *ma*									X										X
Functional properties																			
jointness	X									X									
successiveness		X	X		X						X	X	X		X				
Focus on successiveness		X	X								X	X	X						
summary		X	X								X	X	X						
aside remark		X		X							X			X					
past time			X										X						
different tenses		X			X						X	X		X					
simultaneity					X										X				
similarity						X										X			
disjunctive							X										X		
adversative								X	X									X	X

both have several coordinators, two in Nenge(e) and three in Matawai. They have similar structural and functional properties, but occur in different temporal contexts and do not derive from English *and*. The temporal difference has, to date, not been documented for other P/Cs outside of Suriname. The overt verbal coordinators alternate with juxtaposition, which may express successiveness and, unlike the overt coordinators, also simultaneity. However, unlike Cape Verdean, juxtaposition is not the dominant strategy nor can it express disjunction and adversativeness. Both Matawai and Nenge(e) also use a reflex of "and" in European languages, *en*. However, this marks additional or aside remarks rather than conjoining equal conjuncts. Finally, coordinators can also be combined with each other either to strengthen their meaning (*ma toku*) or perform related functions (*da neen*). A summary of the properties of the different coordination constructions is given in Table 7.2. The discussion in this chapter suggests that coordination in the two Suriname Creoles differs in several respects from that in English, Dutch or Portuguese, and also does not fully cohere with established patterns for other P/Cs. Further research on coordination in the relevant substrate languages may help to clarify the origins of these divergent features.

References

Aboh, E. O. 2006. Complementation in Saramaccan and Gungbe: The case of C-type modal particles. *Natural Language and Linguistic Theory* 24: 1–55.
https://doi.org/10.1007/s11049-005-1815-4

Byrne, F. 1987. *Grammatical Relations in a Radical Creole: Verb Complementation in Saramaccan* [Creole Language Library 3]. Amsterdam: John Benjamins.
https://doi.org/10.1075/cll.3

DeGraff, M. 2007. Kreyòl Ayisyen, or Haitian Creole (Creole French). In *Comparative Creole Syntax: Parallel Outlines of 18 Creole Grammars,* J. Holm & P. Patrick (eds), 101–126. London: Battlebridge.

Goodall, G. 2017. *Coordination in Syntax* [Linguistics: Oxford Research Encyclopedias]. <http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-36> (4 April 2019).

Goury, L. & Migge B. 2003. *Grammaire du Nengee: Introduction aux Langues Aluku, Ndjuka et Pamaka*. Paris: Editions IRD. (Revised version published in 2017).

Haspelmath, M. 2005. Coordinating constructions: An overview. In *Coordinating Constructions* [Typological Studies in Language 58], M. Haspelmath (ed.), 3–40. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.58.03has

Haspelmath, M. & APiCs Consortium. 2013. Nominal and verbal conjunction. In *The Atlas of Pidgin and Creole Language Structures*, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 284–287. Oxford: OUP.

Haspelmath, M., Dryer, M. S., Gil, D. & Comrie, B. 2005. *The World Atlas of Language Structures*. Oxford: OUP. <https://wals.info/> (4 April 2019).

Hassamal, S. 2016. Two AND-conjunctions in Mauritian. In *Coordination and Subordination: Form and Meaning, Selected Papers from CSI Lisbon 2014*, F. Pratas, S. Pereira & C. Pinto (eds), 135–156. Newcastle upon Tyne: Cambridge Scholars.

Holm, J. 2000. *An Introduction to Pidgins and Creoles*. Cambridge: CUP.
https://doi.org/10.1017/CBO9781139164153

Huttar, M. 2007. Ndyuka (Creole English). In *Comparative Creole Syntax*, J. Holm & P. Patrick (eds), 217–236. Westminster: Battlebridge.

Huttar, M. L. & Huttar, G. L. 1994. *Ndyuka*. London: Routledge.

Lefebvre, C. 1991. *Serial Verbs: Grammatical, Comparative and Cognitive Approaches* [Studies in the Sciences of Language Series 8]. Amsterdam: John Benjamins.
https://doi.org/10.1075/ssls.8

Lefebvre, C. 2015. Coordinating constructions in Fongbe with reference to Haitian Creole. In *Functional Categories in Three Atlantic Creoles: Saramaccan, Haitian and Papiamentu* [Creole Language Library 50], C. Lefebvre (ed.), 69-108. Amsterdam: John Benjamins.
https://doi.org/10.1075/cll.50

Maurer, P. & APiCS Consortium. 2013a. Comitatives and instrumentals. In *The Atlas of Pidgin and Creole Language Structures*, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 276–279. Oxford: OUP.

Maurer, P. & APiCS Consortium. 2013b. Noun phrase conjunction and comitative. In *The Atlas of Pidgin and Creole Language Structures*, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 280–283. Oxford: OUP.

McWhorter, J. H. 1992. Substratal influence on Saramaccan serial verb constructions. *Journal of Pidgin and Creole Languages* 7: 1–53. https://doi.org/10.1075/jpcl.7.1.02mcw

McWhorter, J. H. & Good, J. 2012. *A Grammar of Saramaccan Creole*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110278262

Mendes, M., Quint, N., Ragageles, F. & Semedo, A. 2002. *Dictionário Prático Portugués-Caboverdiano*. Lisbon: Verbalis.

Michaelis, S., Maurer, P., Haspelmath, M. & Huber, M. (eds). 2013a. *The Atlas of Pidgin and Creole Language Structures*. Oxford: OUP.

Michaelis, S., Maurer, P., Haspelmath, M. & Huber, M. (eds). 2013b. *The Survey of Pidgin and Creole Languages*. Vols. 1-3. Oxford: OUP. <https://apics-online.info/> (3 January 2019).

Michaelis, S., Maurer, P., Haspelmath, M. & Huber, M. (eds). 2013c. *The Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://apics-online.info/> (03 January 2019).

Migge, B. 1998. Substrate influence in creole formation: The origin of *give*-type serial verb constructions in the Surinamese Plantation Creole. *Journal of Pidgin and Creole Languages* 13: 215–265. https://doi.org/10.1075/jpcl.13.2.02mig

Migge, B. 2013. Nengee. In *The Survey of Pidgin and Creole Languages*. Vol. 1, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 39-48. Oxford: OUP.

Migge, B. 2017. Putting Matawai on the Surinamese linguistic map. *Journal of Pidgin and Creole Languages* 30: 232–262. https://doi.org/10.1075/jpcl.32.2.02mig

Migge, B. & Léglise, I. 2015. Assessing the sociolinguistic situation of the Maroon Creoles. *Journal of Pidgin and Creole Languages* 30: 63–115. https://doi.org/10.1075/jpcl.30.1.03mig

Migge, B. & Winford, D. 2013. Fact-type complements in Gbe and the Surinamese Creoles. *Lingua* 129: 9–31. https://doi.org/10.1016/j.lingua.2012.03.006

Mufwene, S. S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP.
https://doi.org/10.1017/CBO9780511612862

Plag, I. 1993. *Sentential Complementation in Sranan: On the Formation of an English-Based Creole Language*. Tübingen: Max Niemeyer.

Quint, N. 2008. Coordination et parataxe en capverdien moderne (dialecte santiagais ou badiais). In *Surbordination, Dépendance et Parataxe dans les Langues Africaines*, B. Caron (ed.), 29–48. Louvain: Peeters.

Richardson, L. A. 1977. The phrasal conjunctor and the comitative marker in Papiamentu. *Amsterdam Creole Studies* 1:55–68

Sebba, M. 1987. *The Syntax of Serial Verbs* [Creole Language Library 2]. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.2

van den Berg, M. & Smith, N. S. H. 2013. Early Sranan. In *The Survey of Pidgin and Creole Languages*. Vol. 1, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 3–14. Oxford: OUP.

van Stipriaan, A. 2015. *Maroons and the communications revolution in Suriname's interior*. In *In and Out of Suriname: Language, Mobility and Identity*, E. Carlin, I. Léglise, B. Migge & P. Tjon Sie Fat (eds), 139–163. Amsterdam: Brill.

Veenstra, T. 1996. *Serial Verbs in Saramaccan: Predication and Creole Genesis*. The Hague: HIL.

Winford, D. 2008. Atlantic Creole syntax. In *The Handbook of Pidgin and Creole Studies*, S. Kouwenberg & J. V. Singler (eds), 19–47. Malden MA: Wiley-Blackwell. https://doi.org/10.1002/9781444305982.ch2

Winford, D. & Plag, I. 2013. Sranan. In *The Survey of Pidgin and Creole Languages*. Vol. 1, S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds), 15–26. Oxford: OUP.

Yakpo, K., van den Berg, M. & Borges, R. 2015. On the linguistic consequences of language contact in Suriname: The case of convergence. In *In and Out of Suriname: Language, Mobility and Identity*, E. Carlin, I. Léglise, B. Migge & P. Tjon Sie Fat (eds), 164–195. Amsterdam: Brill.

CHAPTER 8

Reflections on Darwin's natural selection

A lens on variation, competition and change

Marlyse Baptista
University of Michigan

In this chapter, I adopt Mufwene's (2001, 2008) concepts of *feature pool* and *feature selection* and *competition* and compare it to Darwin's theory of *natural selection* (Darwin 1859). Such a comparison brings to light both the descriptive and explanatory adequacy of Mufwene's theoretical framework, showing how his work has undoubtedly furthered our understanding of language emergence and development. I will examine more specifically how, given Mufwene's biological model of language evolution, variation and diversity can be seen as key factors driving feature competition and selection, ultimately leading to language change. I propose that the ways in which linguistic features emerge, coexist, compete and get selected can find parallels in processes of natural selection in nature.

Keywords: feature pool, natural selection, competition, language variation, language diversity,

1. Introduction

In honour of Salikoko Mufwene's invaluable contributions to the field, this chapter focuses on the relevance of Mufwene's "competition and selection" framework (2001, 2008) to Darwin's theory of natural selection (Darwin 1859).[1] Such a comparison brings to light both the descriptive and explanatory adequacy of

---

[1] It is a privilege to be part of this volume honouring the monumental contributions that Salikoko Mufwene has made to the field of linguistics in particular and to science in general. Sali is undoubtedly one of the greatest thinkers and linguists of our times. I owe him a personal debt of gratitude for being an unofficial mentor ever since I was in graduate school, and an unflinching source of support. More importantly, Sali's impactful work has consistently been a guiding light in my and many scholars' research, as well as a source of inspiration for

Mufwene's theoretical framework, showing how his work has undoubtedly furthered our understanding of language emergence, development and change. More precisely, this comparison draws parallels between Darwin's study of how the selection process operates in nature and Mufwene's analysis of how linguistic features compete and are selected in a feature pool. In sum, the objective of this chapter is to explore how, given a biological model of language evolution, variation and diversity can be seen as key factors driving feature competition and selection, ultimately leading to language change. I propose that the ways in which linguistic features emerge, coexist, compete and get selected can find parallels in how Darwin has hypothesized natural selection processes in nature.

This chapter is organized as follows: in the second part, I show how the notion of competition and selection has been used in linguistics to study first and second language acquisition, bilingual speech production and language contact, highlighting how Mufwene's framework has proved to be a highly valuable heuristic tool for explaining how features are selected when new languages emerge, and how such features ultimately contribute to language change. In the third Section, I elaborate on Mufwene's examination of idiolects as driving engines of change, and in the fourth, I summarize Darwin's view of how natural selection operates in nature and transpose his observations on the importance of minute variation in nature (driving evolution) to the importance of idiolectal variation in Mufwene's work. As a working definition, we can adopt the idea that an idiolect is an individual's distinctive and unique use of language that affects their vocabulary, grammar and pronunciation. I highlight the extent to which every little bit of variation matters in contributing to language development and change. Finally, I examine a Swadesh list representative of five distinct varieties of the Cabo Verdean language to show how the observed variation involving many of the lexical items is clear evidence in support of Mufwene's framework of feature competition and selection.[2] In 2010, I taught a course on field methods at the University of Cabo Verde in Praia. The students in that course came from several islands in the archipelago, thus presenting us with the unique opportunity of building a Swadesh list to illustrate dialectal variation between islands.[3] As it turned out, the Swadesh list

---

2. A Swadesh list is a compilation of basic words and concepts in a language. It was originally designed by Morris Swadesh for cross-linguistic comparative purposes.

3. I am deeply grateful to the first cohort of Master's in Creolistics students who attended the Field Methods course I taught at the University of Cape Verde in December 2010. I would not have been able to compile this Swadesh list without their valuable contribution and insightful

clearly revealed variation within each island, and within the groups of speakers representative of each island. This highlights Mufwene's notion of inter-idiolects as the loci of language variation and change (Mufwene 2008). It is also reminiscent of Chomsky's notion of I-language (Chomsky 1995) and the use of this concept in DeGraff's (1999) analysis of "I-creolization", with I-language being the main locus of language creation and change.[4] Using Mufwene's framework to analyze the Cabo Verdean Swadesh list featuring lexical variation ultimately demonstrates the benefits of his model in furthering our understanding of how variation, competition and selection operate in language.

2.   Competition as a valuable heuristic tool in linguistics

The notion of feature competition has been used productively in the domains of child language acquisition, second language acquisition, bilinguals' speech production and situations of intense language contact, as in the case of Creole formation.

In child language acquisition, Bates and MacWhinney (1987) and MacWhinney and Bates (1989) developed a psycholinguistic model of language acquisition and sentence processing called the competition model. The model proposes that when acquiring language, the language learner first compares a number of linguistic cues endowed with specific functions before settling on a particular one. Bates and MacWhinney examine competition within the same language and focus, for instance, on how competitive processing in language acquisition involves the child choosing between the transitive "sink" (as in "the pirate sank the boat") and the intransitive "sink" ("the boat sank"). MacWhinney and Bates consider a variety of cues in an attempt to account for which one of these two forms wins out when lexical activation takes place; their model proposes that when interpreting a given sentence, individuals consider contextual, syntactic, morphological and semantic cues in order to select the interpretation which is probabilistically the most likely to be correct.

---

discussion. Their names are listed here: Diamantino Tavares Freire, Adelica Teixeira Barbosa Goncalves, Amélia Gomes, Gil Vaz, Delcy Sena, Elisangela Spencer Coelho, Karina Moreira, Lurdes Lima, Maria do Céu Santos Baptista, Eliane Vieira, Emanuel de Pina, Risete dos Santos, Raimundo Lopes, Neidina Lopes, Felisberto Veiga, Angela Cidário, Angélica Cardoso, Hulda Costa, Dysena Pereira, Manuel Rosa, Neidina Lopes, Muniz Afonso, Carmen Oliveira Semedo, Ligia Herbert Robalo, Sérgio Costa, Djito Lopes.

4.   Mufwene's (2008) treatment is indeed very similar to DeGraff's (1999) insights. I refer the reader to DeGraff's "Language Creation and Language Change: A prolegomenon" (1999: 9) and "Language Creation and Language Change: An epilogue" (1999: 474–485).

For other linguists like Yang (2002), who also examines child language acquisition, competition takes place between the I-grammars available to a child from the very start. Yang proposes that child language acquisition involves a statistical combination of multiple possible grammars that are all allowed by UG, but only some of which are retained by the time children acquire their L1 (Yang 2002: 12). His model of language acquisition argues for parametric variation (Yang 2002: 26) and is schematized in example (1). This model is anchored in the notion of multiple competing grammars.

(1) Each grammar Gi is associated with a weight Pi, which reflects the degree of dominance of Gi in the learner's language faculty. Evolving in a linguistic environment E, three factors contribute to determining the weight Pi (E, t) of a given grammar: the learning function L, the linguistic evidence in E, and the time variable t (t is the time that has elapsed since language acquisition started).

Yang's learning model, which argues for competition between grammars and which ultimately leads to the selection of one, is schematized in (2):

(2) Upon the presentation of an input datum s, the child
    a. selects a grammar Gi with the probability Pi
    b. analyzes s with Gi
    c. – if success, reward Gi by increasing Pi
      – otherwise, punish Gi by decreasing Pi.   (Yang 2002: 26–27)[5]

The schema in (2) is meant to illustrate that in a set of competing grammars, those that succeed in analyzing a given utterance gain more prominence in a learner's hypothesis space, whereas grammars that fail are penalized and are gradually eliminated from the competition. Crucially, this schema involves competition and elimination of failing grammars.

While Bates and MacWhinney (1987), MacWhinney and Bates (1989), and Yang (2002) examine the notion of competition in child language acquisition, other scholars have extended the competition model to L2 acquisition and bilingual speech production.

For instance, the system of cue weights was recently implemented for L2 acquisition, as the competition model postulates that different languages use different cue weights that speakers need to learn in order to derive the correct meaning of a given sentence (Kroll & de Groot 2005). Kroll and de Groot propose that a set of cognitive mechanisms is responsible for the activation of competing

---

5. As pointed out by an anonymous reviewer, Yang's "learning by parsing failure" approach is reminiscent of Clark & Roberts (1993).

representations whose weights are computed in the target language in the course of its acquisition and usage.

Scholars like Costa et al. (2003) explore the same competition model as it operates in bilingual speech production. They explore lexical and phonological factors in cross-language competition in bilingual speech production, focusing specifically on lexical selection. More precisely, their study considers how bilingual speakers retrieve the lexical item that corresponds to the concept they wish to convey. Based on the assumption widely shared in speech production models that there is cross-language competition in bilingual speech, Costa et al.'s (2003) study aims at showing that the targeted word that is accessed for retrieval competes with other activated lexemes. According to such models, when retrieving a lexeme (for instance, "dog"), its semantics are activated at the same time as other semantically related words (for instance, "cat"), and the degree of difficulty in selecting the targeted lexeme will depend on both that item's activation level and the activation of the other competing semantically related lexemes.

In their study using highly proficient bilinguals of Spanish and Catalan, Costa et al.'s (2003) subjects were instructed to name pictures in their L2 (Catalan), while ignoring distractor words presented in their L1 (Spanish). The aim of their two experiments was to explore whether, in L2 naming, distractor words that are phonologically related to the target's translation (this is called the phono-translation condition) slow down naming latencies (mean time from presentation to articulation of symbol) when compared to distractors that are not phonologically related. In other words, their objective was to examine whether Spanish-Catalan highly proficient bilingual speakers can prevent interference from their L1 (Spanish) when producing speech in their L2 (Catalan). The most relevant condition in the two experiments in their study was the one in which the distractor word was phonologically related to the target's translation – the phono-translation condition. In support of a cross-linguistic competition view of bilingual speech production, the results of their two experiments clearly showed that phono-translation distractors interfere more than phonologically unrelated distractors in the course of picture naming.

The notion of competition and selection has also proved heuristically very useful in cases of "language creation", especially given a scenario where contact between multiple languages gives rise to a new linguistic system, i.e., a pidgin or a Creole. Mufwene (2001, 2008) offer the first rigorous articulation of competition and selection as these processes contribute to the emergence of new contact varieties. Mufwene's concept of *competition* refers to linguistic features in a feature pool representative of diverse grammatical features competing for survival in the newly emerged language. In regard to the term "competition", Mufwene (2008) makes it explicit that this involves the participation of features from *both* substrates and superstrates to the feature pool. As he puts it,

> in the case of creoles, features of the substrate languages that new appropriators of the target language had spoken before were also contributed to the feature pool. The presence of these xenolectal features in the pool changed the balance of power among the native variants, making it possible for the selection into creoles of features other than those selected into the other, non-creole varieties. It also made it possible for features from the substrate languages to make their way into the idiolects of creole children …   (Mufwene 2008: 118)

Such statements clearly assume not only that both substrates and the lexifiers have a role to play in the emergence of new contact varieties, but also that idiolects have an important role to play in language emergence and change; this is a point that I capitalize on in this chapter when drawing parallels between feature selection and competition in language development on the one hand, and natural selection among species on the other.

In the next section, I elaborate on Mufwene's insightful observation that idiolects are the locus of language variation and change, before comparing his stance to Darwin's position on the importance of small variations and their significant long-term effects on natural selection, in Section 4.

3.   On the role of idiolects, infinitesimal variations and their powerful effects on language change

What is an idiolect? What counts as a variant? How do variants participate in language change? Mufwene's framework of selection and competition is effective in answering these questions. Indeed, when examining the very nature of the variants present in the feature pool, Mufwene argues that the variants that are in competition may differ not only lexically (choice of one word over another) but also phonologically, morphologically, syntactically, semantically and pragmatically. In brief, these variants occur and compete with each other in all areas of grammar. Distinct environments of interaction between learners within and across diverse speech communities can lead speakers to select different subsets of variants into their *idiolects* (Mufwene 2008). Mufwene fine-tunes the notion of *inter-idiolectal variation* in the following passage:

> Variation among individual speakers' personalities and histories of social interactions produces inter-idiolectal variation of the kind where no speaker replicates anybody else's idiolect, although those of speakers who evolve in the same communication networks are more similar to each other. That is, although all speakers/learners select their features (albeit with some modification) from the same pool, they do not select exactly the same variants. Since selection is not always in exclusive terms, even if they select the same features, these are not recombined the same way in their idiolects, because those that wind up as dominant are not always the same.   (Mufwene 2008: 117–118)

With such statements, Mufwene clearly identifies the idiolect as being an important level of variation, as each individual recombines the selected variants in a unique, distinctive and idiosyncratic way, the nature of such recombinations depending on interaction settings and experiences. This particular lens on the idiolect as a powerful factor and ultimate locus of language variation (and change) makes it akin to an individual's "linguistic fingerprint", with long-term ripple effects on the development of dialects and languages. Mufwene's view of change with the idiolect as the starting point is made explicit in the following statement:

> The variation that *matters* to evolution really begins at that inter-idiolectal level, before reaching the next higher level of cross-dialect and/or cross-language differences. Contact, which has been dealt with primarily at the level of dialects or languages, really begins at this level of idiolects. Since the locus of dialect or language contact is the mind of the individual speaker, the difference between idiolect contact and language contact or dialect contact is more quantitative than qualitative. (Mufwene 2001: 149–150, my emphasis)

Mufwene's view of inter-idiolectal variation sets the stage for a scenario where language change starts in the individual before propagating to the dialect and to the language as a whole.[6] This view also has major implications for language acquisition and language transmission. Indeed, in other work, Mufwene (2006) proposes that language is never transmitted wholesale and that language transmission necessarily entails a reconstructive process that proceeds one piece at a time, which supports the generally agreed upon idea that there is never perfect replication in language transmission/acquisition (Mufwene 2006: 463).[7]

This perspective on the role of the idiolect is in keeping with a view of language as a species, due to variation and different evolutionary paths across languages, different speeds at which they develop and different stages and processes in their

---

6. As one of the reviewers of this chapter points out, this is not always clear when one is dealing with a language versus a dialect, as ideological considerations can blur the lines between the two. The reviewer draws a likely parallel between the spread of features that is discussed here and various sociolinguistics models (e.g., Labov 2007) that refer to feature propagation as dialect levelling or koineization.

7. This is reminiscent to some extent of Crain et al. (2006), who support the continuity hypothesis according to which children's speech differs systematically from the language spoken in their environment. In early stages of language acquisition, children reproduce patterns found in other adult grammars in the world; such divergences are observed as the children experiment with different constructions that differ from those of the local language but are attested in other languages of the world. However, such drastic divergences are only temporary in child language, as eventually the child settles on the patterns found in the local language, although one needs to keep in mind that they may bring modifications to the local language that may ultimately lead to language change.

development (thriving versus dying), all evidence pointing to individual speakers as the ultimate agents of language evolution (Mufwene 2001: 148–149). Mufwene is among the first to adopt a biological approach to the evolution of languages as species, while acknowledging the limits of such a comparison: within the limits of the scholarship on biological evolution, he proposes that languages can be posited as full-fledged species which do share properties that are found in other biological species, but can also be viewed as distinct entities from other species, based on how they emerge and develop over time (ontogenesis) and on their architectural properties (how they are designed) (Mufwene 2008: 11).[8] Given Mufwene's evolutionary perspective on language competition and selection, I now turn to Darwin's observations on natural selection and transpose some of his insights to how variation, feature competition and selection operate within language and language development.

4.   Darwin's perspective on variation in natural selection

Mufwene's notion of the idiolect as the incipient and most important locus of language variation is in keeping with Darwin's observation that in natural selection, variation leading to change always starts very small.[9] This point is clearly captured in the following quotes drawn from *On the Origin of Species* (Darwin 1859) and reported in Duzdevich's latest (2014) rendition of Darwin's transformative work:

> Natural selection can act only by the preservation and accumulation of *infinitesimally small* inherited modifications, each profitable to the preserved being.   (Darwin 1859 in Duzdevich 2014: 59, my emphasis)[10]

For Darwin, any small degree of change (modification) in the natural world that is added cumulatively onto previous changes of its kind contributes in important ways to natural selection. He locates the source of such small changes in individuals, and argues that such small modifications can be selected and survive when

---

8.   As pointed out by a reviewer, it is worth noting here the relevance of August Schleicher's (1863) study, which adapted Darwin's "tree of life" to the Neo-Grammarian treatment of *Stammbaumteorie*.

9.   As noted by a reviewer, the term "small" remains vague in this context. As one cannot really measure it, I would simply characterize it as surface-level change that does not entail any deep parametric change.

10.   As I show in Section 5 below, this important observation pointing to the very gradual and imperceptible language change that occurs through minute and incremental variation is supported by the examination of a Swadesh list of Cabo Verdean.

they benefit their environment; one condition for their survival, according to Darwin, is for them to carve a specific role for themselves. Hence, each modification (or variant), no matter how small, can be important when it contributes long-term to the survival of its host. The variants that do not survive or do not get selected are those that, due to their competing with other variants too close to them in nature, cannot find their place in the host species and are thereby eliminated. On this topic, the following quote from Darwin highlights the importance of individual differences (variants), the key role they can play in natural selection, and how every organism needs to compete for a place in nature in order to survive.

> The circumstances favorable to natural selection present an extremely intricate subject. A large amount of heritable and diverse variability is favorable but *mere individual differences suffice for natural selection to work*. An important point is that the existence of a large number of individuals increases the chance that a profitable variation will appear in a given period of time, compensating for low levels of variability in individuals … Every organism strives to *seize a place in nature*, and a species is quickly exterminated if it fails to become modified and improved as its competitors do.
>
> (Darwin 1859 in Duzdevich 2014: 63, my emphasis)

The fact that each genetic variant among a progeny needs to find its place and purpose in nature in order to be selected, and hence survive, and the pace at which the selection may occur are made explicit in Darwin's statement that "[n]atural selection always acts extremely slowly. Its action depends on *available niches* that may be better occupied by inhabitants undergoing modification" (Darwin 1859 in Duzdevich 2014: 66, my emphasis).

One should note that elsewhere in his work, Darwin also makes insightful observations about the properties of the competing forms, the ecological environment in which such forms compete and the conditions that must be met in order for them to survive: he remarks that forms that are too much alike struggle to find their place and purpose in a given environment, and hence are likely to disappear. This points to the benefits of diversification, as the more distinctive a form is, the more likely it is to carve a place for itself in a given ecology. These insights are captured in Darwin's statement that

> [f]orms in the closest competition with those undergoing modification and improvement naturally suffer the most … Closely related forms (varieties of the same species and species of the same genus or related genera) generally engage in the most severe competition because they have nearly the same structures, constitutions and habits.   (Darwin 1859 in Duzdevich 2014: 68)

The observations that Darwin made in relation to species in nature, the modifications they undergo, and the competition that such modifications are subject to in

the course of their development, could be to some extent transposed to language, adopting Mufwene's view of language as a species. Transposed to language, one could argue that language variants (modifications in Darwin's terms) that are too close to be teased apart are less likely to survive in the feature pool, in contrast to more distinctive, salient variants that have managed to gain indexicality in the grammar of the language. Indeed, such distinctiveness could be due to the variant indexing a distinct geographic location, age, social class, level of education, and so on for the speaker, in the tradition of variationist sociolinguistics. Following Darwin's line of reasoning, variants that are selected are the ones that succeed in finding "an available niche", "seizing a place" in the linguistic ecology in which they evolve. Hence, the closest forms could be viewed as the fiercest competitors with each other, because they are competing for the same niche. In contrast, the variants that are phonologically, morphologically, semantically and syntactically more distinct are the most likely to find their own niche and survive. As mentioned earlier, Darwin views the process of natural selection as very slow and imperceptible, which could account for the temporary coexistence of even very close forms, although in the long term they are bound to disappear and give way to more distinct forms. The more distinct the forms, the more likely they are to be selected and survive by finding the right niche in their ecological environment.

If one follows Mufwene's proposal that languages are comparable to species, then it stands to reason that varieties are distinct from languages, though the line between the two may be at times too fine to discern. Regarding the distinction between varieties and species in the natural world, Darwin has the following to say: "Even strongly marked varieties having somewhat the character of a species … differ less from one another than distinct species differ from one another. Varieties are species in the process of formation – incipient species, according to my theory" (Darwin 1859 in Duzdevich 2014: 68). Reinforcing this point, he states elsewhere that "varieties when rendered very distinct from one another take the rank of species" (Darwin 1859 in Duzdevich 2014: 70).

Transposed to the linguistic domain, languages can be seen as species that have evolved from varieties, and varieties can be viewed as developing from idiolects whose competing features have won and propagated.[11]

---

11.   As a reviewer points out, there are of course limitations to the comparison between biology and linguistics. The reviewer observes that there is a clear definition of a species in the sense that we know that we are dealing with a "species" when two of its members can interbreed; but in contrast to the notion of a species, there is no uniform definition of a "language" that all linguists can agree on. Another limitation lies in the difficulty of defining the role of vertical versus horizontal transmission in shaping language continua.

I now turn to examining a Swadesh list of the Cabo Verdean language, which I use to illustrate how variation, competition and selection operate in language.

5.   On the relevance of Swadesh lists: Insights into variation, competition and selection

First, before introducing the Swadesh list I will use in this chapter, I should clarify the methodology I adopt in my analysis of it. As discussed in the introduction, one of the objectives of this chapter is to support Mufwene's proposal that the locus of language change is in the idiolect, putting idiolectal features firmly at the centre of language variation and at the very source of language change. According to the working definition of idiolect that was presented in the introduction, an idiolect is an individual's distinctive and unique use of language that affects their vocabulary, grammar and pronunciation. Taking this one step further, we can argue that idiolectal features are conditioned by a variety of factors that can include, but are not limited to, gender, age, degree of education and geographic area. The Swadesh list presented in this section surveys variation that exists across dialects spoken on five islands of Cabo Verde. As I will demonstrate below, the Swadesh list displays variation in vocabulary and pronunciation. However, not all the competing lexemes have the same semantics; in the context where the items were collected, the students contributed items that they felt were equivalent, though not necessarily identical, in meaning. This is an important caveat to highlight, but I will argue that because the students volunteered lexemes they viewed as equivalent, this type of variation still bears on Mufwene's (2001: 149) proposal that "the variation that matters to evolution really begins at that inter-idiolectal level" and that "the locus of dialect or language contact is the mind of the individual speaker" (Mufwene, 2001: 150).

In this section, I analyze a Swadesh list of the Cabo Verdean language, spoken on the archipelago of Cabo Verde, off the coast of Senegal. The Cabo Verdean language is well known to linguists for the drastic variation it displays across the nine inhabited islands where it is spoken.[12] The observed variation may be due to

---

12.   A reviewer points out that all languages necessarily exhibit a lot of variation and they remind us that how much variation we detect depends largely on what the linguist views as being part of the same language.

Cavalli-Sforza and Wang (1986) wrote a very interesting paper in which they applied the stepping-stone model which tests the isolating effect of distance in order to study both genetic and linguistic similarity in a chain of Micronesian islands. The authors chose those islands as they are known for their remarkably high level of linguistic diversity due in part to

three main factors, among others. The different histories of settlement of the various islands may be a major factor: for instance, among the leeward islands, Santiago was the first to be settled in 1461 by the Portuguese and by slaves from the Guinean coast (Santos et al. 2007: 71), whereas among the windward islands, São Vicente was settled in 1874 and involved different populations. Some of these populations were internal to the archipelago and came from islands like Maio, whereas others were external and originated from countries like England. A second factor is that for centuries and until very recently (the past two decades or so), there was not much widespread travel and migration between the islands, a situation exacerbated by the large distances between some of them (between Santiago and São Vicente, for instance); this may have contributed to the preservation of idiosyncratic features in the distinct varieties of the language, reinforced by the isolation of the archipelago.[13] A third factor is that until 1998, there was no official standard orthographic convention for the language, perpetuating different representations of its distinct varieties. Until then, it was standard practice for speakers to represent the language based on the forms that they used as individuals, with meaning based on their own idiolects.[14] It was indeed only on December 31st, 1998 that the Cabo Verdean government passed a law decreeing that the Alfabeto Unificado para a Escrita do Cabo-verdiano was to become the official script for the writing of the Cabo Verdean language.[15]

---

the great distances separating the islands, as well as to the sparse populations on the islands. Their expectation was that there would be greater similarity between two languages spoken at a shorter geographic distance, with more differences obtaining with increasing distance between the islands. Their study is quite relevant to the present one, as they were preoccupied with measuring as accurately as possible the degree of similarity between close communities, using the method of Pearson's correlation coefficient *r*. Comparing replacement rates of individual words across three language groups (Indo-European, Philippine and Cushitic), they found that each word seemed to have a particular replacement rate, that the replacement rates in space correlated with those in time, and that the differential rate of change seemed to correlate with its frequency of use.

13. It would be reasonable to predict that the recent frequency and intensity of movements between islands, due to the availability of affordable inter-island flights and fast and frequent ferries, may contribute to reducing the current variation by promoting levelling of the current variants within and across the island varieties. It is reasonable to assume that more frequent exchanges may also promote intermarriages, contributing to the levelling.

14. It is worth noting that although Portuguese is the official and the most widespread written language, Cabo Verdean has a long literary tradition of prose and poetry written in the language.

15. One could predict that the long-term use of this unified script could result in a homogenization of the varieties.

The Swadesh list that is discussed in this section is representative of the currently observable acute dialectal variation across five varieties spoken in Cabo Verde. The three varieties spoken on the leeward islands are Santiago, Brava and Fogo, and the two spoken on the windward islands are São Vicente and Santo Antão.[16] Methodologically, the Swadesh list of forty-nine words in Table 8.1 was collected from Cabo Verdean students who were all native speakers of Cabo Verdean Creole and who were representative of the five islands mentioned above and listed in the table. These students belonged to the first cohort of the Master's degree in Creolistics offered at the University of Cabo Verde in 2010. They were enrolled in the course on field methods for which I was the instructor.

Table 8.1 A Swadesh list of Cabo Verdean across five varieties

	Santiago	Fogo	Brava	Santo Antão	São Vicente
1	tudu, moku, fepu sima sta "all"	tudu, moku, fepu	tudu, moku, fepu	Tud	Tud
2	ku, y "and"	y, ku	y, ku	y, ma	y, ma
3	limaria, lumaria, bitxu, animal "animal"	Limaria	limaria, bitxu	nimal, animal, bitx, elmara	bitx, animal
4	sinza, sinsa "ash"	Sinza	sinza	Sinz	Sinza
5	na, pa "in"	na, pa	na, pa	ne, pe	na, pa
6	kosta "back"	Kosta	kosta	koxta, tras	Koxta
7	kasabi, kadretu, mariadu, mau, runhu, malvadu, kabala, mufinu, bidjaku "bad"	mau, kasabi, mariadu, mofinu, fedi	mau, mariadu, run	mariod, mofnod	mau, mariod
8	kaska di po, laska, fiasa di po "trunk"	kaska di po	kaska di po	Rom	Rom

(*Continued*)

<hr>

16.   As a reviewer correctly observes, this Swadesh list clearly shows that the Santiago variety tends to display consistently more variants than the other varieties. It remains unclear whether this is an effect of the sample or could be attributed to there being more variants in Santiago, given its early period of settlement. As the reviewer suggests, Santiago being the first island to be settled, one could entertain the possibility that this difference could be due to a bottleneck effect during the expansion of the language to the other islands. Other reasons could involve more complex networks of speech groups, competition from recent loans, and possibly competition between recent and old forms.

Table 8.1   (*Continued*)

	Santiago	Fogo	Brava	Santo Antão	São Vicente
9	pamo, pamodi, pabia, parabia, parabias, purkauza, purkazu "because"	Pamodi	pamodi	Mod	Purke
10	bariga, ventri "belly"	Barriga	bariga	Berig	Beriga
11	gordu, grandi, tamanhu, gros "fat"	gordu, grandi, forti	gordu, grandi, forti	grand, tmonhon	Grande
12	pasu, avi, txota "bird"	Pasu	pasu	Pos	Pos
13	mordi "to bite"	Morde	morde	Morde	morde, mukna
14	pretu, negru, sukuru "black"	Pretu	pretu	Pret	Pret
15	sangi "blood"	Sangi	sangi	Seng	Sang
16	sopra "blow"	Sopra	sopra	Sopra	Sopra
17	osu "bone"	Osu	osu	Os	Os
18	kema, txema "burn"	Kema	kema	kema, psi	kema, txumska
19	mininu, kriansa "child"	Mininu	mininu, minizu	Mnin	Mnin
20	nubri, nuven "cloud"	Nuven	nuben	Neva	Nuven
21	friu "cold"	Friu	friu	Fri	Friu
22	ben, txiga "come/ arrive"	txiga, ben	txiga, ben	ben, txega	ben, txega
23	konta "to tell/narrate"	Konta	konta	Konta	Konta
24	sapa, korta, pika "to cut"	korta, sapa, pika	korta, sapa	tsapa, korta, pika	korta, sapa
25	dia, dia "day"	Dia	dia	Dia	Dia
26	more "to die"	More	more	Mre	More
27	koba "to dig"	Koba	koba	Kova	Kava
28	xuxu, suxu, suju, mariadu "dirty"	Suju	suju	xuj, kerdid, inkorod	suj, xuj, inkardid
29	katxor "dog"	Katxor	katxor	Katxor	Kotxor
30	bebi "to drink"	Bibe	bebe	Bebe	Bibe
31	seku "dry"	Seku	seku	sek, storod	Sek
32	segu "blind"	Segu	segu	seg, mok	seg, mok
33	poera, po, tera "dust"	puera, tera	puera, tera	tera, po, puera	tera, po, puera
34	oredja "ear"	Oredja	oredja	Orea	Orea

Table 8.1 (*Continued*)

	Santiago	Fogo	Brava	Santo Antão	São Vicente
35	txon, tera "ground"	tera, txon	tera, txon	txon, tera	txon, tera
36	kumi, kume, nhemi "to eat"	Kume	kume	Kme	Kme
37	obu, ovu "egg"	Obu	obu	Ov	Ov
38	odju "eye"	Odju	odju	Oi	Oi
39	kai, kei "to fall"	Kai	kai	Kei	Kei
40	lonji, distanti, afastadu "distant"	Lonji	lonji	foxtod, lonj	Lonj
41	gordura, gurdura, seti, untu, banha "fat/grease"	gurdura, banha	gurdura, banha	gurdura, banha	gurdura, banha
42	pai, papa, nha grandi "father"	papa, pai, nha pa	papa, pai	Pe	Pai
43	xinti medu, tene medu, meda "to fear"	ten medu	ten medu	treme di med, te k'med, treme d'med	ten med
44	pena "feather"	Pena	pena	Pena	Pena
45	poku, pitada, faiska, kusinha "little"	poku, kuzinha, bokadinhu	poku, kuzinha, bokadinhu	pok, kzukuk, kzinha, bkedin	pok, kzinha, bkedin
46	briga, luta, gera "to fight"	Briga	briga	briga, gera	Briga
47	lumi, fogu "fire"	Lumi	lumi	Lum	Lum
48	pexi "fish"	Pexi	pexi	Pex	Pex
49	sinku "five"	Sinku	sinku	Sink	Sink

The purpose of the brief analysis of this list presented in this section is not only to demonstrate the dramatic variation that occurs within the same language, but also to illustrate the competition and selection that even the lexicon of a single language can display, as in the case of Cabo Verdean. As we see below, some of the words only have one variant across the five islands of Cabo Verde, whereas others have multiple variants. The variation manifests itself at both the phonological and morphological levels, and for some of the lexemes, some of the variants may be quite old while others are quite new, attesting to long-term competition between some of them. This shows that variants can compete for a long time without necessarily resulting in the prompt elimination of those variants that could be viewed as redundant and superfluous. As I mention below, it could be possible that the observable competition may not have resulted yet in the disappearance of variants, because they fulfil a specific function in the language by indexing social meaning; this would support Darwin's notion that survival depends on finding "a niche".

I must add important caveats to the statements above.[17] I would like to make it clear that in the absence of diachronic comparative data, if one assumes a long-term competition between some of these items, it is not always clear which ones are new and which ones are older. Also, it is not at all the case that different forms in each variety or across varieties exhibit exactly the same semantics. Recall that this Swadesh list was collected in a classroom from students who were representative of these varieties; hence some of the terms may not be variants proper but may have been added to that category due to some mental associations that students had with the elicited items. Indeed, it is important to keep in mind that given the method of elicitation of this particular Swadesh list, cognacy between variants used in different idiolects does not imply that these variants share exactly the same meaning. In other cases, some of the variants may be filling a niche, as the idea of a niche does not entail elimination, but instead adaptation and, at times, elaboration. Finally, since competition and selection require a historical dimension, it is worth investigating in future work whether shared lexemes (that are phonetically identical or similar) between islands are symptomatic of a generalized older (proto-)layer whose variants started competing, locally, with new items in more recent times.

I group the lexemes in Table 8.1 in four categories based on the type of variation they display:

a. Some words only show one form with the same meaning across all varieties.
b. Other words display phonological change involving one vowel or one consonant. The change can be apocope or epenthesis but every time affecting only one phoneme.
c. Other words may involve morpho-phonological change in that a morpheme can undergo a change via deletion or alteration.[18]
d. In a fourth category, some variants have given way to new words entirely.

By way of illustration, in group (a) one can find rare cases of lexemes that do not change within and across varieties, as with *sopra* "to blow" (16), *dia* "day" (25) or *pena* "feather" (44). In contrast, in the (b) group, some words undergo a vowel change from /i/ to /u/, as in *limaria*/*lumaria* "animal" (3), /i/ to /e/ as in *mordi*/

---

17. I am indebted to a reviewer for the following shrewd observations which I intend to pursue in future work.

18. As a reviewer suggests, a promising future avenue of investigation would be to consider whether competition, selection and adaptation proceed in similar ways for phonemes, morphemes and lexemes, the prediction being that the lexicon and semantics may take a different developmental path from the syntax, for instance.

morde "to bite" (13), /a/ to /e/, as in *bariga/beriga* "belly" (10); others undergo apocope as in *beriga/berig* "belly" (10), *pretu/pret* "black" (14), *sangi/sang* "blood" (15) or *osu/os* "bone" (17). In the domain of consonants, a frequent alternation involves /v/ and /b/ as in *nuven/nuben* "cloud" (20) or *kova/koba* "to dig" (27). In the (c) group, one finds words that undergo morphological change via alteration as in *mariadu/mariod* "bad" (7) or deletion as in *pamodi/pamo* "because" (9).[19] In the (d) group, the same concept can be rendered by numerous lexemes (though not all of them may be semantically equivalent). For instance, in the Swadesh list under examination, the equivalent of English "because" is rendered by the semantically equivalent *pamo, pamodi, pabia, parabia, parabias, purkauza, purkazu* within the same Santiago variety and by *mod, purke and pamodi* (9) across all varieties. In contrast, the equivalent to the English word "bad" is rendered by lexemes that are not all semantically equivalent, although students in the class lumped them in the same general semantic category. Indeed, the terms *kasabi, kadretu, mariadu, mau, runhu, malvadu, kabala, mufinu* and *bidjaku* (7) were elicited from the class as meaning "bad", but to mention just two examples, *kasabi* is closer to meaning "unpleasant" and *runhu* to "ill-mannered". This is the case of an example where the lexemes are not semantically identical but are viewed by the students as equivalent.

Recall that although the original motivation to collect this Swadesh list was to illustrate dialectal variation between islands (based on the presence in the class of speakers of five varieties), the Swadesh list clearly revealed variation within each island and within the groups of speakers representative of each island. This highlights Mufwene's notion of inter-idiolects as the loci of language variation and change. Beyond the natural geographic boundaries of each variety, there can be a number of reasons for the observed variation including age of speakers, gender, level of education and sense of identity, in the sense that the speech community that a speaker identifies with may lead them to emulate or accommodate to a particular type of pronunciation. This is how each variant can find its "niche". In addition, there can be old and new variants of the type noted in Aboh & DeGraff (2017: 23), where they note that the -è/ε/ alternation in Haitian Creole words like *vòlè/vòlò* "thief" or *mantè/mantò* "liar" results from earlier pronunciation of French *voleur* and *menteur* and is attested to this day in Franco-Provençal. Again, if we refer back to the earlier definition of idiolect which includes individual use of vocabulary, grammar and pronunciation, and if we agree that distinct pronunciation could be inherited from geographic origin, we can still maintain that the

---

19. Note that the example *mariadu/mariod* "bad" fits category (c), because the masculine gender agreement marker *-u* present in *mariadu* is overtly absent in *mariod*, but it could also fit category (b), as it marks a phonological change.

study of dialectal variation bears on Mufwene's proposal about the idiolect as the main locus of variation.

Why is there so much variation featuring competing variants over a long period of time on islands like those of Cabo Verde and Haiti? Darwin would attribute the cause squarely to the topography of the place. In his view, small isolated areas are propitious for the rise of new species, but over time modification is *slower*. Darwin's position on the matter is clearly expressed in the following statement:

> A small oceanic island exemplifies a "small isolated area." The total number of species on oceanic islands is small … but most of these species are endemic, meaning they are produced there and nowhere else. An oceanic island may therefore seem highly favorable to generating species, but to see if this is actually the case, a comparison with equal times is necessary to ascertain if it is more favorable than a large open area like a continent … Indeed, I believe that, on the whole, a large open area is more important than isolation for the generation of new species, especially those that will endure and spread widely … I conclude that although small isolated areas have probably been favorable to the generation of new species, *modification is more rapid in large areas, and its new forms – already victorious over many competitors –* play an important part in the changing history of the organic world by spreading widely and giving rise to new varieties and species.
>
> (Darwin 1859 in Duzdevich 2014: 65, my emphasis)

Transposed to language, this could mean that in spaces like islands, new variants may emerge and coexist with older ones for a very long time before being eliminated. Note that although I am arguing here that spaces like islands promote the emergence and coexistence of older variants with new ones for a long period of time before they disappear, I do not deny that other places on the planet may act like islands, for instance tropical forests or mountainous areas. Such geographic areas create isolation of communities of speakers and, as a result, are well known to dramatically limit opportunities for horizontal transmission (via contact) and diffusion, and can as a consequence inhibit the rise of competitors. In the particular case that concerns us, the Cabo Verdean language represents a complex situation in which a newborn language spread throughout the archipelago over the centuries, cementing a "genetic" bond between the varieties that developed across the islands; these varieties undoubtedly contain traces of the different populations that contributed to their formation, such populations coming both from other islands (genetic) in the archipelago and from external sources (contact).[20]

---

20. I thank a reviewer for this insightful observation.

6.   Conclusion

Darwin (1859) in Duzdevich (2014: 51) states that

> [n]atural selection constantly scrutinizes every variation, rejecting the bad while preserving and accumulating the good; whenever and wherever opportunity affords, nature works silently and unnoticed to improve every organism with respect to its environment. We see nothing of these slow changes in progress until the hand of time has marked the long lapse of ages.

When transposed to variant selection in language, one sees parallelisms between natural selection and variant selection as the variants fit for the communicative system are selected, and the smallest degree of variation, at the phonological level for instance, could be easily dismissed as insignificant. However, based on the discussion in this chapter, the smallest variation (whether phonological, morphological or lexical proper) actually has its role to play in the march towards language change. The selected variants are significant in ensuring the continuing efficiency of language that remains remarkably fit for communication in spite of being subjected to constant change and flux.

Acknowledgement

References

Aboh, E. O. & DeGraff, M. 2017. A null theory of Creole formation based on Universal Grammar. In *The Oxford Handbook of Universal Grammar*, I. Roberts (ed.), 1–67. Oxford: OUP.

Bates, E. & MacWhinney, B. 1987. A functionalist approach to the acquisition of grammar. In *Functionalism in Linguistics* [Linguistic and Literary Studies in Eastern Europe 20], R. Dirven & V. Fried (eds), 209–264. Amsterdam: John Benjamins. https://doi.org/10.1075/llsee.20.12bat

Cavalli-Sforza, L. L. & Yang, W. 1986. Spatial distance and lexical replacement. *Language* 62: 38–55. https://doi.org/10.1353/lan.1986.0115

Chomsky, N. 1995. Language and nature. *Mind* 104: 1–61. https://doi.org/10.1093/mind/104.413.1

Clark, R. & Roberts, I. 1993. Computational model of language learnability and language change. *Linguistic Inquiry* 24: 299–345.

Costa, A., Colomé, A., Gomez, O. & Sebastián Gallés, N. 2003. Another look at cross-language competition in bilingual speech production: Lexical and phonological factors. *Bilingualism: Language and Cognition* 6: 167–179. https://doi.org/10.1017/S1366728903001111

Crain S., Goro, T. & Thornton, R. 2006. Language acquisition is language change. *Journal of Psycholinguistic Research* 35(1): 31–49. https://doi.org/10.1007/s10936-005-9002-7

Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

DeGraff, M. 1999. *Language Creation and Language Change: Creolization, Diachrony, and Development*. Cambridge MA: The MIT Press.

Duzdevich, D. 2014. *Darwin's On the Origin of Species: A Modern Rendition*. Bloomington IN: Indiana University Press.

Kroll, J. M. & de Groot, A. M. B. (eds). 2005. *Handbook of Bilingualism: Psycholinguistic Approaches*. Oxford: OUP.

Labov, W. 2007. Transmission and diffusion. *Language* 83: 344–387.
  https://doi.org/10.1353/lan.2007.0082

MacWhinney, B. & Bates, E.. 1989. *The Crosslinguistic Study of Sentence Processing*. Cambridge: CUP.

Mufwene, S. S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP.
  https://doi.org/10.1017/CBO9780511612862

Mufwene, S. S. 2006. Multilingualism in linguistic history: Creolization and indigenization. In *The Handbook of Bilingualism*, T. K. Bhatia & W. C. Ritchie (eds), 460–488. Oxford: Blackwell. https://doi.org/10.1002/9780470756997.ch18

Mufwene, S. S. 2008. *Language Evolution: Contact, Competition, and Change*. London: Continuum.

Santos, M. E. M., Torrão, M. M. F. & Soares, M. J. 2007. *História Concisa de Cabo Verde: Resumo da História Geral de Cabo Verde*. Lisbon & Praia: IIPC.

Schleicher, A. 1863. *Die Darwinsche Theorie und die Sprachwissenschaft: Offenes Sendschreiben an Herrn Dr. Ernst Haeckel*. Weimar: H. Boehlau.

Yang, C. 2002. *Knowledge and Learning in Natural Language*. Oxford: OUP.

CHAPTER 9

Building grammar in the early stages of development of French Creoles

Insights from Second Language Acquisition

Georges Daniel Véronique
Université Aix Marseille, CNRS, LPL

This chapter contributes to the research on naturalistic adult Second Language Acquisition (SLA) and the understanding of the early stages of Creole grammar formation. The two sets of data discussed here relate to the acquisition of French as a second language (FSL), and the acquisition of French related Creoles (FRC). Studying developmental sequences in these two acquisition contexts helps understand the internal reconstruction of grammar development in FRC. Despite various methodological challenges (e.g., scarcity and the linguistic nature of the FRC data, how comparable FSL and FRC are, given their respective ecologies), this comparative approach suggests that the linguistic factors that produce pre-basic and basic varieties in SLA are equally at work in the development of creole languages.

Keywords: French related Creoles, French as a second language, basic variety, developmental sequence, internal reconstruction

1. Introduction

Language acquisition has been seen as a factor in the development of pidgins and creoles since the late nineteenth century and during the first half of the twentieth century (Bloomfield 1933 and Hall 1966). In the early 1980s, Schumann (1978) and Andersen (1983) drew a parallel between stages in the dynamics of creole in Guyana (Bickerton 1975) and the acquisition of English by adult Spanish L1 naturalistic learners. Independently, Chaudenson (1974) claimed that French-related Creoles (FRC) started as adult L2 learner varieties. During the past forty years, several volumes have addressed the relations between creolization, acquisition

and language change (Aboh 2015; DeGraff 1999; Kouwenberg & Patrick 2003; Lefebvre et al. 2006, *inter alia*).

Taking stock of previous work, including Mufwene's feature pool theory (Mufwene 2001, 2005) and Aboh's hybrid grammar (Aboh 2015), this chapter attempts to assess the contribution of research on naturalistic adult second language acquisition (SLA) to the understanding of the development of grammar in the early stages of creole languages, through a comparison of data on the acquisition of French as a second language (FSL) and on FRC. Section 2 of the chapter explores some of the issues involved in approaching Creole language development as a case of SLA. Section 3 discusses some of the challenges of recourse to early FRC texts in order to reconstruct the development of grammar in FRC. It also provides a brief sketch of the external ecology of FRC in the Atlantic and the Indian Oceans. Section 4 explains the rationale of the chapter. It presents the functionalist approach to SLA which provides the theoretical backbone of the chapter and the Basic Variety (BV) model developed by Klein and Perdue (1997). Section 5 introduces the L1 Arabic informants of FSL. Section 6 analyzes some aspects of the development of grammar in the naturalistic acquisition of FSL by adult learners. Section 7 explores similar aspects of the building of grammar in FRC. Finally, Section 8 offers a discussion of the findings and assesses the value of the BV model as an explanation of the development of grammar in FRC.

2.   SLA and Creole language development

The existence of a possible relation between SLA and Creole language development has been fiercely disputed during the past forty years. In this section, some of the arguments *pro* and *contra* the relevance of SLA research for explaining the development of Creole languages are discussed. An attempt is made to analyze the underlying epistemological issues of the debate.

2.1   Creole language development as SLA: Previous work

At least three trends can be identified in the contributions that have sought to relate SLA to the development of Creole languages. In the first, the pidgin-creole life cycle (Hall 1966) and the creole continuum hypothesis (Bickerton 1975) have been applied to the dynamics of interlanguage (Andersen 1983). In the second, Chaudenson (1974, 1978) and Valdman (1983) have argued that features of the early development of adult learner languages in FSL provide an illustration of the first stages of FRC languages, not so well documented. Valdman (1983) faced a firm rebuttal from Bickerton (1983), and Chaudenson (2001, 2003) has been

thoroughly discussed *inter alia* by Aboh (2015: 62–73). In the third trend, it has been argued that development in SLA parallels language change and creole language development (Baron 1977).

In recent years, Klein and Perdue's BV model (Klein & Perdue 1997), Pienemann's processability theory (Pienemann 1998), and Schwartz and Sprouse's full access/ full transfer hypothesis (Schwartz & Sprouse 1996) have rekindled interest in SLA, in the domain of Creole studies. Despite Sprouse's complaint (Sprouse 2010: 261) about the "invisibility of SLA theory in mainstream creole linguistics", the controversy between Plag (2008a, 2008b, 2009a, 2009b, 2009c), who refers to Pienemann's processability theory, and Sprouse (2009) concerning the possibility of considering Creoles and pidgins as early varieties of interlanguage, and the validity of relying on Pienemann's theory or on the full access/ full transfer theory, has updated the discussion.

Bickerton (1983), Mufwene (2001) and Aboh (2015), who reject the relevance of SLA findings for studies of Creole development for different reasons, share at least the following arguments, which need to be stated to circumscribe the debate. According to these authors, no parallel can be drawn between SLA and Creole language development because SLA describes individual idiosyncratic trajectories in L2 whereas the development of Creole languages is a social process. L2 learners aim at a specific target language (TL), whereas the Creole language makers had no such target in view. Although these arguments are quite cogent, it must be emphasized that sequences of development in a given L2, bar individual variation, are shared by a majority of learners of that language (Wode 1981).

2.1.1   *Mufwene and SLA research*

Mufwene has regularly discussed the impact of SLA research on Creole linguistics. At various points, following Chaudenson (1992), he has vindicated the argument that some form of "approximation of approximation" is at work in creole development (Mufwene 1997). Mufwene (1996) claims that the communicative strategies of creole language makers in language contact situations provide the founding mechanisms of Creole development. According to him, when communicative strategies compete, the least marked and the least costly, in both structural and cognitive terms, win. Thus markedness, resulting from such factors as regularity, invariability, frequency, extension, semantic transparency and perceptual salience, together with basilectalization (Baker 1984), account for the restructuration, the innovations and the grammaticizing that produced Creole vernaculars. Mufwene (1996) posits that Creole languages developed by chance through restructuring and adjustments during the communicative process proper, in specific socio-historical conditions.

According to Mufwene (2008), the emergence of interlanguage in adult learners of L2 is closer to some form of pidginization than to Creole development, given

that the bilingual learners continue to communicate and interact in L1. Mufwene notes that none of the varieties developed by L2 learners have been adopted as language vernaculars, replacing the L1 of their speakers, and becoming in turn the L1 of their offspring. In light of these arguments, L2 learner varieties seem to be at best a means of inter-ethnic communication with other speakers. Mufwene insists on the fact that the emergent L2 varieties exhibit no creation of specific norms distinct from the native speaker usage, as opposed to the development of Creole languages.

Although Mufwene (2010) believes that cross-pollination between the fields of SLA research and Creole studies could be fruitful, he holds that naturalistic SLA cannot contribute directly to the explanation of Creole development because of the intricacies of its social and linguistic matrices.

2.1.2   *Aboh's rejection of Creoles as interlanguages*

For Aboh (2015), the development of Creole languages provides an interesting case of language change and language acquisition. Aboh rejects various hypotheses on the emergence of Creoles that posit that these languages arose out of incomplete SLA by adult learners (Aboh 2015: 60–112). This includes both Chaudenson's approach (2001: 58) to Creole development as resulting from approximation of the TL and Plag's hypothesis of Creoles as conventionalized interlanguages (Plag 2008a: 116). *Contra* Plag's hypothesis, Aboh observes both conceptual flaws – such as the non-plausibility of the implicational scale of development postulated by Pienemann's processability theory (Pienemann 1998) used by Plag – and methodological shortcomings. Aboh (2015: 99) asserts that "interlanguages are not full-fledged systems" that may constitute "entities of linguistic comparison". As "transitional stages learners go through", interlanguages cannot be used for linguistic comparison. Besides, the "structural characterization of interlanguages" necessarily implies a specific TL. The fact that a particular interlanguage feature develops in specific learner varieties because of language universals, does not explain its instantiation in a particular Creole language as well, through identified language contact. To demonstrate the empirical shortcomings of Plag's approach to creole development, Aboh resorts to Klein and Perdue's BV model (Klein & Perdue 1997), which evolved out of SLA research (see Perdue 1993). This model will provide the theoretical framework of this chapter.

2.2   What is at stake in the current debate on Creole development and SLA?

The discussion conducted *inter alia* by Mufwene and Aboh on the second language development hypothesis as offering an explanation for the development of Creole languages seems to have settled matters: no parallel may be drawn between

interlanguage development and Creole development. The present chapter is an attempt to re-examine the issue by first acknowledging the epistemological and methodological questions that need to be answered for such a comparison to be fruitful, and then by plotting what a comparison between SLA and Creole data could tell us about the linguistic processes involved in Creole development. This endeavour is close in intent to Mather's discussion of SLA and creolization (Mather 2006).

What is at stake in the ongoing discussion about Creole language development and SLA is whether some form of interlanguage or learner variety represents in any way the initial state of Creole languages, and whether the specific interlanguage development observed for the acquisition of features of a given TL can provide information about the way a Creole language related to that same language developed. Providing answers to these questions is incumbent on the theoretical stance adopted about language change, language evolution and language acquisition. In effect, ideas vary as to the status and importance of initial states in language development, and about the factors and processes that contribute to the inception of Creole languages. Differences in theoretical standpoints partly explain the objections raised by opponents of the view that comparing second language and Creole development processes is a worthwhile enterprise. These differences foster the existing *dialogue de sourds* that prevents possible cross-fertilization.

Chaudenson's approach to Creole language development (Chaudenson 2001) is implicitly based on Saussurean *langue*, albeit under the guise of "*français zéro*" (a theoretically "neutral" state of French) modified by dialectal variation. His theory of "approximation of approximation" attempts to explain why and how FRC languages broke loose from their lexifier language, under the pressure of various factors including transfer from L1, to form new linguistic systems. In Chaudenson's perspective, internal linguistic factors and external social factors affect all levels of language contact on a par; he places no specific emphasis on grammar. Chaudenson makes no difference between the development of individual internal grammar and the Saussurean *langue* (language)-*parole* (speech) distinction, as instantiated in the emergence of Creole languages. Neither does he oppose external and internal linguistic ecology.

Pienemann's processability theory (Pienemann 1998) is a cognitive theory of SLA. It is based on the idea that each linguistic stage in the incremental development of interlanguages – from the lemma to higher-order syntactic combinations – is supported by specific cognitive parsing procedures. The lexical-functional linguistic approach backing the claims of processability theory is geared towards explaining the staged development of *i-grammar* in L2: the initial state of interlanguage is lexical in nature; later, clauses of varying syntactic complexity develop in an orderly manner. Processability theory is not focused on the emergence of *e-grammar*.

Hailing from Minimalism, Aboh's theory of hybrid grammar is also concerned with language change at an individual level. What matters most in his approach is to explain how semantic and syntactic features of languages in contact recombine to form hybrid items and hence, hybrid grammars. Mismatches and variation may occur at either the semantic or syntactic level to coin new items and grammatical rules. To explain the emergence of *e-grammar*, Aboh relates hybrid *i-grammar* to the external ecology of language contact, in the wake of Mufwene (2001).

In turn, as Plag (2011) notes in his criticism of the role of linguistic feature pools and typological predictions in creolization, processing constraints, including parsing of input, as evidenced in SLA, must not be neglected. Language processing and transfer in contact situations are to be envisioned as dynamic processes which do not affect all levels and subsystems of the languages concerned at the same time. The linguistic outcome of such processes needs to be stabilized through social interactions and acts of identity in order to emerge as full-fledged languages (Le Page & Tabouret-Keller 1985).

The theoretical underpinnings of the conflicting approaches sketched in this section partly explain the scope and nature of the debate at hand. Arguments are sometimes used at cross-purposes. Thus, it may seem incorrect to argue against an approach where language contacts are viewed as Saussurean linguistic systems in contact from the vantage point of a study devoted to the development of *i-grammar*. Although the perspective defended in this chapter is based on a functionalist usage-based approach to SLA and Creole development, it aims to show, in a similar vein to Mather (2006), that the processes governing *i-grammar* in Creole and second language development are identical even if their *e-grammars* may diverge.

3.   Analyzing early creolization in FRC

As has been pointed out by Arends (1995a: ix), two interconnected directions may be taken to explore early creolization: the study of early Creole documents and the "investigation of the sociohistorical and demographic matrix of creole genesis". As the volume edited by Arends (Arends 1995b), which includes papers examining both directions, shows, the development of Creole languages was a gradual process sparked by the inception of agro-business in the colonies, especially the sugar industry. As claimed by Arends 1995a (but see also Baker 1984 and Chaudenson 1992), the development of Creole languages is a gradual process, which results mainly from the communicative activities and the incidental learning of the dominant language in the colonies by adult L2 speakers.

According to Alleyne (2000), the inception and development of English-related Creoles (ERC) and of FRC have followed different paths. Following this

author, ERC, with Saramaccan as their paragon, started with maximal divergence from their lexifier, and later in the course of creolization tended to converge. In the case of FRC, the picture is different: the Atlantic and Indian Ocean FRC were closer to each other and to their lexifier at their inception, but diverged strongly during their development. Alleyne's cautionary note about the historical development of FRC should be borne in mind.

3.1   The external ecology of FRC

The FRC examined in this chapter, produced during the French colonial expansion that started around 1625 in the West Indies, have all undergone the homestead and plantation periods typical of the French colonies in the Atlantic and Indian Oceans. They grew up within similar social organizations, the Antillean colonies providing the initial template (Ludwig 2018). Although the colonial settings are close – compare Martinique and Guadeloupe on the one hand, and Bourbon (Réunion) and Île de France (Mauritius) on the other – the available language pools in the two areas were different. The languages spoken by the founder populations in the Atlantic and Indian Ocean colonies have been at least partially identified (Adam 1883 used the term "*negro-aryen*" (negro-aryan) to characterize the Atlantic French creoles and "*malayo-aryen*" (malayo-aryan) for the Indian Ocean creoles; see also Aboh 2015). Besides Malagasy, Bantu languages were present in the Indian Ocean islands during the period of slavery (1665 – ca. 1848). For ten years (1727–1735), after the settlement of Île de France in 1721, slaves speaking Niger-Congo languages constituted the majority of the workforce there (Baker 2007). As for the slave workforce in the French West Indies, it came mainly from West Africa, speaking languages from such different language families as Ewe-Fon, Mande and Fula (Singler 1993, 1995)

The development of Atlantic and Indian Ocean FRC offers a test bed to understand both the ecology of language change and the relation between external and internal ecology, as defined by Mufwene (2001). Similarities between the two sets of FRC may be accounted for by the fact that the founder slave populations lived in similar societies, and were eventually exposed to the same or to close French colonial linguistic varieties; and dissimilarities may be explained by the language pools involved and the linguistic features available. The contention of this chapter is that early FRC grammars that developed in the various colonies were produced by the same processing strategies responsible for SLA, despite differences in the language contact situations.

3.2   Early texts in FRC

Some written linguistic data are available to document the early stages of FRC development. The first written texts were produced at some period after the

language had started to break away from French. The mean time interval between the outstart of plantation societies in the colonies and the time when written Creole texts were produced is ca. thirty years (see Baker & Fon Sing 2007; Bollée 2007; Chaudenson 1981; Hazaël Massieux 2008; Neumann-Holzschuh 1987). Early texts were mainly produced by bilingual white speakers – the speech of black slaves being at best reported – who used various *scriptae* derived from French orthography to transcribe language varieties that they tended to overlook.[1] As signalled by Adone (1994: 29) and Hazaël-Massieux (2008: 16 et seq.), early texts available must be handled with care. Besides the modified French orthography used in the texts, their genre (religious writings, catechisms, extracts from the minutes of trials, etc.) probably influenced their content and style, if not their authenticity. However, despite these limitations, my contention is that the texts available provide information about the development of early grammar in FRC.

It is difficult to assign the specific texts that will be quoted in this chapter to particular stages of development of FRC. For one thing, FRC, contrary to ERC, cannot be ascribed to a single source, Saramaccan or some West African pidgin (Winford 1993); the polygenetic emergence of FRC seems to be a well-established fact (Baker 1984). Thus, no linguistic continuum can be drawn for FRC *in toto*. In fact, the situation of FRC is the reverse of, say, ERC; it is the study of available texts that may eventually yield sequences of language change. An external and rather crude criterion for equating available data and stages of development of the various FRC involved is to posit a direct relation between the date of production of a given text and the stage it represents within a specific language change continuum, i.e. early Creole texts exhibit early stages of development of the Creole language examined. An internal criterion may also apply, based on the principled and logically sequenced unfolding of layers of grammaticalization as described by Heine and Kuteva (2007), or through the development of linear grammar as defined by Jackendoff (1999). External evidence and internal reconstruction can probably help to identify what stage of development a given FRC text pertains to. However, it is to be expected that a given text may exhibit features belonging to various stages of development of the Creole language being examined.

3.3 A historical account of the development of FRC grammar

McWhorter (2018), in a dispute with DeGraff (2001), accuses the latter of ignoring the principles of historical linguistics. As a matter of fact, as emphasized by Bollée

---

1. Texts are quoted in the orthography used in the manuscripts throughout this chapter.

and Neumann-Holzschuh (1993), historical grammars of French Creoles still need to be written. Alleyne's *Syntaxe historique* (1996) provides a good illustration of FRC, but retracing the grammatical evolution of FRC, given their polygenesis, is still a task to be undertaken.

4.   Rationale

This chapter compares and contrasts some aspects of the emergence of predication in adult naturalistic acquisition of French as L2, and in the early phases of the development of grammar in various FRC. It is hypothesized that the early stages of the emergence of the grammar in L2 will inform *ceteris paribus* the development of early grammar in Creole languages. Because of the nature of FRC early texts, and because of its polygenesis, any comparison between FRC texts and FSL needs to be indirect. The comparative approach adopted in this chapter is based on a functionalist perspective on language and SLA. It claims that the BV (Klein & Perdue 1997) identified as a stage in the development of grammar in SLA – akin to the linear phase of grammar development postulated by Jackendoff (1999), or to the first two layers (out of five) of the genesis of grammar of Heine & Kuteva (2007) – provides an adequate basis for exploring the early grammatical development of FRC languages.

4.1   A functionalist account of the development of grammar in SLA

A functionalist perspective on language posits that syntactic and semantic constraints are generated through discourse; no distinction is made between *i-* and *e-grammars*. A functionalist account of SLA envisions the development of L2 knowledge as resulting from the interplay and competition of language contacts, general learning and communicative skills, and principles of information organization.

4.1.1   *Stages in L2 acquisition*
A functionalist account of SLA posits that the development of proficiency in L2 is a gradual process shared by all learners, bar some measure of individual variation. According to Klein and Perdue (1992, 1997), three major stages can be identified in the development of L2 learner varieties. At stage 1, called Nominal Utterance Organization (NUO), utterances are mainly lexically based and ordered following pragmatic (focus last) and semantic constraints (controller first). Interlanguages exhibiting such properties are described as pre-basic learner varieties (Perdue 1996). At stage 2, the Infinite Utterance Organization (IUO), utterances tend to be organized around the

non-inflected verb; basic learner varieties emerge at that stage.[2] At stage 3, the Finite Utterance Organization (FUO), verbs acquire inflection and finiteness. At NUO and IUO stages, utterances are organized (i) by pragmatic and semantic constraints and (ii) through the semantic domains of reference to entities, to time, to space, to agentivity, etc. involved in the construal of various types of texts. Perdue (1996) mentions that the pre-basic and basic learner varieties are shaped *inter alia* by:

– dependence on immediate context;
– maximising use of linguistic, mainly lexical, means available;
– search for transparency and simple form-to-function relation.

4.1.2 *The pre -Basic Variety (pre-BV) in SLA: Properties and constraints*

Perdue (1996: 143) shows that at the NUO stage, pre-basic learner utterances are highly structured and that "this structuring is the very basis for further development". He shows that in pre-BV utterances, alignment is determined by Topic – Focus order. In referential expressions, the theme precedes the Relatum and the scope of particles (such as negators, connectors and adverbs) goes to the right.

4.1.3 *The Basic Variety (BV) in SLA: Properties and constraints*

Based on the same semantic and pragmatic principles as pre-BV, BV is largely independent of the target language or the first language, reflecting the "'construction faculty' more than … [the] 'copying faculty' (Klein 2009: 335)" of the language acquirers (Dimroth 2018: 3). In the terms proposed by Jackendoff and Wittenberg (2014: 69), the BV is a "part-of-speech simple phrase grammar". An adaptation to the properties of the target language, as well as more substantial L1 influence, is assumed to happen only when this simple and self-contained system evolves because of communicative shortcomings and starts to develop finiteness (Klein & Perdue 1997). According to the BV model, L1 transfer is rare at early stages of L2 acquisition. According to Klein and Perdue (1997), L1 transfer as a contributive factor to the development of L2 grammar is observed mainly at the IUO stage. Even then, L1 influence varies according to the linguistic domains involved (NP, VP, PrepP etc.).

4.2 Grammatical categories and phrasal constraints in early SLA

According to Klein and Perdue (1993: 260–261):

---

2.   In pre-basic and basic varieties of FSL, the verb is not inflected for tense and bears no agreement. In FSL, the most frequent form of the non-inflected verb is the verb stem (*VØ*) or the verb lexeme or verb stem + E / I, e.g., [parte] / [parti] (*go*) (*Ve*).

> ALA (Adult Language Acquisition) is a halfway house between language change and creolisation … In language change, lexical items functioning within an already grammaticalized system get bleached … whereas in creolisation, the learner creates phrasal constraints in the absence of input … In both these cases, what needs to be explained is the choice of categories which speakers grammaticalize, and in the latter case, their respective order. In ALA on the other hand, the learner is almost always dealing with grammatical input (as opposed to foreigner talk) and what needs to be explained is both the order of the TL phrasal constraints that *are* acquired, and also why some grammaticalized categories of the TL are *not* acquired …

Klein and Perdue's caveat is congruent with what Heine and Kuteva (2007: 168) have identified as a "stripping" process in pidginization. In a similar vein, Bickerton (1981, 1988) provided a list of features and markers that are restored if lost during the development of creole languages. This includes such items as: articles, tense mode aspect (TMA) markers, interrogative words, plural markers, personal pronouns, locative prepositions, irrealis markers, relative markers and reflexive and reciprocal markers. For both Bickerton and Heine and Kuteva, the "stripping" process which corresponds to "protolanguage" or "pidginization" is followed by a restructuring process which affects layers three to five (from adjectives and adverbs to agreement and subordination) in Heine & Kuteva (2007), and produces the FUO stage in Klein & Perdue (1997)'s developmental sequence for ALA.

4.3   *A tertium comparationis:* Comparing and contrasting early FSL and FRC data

The purpose of the comparative approach to FSL naturalistic data and FRC data is to identify the "inner" developmental sequences in the processing of the linguistic input available, if any. External ecology is starkly contrasted in naturalistic adult acquisition and the development of Creole languages, but the manner in which lexical items and strings of the input are handled in both cases is shaped by the caveats underlying pre-BV and BVs. It is expected that this framework will foster comparison and enable internal reconstruction of FRC evolution.

The following features that pertain to pre-BV and BVs of FSL will be examined in both sets of data:

– verb forms and negation placement;
– presentationals.

At pre-BV, besides chunks, naturalistic learners produce one- or two-word utterances. Moving from pre-BV to BV implies alignment of the constituents of the utterances organised by the verb.

The three main strings identified at the pre-BV and BV levels in French L2 learner data are:

A. NP^3_1 V NP_2
(1) Abdelmalek (A) (Cycle 3^4): le^5 elle /ilmavole[6]/ un baguette de pain

("the she stole a loaf of bread")

B. NP_1 Cop SN_2/ Adj. / Prep. P
(2) A. (Cycle 1): France /se/ difficile

("France it's difficult")

C. V NP_2
(3) A. (Cycle 1): après /evini/ administration ("after comes administration")
(4) A. (Cycle 1): /jãna[7]/ cinq heures ("there's five hours")

In the chapter, the focus will be mainly on type B and type C utterances, where copulas and presentationals are used.

4.4 Research questions

According to Starren (2001), the development of temporality marking in FSL follows a specific path. At the lexical stages (NUO and IUO), i.e. pre-BV and BV, the verb bears no inflection and temporality is expressed *inter alia* by temporal adverbials and various discourse organisational principles (for instance the "natural

---

3. This notation does not imply that the linguistic categories used by learners are identical with the TL linguistic categories. Indices on the categories indicate order of alignment.

4. In the course of the ESF project, data have been collected from each informant on a monthly basis. Three cycles of nine encounters were organized, lasting for three years. All the tasks were repeated three times at ca. one-year intervals. Learner utterances analyzed in the chapter refer to the cycle and encounter where they occurred (e.g., Cycle 1, encounter 1); see Giacomi et al. (2000). In the study of Egyptian Arabic L1 learners (Guénédi 2014), data collection took place during one cycle of data collection (ca. one year) comprising 7 encounters.

5. Main conventions for transcription are:

– VPs are transcribed in IPA;
– +, ++, +++ indicate the length of pauses.

6. /ilmavole/ (stole) is considered to be an uninflected verb form.

7. /jãna/ and variants /jana/, /jan/ may have both an existential/ presentational reading (translated by *there is*) and a possessive reading (translated by *have*). By convention, /jãna/ and variants will be translated by *there is*.

order principle": "what comes first is narrated first"). Verbal inflection *per se* never develops in pre-basic and basic learner varieties of French. At the morphological (FUO) stage, i.e. post-BV, auxiliaries and copulas develop. This second stage witnesses the beginning of the development of finiteness on the verb.

To describe the use of verbs in the data, Klein's (1994) classification of the inherent temporal features of verbs will be used. Klein proceeds to a tripartite distinction between (a) 0-state verbal items which present no temporal contrast, usually refer to states and exhibit the following features: (+ durative, +atelic); (b) 1-state verbal items which contain at least one temporal contrast, refer to activities and exhibit the following features: (+ dynamic + atelic); and (c) 2-state verbal expressions which contain a temporal contrast between a source and a target state, refer to achievements and accomplishments, and exhibit the following features: (+ dynamic + telic).

The comparative study of FRC and FSL explores the following topics:

1. the emergence of uninflected verb forms in both sets of data, in relation to Klein's classification of verbal items, and the path towards the acquisition of inflection;
2. the initial placement of negation vis-à-vis light verbs, presentationals and lexical verbs;
3. the development of presentationals in both sets of data.

5. The FSL informants

The oral FSL data to be compared and contrasted with FRC data have been collected from Arabic L1 naturalistic adult learners of French.[8] Data from two

---

Table (i) Moroccan informants from the ESF project (Giacomi et al. 2000)

	Zahra (Z)	**Abdelmalek (A)**
Civil status	Maried (3 Children)	Batchelor
Education	None	Primary
Period of stay in France before investigation	13 months	14 months

Moroccan Arabic L1 informants of the European Science Foundation (ESF) project (Giacomi et al. 2000), Abdelmalek and Zahra, and two Egyptian Arabic L1 informants, Muhammad and Ali (Guénédi 2014), will be examined. An independent longitudinal analysis of the learner varieties collected showed that these learners belonged to the pre-basic and basic stages of development of French as L2.

6.   Aspects of the development of early grammar in FSL

This section describes the emergence of verb forms, the marking of negation and the use of presentationals in pre-BVs and BVs of French as used by Arabic L1 naturalistic learners of French L2.

6.1   Verb forms at early stages of FSL

The following narrative by Abdelmalek, a Moroccan informant, provides a good illustration of the type of text produced at the NUO stage, i.e., at the lexical stage (Starren 2001).

(5)   (Abdelmalek narrates his arrival in France: Cycle 1, encounter 1)

	Zahra (Z)	Abdelmalek (A)
Age at the time of investigation	24	20
Courses in French	1h/week (7 months)	1h/week (7 months)
Other languages	None	Literate in Arabic + elementary Spanish
Occupation	Textile worker	Fisherman

Table (ii) Egyptian Arabic learners of French (Guénédi 2014)

	Ali	Mohamed (Moh.)
Civil status	Married (2 children)	Married (1 child)
Education	Primary school in Egypt	Primary school in Egypt
Period of stay in France before investigation	6 years	7 years
Age at the time of investigation	31	30
Courses in French (France)	2 semesters	-
Other languages	English	English
Occupation	Worker	Worker

    I.   tu peux/ oui alors tu peux me dire quand c'est/ depuis combien de temps tu es là?
("Can you tell me when /since when you have been here?")

        1.   A. comment le problème comme /ãtre/ la France?
("how the problem as / how enter France")

    I.   ouais par exemple ouais ("yeah for instance yeah")

        2.   A. ah ouais parce que moi /liãtre/ la France/
("ah yes because me enter France")
jana/pas de passeport /jana/pas de rien
("there's no passport there's nothing")

    I.   ouais
("yeah")

        3.   A. parce que /ãtre/ la France /e/ la montagne
("because enter France is/ and the mountain")

    I.   tu es passé par la montagne?
("you went through the mountain?")

        4.   A. ouais
("yeah")

    I.   ah
("ah")

        5.   A. /jana/ cinq jours /e/ la montagne après /lãtre/ la France /lepase/ la douane de France
("there's five days/ is/ and the mountains after / enter/ France")
/komjes/? quinze kilomètres
("how much? fifteen kilometres")

    I.   à pied?
("on foot")

        6.   A. ouais /lapje/ /e/ après /ilaparte/ l'autoroute /jana/ pas des sous /jana/ rien après /jana/ /e/ le stop après /leveny/ le gendarme
("yeah foot and/is after go/ went/ motorway there's no money there's nothing after there's the hitchhiking after come / came the policemen")

In this extract, background information is introduced by presentational /*jana*/ ("there's"). Connectives *après* ("after") and *parce que* ("because") link the various utterances in the foreground (lines 1 to 6). Verbal items /*liãtre/, /ãtre/, /lãtre/, /lepase/, /ilaparte*/ and /*leveny*/ are long verb forms (*Ve*), preceded by an element /*li-*/, /*l*/, /*le*/ and /*ila*/ whose linguistic status is unclear at this stage. The vocalic element in these clusters preceding the lexeme forms a possible base for the pre-verbal development of a precursor auxiliary called proto-aux by Starren (2001:159–160).

6.1.1   *Basic verb forms:* VØ *and/or* Ve

As illustrated by Example 5 above, tokens of French verb forms in pre-basic and basic learner varieties of FSL are often preceded by a consonant + vowel cluster /l+e/ or /l+i/, or by a sole vowel /e/i/ and followed by a vocalic ending *e/i*, hence /lekase/ ("broken / has broken"), /lesini/ ("signed / has signed") etc. The initial cluster of the verb can neither be analyzed as a clitic pronoun nor as an auxiliary. In some contexts, the initial vowel or cluster alternates freely with Ø (see Véronique 1983).

As shown in Tables 9.1 and 9.2, in the early stages of the acquisition of FSL, verb forms may be either in the verbal stem or short form (*VØ*), or in a long form (*Ve*). Only a handful of verbs tend to alternate between *VØ* and *Ve*.

Table 9.1 Tokens of *VØ* and *Ve* in Mohammad's learner variety

	En(counter) 1	En. 2	En. 3	En. 4	En. 5	En. 6	En. 7	Total
VØ	40	31	69	61	92	79	72	444
	32.5%	25%	48.5%	39%	51.6%	42%	40%	40%
Ve	25	27	41	43	48	61	86	331
	20%	21%	28%	27.5%	26%	32%	47%	30%
Total of tokens of French verbs	65	58	110	104	140	140	158	775
Tokens of Arabic verbs	21	54	24	47	33	45	95%	233
	17%	43.5%	16%	30%	18.5%	23%		21%
Tokens of English verbs	37	12	8	5	0	3	0	65
	30%	9.6%	5.6%	3%	0%	1.5%	0%	5.9%
Total of tokens of all verbs	123	124	142	156	173	188	167	1073

Table 9.2 Tokens of *VØ* and *Ve* in Ali's learner variety

	En(counter)1	En.2	En.3	En.4	En.5	En.6	En.7	Total
VØ	37	85	125	44	79	82	36	488
	49.3%	59%	57.3%	37.9	47.8	47.1%	33.9%	48.8%
Ve	25	50	86	56	55	79	52	403
	33.3%	34.7%	39.4%	48.2%	33.3%	45.4%	49%	40.3%
Total of tokens of French verbs	62	135	211	100	134	161	88	891
Tokens of Arabic verbs	0	0	3	2	0	3	0	8
				1.7%		1.7%		0.8%
Tokens of English verbs	0	0	1	3	1	0	1	6
			0.4%	2.5%	0.6%		0.9%	0.6%
Total of tokens of all verbs	62	135	215	105	135	164	89	905

In the data from Mohammad (Table 9.3) and Ali (Table 9.4), French verbs used are mainly of the form *VØ* or *Ve.* Tokens of verbs from English and Arabic are also observed; their use tends to decrease during the data collection period.

Starren (2001) shows that the use of *VØ* and *Ve* forms is not directly related to the inherent semantic properties of the verbs. In Tables 9.3 and 9.4, she provides a list of the forms found for six 1-state verbs and six 2-state verbs in retellings of extracts of the film *Modern Times* produced by Abdelmalek and Zahra, during Cycle 2 and Cycle 3 of the data collection period (i.e., approximately three and four years after the arrival of the informants in France). Note that in other types of data, Abdelmalek produces /kon/ ("know") (Example 14, during Cycle 1) and /ʃerʃe/ (Example 17, during Cycle 2) and Zahra /ʃerʃ/ and /ʃerʃe/ (Example 9).

In the data from Abdelmalek, there is no relation between the inherent temporal properties of the verb and the form of the verb. Some pre-verbal type of inflection starts to develop as the learner variety moves out of BV.

Table 9.3 *VØ* and *Ve* with 1-state verbs and 2-state verbs in *Modern Times* (Starren 2001) in Abdelmalek's learner variety

1-state verbs	chercher ("look for")	rester ("stay")	regarder ("look at")	manger ("eat")	dormir ("sleep")	connaître ("know")
Cycle 2	ʃerʃ	reste	rogard	mã??e	dor	kone
Cycle 3	meʃerʃ	reste	rogarde	mãʒe	dor	kone
2- state verbs	partir ("go")	voler ("steal")	tomber ("fall")	venir ("come")	ouvrir ("open")	retourner ("return")
Cycle 2	eparti	avole	tõbe	evini	uvr	turn
Cycle 3	maparti	mavole	matõbe	mevjê		iturn

Table 9.4 *VØ* and *Ve* with 1-state verbs and 2-state verbs in *Modern Times* (Starren 2001) in Zahra's learner variety

1-state verbs	chercher ("look for")	parler ("speak")	regarder ("look at")	manger ("eat")	dormir ("sleep")	connaître ("know")
Cycle 2	ʃerʃ	parl	rogard	mãʒ	dor	kone
Cycle 3	ʃerʃ	tparl	rogard	mãʒ	dor	kone
2-state verbs	partir ("go")	voler ("steal")	tomber ("fall")	venir ("come")	sortir ("go out")	retourner ("return")
Cycle 2	eparti	vole	trõbe	vjê	–	–
Cycle 3	eparti	tvole/evole	trõbe	ilvini	isort	iturn

Although Zahra's list of verb forms differs from Abdelmalek's, in her data too no relation can be established between the lexical properties of the verbs and the form of the verb (*VØ* or *Ve*). At the end of Cycle 3 of data collection, she develops some form of pre-verbal marking and agreement.

6.1.2 *The semantics of VØ and Ve in pre-basic and basic learner varieties*

According to Klein and Perdue (1993), in early varieties of FSL (pre-BVs and BVs), the form chosen for each verb is determined by multiple factors. *Ve* tends to be associated with 1-state verbs (activities) and 2-state verbs (accomplishments), whereas *VØ* is associated with 0-state (or stative) verbs. However, as shown in Tables 9.3 and 9.4, this is not a categorical phenomenon. *VØ* refers also to the imperfective or prospective aspect, whereas *Ve* tends to refer to the perfective. Finally, *VØ* represents generics and habituals as in (7), whereas *V+ e* refers to individuated events in the past as in (8).

(7) A. moi je /demãd/ 160 francs il /don/ rien + l'autre /ilmadi/ "non je /don/ 140" moi je /di/ "non 160 francs" et je /vãdr/ 150
("me I ask for 160 francs he gives nothing + the other man he tells me I give 140 me I say no 160 francs and I sell for 150")   (Cycle 2, encounter 3)

(8) A. il /vjê/ il /done/ la clé
("he comes he gave (me) the key")   (Cycle 2, encounter 3)

Factors determining verbal alternation on a limited set of verbs in FSL BV include the marking of aspect as in (9), and the scope of the negator in (10):

(9) Z. /e/ après /ʃerʃ/ celui-là la maison grand /e/ après /raʃte/ toutes les choses /e/ après moi /parti/ le Maroc /ʃerʃe/ les enfants avec la carte jaune à la Paris
("and afterwards [we] looked for a big house and bought everything and then I left for Morocco to fetch the kids for Paris with the yellow card")   (Cycle 2, encounter 1)

In (9), /ʃerʃ/ refers to a (+ dynamic – telic) activity whereas /ʃerʃe/ refers to a (+ perfective) activity.

(10) Z. /tikompri/ /e/ je /kõprã/ pas /parle/   (Cycle 1, encounter 2)
(= "you understand and I do not understand / know [how to] speak")

In (10), long form */kompri/* alternates with short form */kõprã/*, which is under the scope of the negator. However, the effect of the negator on the form of the verb is variable, as illustrated in Examples (14), (15) and (16) from Abdelmalek below.

6.1.3 *Summary*

In pre-BVs and BVs of FSL, the French verb usually exhibits one of two forms, *VØ* or *Ve*. Only a few verbs exhibit verbal alternation (*VØ* and *Ve*) determined by

various factors, including the marking of aspect. Verbs are inflected neither for agreement nor tense. Temporality is marked mainly through discoursal organization and time adverbials. According to Alleyne (1996), the use of the verb forms *VØ* or *Ve* is the foundation of what may constitute the ontological basis of any verbal system, i.e. the relation between the achievement of a process or event, the attainment of its end state and the resulting state or event.

6.2 Negation in FSL

The path of development for the acquisition of the negator in FSL is the following: *non* (no) is the first negative item acquired; *pas* is mainly found in un-analyzed chunks, with presentationals. At a second stage, *pas* extends to the context of lexical verbs. The next stage, which is related to the development of finiteness on the verb, is the acquisition of placement of the negator on the VP.

In pre-basic learner varieties, *non* is by far the most productive item, and *pas* occurs in some rote-learned expressions such as */se/ pas* ("know not") or *ça va pas* ("not OK"). Anaphoric *non* in focus position is often associated with a lexical topic both in pre-BV and BV, as in (11) and (12):

(11) I. Vous êtes mariée ou célibataire?   ("you are married or single?")
      Z. non mariée ("no married")   (Cycle 1, encounter 3)

At the basic variety stage, *pas* is first observed in chunks such as *je /kõprã/ pas* ("I don't understand"), *je/se/ pas* ("I don't know"), or with presentational verbs as */jãna/* ("there is") and */se/* ("it is") as in (12),

(12) Z. heureusement euh les enfants /jãnapa/ de l'école aujourd'hui
          ("Fortunately the children there is no school today")
                                                                        (Cycle 1, encounter 2)

Later the use of *pas* extends to lexical verbs:

(13) Z. /ilafe/ pas les chaussures   ("he has made not the shoes")
                                                                        (Cycle 3, encounter 3)

Differently from Zahra, Abdelmalek places *pas* in pre-verbal position in the early phases of his learner variety, except in rote-learned expressions such as */janpas/* ("there is").

(14) I. tu connais Aix?   ("You know Aix?")
      A. non pas /kon/ Aix ("no not know Aix = no I don't know Aix")
                                                                        (Cycle 1, encounter 1)

By the end of the data collection period, *pas* is placed in post-verbal position following TL norm. In the early stages of Abdelmalek's BV, */napa/ X* and *pas X* alternate freely, as in (15):

(15) A.   /saje/ ça /napakompri/ ça pas /kompri/ ("OK not understood that not understood" = "OK that I have not understood that I have not understood")   (Cycle 1, encounter 2)

The following alternation is also observed: post-posed *pas* occurs with *VØ* whereas pre-posed *pas* is found in the context of *Ve*.

(16) A.   non je /sufl/ pas non ("no I am not breathless") versus non je pas /sufle/ ("no I am not breathless")[9]   (Cycle 2, encounter 6)

However, this may be an idiosyncratic variation as illustrated in (14) above. A certain measure of inter-individual variation may be observed in the early expression of negation in FSL. The placement of negator *pas* is variable depending on the type of verb involved; it is always post-posed to presentationals.

6.3   *Il y a* and *se* ("There is") in FSL

6.3.1   Il y a *("There is")*

One of the strings often produced at BV is Verb / Presentational in initial position followed by NP (Klein & Perdue 1997). *Il y a* ('there is / it is')[10] is found in position 1 of many utterances in FSL data. It fulfills the following propositional functions:

– clefting;
– providing support for negation and temporal marking, and discursive functions;
– topic introduction;
– marking contrast between foreground and background in texts.

il y a or rather *jāna* is found in all learner varieties. It expresses possession as well as existence (Duff 1993). In (17) the first token of *jāna* is an existential and the second one expresses possession:

(17) A.   voilà /jāna/ deux personnes /iveni/ /ʃerʃe/ avec quelqu'un /jāna/ des drogues + voilà /eveni/ euh voilà /ja/ rien /itrap/ avec l'autre qui /mã??/ à côté de   (Cycle 2, encounter 5)
            ("OK there are two persons [they] come search with somebody has drugs OK come euh OK there's nothing catch the other who is eating besides")

---

9.   One reviewer notes that the alternation /sufl/ ~/sufle/ could be proof of the emergence of finiteness on VP. I do not share this opinion. Alternation of forms is related to the placement of the negator.

10.   *il y a* occurs under various phonetic guises in the data: /ja/, /jāna/.

6.3.2   C'est *("It is")*

Like *il y a*, *c'est* combines with a wide range of categories and fulfills similar discoursal and pragmatic functions. *C'est* introduces specific referential expressions, as in:

(18)   Z.   /se/ pas un studio une chambre ("It is not a studio, it's a room").

C'est acts an an existential in competition with *jãna* as in (19):

(19)   A.   loin de Marseille /se/ grande fabrique de sardines ("Away from
                     Marseille, it is / there's a big sardine factory")   (Cycle 1, encounter 3)

6.3.3   *Summary*

C'est and *il y a* are topics markers (Véronique 1994) and mark background information in narratives (Giacomi 1986). As presentationals, these units introduce negators, modals and temporal adverbs (Gajo 1993). They predicate existence and possession and identity.

7.   **Aspects of the building of grammar in FRC**

As in Section 6, this section explores the same aspects of the development of predication in various old texts from Atlantic and Indian Ocean Creoles, i.e., the emergence of verb forms, negation and presentationals. Despite some ecological differences, texts from French Guyana and Louisiana have been considered to pertain to Atlantic Ocean FRC, as well as texts from Haiti and the Antilles. As in Section 6, the extracts of FRC texts quoted represent a stage of FRC where TMA markers have not yet developed.

7.1   Verb forms in FRC

In all the early texts available from Atlantic and Indian Ocean Creoles, verbs are in the non-finite *VØ* and/or *Ve* forms, similar to FSL *VØ* and *Ve*.

a.   *Atlantic Ocean FRC*

In examples (20), (20') and (20") from Prudent (1993: 125), and (21) from Hazaël-Massieux (2008: 30) – these texts quote a 1671 report – and (22) from Jennings (1995), verbs are not inflected. All verbs in italics are in one of the basic verb form available, the *VØ* or *Ve* forms, as in FSL.

(20)   ly *teny* trois fois sur l'eau & *gardé* toujours avec ses yeux gris; my *teny*
                mouche per ly *faire* … puis pour *voir* lui ("he stands /stood three times on
                water and looks /looked with its grey eyes; me was not afraid [of] what he
                did … then to see him")

(20')   moy *teny* petit peur, non pas grand, non …   ("I was a bit afraid not much, no")

(20")   moy non *miré* bien luy, parce que lui *étoit* dans l'yau, lui *sembloit* poisson, moi *teny* peur ("me not see him well because it was in the sea, him looked like a fish I was afraid")

(21)   moi *mirer* un homme en mer du Diamant. moi *voir* li trois fois. Li *tini* assés bon visage et zyeux comme monde. li *tini* grande barbe grise ("me see a man in the Diamant sea. Me saw him three times. He had a quite good face and eyes as everybody. He had a grey beard")

(22)   Seigneur, toy bien *sçave*[11] que mon frère luy point *mentir*, point luy *iurer*, point *dérober*, point *aller* luy à femme d'autre ("Lord you well know that my brother [does] not lie, [does] not swear, [does] not go after the wife of others")   (1655 in Jennings 1995: 74)

b.   Indian Ocean FRC

These two extracts of Île de France Creole, quoted from Chaudenson (1981), show that verb morphology in the Indian Ocean FRC is also of the type *VØ* or *Ve*.

(23)   Si nou n'a pas *gagné* malheur, ça bon (1768)

("If we have no problem, it's OK")

(24)   Moi *voulé baiser* ça négresse la (1777)   ("I want to fuck this negress")

In (24), a modal verb precedes the full lexical verb. Stein (2007) has surveyed verb forms in the whole set of data provided by Baker & Fon Sing (2007) for Mauritian Creole. He notes that light verbs and modals (attested as early as 1734; the island was settled by the French in 1721) are probably the first explicit markers to be associated with the basic verb form. Forms derived from periphrastic French forms such as *était à* ("to be in the process of") appear later. In Mauritian Creole, Stein (2007) shows that verb alternation between a long and a short form of the verb is also a late development.

7.1.1   Summary

Verb forms in early FRC texts are mainly either *VØ* or *Ve*. Alternating verbs seem to be more frequent in the Indian Ocean Creoles than in Atlantic Creoles, but verbal alternation is a late development. Lexical aspect as conveyed by the verb is the only means available to mark aspectual information on the verb. Modal verbs are the first forms to be associated with lexical verbs (Hazaël-Massieux 1996: 209ff; Véronique 1995). Pre-verbal TMA markers develop later in both Atlantic and Indian Ocean Creoles (Véronique 1995).

---

11.   *Sçave* (know) could be interpreted as a short form [sav] because of the inconsistencies of French orthographic *e*.

7.2 Negation in FRC

a. Atlantic Ocean FRC

In early texts of Atlantic Ocean FRC, pre-posed negation is frequent. Negators *non, point, pas* are found in pre-verbal position:

(25) Luy mouche manigat. Mouche manigat mon compère moy *non* faché à toy
("He is much skillful. You are much skillful mate. I am not angry with you")
(1640, in Prudent 1993: 146)

(26) Seigneur toi bien savé que mon frère lui *point* mentir, *point* lui jurer ("Lord
you know well that my brother he [does] not lie, [does] not swear')
(1682, in Prudent 1993: 119)

(27) Toi *pas* connaître moi ("you don't know me")   (1682, in Prudent 1993: 122)

Speedy (1995) and Valdman (1996) quote some early examples of negative items in Louisiana FRC where the preposed negator includes the TL auxiliary or a single *pas*:

(28) Loui. Cela *n'est pas* bon, s'y toy mourir seul et *n'y a pas* faire mourir monde
qui *n'y a rien faire* avec toy ("It's not good if you die alone and do not get
people killed who had nothing to do with you")   (1748, in Speedy 1995: 105)

(29) Loui. Vous *pas* mire donc Maître à moi, ça Caïman qui mange monde?
("You have not seen my Master, the caiman that eats everybody?")
(1777, in Speedy 1995: 105)

The expression of negation in Atlantic Ocean FRC is initially based on the re-analysis of French *Neg. + Auxiliary* to form a compound pre-verbal negator. At a first stage, *pas* and *n'a pas / n'est pas* seem to alternate. Further, there does not seem to be any difference between negator *n'a pas* and the negative verb *n'a pas* ("it does not exist"):

(30) Compère na pas tenir peur, si canot tourner toi tenir coeur fort ("Friend do
not have fear, if the boat overturns you keep [your] heart steady")
(1722, in Prudent 1993: 191)

pas is often post-posed with modals and *(é)té* as a copula:

(31) Guy(ane). Vou *té pas*[12] tiré aucun profit di vou travail ("You didn't benefit
from your work")   (1797, in Hazaël-Massieux 1996: 175)

---

12.   One reviewer notes that in 31, 32 and 33, pre-verbal marker *té* is present. In Atlantic as in Indian Ocean FRC, pre-verbal TMA markers develop after modal and light verbs. In Antillean FRC, the first markers such as *té* are found in *La passion en langage nègre* ca. 1720–40. In Mauritian Creole *va* is present in a text from 1777 and *été* in a text from 1779 (Stein 2007). In examples such as 31, 32 and 33, it may be argued that *té* is a copula because of the post-position of the negator, rather than a TMA marker.

(32)   Guy. Mo *té pa* briga ("I was not a thief")

(1872, in Hazaël-Massieux 1996: 175)

(33)   Loui. Li *té pa* gê sulyé ("he had no shoes")   (Hazaël-Massieux 1996: 176)

b.   Indian Ocean FRC

Early negation in Indian Ocean Creole develops along a path similar to that of Atlantic Ocean FRC, at least in the case of Île de France (Mauritius). The early form of the negator for Île de France FRC is *n'a pas/ napas* preposed to the verb:

(34)   Mau. ça n'a pas bon, Monsié ("It is not good Sir")

(1773, in Chaudenson 1981: 77–78)

(35)   Mau. moy n'apa été batté ça blanc là ("I didn't hit that white man")

(1779, in Chaudenson 1981: 78)

In the case of Réunion FRC, according to Chaudenson (1981) and Bollée (2007), various forms of negation may be observed. In Caulier ca. 1770, both pre-verbal *pas* (36) and *n'a pas* (37) are found along with French *ne … pas:*

(36)   Pas jurer jamais   ("never swear")

(37)   N'a pas touzours   ("not always")

7.2.1   *Summary*

In FRC, the development of a pre-verbal negator is linked to the re-analysis of French *Neg + Auxiliary* and to the simplification of verb morphology. From another theoretical perspective, it could be asserted, following Mather (2006: 264), that the uninflected verb remains to the right of the negator, in base position. No movement of the verb may be observed until inflection develops as in 31, 32 and 33 above. Later as pre-verbal markers and auxiliaries develop (see 6.3 below), the placement of the negator is modified. In the domain of negation, Atlantic and Indian Ocean Creoles partly differ in the use of negators; *mouche* and *point* are found only in Atlantic Ocean FRC.

7.3   *Il y a* and *C'est* ('There is') in FRC

The realm of presentationals is probably the area where Atlantic and Indian Ocean FRC differ the most. The lexical items for predicating existence, possession and identity differ from one area to the other, as does the use of presentationals and copulas. Antillean FRC differs from Haitian, Louisiana and Guyanese Creoles by using (old *tenu*) modern *tini* for existentials. Louisiana Creole uses forms such as *ya, yan na* and *navé* (Neumann-Holzschuh 1987: 21) for predicating existence and possession, reminiscent of corresponding Indian Ocean items.

a.   Atlantic Ocean FRC

As illustrated in (20), repeated below, and in (22) above:

(20)   ly *teny* trois fois sur l'eau & *gardé* toujours avec ses yeux gris; my *teny*
        mouche per ly *faire* … puis pour *voir* lui ("he stood three times on water
        and always looked with his grey eyes; me was not afraid [of] what he did …
        then to see him"),

teny/ tini is used both as a lexical verb and as an existential in Antillean FRC. Besides,
Atlantic Ocean FRC have maintained (*((é)té) / C'est/ C'étoit* as a copula and presenta-
tional during the early stages of their development, hence (38) (38') and (38"):

(38)   Comme lendimain *c'étoit* grand fête pâque ("as the following day it was
        Easterday")   (1720–40, in Hazaël-Massieux 2008: 66)

(38')   *C'est* li même qui va trahi moé ("He is the one who will betray me")
                                                       (1720–40, in Hazaël-Massieux 2008: 63)

(38")   Loui. *Cé* li ki apé volè vou moutons ("He is [the person] who is stealing
        your sheep")   (1878, in Neumann-Holzschuh 1987: 69)

In some Atlantic Ocean FRC, *gain* (to have) is also present (see texts from Guy-
ana in Wiesinger 2017: 493–494 and from Louisiana in Neumann-Holzschuh
1987: 19). According to Hazaël-Massieux (1996: 177), (*((é)té) / C'est/ C'étoit* may
be considered as exemplars of copulas, because of the post-position of *pa* to *té*
(see Examples 31–33 above). Later, two forms emerge: the TMA marker *te* and
th*e* copula *se*.[13]Fattier (1994: 68) observes that in the colonial creole of the white
planter Ducoeurjoly (1802), copulas are Ø and *c'est*.

b.   Indian Ocean FRC

In early texts from Île de France, the marker expressing existence and possession
is *y en a* (there is / have):

(39)   Moi *y en a* femme ("I have a woman")
                                    (1804–05, in Chaudenson 1981: 81) (possession)

(40)   *Y en a* ça qui bon *y en a* ça qui mauvais ("There are those who are good,
        there are those who are bad")
                                    (1804–05, in Chaudenson 1981: 81) (existence)

Early data from Réunion creole show that the item expressing existence in that
creole is *n'en a*:

(41)   *Mon ami n'en a un travail que nous y faut faire ce coup ci* ("My friend there's
        a job we must do this time"   (1799, in Chaudenson 1981: 59). (existence)

---

13.   DeGraff (1992) analyzes *se* as a resumptive pronoun.

7.3.1   *Summary*

Atlantic Ocean FRC present a wider range of forms of presentationals and copulas than Indian Ocean FRC. The copula *c'est / se* etc. seems to be used in more contexts than *ete* in Indian Ocean FRC. *Tini* is specific to Antillean Creoles but *gain* and forms derived from *il y a / il y en* are re-analyzed along similar lines in both early Atlantic and Indian Ocean FRC texts. These items produce a full verb meaning possession and ownership on the one hand, and a light verb marking existence on the other.

8.   Discussion

The claim of this chapter is that early phases of naturalistic SLA yield insights into the early development of Creole languages. As Mather asserts (2006: 258), the claim is not that "the interlanguage of L2 learners of French is structurally identical to creoles" but rather that the restructuring/re-analysis of data in creole languages obeys the same linguistic and cognitive laws as in SLA. More precisely, it is claimed that *ceteris paribus*, a comparison of developmental sequences in the acquisition of French as a second language and of the development of grammar in FRC will aid the internal reconstruction of grammar development in FRC. Besides, this comparative approach is expected to yield some insights into the cognitive and linguistic processes involved in Creole language development. In this final section, problems that arise when comparing SLA and Creole development will be discussed. It will also be shown that the BV model provides interesting insights into Creole development, although it does not yield the whole picture.

The comparison of FSL and FRC data is fraught with theoretical and methodological problems springing from the scarcity and the nature of the data available for FRC. The external ecologies of FRC and FSL are strikingly different, and it is difficult to establish whether the extracts selected for the sake of comparison represent equivalent stages in their specific continua of development. On top of these factors, the language pools involved in the development of Atlantic and Indian Ocean FRC are partially different. However, at the same time we stand in real need of understanding how creole languages originated. The explanation offered in this chapter for the emergence and development of grammar in FRC is based on the idea that Creole language development constitutes a particular case of adult SLA. The functionalist BV model of SLA propounded by Klein and Perdue (1997) has been chosen to vindicate this position. It is hypothesized that the cognitive and linguistic factors that conspire to produce pre-basic and basic varieties in SLA are the very factors at work in the development of Creole languages. As a usage-based theory, the BV model is fully compatible with various theories of

Creole development, including the feature pool theory of Mufwene. The BV model holds that semantic, pragmatic and discourse constraints are the shaping factors of elementary learner varieties. Language transfer represents an intervening factor later on, when these elementary systems develop finiteness. The BV model illustrates the development of the first two layers of grammatical development – nouns and verbs – postulated by Heine and Kuteva (2007) and is fully compatible with Jackendoff's (1999) linear grammar.

In light of the BV model, the following empirical findings have been obtained for the development of grammar in FSL and in FRC. During the development of predication in FSL and FRC, lexical verbs are acquired and put to use under one of the following base forms: *VØ* or *Ve*. The verb forms (*VØ* and /or *Ve*) are not chosen in FSL and FRC grammars on the basis of their inherent temporal features (0-state, 1-state or 2-state verbs). However, both *VØ* and *Ve* bear various aspectual meanings. Few verbs alternate *VØ* and *Ve* in verb predication in FSL and FRC. The functional use of alternation in FRC as described by Becker and Veenstra (2003) is a late development. The development of pre-verbal inflection in FSL (Starren 2001) and of TMA markers in FRC is also a late event in both continua.

Presentationals and light verbs such as modals emerge quite early in both FSL and FRC, before any token of the marking of tense via pre-verbal TMA or inflection is observed. Negation is marked at the sentential level before spreading to the predicate level. Atlantic and Indian Ocean FRC differ partly in the expression of negation and in the use of presentationals and copulas. In Indian Ocean Creoles, *jāna* ("there is") and its equivalents express both existence and possession. However, this is also the case for *gen* ("there is") in Haitian, Louisiana and Mauritian Creoles. These differences may be partly due to the language pools involved.

The fact that *VØ* and/or *Ve* are the first verb forms used in FLS and FRC shows that they constitute the basis of any verbal system. I shall follow Alleyne (1996), who considers that the ontological basis of any verbal system is the relation between the achievement of a process or event, the attainment of its end state and the resulting state or event. According to this author, *Ve* or long form, which is present in all FRC, is a perfect ("Topic Time" after "Time of Situation" [Klein 1974: 91]) which conveys both a resultative and a completed (*accompli*) meaning. This is consistent with results from the acquisition of French as L1 and L2 – according to which from the start two verbal forms are available.

A comparison of the evolution of FSL with that of FRC in the domain of grammar construction, as has been attempted here, illustrates the internal mechanisms that are responsible for the development of creoles, although it does not provide a full picture of creolization *per se*. This comparison enables the determination of the onset of specific linguistic categories and grammatical constraints. Incipient grammatical development of Creoles, and possible transfers from source

languages (Mufwene 1990) or hybridization of grammar (Aboh 2015), begin, as in SLA, when linguistic categories acquire grammatical status and are constrained, the major feature to develop being finiteness on the verb.

Acknowledgement

References

Aboh, E. O. 2015. *The Emergence of Hybrid Grammars: Contact, Change and Creation*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139024167

Adam, L. 1883. *Les Idiomes Négro-Aryen et Maléo-Aryen. Essai d'Hybridologie Linguistique*. Paris: Maisonneuve & Cie.

Adone, D. 1994. Creolization and language change in Mauritian Creole. In *Creolization and Language Change*, D. Adone & I. Plag (eds), 23–43. Tübingen: Niemeyer. https://doi.org/10.1515/9783111339801

Alleyne, M. C. 1996. *Syntaxe Historique Créole*. Paris: Karthala Éditions.

Alleyne, M. C. 2000. Opposite processes in "creolization". In *Degrees of Restructuring in Creole Languages* [Creole Language Library 22], I. Neumann-Holzschuh & E. W. Schneider (eds), 125–133. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.22.08all

Andersen, R. W. (ed.). 1983. *Pidginization and Creolization*. Rowley MA: Newbury House.

Andersen, R. W. 1983. Introduction: A language acquisition interpretation of Pidginization and Creolization. In *Pidginization and Creolization as Language Acquisition*, R. W. Andersen (ed.), 1–56. Rowley MA: Newbury House.

Arends, J. 1995a. Introduction. In Arends (ed.), ix–xv.

Arends, J. (ed.). 1995b. *The Early Stages of Creolization* [Creole Language Library 13]. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.13

Baker, P. 1984. Agglutinated French articles in Creole French: Their evolutionary significance. *Te Reo* 27: 89–129.

Baker, P. 2007. Elements for a sociolinguistic history of Mauritius and its creole (to 1968). In *The Making of Mauritian Creole. Analyses diachroniques à partir de textes anciens*, P. Baker & G. Fon Sing (eds), 307–333. London: Battlebridge.

Baker, P. & Fon Sing, G. (eds). 2007. *The Making of Mauritian Creole. Analyses diachroniques à partir de textes anciens*. London: Battlebridge.

Baron, N. S. 1977. *Language Acquisition and Historical Change*. Amsterdam: North Holland.

Becker, A. & Veenstra, T. 2003. The survival of inflectional morphology in French-related creoles. *Studies in Second Language Acquisition* 25: 283–306. https://doi.org/10.1017/S0272263103000123

Bickerton, D. 1975. *Dynamics of a Creole System*. Cambridge: CUP.

Bickerton, D. 1981. *Roots of Language*. Ann Arbor MI: Karoma Press.

Bickerton, D. 1983. Comments on Valdman's "Creolization and Second Language Acquisition". In *Pidginization and Creolization*, R. W. Andersen (ed.), 235–240. Rowley MA: Newbury House.

Bickerton, D. 1988. Creole languages and the bioprogram. In *Linguistics: The Cambridge Survey,* Vol. 2: *Linguistic Theory: Extensions and Implications*, F. J. Newmeyer (ed.), 268–284. Cambridge: CUP. https://doi.org/10.1017/CBO9780511621055.016

Bloomfield, L. 1933. *Language*. New York NY: Holt.

Bollée, A. 2007. *Deux textes religieux de Bourbon du 18e siècle et l'histoire du créole Réunionnais*. London: Battlebridge.

Bollée, A. & Neumann-Holzschuh, I. 1993. Pour une grammaire historique des créoles. In *Sprachwandel und Sprachgeschichte. Festschrift für Helmut Lüdke zum 65*, H. Lüdtke, J. Schmidt-Radefeldt & A. Harder (eds), 9–21. Tübingen: Gunter Narr.

Bossu, J.-B. 1777. *Nouveaux voyages dans l'Amérique septentrionale*. Amsterdam: Changuion.

Chaudenson, R. 1974. *Le lexique du parler créole de la Réunion*, Vols. 1 & 2. Paris: Librairie Honoré Champion.

Chaudenson, R. 1978. Créole et langage enfantin: Phylogenèse et ontogenèse. *Langue Française* 37: 76–90. https://doi.org/10.3406/lfr.1978.4852

Chaudenson, R. 1981. *Textes créoles anciens (La Réunion et Île Maurice). Comparaison et essai d'analyse*. Hamburg: Helmut Buske.

Chaudenson, R. 1992. *Des îles, des hommes, des langues. Essai sur la créolisation linguistique et culturelle*. Paris: L'Harmattan.

Chaudenson, R. 2001. *Creolization of Language and Culture*. London: Routledge.

Chaudenson, R. 2003. *La créolisation. Théorie, applications, implications*. Paris: L'Harmattan.

DeGraff, M. 1992. Creole Grammars and the Acquisition of Syntax. The Case of Haitian. PhD dissertation, University of Pennsylvania.

DeGraff, M. (ed.). 1999. *Language Creation and Language Change. Creolization, Diachrony and Development*. Cambridge MA: The MIT Press.

DeGraff, M. 2001. On the origin of creoles: A Cartesian critique of Neo-Darwinian linguistics. *Linguistic Typology* 5(2-3): 213–310.

Dimroth, C. 2018. Beyond statistical learning: Communication principles and language internal factors shape grammar in child and adult beginners learning Polish through controlled exposure. *Language Learning* 68: 863–905. https://doi.org/10.1111/lang.12294

Duff, P. 1993. Syntax, semantics, and SLA: The convergence of possessive and existential constructions. *Studies in Second Language Acquisition* 15: 1–34.
https://doi.org/10.1017/S0272263100011645

Fattier, D. 1994. Un fragment de créole colonial: Le "Manuel des habitants de Saint-Domingue" de S.J. Ducœurjoly, 1802. Réflexions sur l'apprentissage et la créolisation. In *Créolisation et acquisition des langues*, D. Véronique (ed.), 53–87. Aix-en-Provence: Publications de l'Université de Provence.

Gajo, L. 1993. L'acquisition de "c'est", "c'est...qui/ que". *Étude de cas. Cahiers de l'ILSL* 4: 207–224.

Giacomi, A. 1986. Processus de structuration de l'énoncé en acquisition et en interaction. In *Acquisition d'une langue étrangère: Perspectives et Recherches*, A. Giacomi & D. Véronique (eds), 285–303. Aix-en-Provence: Publications de l'Université de Provence.

Giacomi A., Stoffel, H. & Véronique, D. (eds). 2000. *Appropriation du Français par des Marocains Arabophones à Marseille*. Aix-en-Provence: Publications de l'Université de Provence.

Guénédi, M. 2014. Étude sur l'appropriation du Français par des apprenants Égyptiens: L'expression du temps. PhD dissertation, Cairo University.

Hall, R. 1966. *Pidgin and Creole Languages*. Ithaca NY: Cornell University Press.

Hazaël-Massieux, G. 1996. *Les créoles. Problèmes de genèse et de description*. Aix-en-Provence: Publications de l'Université de Provence.

Hazaël-Massieux, M. C. 2008. *Textes anciens en créole Français de la Caraïbe, histoire et analyse*. Paris: Publibook.

Heine, B. & Kuteva, T. 2007. *The Genesis of Grammar. A Reconstruction*. Oxford: OUP.

Jackendoff, R. 1999. Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences* 3: 272–279. https://doi.org/10.1016/S1364-6613(99)01333-9

Jackendoff, R. & Wittenberg, E. 2014. What you can say without syntax: A hierarchy of grammatical complexity. In *Measuring Grammatical Complexity*, F. Newmeyer & L. Preston (eds), 65–82. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199685301.003.0004

Jennings, W. 1995. Saint Christophe: Site of the first French creole. In *From Contact to Creole and Beyond*, P. Baker (ed.), 63–80. London: University of Westminster Press.

Klein, W. 1994. *Time in Language*. London: Routledge.

Klein, W. & Perdue, C. 1992. *Utterance Structure. Developing Grammars Again* [Studies in Bilingualism 5]. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.5

Klein, W. & Perdue, C. 1993. Concluding remarks. In *Adult Language Acquisition: Cross-Linguistic Perspectives*. Vol. 2, C. Perdue (ed.), 251–272. Cambridge: CUP. https://doi.org/10.1016/0003-4975(93)90557-X

Klein, W. & Perdue, C. 1997. The Basic Variety (or Couldn't natural languages be much simpler?). *Second Language Research* 13: 301–347. https://doi.org/10.1191/026765897666879396

Klein, W. 2009. Finiteness, universal grammar, and the language faculty. In *Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Isaac Slobin*, J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura & S. Özcaliskan (eds.), 333–344. New York NY: Psychology Press.

Kouwenberg, S. & Patrick, P. L. 2003. Introduction. *Studies in Second Language Acquisition* 25: 175–184. https://doi.org/10.1017/S0272263103000081

Lefebvre, C., White, L. & Jourdan, C. (eds). 2006. *L2 Acquisition and Creole Genesis: Dialogues* [Language Acquisition and Language Disorders 42]. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.42

Le Page, R. B. & Tabouret-Keller, A. 1985. *Acts of Identity. Creole-Based Approaches to Language and Ethnicity*. Cambridge: CUP.

Ludwig, R. 2018. Diachronies française et créole: Rapports épistémiques. In *Nouvelles voies d'accès au changement linguistique*, W. Ayres-Bennett, A. Carlier, J. Glikman, T. Rainsford, G. Siouffi & C. Skupien Dekens (eds), 143–177. Paris: Classiques Garnier.

Mather, P. A. 2006. Second language acquisition and creolization. Same (i-) processes, different (e-) results. *Journal of Pidgin and Creole Languages* 21: 231–274. https://doi.org/10.1075/jpcl.21.2.01mat

McWhorter, J. 2018. *The Creole Debate*. Cambridge: CUP. https://doi.org/10.1017/9781108553308

Mufwene, S. S. 1990. Transfer and the substrate hypothesis in Creolistics. *Studies in Second Language Acquisition* 12(1): 1–23. https://doi.org/10.1017/S0272263100008718

Mufwene, S. S. 1996. The founder principle in Creole genesis. *Diachronica* 13: 115–168. https://doi.org/10.1075/dia.13.1.05muf

Mufwene, S. S. 1997. Jargons, pidgins, creoles and koines: What are they? In *The Structure and Status of Pidgins and Creoles* [Creole Language Library 19], A. K. Spears & D. Winford (eds), 35–70. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.19.05muf

Mufwene, S. S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511612862

Mufwene, S. S. 2005. *Créoles, écologie sociale, évolution linguistique*. Paris: L'Harmattan.

Mufwene, S. S. 2008. *Language Evolution: Contact, Competition and Change*. London: Continuum.

Mufwene, S. S. 2010. Second language acquisition and the emergence of creoles. *Studies in Second Language Acquisition* 32: 1–42. https://doi.org/10.1017/S027226311000001X

Neumann-Holzschuh, I. (ed.). 1987. *Textes anciens en créole Louisianais*. Hamburg: Helmut Buske.

Perdue, C. 1996. Pre-basic varieties: The first stages of second language acquisition. In *Eurosla 6. A Selection of Papers*, E. Kellerman, B. Weltens & T. Bongaerts (eds). *Toegepaste Taalwetenschap in Artikelen* 55: 135-149.

Perdue, C. (ed.). 1993. *Adult Language Acquisition: Cross-Linguistic Perspectives,* Vol. 1: *Field Methods,* Vol. 2: *The Results*. Cambridge: CUP.

Pienemann, M. 1998. *Language Processing and Second Language Acquisition. Processability Theory* [Studies in Bilingualism 15]. Amsterdam: John Benjamins.
https://doi.org/10.1075/sibil.15

Plag, I. 2008a. Creoles as interlanguages: Inflectional morphology. *Journal of Pidgin and Creole Languages* 23: 109–130. https://doi.org/10.1075/jpcl.23.1.06pla

Plag, I. 2008b. Creoles as interlanguages: Syntactic structures. *Journal of Pidgin and Creole Languages* 23: 307–328. https://doi.org/10.1075/jpcl.23.2.06pla

Plag, I. 2009a. Creoles as interlanguages: Phonology. *Journal of Pidgin and Creole Languages* 24: 121–140. https://doi.org/10.1075/jpcl.24.1.06pla

Plag, I. 2009b. Creoles as interlanguages: Word formation. *Journal of Pidgin and Creole Languages* 24: 339–362. https://doi.org/10.1075/jpcl.24.2.05pla

Plag, I. 2009c. Creoles, interlanguages, processes and transfer: A response to Sprouse. *AILE* 2: 165–170. https://doi.org/10.4000/aile.4545

Plag, I. 2011. Creolization and admixture: Typology, feature pools and second language acquisition. *Journal of Pidgin and Creole Languages* 26: 89–110.
https://doi.org/10.1075/jpcl.26.1.04pla

Prudent, L.-F. 1993. Pratiques Martiniquaises. Genèse et fonctionnement d'un système créole. PhD dissertation, Université de Haute-Normandie.

Schumann, J. H. 1978. *The Pidginization Process. A Model for Second Language Acquisition*. Rowley MA: Newbury House.

Schwartz, B. D. & Sprouse, R. 1996. L2 cognitive states and the full transfer/full access model. *Second Language Research* 12: 40–72. https://doi.org/10.1177/026765839601200103

Singler, J. V. 1993. African influence upon Afro-American language varieties: A consideration of sociohistorical factors. In *Africanisms in Afro-American Language Varieties*, S. S. Mufwene & N. Condon (eds), 235–253. Athens GA: University of Georgia Press.

Singler, J. H. 1995. The demographics of creole genesis in the Caribbean: A comparison of Martinique and Haiti. In Arends (ed.), 203–232.

Speedy, K. 1995. Mississippi and Tèche Creole: Two separate starting points for Creole in Louisiana. In *From Contact to Creole and Beyond* [Westminster Creolistics, Series 1], P. Baker (ed.), 97–114. London: University of Westminster Press.

Sprouse, R. A. 2009. Creoles and interlanguages, conclusions and research desiderata: A reply to Plag. *Acquisition et Interaction en Langue Étrangère. Langage, Interaction et Acquisition* 1: 283–301. https://doi.org/10.4000/aile.4527

Sprouse, R. A. 2010. The invisibility of SLA theory in mainstream creole linguistics. *Second Language Research* 26: 261–277. https://doi.org/10.1177/0267658309337758

Starren, M. 2001. *The Second Time. The Acquisition of Temporality in Dutch and French as a Second Language*. Utrecht: LOT.

Stein, P. 2007. L'absence de marqueurs préverbaux et les fonctions du marqueur zéro. In *The Making of Mauritian Creole. Analyses diachroniques à partir de textes anciens*, P Baker & G. Fon Sing (eds), 157–172. London: Battlebridge.

Valdman, A. 1983. Creolization and second language acquisition. In *Pidginization and Creolization*, R. W. Andersen (ed.), 212–234. Rowley MA: Newbury House.

Valdman, A. 1996. La diffusion dans la genèse du créole louisianais. *Communication au 8ème Colloque International des Études Créoles*. Guadeloupe: Pointe-à-Pitre.

Véronique, D. 1983. Observations préliminaires sur *li* dans l'interlangue d'Abdelmalek. *Acquisition du Français par des Travailleurs Marocains. Papiers de Travail* 1: 55–80.

Véronique, D. (ed.). 1994. *Créolisation et acquisition des langues*. Aix-en-Provence: Publications de l'Université de Provence.

Véronique, D. 1995. Acquisition des modalités en français langue étrangère et développement des modalités dans les créoles français. In *From Pragmatics to Syntax. Modality in Second Language Acquisition*, A. Giacalone Ramat & G. Crocco Galèas (eds), 59–82. Tübingen: Gunter Narr.

Winford, D. 1993. *Predication in Caribbean English Creoles* [Creole Language Library 10]. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.10

Wiesinger, E. 2017. *Le syntagme nominal en créole Guyanais. Une étude synchronique et diachronique du marqueur LA*. Hamburg: Buske.

Wode, H. 1981. *Learning a Second Language. An Integrated View of Language Acquisition*. Tübingen: Gunter Narr.

Foundings and futures

How to live like a Peranakan in the post-digital ecology

Lisa Lim & Umberto Ansaldo

Curtin University / University of Sydney / University of Hong Kong

This chapter contemplates the positioning of the Peranakans (also known as Straits-born Chinese or Babas) – descendants of southern Chinese seafaring traders to the Malay archipelago who married local women, and settled in the region – across different eras and ecologies. We provide a critical digest of the contribution of scholarship on the Peranakans to Creole studies, World Englishes, and language endangerment – including a consideration of the *Founder Principle* in the ecology paradigm, as underscored by Mufwene, in establishing the Peranakans' role as early adopters in the spread and evolution of English in the region. We examine sociological and communicative factors in their language practices in local, transnational, and digital ecologies, highlighting issues of postvernacular practice, authenticity indexing, and identity branding.

Keywords: language ecology, transnational communities, postvernacular, authenticity, heritage branding

1 "Live like a Peranakan"

"Live like a Peranakan in Singapore", proclaims a feature in an online magazine (AccorHotels 2016), outlining five ways to "experience this unique culture", namely: eat Peranakan (by tasting typical dishes and cakes or learning to cook the cuisine), know Peranakan (by viewing artefacts in museums and houses), wear Peranakan (by procuring distinctive clothing, beaded slippers and jewellery), feel Peranakan (by visiting a heritage district for its conserved architecture), and live

Peranakan (by furnishing one's home with typical interior decor and furniture).[1] It is, in short, an experience of the Peranakans' tangible cultural heritage – food, history, artefacts, dress, decor. Notably absent is any mention of the Peranakan vernacular, whether simply listening to or experiencing it, let alone learning to speak it. One interpretation of this omission is simply that intangible cultural heritage is often the dimension of a community that receives less attention where preservation is concerned – though this is of course not to downplay the increasing resources in scholarly research and international initiatives directed to documenting and preserving intangible cultural heritage in recent years. The related explanation for this, in the case of the Peranakans, is simply that today, to "speak like a Peranakan" would not go beyond knowing the names for material artefacts, or symbolically using greetings and stock phrases, the Peranakan community in Singapore having long since shifted from their heritage language, Baba Malay, to English.

Most research has naturally tended to focus on documenting the vernacular language Baba Malay in the different original heritage ecologies of Singapore, Malacca and Penang (e.g., Ansaldo & Matthews 1999; S. Lim 1988; Pakir 1986; Tan 1988a, 1988b), noting the characteristics of the Peranakans as a creole community (Ansaldo et al. 2007), and recent work has reiterated the urgent need to represent the Baba Malay varieties in the different ecologies (Lee 2018). Other recent scholarship has offered an alternative recognition of the language practices of contemporary reality – namely, the Peranakans' multilingual practices and their emergent variety of Peranakan English (Lim 2010, 2016a, 2016b).

In this, we first provide an overview of the sociohistory of the Peranakans and summarise the contributions that scholarship on the community and their languages has made in the areas of creole studies, World Englishes, language endangerment, and sociolinguistics more broadly. We then move on to examine the evolution of language and culture in Peranakan communities in the context of the post-digital age (also referred to as the experience age; see final section).

---

1. To have been able to introduce Sali to the Peranakan culture in Singapore when we first met two decades ago was an enormous professional and personal pleasure. Little did we suspect then that the parallels in sociohistory and evolution of the creole communities in our respective areas of expertise would fire up inspiration, bring us together to co-author a paper soon afterwards, and, together with other interactions at the National University of Singapore and in Singapore, form the basis of our professional relationship and friendship over the years. We are grateful to the editors, Enoch, Michel and Cécile, for having invited us to be part of this volume for Sali, where we continue to explore in our contribution the ever-evolving situation of the Peranakans in new ecologies; and we thank the anonymous reviewers for their comments, which have helped to sharpen the ideas in this chapter. It would be remiss if we didn't mark this special occasion in the authentic Peranakan manner, by wishing Sali *panjang-panjang umor, badan kuat-kuat*, 'very long life [and] good health'!

While 'post-digital' is interpreted in several different ways, we use it here to refer to the state of things after digitisation, an existence where the digital is ubiquitous, commonplace and accepted, and here to stay (Cramer 2014; Davies 2009; Negroponte 1998). To this end, we undertake a preliminary exploration of the representation and language practices of the communities in the particular ecology of social media. We consider the communities in the traditional Southeast Asian Peranakan heartlands of Melaka, Penang and Singapore, and also transnational communities of Peranakans who emigrated and settled in new ecologies – taking the Australian cities of Perth, Melbourne and Sydney as a case in point – and look specifically at their language practices on social media.

2.   The making of the Peranakans: A snapshot

Reams have been written on the sociohistory of the Peranakans (the most notable modern works include Ansaldo et al. 2007; L. Lim 2010, 2014, 2016a, 2016b; S. Lim 1988; Rudolph 1998; Tan 1988a, 1988b); the following snapshot is distilled from these accounts. The Peranakans are descendants of southern Chinese seafaring traders who settled in the Malay Archipelago from at least the seventeenth century and who married non-Muslim natives of the region, such as Balinese or Batak slaves. Settling primarily in Malacca and Penang in peninsular Malaysia, and in Singapore – British colonies on the Strait of Malacca and the Strait of Singapore which were amalgamated in 1826 to form the Straits Settlements – the Peranakans, also known for this reason as the Straits Chinese, comprised one of the earliest and largest groups of the influential class of Chinese capitalists in the region. In contrast to the Chinese who returned to China, the Peranakans, even with their trading movements between the ports of southern China and Southeast Asia, always returned to the Straits Settlements and considered Malacca and Singapore their home. Until as recently as the 1950s, only the Peranakans could be considered "permanent", "native", or indigenised Chinese communities in the region (Song 1923 cited in Kwok 2000: 205).

The novel hybrid culture that emerged in this context shows unique traits that set the Peranakans apart from other Chinese, the more indigenous local populations, and other ethnically mixed groups (Rudolph 1998; Tan 1988a). Non-linguistic examples include a mixed *nyonya* cuisine consisting of Chinese culinary practices largely influenced by Malay traditions, and the wearing of the Malay/Indonesian *sarong* and *kebaya* (traditional female outfits), instead of the Chinese dress, by the women. These contrast with the retention of Chinese rituals, such as religious practices mentioned above and traditional wedding customs involving imperial-era wedding costumes (Tan 1988a: 299). According to some observers,

the Peranakans had "lost touch with China in every respect, except that they continued to uphold Chinese customs, and to practice, in variously modified forms, the social and religious practices of the forefathers" (B. K. Lim 1917, cited in Kwok 2000: 202; Tan 1988b: 47). Scholarly attention has been paid to their sociocultural identity, and their characteristics as a creole community (e.g., Ansaldo et al. 2007; Rudolph 1998; Tan 1988a, 1988b), as well as to their vernacular, Baba Malay, a restructured variety of Malay with substantial southern Chinese (primarily Hokkien) influence, both in creole studies as well as in the language endangerment discourse (e.g., Ansaldo & Matthews 1999; Lim 1988; Pakir 1986; Tan 1988a, 1988b).

Due in no small part to the fact that they had been in the region longer, more continuously and more permanently than the other Chinese immigrants, the Peranakans formed the larger proportion of the influential class of Chinese capitalists in the Straits Settlements. They established themselves in the mining of gold and tin; the large-scale commercial agriculture business in gambier, pepper, tapioca, and especially rubber; the import-export business; and other economic enterprises that had been drawing Chinese to Malacca for years (Tan 1988b: 48). By the time of the European exploitation and colonisation of the region in the nineteenth century, most Babas – the term for Peranakan males, but also used to refer to the community as a whole, synonymous with "Peranakans" – in Malacca had accumulated much wealth and become prestigious subgroups in the region, forming separate communities of their own. In particular they distinguished themselves from the later Chinese immigrants, referring to them derogatorily as *sinkeh* "new guests", i.e., "new arrivals", whom they considered poor and of low social status (Tan 1988b: 45). In Singapore as well, the Babas were a class apart from the other ethnic groups.

Although small in number ("Malacca men" comprised only 2.5% of the Chinese population in 1848, growing to just 9.5% in 1881), their social and economic influence was disproportionately strong in comparison to their size, and they formed an important sector of the local elite (Kwok 2000: 202–204). By the 1920s, Singapore-born Peranakans controlled the pineapple industry and most of the rubber that was cultivated; this comprised, at one time, more than 8,000 hectares in Singapore as well as in Malaya – and, together with tin, drove Singapore's prosperity in the late nineteenth and twentieth centuries (Liu 1999: 98). In Malacca, the well-off Babas were able to take over the houses of the great Dutch merchants in Heeren Street, which then became "the fashionable and aristocratic resort of the Chinese" (Braddell 1853: 74). In Penang, it was also noted that the Chinese "who have long been settled in the place, and who have wedded native wives, dwell in large and elegant houses environed with fruit and flower-gardens" (Thomson 1875: 13). In Singapore, the Peranakans were wealthy enough to afford

weekend retreats or second homes in the form of seaside bungalows – some with swimming enclosures – on the east coast of the island, an increasingly attractive residential area from the end of the nineteenth century (Liu 1999: 148).

In the politics of segregation introduced by the Dutch in Southeast Asia (Reid 2000) and continued by the British, individuals of mixed origin were used as middlemen, merchants and interpreters between the colonial administration, and the local population and newer arrivals. Many Peranakans worked for the British (as well as, in earlier days, the Dutch) East India Company (Tan 1988b: 51ff.), and their command of the English language afforded them greater interaction with British administrators and merchants (Nathan 1922: 77). Furthermore, their multilingual repertoire, which comprised Baba Malay, Bazaar Malay, Hokkien and possibly one or two other Chinese languages, as well as English, and their knowledge of local ways afforded them a significant role as intermediaries between Europeans, locals, and Asian newcomers (Kwok 2000; Lim 2016a, 2016b; Tan 1988b). All this, together with their business acumen, gave them predominance in the commercial sectors (see also Ho & Platt 1993: 8–9), and they were considered the best-educated, wealthiest and most intelligent section of the Chinese community (Nathan 1922: 77).

English was already becoming an increasingly important language of Southeast Asia, especially British Malaya, from the early nineteenth century. Being wealthy merchants of high social standing, the Peranakans not only had a high regard for English-medium education but, crucially, were one of the earliest and privileged few communities in Singapore to have access to it, and they sent their children, including girls – a rare occurrence in that era – to English-medium schools. By the early 1800s, members of the community had established four educational institutions in Malacca and Singapore that were especially important to the development of the community (Tan 1988b: 52). The establishment of the Queen's Scholarship in 1885 for British subjects in the Straits Settlements further enabled a few Peranakans to be educated in higher education institutions in Britain, producing scholars and leaders (Tan 1988b: 65, 82). Already in earlier days the Peranakans had been noted to speak English "tolerably well" (Earl 1837, cited in Tan 1988b: 50). By the mid-nineteenth century their ability to converse in this colonial language had strengthened their prominent socioeconomic position within other local communities in relation to the British, to the point where they were in fact sometimes referred to as the "King's Chinese" (Tan 1988b: 53), referring to the King of England.

In other ways they realigned themselves culturally, distinguishing themselves from the continuously increasing population of China-born immigrants by their local (Malayan) orientation and their pro-British sentiments (Tan 1988b: 54ff.). In their social clubs "to which they will admit no native of China … they [played]

billiards, bowls, and other European games, and [drank] brandy and soda ad libitum" (Tan 1988b: 54ff.). Not an uncommon observation then was for Peranakans "on being asked if they were Chinamen [to] bristle up and say in an offended tone 'I am not a Chinaman, I am a British subject'" (Vaughan 1879). Identifying politically with the British (Kwok 2000: 205), they formed the Straits Chinese British Association in August 1900, with an avowed aim to promote trade with, and foster loyalty to, the British Empire (Song 1923: 319).

3.   A study for the paradigms

As already alluded to above, the Peranakans and their languages bring not insignificant contributions to the understanding of language contact dynamics and evolution in both creole studies and World Englishes. They also comprise a rich case study in phenomena associated with language shift in the field of language endangerment. Further, the recognition of their emergent English variety as equally integral to their language repertoire also offers fresh insights for broader sociolinguistic studies. The following sections provide a synthesis of the contributions that have been recognised in different fields.

3.1   The persistence of founders

Baba Malay, in addition to having been valued within creole studies, has also more recently been recognised for its significance in the evolution of New Englishes (Lim 2009, 2011, 2014, 2016a). An observation of the prosody of Singapore English (SgE) reveals a systematic pattern of word- or phrase-final prominence – this, notably, contrasts with the prosodic patterning found in all other contact varieties in which tone has evolved, such as Nigerian English and Hong Kong English, where the general pattern locates High (H) tones on what would be stressed or accented syllables at word- or phrase-level, and Low (L) tones on unstressed ones. SgE's uncharacteristic prosody can be explained (a) if we consider the Founder Principle in the ecology paradigm (Mufwene 2001, 2008), viz., that the founder population in an ecology exerts a strong influence on features, an influence which persists in the emergent variety; and (b) if we recognise the Peranakans as a founder population in Singapore's ecology, one with significant economic and social prominence in this ecology. Structural analysis aligns with evolutionary and sociolinguistic facts. In the Peranakans' vernacular, Baba Malay, we find word- and phrase-final prominence; similar prosodic patterning of right-edge phrasal prominence is widespread in many Malay varieties (see e.g., Van Zanten & Goedemans 2009; also chapters in Gensler and Gil forthcoming). This feature also developed in the

Peranakans' emergent variety of English, Peranakan English (Lim 2011, 2014, 2016a), a result of language contact in the Peranakans' multilingual repertoire. The Peranakans, though a small minority, were clearly dominant in the ecology of Singapore, as outlined earlier, due to their political, economic and social status, and their position as intermediaries, and later as teachers. As early English adopters, crucially during the British colonial period, theirs would have been the early features influencing the emerging variety of SgE. Clearly, the Peranakans and their vernacular together have been significant for informing studies of language contact, language evolution, and cultural creolisation.

3.2   Multilingualism matters

In the discourse of language endangerment, one line of research has viewed the shift that has taken place from Baba Malay to English in current generations (particularly in Singapore) not as a loss, as in the traditional literature. Rather, long before the translingual turn in sociolinguistics took hold, recognition was made (Ansaldo 2009; Ansaldo et al. 2007; Lim 2016a) of the fact that communities such as the Peranakans – and the same is noted for other such communities in Asia, such as the Macanese and the Sri Lanka Malays (see Ansaldo 2009, 2010; Lim 2016b; Lim & Ansaldo 2007) – not only spoke their creole vernacular, but also, by virtue of their mixed heritage and their position in the colonial set-up, usually as mediators, were characterised by their multilingual repertoire. The languages shifted to are accepted by the community as legitimate and valued resources, and not seen as threats. For the Peranakan community in the early twenty-first century, such a shift is no longer even viewed as problematic. As is evident in the quote below, from an editorial in the newsletter of the Peranakan Association of Singapore (TPAS) (July/September 2002), the actual code used by the community is not the issue, and the Peranakan identity and identification are clearly and positively recognised, whatever language is used.

> We Peranakans have our own way or style of speaking that has become our trademark, which those outside the community recognise immediately, be it in English, Malay or Chinese. One Nyonya, for instance, tells me she is never surprised when people she meets for the first time straightaway say "Ah, you are Peranakan, right"? (TPAS 1994–2008)

This description explicitly reinforces the observation that it has been, and still is, the multilingual repertoire of the community that has been significant for their existence and cultural identity, from the colonial era to the present day, rather than any specific language that is critical to their vitality, whether ancestral or emergent (Ansaldo & Lim 2014, 2018; Lim 2016b).

3.3   The practice of postvernacularity, and beyond

The Peranakan situation in Singapore in the new millennium may be best understood, as first discussed in Lim (2016b), using the notion of *postvernacularity* (Shandler 2006) – situations where a language serves the purpose of identity-building within a community even after it has ceased to be used as a vernacular for daily communication. Postvernacularity has been receiving attention from several scholars in recent years, for example, with regard to Yiddish in the US (Shandler 2006), Low German in Northern Germany (Reershemius 2009), and Breton (Hornsby & Vigers 2013), where, for the most part, studies document the use of the postvernacular language in a number of cultural practices, such as amateur theatre, music and folklore, translation, and attempts to learn the language in evening classes, and in its primarily symbolic value the tendency to preserve only the language's most colourful or evocative elements.

In the Peranakan community in Singapore, this is evidenced in several areas.[2] A first such impetus has involved the documentation of the vernacular, Baba Malay, and its maintenance in cultural and ritual practices. These include publications such as a Baba Malay dictionary and a collection of Baba Malay idioms (Gwee 1993, 2006); plays written and performed regularly in Baba Malay; one cultural group's regular performance of traditional Baba Malay songs; and a church in the traditional Peranakan district of Katong holding services in Baba Malay for several years (until the Baba Malay-speaking priest passed away on 1 June 2013). A second thrust relates to symbolic revitalisation and reinvention of the vernacular, such as the TPAS's youth group holding regular gatherings that involve engaging in Peranakan cultural activities; Baba Malay classes started in 2017 by the cultural group Gunong Sayang Association, due to increased interest from younger Peranakans; and the use of Baba Malay in popular culture – in rap and hip hop – by a group who set out to modernise Peranakan entertainment, as a way to keep the heritage alive through innovation and to reach out to as many people as possible. And in a third area, there has been increased formal institutional recognition of the Peranakans as an important cultural group, and increased support by the state. This has resulted, for example, in the creation of a dedicated Peranakan Museum which primarily features traditional

---

2.   As noted in Lim (2014, 2016b), the Peranakan community in Singapore experienced an unprecedented surge in cultural vitality towards the end of the first decade of the twenty-first century, due to a combination of several factors, most crucially the fact that a new group of leaders were elected to the community's association, who, as relatively prominent individuals in Singapore, garnered significant state support, both economic and symbolic, as well as presence in local media; together these led to a level of cultural vitality in the community not seen in modern times.

artefacts and practices, and the restoration of the traditional ancestral home of a Peranakan family as a heritage house, opened in 2008 and 2009 respectively.

The case of the Peranakans develops theorising in postvernacularity even further, as discussed in Lim (2016b): what is observed is a maintenance – even an increase – in cultural vitality, going beyond the use of the vernacular for symbolic purposes, as in postvernacular practice, in the use of the new English variety. This is evidenced in several diverse initiatives. In a traditional showcase of the community's culture, a Sarong Kebaya Exhibition at the Peranakan Museum in 2011, the storytelling event was conducted in English. An exhibition of Peranakan artefacts comprising furniture, beaded and embroidered textiles and porcelain, curated by Singapore's Asian Civilisations Museum and hosted at the Musée du Quai Branly in Paris in 2010, was entitled *BabaBling* – using in its title the originally African American Vernacular English hip hop term *bling*, now mainstream. In the performing arts, recent compositions for the Peranakan community include songs composed in English like "*Bunga Sayang*" ("Flower of Love"), by local singer-songwriter Dick Lee, which has become the community's favourite contemporary tune and which, more significantly, is used as the theme song for the heritage showcase in Singapore Changi Airport's newest terminal. While acknowledged initially as a Peranakan song, "My Old Hometown, Katong", composed by Singapore a cappella group Vocaluptuous, was subsequently adopted as a Singapore song, and performed at significant events including the Asia-Pacific Economic Cooperation (APEC) meeting in Singapore in 2009 and the country's National Day Parade in 2013. It is obvious that the use of English in such contexts serves to create maximal inclusivity, affording a wider, more national and international audience (even if it may be noted that these examples comprise commercial activities). Nonetheless, the situation may be analysed as the action of a community no longer needing to be constrained by the ideology of purity, as noted in Lim (2016b), whereas in the past, for example, the desire for "true-blue" Peranakans who spoke "pure" Baba Malay used to be regularly expressed. In contrast, today Peranakan musical genres do not have to be restricted to traditional *bangsawan* (a form of traditional Malayan operatic theatre) and *dondang sayang* (love ballads) in Baba Malay; the contemporary compositions in English are just as – if not more – enthusiastically embraced and accepted by the community as the traditional ones. We also see here implications for commodification and branding, discussed in a later section of this chapter.

3.4 Maintenance of culture in the language of wider communication

A further dimension of a shift to a language of wider communication is the traditional argument that the loss of a language is usually accompanied by the diminishing of cultural diversity and loss of intangible cultural heritage (e.g., Hale et al. 1992), a discourse which receives its share of criticism (including e.g., Mufwene 2002,

2004). Indeed, this is not always the case. In some situations of language shift, the language of wider communication – such as English – with which the vernacular or ancestral language comes into contact, is, as noted by Woodbury (2005), at times adapted culturally, if not always structurally, to communicative ends which are continuous with those earlier fulfilled by the ancestral language. This is observed, for example, in Aboriginal communities in south-eastern Queensland, Australia (Eades 1988: 97, 101):

> While many Aboriginal people [in south-eastern Queensland] may speak English as their first language, the context of conversation has significant Aboriginal cultural and social aspects which lead to distinctively Aboriginal interpretations and meanings. While the chosen language code is frequently English, there are important continuities in the ways language is used. … The Aboriginal priority on developing, maintaining, and strengthening social relationships is both reflected in, and created by, the way people speak to each other, whether the language variety is English, Aboriginal English, or Lingo [any Aboriginal language].

Similarly, in Koyukon communities in Alaska's interior, even in the rapid shift from Koyukon to English, cultural patterns are transferred (Kwachka 1992: 70–71):

> The Koyukon people have been able to transfer and permute a very important cultural pattern at the discourse level, the tradition of narrative … Although [stories from a distant time] are rarely told today, the narrative, as a social and rhetorical structure, has not only persisted but flourished.

In the Peranakan community in Singapore, the language shifted to is a distinct contact variety of English, Peranakan English, first documented in Lim (2010). In written mode (based on a corpus of newsletters published by TPAS from 1994 to 2008), Peranakan English encompasses numerous lexical items of Baba Malay and Hokkien origin, conveying cultural elements (1a, b), naming and address practices (1c), and exclamations, greetings, wishes and thanks (1d, e), samples of which are shown in the examples below.

(1)   a.   *dondang sayang* [Malay]: love ballads, originating in Malacca in the fifteenth century, influenced by traditional Portuguese folk music, now a traditional form of entertainment for Malays and Peranakans involving a violin, two Malay *rebana* "drums" and a *tetawak* "gong", in which singers exchange Malay pantun "poetry" in lighthearted and sometimes humourous style.

b.   *biji saga seeds* [Malay *biji* "seed"]: small, hard, bright red seeds from the fruit pods of the red sandalwood, also known as the coral tree or saga tree, a deciduous tree found in tropical and subtropical regions of the world; the seeds are used as beads in jewellery, leis and rosaries; they were also used in ancient India for weighing gold; the word *saga* is traced to the Arabic term for "goldsmith".

    c.   *kimpoh choh* [Hokkien]: maternal great-grandaunt.

    d.   *May we extend to all readers a Selamat Tahun Baru* [Malay "happy new year"] *and may you all enjoy panjang panjang umor* [Malay "very long life"] *in the Year of the Goat.*

    e.   *The Main Wayang Company would like to say a big KAMSIAH* [Hokkien "thank you"].

It should be noted here that literacy in Baba Malay was not absent. The Peranakans – displaying their "highly eclectic, cosmopolitan" outlook – were one of the early and significant communities in the region involved in the contribution to Malay literary heritage, who, from the mid-nineteenth through the mid-twentieth century, sponsored and produced romanised Malay and English newspapers, hundreds of translations of epic Chinese classic literature in romanised Malay, and writings in both Malay and English in their association's *Straits Chinese Magazine*; books in Baba Malay comprised some 8% of all Malay-language books printed in the Straits Settlements between 1886 and 1920 (Clammer 1981; Leow 2016: 82; Proudfoot 1986: 108). In the twentieth century, even with the Singapore Peranakans' shift to English, Baba Malay literacy was still maintained in postvernacular practice, as discussed in Section 3.3. above.

    In spoken form, Peranakan English is even more clearly a single English-Baba Malay code (Lim 2010), as illustrated in example (2) (Lim 2010: 336).[3]

(2)   Like drugs you know when you're under drugs … The babies come out crying crying. *Apa dia mo? mo?* Drugs. *Nanti* alcohol the same. They get into their system. *Kita semua tak drin[k]* … Keep yourself clean and healthy. Don[t] drin[k] don[t] drin[k].

    ("Like drugs, you know, when you're under the influence of drugs … The babies are born crying. What do they [the babies] want? Drugs. Then it's the same with alcohol – it gets into your system. We all didn't drink … Keep yourself clean and healthy. Don't drink.")

As noted in Lim (2016b), it is clear that aspects of the ancestral culture involving address practices, food, certain cultural and religious practices, and terms of emotive import and value judgement are still transmitted in the contact variety. In other words, the evolution of Peranakan English, which in this case comprises what may be viewed as a mixed code, affords the maintenance and indexing of culture even if the ancestral language is no longer maintained. Furthermore, this

---

3.   The data presented in Lim (2010) are more English-dominant, since Peranakan English is the focus; in interactions between Peranakans in whom Baba Malay is more active, usually in the older generations, a higher proportion of Baba Malay elements in the mixed code is documented.

variety has the advantages of being one that has evolved for the immediate, current ecology, and of being a variety that is native to the younger generation: this means that it is in a good position, as similarly noted in the Koyukon situation discussed above, not only to survive, but to flourish. And indeed, the linguistic and cultural vitality being experienced by the Peranakan community has been embodied through all their language practices. This challenges the traditional wisdom in endangered languages literature about the link between language, identity and vitality of a community, with the language being viewed as an essential part of a community's cultural identity and heritage; on the contrary, an even stronger cultural vitality than before is observed in the Peranakans' emergent language.

3.5   Commodification and branding in centre-periphery dynamics

In the sociolinguistic scholarship on minority languages, in particular on centre-periphery dynamics, the commodification of authenticity, and the branding of heritage (e.g., Pietikäinen et al. 2016), the Peranakans have proved to have much to contribute, too. In the first decade of the new millennium, there began an evolution in the relationship between the Peranakans, as a peripheral community, and mainstream Singapore culture and identity, as described in Lim (2016b), involving the appearance of Peranakan presence in mainstream media, in the form of two locally produced television series which aired on local television channels in 2009. One of them, *Sayang Sayang*, was a sitcom for the local English television channel, and the other, *The Little Nyonya*, was a drama series produced for the local Chinese television channel. The fact that *The Little Nyonya* series was in Mandarin – one of Singapore's four official languages since the nation's independence, and the major lingua franca amongst the Chinese community since the mid-twentieth century – raised a furore in the Peranakan community because of the use of Mandarin by the Peranakan characters in the series, a language which was never in their repertoire. However, such programming actually meant that the series – and the entire existence of the Peranakans – reached three-quarters of Singapore's population, namely the Chinese-speaking majority, many of whom became acquainted with the culture for the first time through this series. This led, almost overnight, to the appreciation of Peranakan culture outside the Peranakan community, with non-Peranakans in Singapore embracing and consuming its cuisine and material culture. As noted by Lim (2016b), *The Little Nyonya* marked a milestone in terms of the Peranakan community moving into the mainstream.

Once in the mainstream, the value of the Peranakan heritage came to be recognised, and Peranakan culture came to be considered as representing Singapore culture. While they had previously been a marginal community in the territory (subsumed, as they are still, under the category "Chinese" in Singapore), the

Peranakans were proclaimed in 2010 by Singapore's Arts and Information Minister, Lui Tuck Yew, "multiracial emblems of [Singapore's] social mix" (cited in Simon 2010). Peranakan culture and performances are now showcased on the global stage as the essence of being Singaporean: they have been used to represent Singapore at international events, such as the APEC forum in 2009, and at the World Expo in Shanghai, China, in 2010, which featured Peranakan culture in the Singapore pavilion. Singapore's National Heritage Board's range of museum merchandise which "celebrates Singapore and what makes us unique" (TPAS 2013: 18) includes the use of designs from Peranakan ceramic tiles (commonly used in Singapore in the past to accentuate the architectural design of shophouses) in prints "to celebrate the intricate beauty and the unique identity of the Peranakan heritage" (TPAS 2013: 18).

4. Post-digital Peranakan practice

Technological advances, in particular digital media such as websites and social media platforms, have in the past two decades been recognised as affording a new domain for supporting the maintenance and revitalisation of minority and endangered languages. Indeed, as already noted in Lim (2010, 2014, 2016a), the era of Web 2.0 has seen a virtual Peranakan presence, in the virtual world of Second Life, and in the form of the TPAS's website and Facebook page.[4] What is notable is that, apart from cultural items and symbolic postvernacular use of the language akin to the patterns observed in Section 3.4 above, language practices in these virtual spaces are in English – this contrasts with the use of the vernacular in most minority or endangered language communities on such platforms. The TPAS's stated aim is "preserving and promoting Peranakan culture", with no explicit mention of Baba Malay; and most of the contributions on the Facebook Closed Group *We Facebook in Nonya-Baba Peranakan Patois* – even while it "seeks to chat in Peranakan patois" – were primarily in English.[5]

In this section then, we move on to the next obvious step in the scholarship on the Peranakans beyond what has been undertaken thus far, in two thrusts. First,

---

4. https://www.peranakan.org.sg and https://www.facebook.com/theperanakanassociation singapore/ (last access to both links on 16 December 2020).

5. The *We Facebook in Nonya-Baba Peranakan Patois* Group (https://www.facebook.com/ groups/NonyaBaba/) was created in May 2011, and this observation was made c. 2014 (dead link on 16 December 2020). In more recent years there have been no new posts to the group, and as of 2020 the group would appear to be defunct, with only one member remaining.

given the research traditions in diaspora studies and, more recently, transnational communities, we expand our investigation beyond the Singapore Peranakan community, and also examine the communities in the other traditional Peranakan heartlands of Melaka and Penang in Malaysia, as well as the transnational communities of Peranakans who emigrated and settled in new ecologies such as the Australian cities of Perth, Melbourne and Sydney. Such a scope will afford us a contemplation of the distinctions between "traditional homeland", "diaspora" and "transnational community". And as a window on how the twenty-first-century Peranakan lives, we look at Peranakan language practices and representations in social media. Further to this end, posts on Facebook pages/groups of the respective Peranakan associations, over a period of one year, from October 2017 through October 2018, are used as a corpus. A number of intriguing preliminary observations may be noted.

4.1   What's in a name?

The naming of things holds linguistic significance (see e.g., Mufwene 1985), as a performative indexing of affiliation. The names of the various Peranakan associations are indicative in this regard. As seen in (3a), the associations located in Malaysia use Baba Malay: not only is this the community's original vernacular, but Malay is also the official language of the state, and thus dominant and reinforced for those ecologies. Notably, however, the names of the Kuala Lumpur and Melaka associations immediately follow this with the English translation, while for the Penang association the Malay version only comes as the subtitle after the English name. All the other associations have their names in English, as seen in (3b).

(3)   a.   Persatuan Peranakan Baba Nyonya KL Selangor: Official Website of PPBNKLS (Website, n.d.)[6]

Persatuan Peranakan Cina Melaka (Straits Born Chinese Association) (Facebook group, created 2007)[7]

State Chinese Penang Association or Persatuan Peranakan Cina Pulau Pinang

b.   The Peranakan Association Singapore – TPAS (Facebook page and website; n.d.)[8]

Peranakan Association Australia NSW (Facebook page; Association founded in 2011)[9]

---

6.   https://www.peranakan-kl.org (Note that the website is no longer accessible).

7.   https://www.facebook.com/groups/6658837626/ (last access on 16 December 2020).

8.   https://www.facebook.com/theperanakanassociationsingapore/; https://www.peranakan.org.sg/

9.   https://www.facebook.com/PAANSWINC/ (last access on 16 December 2020).

> Melbourne Peranakans/ Peranakan Association Australia Inc
> (Facebook page, n.d.)[10]
>
> The Peranakan Community of Western Australia, INC (Facebook page,
> launched 2012)[11]

It is perhaps of little surprise that, at the outset, in the naming of the associations, the Peranakans' anglophone repertoire and Anglophile affiliation are clearly indexed.

4.2 Culture, not language

An examination of the 'About' section of the Facebook Groups or Pages is similarly revealing, as illustrated in examples (4a–f).

(4)   a.   Mission: We seek to promote and preserve the Peranakan heritage
           through engagement with TPAS members, the public, governmental
           and community organisations, and affiliated Peranakan associations
           throughout Asia. (Singapore)[12]

      b.   A place for people to share and promote the Peranakan Culture.
           (Melaka)[13]

      c.   … and association to cultivate and promote Peranakan culture in
           Sydney. (New South Wales [NSW])[14]

      d.   This is a social page book for Peranakans living in Australia, in
           particular, members of the Peranakan Association Australia Inc. based
           in Melbourne. (Melbourne)[15]

      e.   The Peranakan (Baba Nyonya) culture, its preservation, sharing it with
           the wider cosmopolitan community of Perth & WA and, carefully
           handing down … to share and to disseminate about our genteel
           colourful fused culture. (Western Australia)[16]

---

10.   https://www.facebook.com/babanyonyasMelbourne/ (last access on 16 December 2020).

11.   https://www.facebook.com/The-Peranakan-Community-of-Western-Australia-INC-173911992698690/ (last access on 16 December 2020).

12.   https://www.facebook.com/pg/theperanakanassociationsingapore/about/?ref=page_internal (last access on 16 December 2020).

13.   https://www.facebook.com/groups/6658837626/about/

14.   https://www.facebook.com/pg/PAANSWINC/about/?ref=page_internal

15.   https://www.facebook.com/pg/babanyonyasMelbourne/about/?ref=page_internal

16.   https://www.facebook.com/pg/The-Peranakan-Community-of-Western-Australia-INC-173911992698690/about/?ref=page_internal

> f. The Federation of Peranakan Associations exists to foster close relationships among the Peranakan associations of any country having Peranakan communities which are legally registered as an association in that country; whose activities are to preserve, promote and nurture Peranakan culture and traditions. (Singapore; website)[17]

As evidenced in these explicit statements, the emphasis in the aims of the associations is on the cultivation and promotion of Peranakan heritage or "culture"; while language is of course a dimension of culture, in these cases the reading of the term (based on the websites' content and reported activities) is that it emphasises culture, with little or no attention being paid to language.

4.3   Postvernacular practice

Following on from Section 4.2 then, it is perhaps not surprising to find little of the vernacular being used. As illustrated in the examples below, Baba Malay is only used, as is typical of postvernacular practice, for names and stock phrases.

(5)   a. and enjoy a warm fellowship and *makan* session. (NSW)

   b. *Buah Tangan* – If any member or guest would like to bring along a plate of snacks to add to our Afternoon Tea, you are very welcome to do so.

   c. We will also have a "*joget*" session. (NSW)

   d. As always, our afternoon tea provided a variety of *Nyonya* "*kueh kueh*" – *ondeh*, *apom balek* and *kueh cucor ikan bilis*, specially prepared by our fellow *bibi* and *nyonyas*. (NSW)

   e. Evelyn loves to dance and cook. She loves to teach members how to *joget* (dance) and *ikat sarong* (wrap sarong). (NSW)

The Baba Malay words are sometimes embedded in translingual practices, as with the words *makan* "to eat/ eating" in (5a), and *buah tangan* "fruit [of the] hand" ("a gift brought back from a trip or when visiting someone") in (5b), sometimes given in quotation marks, as with *kueh kueh* "cakes" in (5d), and *joget*, a traditional Malay dance form, in (5c) – of note is the elegance with which the explanations of the vernacular terms are provided in (5b) and (5d) – and sometimes with the English gloss provided in parentheses, as in *joget* (*dance*) and *ikat sarong* (*wrap sarong*) in (5e). Such use of the vernacular may be interpreted as the community conveying authenticity of practice.

---

17.   https://www.peranakan.org.sg/the-federation/ (last access on 16 December 2020).

4.4 Vernacular costs

The minor position that the Peranakans' vernacular plays within their communities' concerns is further exemplified in two dimensions. First, in the one year of corpus data examined, there is only one instance comprising the discussion of language per se – the result of the publication and promotion of a book on Baba Malay. Second, this topic was manifested in the most significant amount of the vernacular itself being used on social media, two examples of which are given in (6a) and (6b) (from the Melaka Association's Facebook page).[18]

(6) a. Jangan tunggu lama lama / Nanti jorang lupa moh chakap Baba (Melaka)

("Don't wait so long / Otherwise they will forget they want to speak Baba")

b. Mo chakap baba ni macam mahal? Mo chakap baba kena ari ari chakap. (Melaka)

("If you want to speak Baba it costs so much? If you want to speak Baba you have to speak it daily.")

In (6a) there is a barely disguised chiding of the community's lack of interest in speaking their vernacular. And in (6b), one of the comments in the thread that develops, a community member, in parallel and poetic form, bemoans the cost of the published book – perceived as being rather expensive – and by extension, lectures on the investment that should be made instead in speaking the language daily, as opposed to the cost of losing it.

4.5 The commodification of authenticity

Widely found throughout Peranakan social media is the discourse of authenticity, explicitly marked – though this applies primarily, even exclusively, to material culture, especially food. A typical example of this is shown in (7).

(7) For our Social Saturday Afternoon Tea, we have invited Christina from Christina's Gourmet to do a demonstration of preparing delicious authentic *Mee Siam*. (NSW)

Other instances of material culture are afforded value through their commodification. In particular, the Facebook page of the Melaka Association appears to be almost exclusively given over to posts advertising the sale of both personal

---

18. https://www.facebook.com/groups/6658837626/

collections and commercial products of Peranakan material culture such as traditional clothing, jewellery, crockery, cutlery and so forth, the text accompanying the images usually indexing authenticity with terms such as "original", "vintage", "unique", as shown in the examples in (8).[19]

(8) a. Vintage *kebaya*.   (Melaka)

b. Straits Chinese beaded (*manik potong*) shoes. In very good condition. The right pair in the 1st pic is unique where the swan is sewn with a 2d effect.   (Melaka)

c. Straits Chinese big spoons. In excellent condition. No chips nor cracks.
(Melaka)

4.6   Heritage branding

Another dimension involving the valuing of the culture is heritage branding, an area of interest not just in marketing and management, but also in diverse fields including history, archaeology, cultural studies and linguistics. Heritage branding in the Peranakan community is exemplified in the discourse found in the Singapore and Kuala Lumpur associations. Example (9a) comes from one of the monthly columns written by the TPAS's president (this one was written shortly after his election), where he explicitly identifies a "TPAS brand" – referring to the association's name – and the importance of the association's committee recognising and revisiting such a branding in their strategic plan.[20] Example (9b), from the Kuala Lumpur Association's website (there being no Facebook page for this association), comprises explicit instructions for interview requests: while there is no explicit mention of branding, implicit in this discourse is the suggestion that the community members' time and quality of knowledge of their heritage (points 1, 3, 4, 6) hold value, comprising a commodity which is to be remunerated (points 5, 7).[21]

(9) a. Letter from Baba Colin Chee, President, The Peranakan Association Singapore   (Singapore)

What's in a Brand?

Last Saturday, 28 July 2018, your new General Committee (GC) held our first Planning Retreat.

Together with several members from our choir, the Peranakan Voices and subcommittees, we revisited the TPAS brand

---

19.   https://www.facebook.com/groups/6658837626/

20.   https://www.peranakan.org.sg/2018/08/letter-from-baba-colin-chee-july-2018/ (last access on 16 December 2020).

21.   https://www.peranakan-kl.org (Note that the website is no longer accessible.)

.…

[TPAS Advisor and long-time GC member in past terms] Baba Peter [Lee] more than ably led our discussions on the TPAS brand and "mystique". It was a very fruitful session.

b. Interview Requests (Kuala Lumpur; website)

Members of the Peranakan association do grant interviews subject to their availability. Interviewers are required to:

1. Send us the reason/topic for the interview e.g. to know more about weddings or food or the family or *kebaya*. We like being prepared.

2. Bring all the equipment you need. Yes, it is okay to ask your interviewer to take your photo with his camera but do not complain if it isn't to your liking.

3. Prepare some questions to ask or topics to discuss. We love discussions but we are not lecturers who give one-way information.

4. Prepare by reading something about Peranakans so you can ask meaningful questions. http://en.wikipedia.org/wiki/Peranakan, http://en.wikipedia.org/wiki/Peranakan_cuisine.

5. Be prepared to treat your interviewee for a lunch or dinner at a simple restaurant or eating place as a way of saying thank you.

6. Always remind your interviewee of the appointment a few days before.

7. You will be required to give our association a copy of your submission paper and you must acknowledge your interviewee. E.g. "I would like to acknowledge Baba Pepper Lim and the PPBNKLS [Persatuan Peranakan Baba Nyonya KL Selangor] for their kind assistance in this project."

5. Experience like a Peranakan

Technopreneurs tell us we are already living in the future (e.g., Vermeulen 2017), where it is no longer the digital or information age, but the experience age[22] (the experience economy already having been predicted decades before; see

---

22. While some may not see these eras as exclusive, it is widely recognised that now, three decades after the introduction of the World Wide Web, the ubiquity of the internet and smart mobile devices and connectivity in the experience age does mean there is a difference in the online experience in, inter alia, how we present ourselves and how we communicate.

Pine & Gilmore 1998).[23] In this chapter, we have probed what it means to live – to experience – like a Peranakan in the current post-digital age, for both communities in the traditional homelands as well as those of the diaspora, and, collectively, as a transnational community, as they perform their linguistic practices – and consume Peranakan culture – on social media.[24]

We turn first to the question of whether a distinction may be observed between the communities considered those of the "traditional homeland" (in Melaka and Penang, as well as Singapore), and the communities of the "diaspora" such as those found in the Australian cities, or whether the Peranakans may be considered a "transnational community". In fact, the language practices of the various communities would appear to be quite comparable. As argued previously (e.g., Ansaldo 2009, 2010; Lim 2016b), for such creole cultures as the Peranakans – and also including groups such as the Macanese, and the Sri Lanka Malays – who settled and evolved in a territory distinct from that of their ancestors, it is both the maintenance of cultural aspects of their origins, together with their assimilation of new cultural features in their new territory, that together define their new identity, with both dimensions needing to continue to coexist in their evolved culture. This may be viewed as an expression or a resolution of the tendencies to change and to maintain, both natural in the evolution of language: there is maintenance to continue to signify the origin; there is change, or transfer from other models of language and culture, as part of the transition and adaptation.

Turning to the digital world, there has been a plethora of research and reports in recent years on the positive contribution of social media as a platform to support the maintenance and revitalisation of minority and endangered languages (e.g., Outakoski et al. 2018). In contrast, there appears to be little need or desire on the part of the Peranakans to use this platform for such a purpose, that is, for engaging in their ancestral language of Baba Malay. But this is not an unforeseen

---

23. In their 1998 article in the *Harvard Business Review*, Joseph Pine II & James H. Gilmore outlined what they saw as the next economic era following on from the agrarian, industrial and service economies of the past, coining for this the phrase "the experience economy", where, the authors argued, businesses would sell memories, i.e., their products would be experiences. The point here is that this may be seen as a defining feature of the current economic era, but one which does not preclude the existence of various other dimensions of the economy, such as the manufacturing and service economies.

24. While the "experience economy" is usually understood to refer to businesses' commodification of "traditions", we have nonetheless not found it inappropriate to use this concept to understand the current state of evolution of the Peranakans' existence, since their representation on social media has as audience both the community itself as well as the non-Peranakan consumer, as suggested at the very outset of this chapter.

outcome: after all, as was already strongly concluded in Lim (2016b) in regard to the Singapore Peranakans, and discussed above, it would appear that the majority of the Peranakan communities have also already moved beyond the situation of postvernacularity; rather, it is (Peranakan) English that is their vernacular now, one which fulfils the purposes of cultural representation and identification in their present ecologies, in both the real and virtual worlds, and through which the strong cultural vitality that is observed in all the associations is consistently manifested.

Authenticity is nonetheless still an issue, even if the discourse of purity and a regard for the "true-blue Peranakan" have not been on the community's agenda for decades. Notwithstanding the reality of (Peranakan) English being the community's vernacular in today's ecology, authentification – the performative discursive construction of authenticity (Coupland 2010: 6) – is observed not just in the recognition and explicit acknowledgement of what are considered authentic dimensions of material culture; it is also manifested in the use of Baba Malay terms and phrases in English, i.e. the mixed Peranakan English code (depending on one's theorisation of multilingualism and mixing), deployed as a resource for indexing authenticity.

Authenticity is often closely linked to the notion of place, in that geographical contexts in which languages were "born" tend to be the places where the most authentic languages are generated, conserved and regarded. The meaning and performance of (linguistic) authenticity in the various Peranakan communities – in their various traditional homelands (i.e., in Southeast Asia), in delocalised settings (i.e., in the geographical diaspora in Australian cities, in this case), and in deterritorialised settings (i.e., in non-geographical loci of the Web 2.0) – would appear to be comparable (the structural variation amongst the Baba Malay and English varieties notwithstanding). This reinforces the point made above about whether there is any distinction to be made amongst these ecologies; further research in this area would be instructive.

Finally, the authentification observed leads us to the question of whether the role of such linguistic practice is for the community's identification – crucial in the deterritorialised digital ecology – or whether, in the globalised new economy, culture, identity and language are invested with value as a source of profit (à la Duchêne & Heller 2012), already evident, as outlined in Sections 3.3 and 3.5. While it is clear that the Peranakans still carry pride in a collective identity, the community has also recognised an element of added value, especially in the current experience economy, that contributes to the maximisation of their cultural capital, both locally within their geographical territory (especially in the Singapore context), and increasingly in the ecology of the virtual, global marketplace.[25]

---

25. See also the Celtic situation in e.g., Brennan & Costa (2013); Tredinnick-Rowe (2017).

In the pre-colonial and colonial eras in Southeast Asia, the Peranakans were the original native compradors, cultural brokers and communicative pioneers, positioning themselves in the ecology so as to maximise their cultural and linguistic resources. In this post-digital experience ecology, the authentification, commodification and heritage branding we observe in the Peranakans' linguistic practice may be interpreted as a contemporary manifestation of the same initiative and positioning, elegantly and successfully adapted for today's ecologies, whether traditional, delocalised or virtual.

References

AccorHotels. 2016. The Peranakans in Singapore: 5 ways to live like one. *The Magazine*. <https://www.accorhotels.com/gb/thailand/magazine/one-hour-one-day-one-week/live-like-a-peranakan-in-singapore-e9116.shtml> (23 January 2019).

Ansaldo, U. 2009. *Contact Languages: Ecology and Evolution in Asia*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511642203

Ansaldo, U. 2010. Identity alignment and language creation in multilingual communities. *The Native Speaker and the Mother Tongue*. *Language Sciences* 32: 615–623. https://doi.org/10.1016/j.langsci.2010.08.003

Ansaldo, U. & Lim, L. 2014. The lifecycle of Sri Lanka Malay. *Language Documentation and Conservation* 7: 100–113.

Ansaldo, U. & Lim, L. 2018. Citizenship theory and fieldwork practice in Sri Lanka Malay communities. In *The Multilingual Citizen: Towards a Politics of Language for Agency and Change*, L. Lim, C. Stroud & L. Wee (eds), 193–220. Bristol: Multilingual Matters. https://doi.org/10.21832/9781783099665-012

Ansaldo, U., Lim, L. & Mufwene, S. S. 2007. The sociolinguistic history of the Peranakans: What it tells us about "creolization". In *Deconstructing Creole* [Typological Studies in Language 73], U. Ansaldo, S. Matthews & L. Lim (eds), 203–226. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.73

Ansaldo, U. & Matthews, S. J. 1999. The Minnan substrate and creolization in Baba Malay. *Journal of Chinese Linguistics* 27: 38–68.

Braddell, T. 1853. Notes of a trip to the interior from Malacca. *Journal of the Indian Archipelago and Eastern Asia* 7: 73–104.

Brennan, S. & Costa, J. 2013. Celtic commodified: The role of minority languages in Scottish and Irish national celebrations. *Babel: Littératures Plurielles* 27(8): 87–104.

Clammer, J. R. 1981. Straits Chinese literature: A minority literature as a vehicle of identity. In *Essays on Literature and Society in Southeast Asia: Political and Sociological Perspectives*, S. C. Tham (ed.), 287–302. Singapore: Singapore University Press.

Coupland, N. 2010. The authentic speaker and the speech community. In *Language and Identities*, C. Llamas & D. Watts (eds), 99–112. Edinburgh: EUP.

Cramer, F. 2014. What is "post-digital"? *Post-Digital*. <http://www.aprja.net/what-is-post-digital> (23 January 2019). https://doi.org/10.7146/aprja.v3i1.116068

Davies, R. 2009. Meet the new schtick. <https://russelldavies.typepad.com/planning/2009/01/meet-the-new-schtick.html> (23 January 2019).

Duchêne, A. & Heller, M. (eds). 2012. *Language in Late Capitalism: Pride and Profit*. New York NY: Routledge. https://doi.org/10.4324/9780203155868

Eades, D. 1988. They don't speak in an Aboriginal language, or do they? In *Being Black: Aboriginal Cultures in "Settled" Australia*, I. Keen (ed.), 97–115. Canberra: Aboriginal Studies Press.

Earl, G. W. 1837. *The Eastern Seas or Voyages and Adventures in the Indian Archipelago in 1832-33-34*. London: Allen & Co.

Gensler, O. & Gil, D. (eds). Forthcoming. *Malay/Indonesian Linguistics*. London: Curzon Press.

Gwee, T. H. W. 1993. *Mas Sepuloh: Baba Conversational Gems*. Singapore: Armour.

Gwee, T. H. W. 2006. *A Baba Malay Dictionary*. Singapore: Tuttle.

Hale, K., Krauss, M. & Watamohigie, L. J. 1992. Endangered languages. *Language* 68: 1–42. https://doi.org/10.2307/416368

Ho, M. L. & Platt, J. 1993. *Dynamics of a Contact Continuum: Singaporean English*. Oxford: Clarendon Press.

Hornsby, M. & Vigers, D. (eds). 2013. *Breton: The Postvernacular Challenge.* Special issue, *International Journal of the Sociology of Language* 223.

Kwachka, P. 1992. Discourse structures, cultural stability, and language shift. *International Journal of the Sociology of Language* 93(1): 67–73. https://doi.org/10.1515/ijsl.1992.93.67

Kwok, K. W. 2000. Singapore. In *The Encyclopedia of the Chinese Overseas*, L. Pan (ed.), 200–217. Cambridge MA: Harvard University Press.

Lee, N. 2018. Baba Malay: Diverging trends in two ecologies. *Journal of Pidgin and Creole Languages* 33: 136–173. https://doi.org/10.1075/jpcl.00004.lee

Leow, R. 2016. *Taming Babel: Language in the Making of Malaysia*. Cambridge: CUP. https://doi.org/10.1017/CBO9781316563007

Lim, B. K. 1917. The Chinese in Malaya. In *Present Days' Impressions of the Far East and Prominent and Progressive Chinese at Home and Abroad: The History, People, Commerce, Industries and Resources of China, Hong Kong, Indo-China, Malaya and Netherlands India*, W. Feldwisk (ed.). 875–882. London: Globe Encyclopedia Co.

Lim, L. 2009. Revisiting English prosody: (Some) New Englishes as tone languages? *English World-Wide* 30: 218–239. https://doi.org/10.1075/eww.30.2.06lim

Lim, L. 2010. Peranakan English in Singapore. In *The Lesser-Known Varieties of English: An Introduction*, D. Schreier, P. Trudgill, E. W. Schneider & J. P. Williams (eds), 327–347. Cambridge: CUP. https://doi.org/10.1017/CBO9780511676529.018

Lim, L. 2011. Tone in Singlish: Substrate features from Sinitic and Malay. In *Creoles, their Substrates and Language Typology* [Typological Studies in Language 95], C. Lefebvre (ed.), 271–287. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.95.16lim

Lim, L. 2014. Yesterday's founder population, today's Englishes: The role of the Peranakans in the (continuing) evolution of Singapore English. In *The Evolution of Englishes. Varieties of English Around the World* [Varieties of English Around the World G49], S. Buschfeld, T. Hoffmann, M. Huber & A. Kautzsch (eds), 401–419. Amsterdam: John Benjamins. https://doi.org/10.1075/veaw.g49.23lim

Lim, L. 2016a. Multilingual mediators: The role of the Peranakans in the contact dynamics of Singapore. In *Multilingualism in the Chinese Diaspora World-Wide*, L. Wei (ed.), 216–233. London: Routledge.

Lim, L. 2016b. The art of losing: From *java* and *patois* to post-vernacular vitality –Repositioning the periphery in global Asian ecologies. In *Endangered Languages and Languages in Danger: Issues of Ecology, Policy and Documentation* [Impact: Studies in Language, Culture and Society 42], M. Pütz & L. Filipović (eds), 283–312. Amsterdam: John Benjamins. https://doi.org/10.1075/impact.42.12lim

Lim, L. & Ansaldo, U. 2007. Identity alignment in the multilingual space: The Malays of Sri Lanka. In *Linguistic Identity in Postcolonial Multilingual Spaces*, E. A. Anchimbe (ed.), 218–243. Cambridge: CUP.

Lim, S. 1988. Baba Malay: The language of the "Straits-born" Chinese. In *Papers in Western Austronesian Linguistics* 3, H. Steinhauer (ed.), 1–61. Canberra: The Australian National University.

Liu, G. 1999. *Singapore: A Pictorial History 1819-2000*. Singapore: National Heritage Board & Editions Didier Miller.

Mufwene, S. S. 1985. The linguistic significance of African proper names in Gullah. *Nieuwe West-Indische Gids /New West Indian Guide* 59: 149–166.
https://doi.org/10.1163/13822373-90002068

Mufwene, S. S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP.
https://doi.org/10.1017/CBO9780511612862

Mufwene, S. S. 2002. Colonization, globalization and the plight of "weak" languages. *Journal of Linguistics* 38: 375–395. https://doi.org/10.1017/S0022226702001391

Mufwene, S. S. 2004. Language birth and death. *Annual Review of Anthropology* 33: 201–222.
https://doi.org/10.1146/annurev.anthro.33.070203.143852

Mufwene, S. S. 2008. *Language Evolution: Contact, Competition and Change*. New York: Continuum.

Nathan, J. 1922. *The Census of British Malaysia, 1921*. London: Waterloo & Sons.

Negroponte, N. 1998. Beyond digital. *Wired*. <https://www.wired.com/1998/12/negroponte-55/> (12 January 2019).

Outakoski, H., Cocq, C. & Steggo, P. 2018. Strengthening indigenous languages in the digital age: Social media-supported learning in Sápmi. *Media International Australia* 169(1): 21–31. https://doi.org/10.1177/1329878X18803700

Pakir, A. 1986. A Linguistic Investigation of Baba Malay. PhD dissertation, University of Hawaiʻi.

Pietikäinen, S., Jaffe, A., Kelly-Holmes, H. & Coupland, N. 2016. *Sociolinguistics from the Periphery: Small Languages in New Circumstances*. Cambridge: CUP.
https://doi.org/10.1017/CBO9781316403617

Pine, J. II & Gilmore, J. H. 1998. Welcome to the experience economy. *Harvard Business Review*. <https://hbr.org/1998/07/welcome-to-the-experience-economy> (12 January 2019).

Proudfoot, I. 1986. A formative period in Malay book publishing. *Journal of the Malaysian Branch of the Royal Asiatic Society* 59: 101–132.

Reershemius, G. 2009. Post-vernacular language use in a Low German linguistic community. *Journal of Germanic Linguistics* 21: 131–147. https://doi.org/10.1017/S1470542709000221

Reid, A. 2000. *Charting the Shape of Early Modern Southeast Asia*. Chiang Mai: Silkworm Books.

Rudolph, J. 1998. *Reconstructing Identities: A Social History of the Babas in Singapore*. Aldershot: Ashgate.

Shandler, J. 2006. *Adventures in Yiddishland: Postvernacular Language and Culture*. Berkeley CA: University of California Press.

Simon, C. 2010. Baba Bling: Domestic signs of prosperity in Singapore – Review. *Guardian Weekly*. <https://www.theguardian.com/artanddesign/2010/oct/19/baba-bling-prosperity-singapore-review> (19 October 2010).

Song, O. S. 1923. *One Hundred Years' History of the Chinese in Singapore*. London: John Murray. (Reprinted 1967, Singapore: University of Malaya Press).

Tan, C. B. 1988a. Structure and change: Cultural identity of the Baba of Melaka. *Bijdragen tot de Taal, Land- en Volkenkunde* 144: 297–314. https://doi.org/10.1163/22134379-90003298

Tan, C. B. 1988b. *The Baba of Malacca: Culture and Identity of a Peranakan Community in Malaysia*. Petaling Jaya, Selangor: Pelanduk Publications.

TPAS (The Peranakan Association of Singapore). 1994-2008. *The Peranakan*. Singapore: The Peranakan Association.

TPAS (The Peranakan Association of Singapore). 2013. *The Peranakan*. Singapore: The Peranakan Association.

Tredinnick-Rowe, J. 2017. The semiotics of minority language branding: A study of the Celtic languages. *International Journal of Marketing Semiotics* 25: 2–16.

Thomson, J. T. 1875. *The Straits of Malacca, Indo-China and China, or Ten Years' Travels, Adventures and Residence Abroad*. London: Sampson Low; Marston: Low & Searle.

Van Zanten-Wervelman, E. A. & Goedemans, R. 2009. Prominence in Indonesian: Stress, phrases, and boundaries. *Wacana. Jurnal Ilmu Pengetahuan Budaya* 11(2): 197–225.

Vaughan, J. D. 1879. *Manners and Customs of the Chinese in the Straits Settlements*. Singapore: The Mission Press. (Reprinted by Oxford University Press, 1971, 1992).

Vermeulen, E. P. M. 2017. We are already living in the future: Why everyone must understand the digital age now. *Hackernoon* <https://hackernoon.com/we-are-already-living-in-the-future-c69443606805> (12 January 2019).

Woodbury, A. C. 2005. Ancestral languages and (imagined) creolisation. In P. K. Austin (ed.), *Language Documentation and Description*, Vol. 3, 252–226. London: SOAS.

Detecting loan words computationally

Liqin Zhang, Franz Manni, Ray Fabri & John Nerbonne
Open University, Netherlands; National Museum of Natural History, Paris;
L-Università ta' Malta; University of Groningen & University of Freiburg

A loanword is a word that is borrowed from one language and adopted into another; examples are the English words *toboggan, skunk,* and *hickory*, all of which were borrowed from Algonquian languages. Among languages that are not (closely) related, loan words are recognizable because they are semantically related and are more similar in pronunciation than one would expect by coincidence. This chapter applies techniques for measuring pronunciation similarity, focusing on edit-distance measures and a sound-class based method. The novel issue in loan-word detection is the circumstance that loan words are normally modified to fit the phonology of the borrowing language, meaning that sensitivity in measuring pronunciation similarity may be deprecated.

Keywords: loan words, automatic detection, edit distance, sound class alignment, language contact

> Neither a borrower nor a lender be;
> For loan oft loses both itself and friend–
> *Hamlet*, Act 1, Scene 3

1. Introduction

Loan words can provide evidence of social, cultural, commercial or other contact, e.g., when we note that Germanic languages owe their words for paved roads to Latin *via strata* "way paved" (< Eng. *street,* Germ. *Straße,* Dutch *straat*), suggesting that not just the word, but also the infrastructural innovation was copied from the Romans. Those trying to reconstruct earlier, undocumented stages of languages must at times wish that languages had heeded Polonius's advice to Hamlet, and avoided borrowing altogether, since loan words confuse the historical record, normally suggesting closer phonetic similarity than is actually warranted. This chapter wishes to contribute to the automatic detection of loan words in non-related

languages, and thereby to the study of language contact which Salikoko Mufwene has so greatly advanced.

The basic idea is simple: if two words from unrelated languages mean roughly the same thing and are similar in pronunciation, then the chances are that one has been borrowed from the other's language. It would be too great a coincidence for the similarity to arise by chance. This very simple characterization reveals, too, that our detection will be symmetric. Words that are semantically and phonetically similar may be classified as involving a borrowing, but we will not attempt to say which language borrowed from the other – or indeed, whether the two languages borrowed from a third.

The restriction to focusing on loan word detection in non-related languages is important because, while loans from related languages may be semantically and phonetically similar, non-loans may also be semantically and phonetically similar due to their historical relatedness. Words are certainly borrowed from related as well as from unrelated languages: the English word *skirt* is a borrowing from the related Old Norse *skyrta* "shirt", arising from the same older common Germanic word that survived in English *shirt*. Note that *skirt* is phonetically more similar to *skyrta* than *shirt* is, but that *shirt* is more similar in meaning to the hypothetical source. The sorts of procedures we examine below would be unlikely to distinguish these two words since both are phonetically similar to the putative Old Norse source, as well as being semantically related. The procedures would not distinguish the two English words well, even though only one of them is a genuine borrowing. In focusing on borrowings from unrelated languages, we avoid this problem. By working on the simpler problem we hope to make progress more likely. Naturally, an approach that detects borrowings without such a restriction will be superior in the long term.

We will operationalize this idea by examining the results of fieldwork carried out in Central Asia (Mennecier et al. 2016). The informants were speakers of either Turkic languages or Indo-Iranian languages, two unrelated language families. The informants were asked to verbalize concepts found in the 200-word Swadesh list (Swadesh 1952), and their answers were recorded acoustically and transcribed in IPA (1999). Reacting to the same concept will be interpreted here as indicating semantic similarity, and we will examine several ways of estimating phonetic similarity based on phonetic transcriptions, focusing on finding the best algorithm for detecting the pronunciation similarity in loan words.

2. Mufwene's perspective

In a series of publications Mufwene has urged that we view linguistic evolution by analogy to biological evolution (Mufwene 2005), but Mufwene more particularly

encourages the view that languages are analogous to species consisting of many, often very different, individual organisms, just as languages may be viewed as collections of many very different idiolects, and he emphasizes the essential role of the environment in understanding how these populations of idiolects develop. This view is sympathetic to that of linguistic variationism, emphasizing the individual variation normally ignored when one compares only the enormous abstractions, the languages. Mufwene (2005) suggests, e.g., that the process of acquisition should not be thought of as a direct transmission from a parent-caretaker to a child, but rather as an active construction of an idiolect by the child, based on the many varieties it is exposed to, emphasizing "the piecemeal way in which speakers develop competence" (Mufwene 2005: 37). The emphasis is on the enormous variation the learner is confronted with, which leads him or her to select linguistic elements, sometimes in parallel, especially in large, heterogeneous societies.

In emphasizing Mufwene's evolutionary perspective we have elsewhere encountered the objection that modern linguists see little or no evidence of primitive languages from which more sophisticated ones have evolved. The objection is correct: evolutionary progress in this sense is not postulated, nor is it necessary to an understanding of how Mufwene sees languages changing.[1] Instead the emphasis is on how language changes in response to its ecology, and in particular, to the other languages in use nearby, where loan words are strikingly interesting. Clearly languages adopt elements from other languages and thereby "adapt" – socially and culturally – to their environments.

Mufwene (2001) applies ideas on population genetics to the analysis of colonial varieties, noting that variation in languages, just as in gene pools in biological populations, is likely to be reduced when a relatively small sample of speakers emigrates to a colony. Succeeding generations in the isolated colony normally have to select their variants from the reduced pool, maintaining (the tendency towards) the initial selections. The so-called *founder effect* in population genetics (Jobling et al. 2004) thus has an analogue in linguistics: early immigrants have an inordinate influence on the populations they engender.[2] Shackleton (2010) traces New England speech to East Anglia, and Virginian speech to the south-west of Great Britain. The largest overall similarities were found between these two pairs of areas.

---

1.   Indeed, evolutionary theorists have often emphasized that this view is too simplistic. Ernst Mayr emphasizes how sophisticated species have sometimes been lost, and how many more primitive forms of life exist than sophisticated ones (Mayr 2002).

2.   See <http://www.blackwellpublishing.com/ridley/a-z/Founder_effect.asp> (last access on 16 December 2020) for an explanation with animation.

This view thus seems congenial with respect to the acquisition and the study of colonial varieties, but it is in the study of pidgin and creole languages where it has been found most convincing. Since pidgin and creole languages arise in a multilingual ecology, it is only natural for their speakers to adopt elements from the different languages they hear. The better established pidgin and creole varieties therefore contain elements selected from their original multilingual environment and transmitted to later generations of language learners.

Mufwene's perspective resonates with the research line from which the present chapter arises for several reasons, so that we think it is more than just another instance of the "contagion of ideas" (Sperber 1996). Most importantly, we detect resonances because Mufwene's theoretical perspective is best served by a quantitative methodology of the sort we employ here. Given that languages are extremely variable, comparisons intended to establish genealogical or areal relations (those arising from contact) must be based on large, representative samples analyzed statistically. Finding a feature *f* in a given variety (a sound, a word, an inflectional affix, or a grammatical construction) that is also found in another, potentially influential variety is always a striking observation. But given that languages consist of dozens of sounds, tens of thousands of words, (often) dozens of morphological elements and hundreds of grammatical constructions, it is incumbent on those wishing to demonstrate a genuine relationship to show that elements have indeed been transmitted from earlier varieties or from others in areal contact. This can be done if large quantities of data are analyzed, preferably from large numbers of varieties. This is exactly our tack in this chapter.

Less importantly, Mufwene's perspective shares a good deal with the variationist programme from which our own work has proceeded. As noted above, his work assumes that there is a great deal of linguistic variation, not only among the dialects or varieties of a language, but also within those. This has consistently been our experience (Nerbonne 2009). Finally, like Mufwene, we have been active in promoting the collaboration between population genetics and (variationist) linguistics (Manni 2017; Manni et al. 2006), e.g., in showing that the number of loan words is proportional to social contact in neighbouring populations (Mennecier et al. 2016).

3.   Previous work

Greg Kondrak has worked regularly on the task of "cognate identification", not only in machine translation (Kondrak et al. 2003) but also in historical linguistics (Kondrak & Sherif 2006). In Kondrak's work on machine translation, "cognates" include what we call loan words, so there is a fairly direct connection. He has compared both linguistically inspired methods, such as the ones we focus on here, but also sophisticated machine learning methods, such as pair Hidden Markov

Models and dynamic Bayesian Networks, with the latter tending to be more accurate, reducing error by approximately 10% (Kondrak & Sherif 2006; Wieling et al. 2007). Our work differs in focusing on what linguists regard as loan words, rather than "cognates", but it is clear that the problem of detecting loan words is quite similar to that of detecting cognates in historical linguistics and to detecting "cognates" in the broader sense of machine translation.

We are also aware of work done in linguistic phylogenetics (Delz 2013) and ancestral state reconstruction (Köllner & Dellert 2016), which undertakes extensive historical reconstruction in order to classify words as borrowings (or non-borrowings). These approaches have the advantage of attempting to detect loans on the basis of language history and may check the plausibility of a native (non-loan) source of a word, but they also require that the language histories be reconstructed. In this chapter we attempt to avoid that step by checking directly for unexpected semantic and phonetic similarity in the synchronic data. Our approach yields less information than the others, but, being less ambitious, may also be more feasible. There are undoubtedly studies in which identifying loan words is itself interesting, even without an account of the entire history of the languages involved.

Finally, Johannes Dellert has applied causal inference techniques to the problem of cognate detection (Dellert 2017, 2018), combining information about pronunciation and semantic similarity with (induced) models of language relatedness. In contrast to the methods presented in this chapter, Dellert's techniques do not require that data be hand-annotated (in order to set a threshold). We leave it to future work to examine his ideas more thoroughly.

4. Data

Mennecier et al. (2016) conducted a survey to explore the language variety of a Central Asian region and then utilized the data to measure the relatedness of languages and to attempt to detect loan words. The data, which are documented and publicly available for the study of loan word detection in this experiment, were collected from twenty-three sites in three Central Asian countries, namely Uzbekistan, Kyrgyzstan and Tajikistan.[3] The sites were chosen for their "complex human and linguistic geography". There were eighty-eight informants from the

---

3. See Mennecier et al. (2016) *Language Dynamics and Change* 6(1): Supplementary Materials, Table S2. <http://booksandjournals.brillonline.com/content/journals/10.1163/22105832 -00601015> (accessed on 24 May 2018, but dead link on 16 December 2020) or Github: <https://github.com/jayliqinzhang/computational-loanword-detection> (4 Jun 2020, last access on 16 December 2020)

three countries. For reasons having to do with genetics, males over forty years old were preferred. Linguistic and genetic sampling proceeded in parallel in order to examine the linguistic and genetic histories of the peoples, and in particular to see whether genetic commonalities were paralleled by similarity in culture (language). This aspect of the work will be reported on separately. The native languages of the informants were Kazakh, Kyrgyz, Karakalpak, Uzbek, Tajik and Yaghnobi, all languages from two language families, Turkic and Indo-Iranian (see Table 11.1). The informants also understood Russian well, since they all went to school during the time of the Union of Soviet Socialist Republics (USSR).

Table 11.1 The languages examined in the study

Turkic	Indo-Iranian
Kazakh	Tajik
Karakalpak	Yaghnobi
Kyrgyz	
Uzbek	

A 200-word extended Swadesh list was first reduced to 178 words, eliminating words that were polysemous or too difficult to understand in the interview context. The reduced list was then presented in Russian to the informants, who were asked to translate the words in the list orally into their native languages. It is clear that presenting the words in Russian introduces a bias in that it facilitates responses involving Russian loan words. The fieldworker attempted to probe in such cases, but one must be aware of the potential for bias. Each word in the Swadesh list represents a concept. The pronunciations were digitally recorded and catalogued in phonetic transcriptions. In total, each informant was asked to produce 178 pronunciations, resulting in more than 15,000 recordings. There are therefore approximately eighty-eight phonetic transcriptions for each Russian word representing a concept, even if some could not be used, for example, when an informant did not pronounce a word clearly enough for transcription. Philippe Mennecier transcribed all of the data.

An expert classification of the words into cognate classes is available in the data set as well, which will make our evaluation straightforward (see below). Within a concept, each pronunciation is marked with a code, so that the pronunciations with identical codes are designated cognates. Hence, a word from a Turkic language (or an Indo-Iranian one) bearing the same cognate designation as another word in the Indo-Iranian family (or, respectively, the Turkic family) means that one of the words is a loan word. Notably, it is common that pronunciations of a concept in one language family are assigned to different cognate classes because there are multiple

languages in a language family, and because informants may know multiple ways to translate a Russian word representing a concept. Besides, each phonetic transcript is coded according to its original language or language family. The pronunciations in the data set originated from Turkic, Iranian, Arabic and Russian.

Example data

To illustrate the procedure more concretely, we provide a sample of the data in Table 11.2. Our procedure will compare all the pairs in T×I, where T is the set of forty-nine Turkic pronunciations and I the set of thirty-nine Indo-Iranian ones. Both in the case of "one" (first column) and in the case of "three" (right-most column), the last few pronunciations, involving [i:] as "one" and [traj] etc. as "three", suggest that a novel lexical item has entered the Indo-Iranian varieties. But note that in neither of these cases do we find closer similarity to the Turkic realizations.

Table 11.2 A sample of the data used in comparing pronunciations in order to detect borrowings. The diagonal slash is introduced to separate pronunciations from different sampling sites. The double slash in the first row of the Turkic pronunciations of "one" indicates missing data.

	"one"	"three"
Turkic	bır / brıw / bırıw / / brıw / brıw / bır / bər / bər / bər / bər / bir / bir / bir / biṟ / biṟ / bir / bir / bir / bir / bır / bər / bıʃ / biʃ / bir / bər / bər / bər / bər / bər / bər / bər / bər / bəṟ / bər / bər / bır / ber / bər / bır / bər / bər / bər / bır / bır / bır / bır / bır / bər /	ɥʃ / ɥʃ / ɥʃjʊː / ɥʃ / ɥʃ / ʃuː / ɥʃ / yʃ / ytʃ / øʃ / ɥʃ / ytʃ / ytʃ / ytʃ / ytʃ / ytʃ / ytʃ / ytʃ / ytʃ / ytʃ / ɥtʃ / ytʃ / ytʃ / ytʃ / ytʃ / ɥʃ / ɥʃ / ɥʃ / ytʃ / ɥʃ / ytʃ / ɥʃ / ɥʃ / ɥʃ / ɥʃ / ɥʃ / ytʃ / ɥtʃ / ɥʃ / utʃ / ytʃ / ytʃ / ytʃ / ɥtʃ / uʃta / ɥtʃ / ɥtʃ / utʃ / utʃ /
Indo-Iranian	jak / jak / jak / jakta / jak / jak / jak / jak / jak / jak / jak / jaktə / jaktɐ / jak / jak / jak / jakta / jak / jak / jak / jak / jak / jak / jaktə / jak / jakta / jak / jak / jak / jaktʋ / jak / jakta / jak / iː / iː / iː / iː / iː /	sʲɛ / se / se / setta / sʲe / se / sɛ / se / sʲe / sʲɛ / se / sʲetɐ / sʲetɐ / se / se / se / se / sʲeta / se / se / se / se / sʲe / se / sʲeta / seː / seˑta / sʲeː / se / se / seta / se / sʲeta / sʲe / saraj / traj / tʲiraj / traj / tɪraj /

5.   Measuring pronunciation similarity

We compare three different algorithms that have been used to gauge pronunciation similarity. In fact, all the algorithms produce dissimilarity measures, but by looking at the pairs of words that are dissimilar to only a small degree, we obtain the best candidate loan words, just as we wish. The first two methods were developed within dialectology, and the third within historical linguistics.

The first method we examine is Heeringa's (2004) modification of the edit-distance or Levenshtein algorithm. The Levenshtein algorithm has been used frequently as a measure of spelling and pronunciation similarity (Nerbonne 2003) and functions by calculating the least costly set of operations needed to transform one string into another, where the operations are normally restricted to insertion, deletion, substitution and sometimes a transpose operation that might model metathesis. Heeringa (2004) modified the algorithm in order to ensure that substituting a sound for a similar one is less costly. He experimented with modifications based on feature systems, but the best performance was obtained from a version in which the similarity of two sounds was determined by measuring how close their spectrographic representations were. He used a demonstration recording of the IPA and measured the distance between the curves at a large number (of combinations) of points in time and frequencies. He used a logarithmic correction of the distance in keeping with psycho-acoustic practice. Heeringa was also able to assign costs to insertions and deletions by measuring the distance between silence and the sound being inserted or deleted. We refer to this manner of determining pronunciation similarity (and dissimilarity) as the SPECTROGRAM METHOD.

Proceeding from the same edit distance algorithm, Wieling et al. (2012) exploited one of the most useful properties of the procedure, namely that, in determining the difference between two strings, the algorithm automatically induces an alignment in which corresponding elements can be identified. For example, the pronunciation of the German *Durst* "thirst" is [tɔʃt] in Vielbrunn and [tʊəʃ] in Aachen (in *Kleine Deutsche Lautatlas*, see Nerbonne & Siedle 2005), and the algorithm produces the following as an alignment:

```
t   ɔ       ʃ   t
t   ʊ   ə   ʃ
```

Wieling et al. used 0/1 substitution costs to align 200 word pairs at 20,000 pairs of sites, and collected the frequencies with which sounds (including the "null sound" in insertions and deletions) appeared in alignment in a large contingency table. They then recalculated the substitution costs, assigning lower costs to all those sound pairs that were frequently aligned. They then iterated the alignment procedure and recalculation of substitution costs until no further alignment changes were noted. This method thus assigns low distances to word pairs with frequent sound correspondences, which is like the importance assigned in historical linguistics to regular sound correspondence. Because the recalculation was performed using POINTWISE MUTUAL INFORMATION, an information-theoretic measure, we refer to this technique as the PMI-BASED METHOD.

Finally, List (2012) developed an alternative measure of pronunciation difference especially focused on application in historical linguistics. We present this

technique based on List et al. (2018), as well. List proceeds from sound classes, e.g., bilabial obstruents {[p], [b]}, which often correspond in historical linguistics. After assigning all sounds to their classes, List aligns the sounds based on their classes, which is why his technique is known as SOUND CLASS ALIGNMENT (SCA). List derived his original classes from Dogopolsky (1964), whose set of ten List expanded to twenty-eight. While we shall note several differences between SCA and the other two methods (Spectrogram-based and PMI-based) below, one difference can be noted immediately. While both edit distance measures discriminate sensitively, SCA ignores at least initially all the differences of sounds in the same classes, e.g., [p] and [b]. The edit distance-based measures are more sensitive, eschewing the equivalence classes of sounds in the SCA approach, which leads to fine distinctions being ignored by SCA. Of course, ignoring fine distinction might be an advantage in detecting loan words, since loan words are often forced into the phonology of the borrowing languages, which may lead to substantial differences. The SCA procedure also assigns weights (contributions to distance) for pairs of classes, and here a second difference emerges. While the edit distance measures are symmetric, SCA aims to model historical development, and therefore assigns different weights, depending on whether one is measuring the likelihood of segment s_1 developing into s_2, or vice versa. Once SCA creates a basic alignment, the overall similarity may be further adjusted depending on the prosodic contours of the words, and finally, also on the exact phonetic realization of the segments (not just the classes). These later steps in processing incorporate levels of sensitivity which suggest that the less discriminating classes may not be so important. The role of the classes is limited to determining alignment, and distance further depends on prosodic contour and the detailed phonetic nature of the phonetic segments.

Software implementing the different measures was provided by Martijn Wieling for the PMI-based method, by Wilbert Heeringa for the spectrogram-based method, and through LingPy for the SCA method. LingPy is a package distributed by Johann-Mattis List.[4]

From previous work (Wieling et al. 2012) we expect the PMI method to be superior to the spectrogram method, but we wish to test the methods on the novel problem of detecting loan words. It is more difficult to predict how well SCA will fare. In particular, neither edit distance-based measure (the PMI and spectrogram-based methods) attempts to incorporate asymmetric substitution costs nor to account for the importance of the prosody of the word. (SCA uses asymmetric

---

4.   LingPy is available at http://lingpy.org (last access on 16 December 2020), and is introduced at http://github.com/lingpy/lingpy-tutorial (last access on 16 December 2020).

costs and also takes prosodic differences into account). It is also difficult to make predictions with respect to this point, because SCA reintroduces segmental sensitivity at a later stage in the processing.

6.   General setup

Recall our basic principle: if two words from unrelated languages meaning roughly the same (are both elicited by the same Swadesh concept and) are similar in pronunciation, then the chances are that one has been borrowed from the other's language (or that they have both been borrowed from a third). We follow Kondrak in assuming that some sort of borrowing is likely (Kondrak et al. 2003; Kondrak & Sherif 2006). To detect this for a given concept in the Swadesh list, we measure the pronunciation difference between every pair of realizations, one from a Turkic language and the other from an Indo-Iranian one. Our hypothesis is that those pairs showing the most similar pronunciations involve a loan.

But how similar do two pronunciations have to be in order to be regarded as loans? We know of no way to answer this question analytically, so we opted for an empirical approach. We had hoped to see a clear break between the distributions of pronunciation distances of pairs of borrowed words and those of words where no borrowing is involved, but no such break emerged from the data. We therefore tested a large number of thresholds empirically and opted for the optimal one. If we keep in mind the prospect of using our approach on new language families, then exhaustively searching for an optimal threshold is impractical. To gauge the likely success of our approach in this situation, we also apply a cross-validation technique.

To gauge the optimal result, i.e. the one using the optimal threshold, we first need to explain how we evaluate a given threshold. As is customary in computational linguistics (CL) when evaluating an automatic process where a human-annotated set is available, we compared the algorithmic results to the "gold truth" of the human annotator (Black et al. 1992). CL converged fairly quickly on a scheme borrowed from information retrieval in which both PRECISION and RECALL play a role. In this sort of evaluation, one analyzes a substantial amount of representative material for which the correct analyses have been noted by human experts, in this case loan words. We refer to the automatic classifications as positive in cases where the procedure deems them a loan, and negative in cases where it does not. We then distinguish:

1.   the genuine loan words correctly classified (true positives, *tp*);
2.   the genuine loans incorrectly classified (false negatives, *fn*);

3. the non-loan words incorrectly classified as loans (false positives, *fp*); and finally

4. the non-loan words correctly classified (true negatives, *tn*).

Precision is then the fraction of classifications that are correct (recognized by human experts), *tp/(tp+fp)*, and recall is the fraction of the humanly recognized loans that the process detects, *tp/(tp+fn)*.

Obviously, we would like to see both scores as high as possible, i.e., as close to 1 as possible, but note that it is trivial to score very well on one score if one disregards the other. Procedures that uniformly classify everything positively will score perfectly on recall. To overcome this difficulty, we examine a combination of the two scores, the so-called F-score (or F1-score), which is the harmonic mean between the two:

$$F1 = 2 \cdot \frac{precision \times recall}{precision + recall}$$

Figure 11.1 shows how precision, recall and the F-score range over 200 different thresholds of pronunciation difference used in the experiments with the PMI method. The blue line shows precision, which is naturally quite high at low thresholds, falling steeply from 0.02 on; the green line with dots and dashes shows recall, which is near zero at low thresholds, but climbs steadily; and the dashed red line traces the F-score, which conveniently shows a single peak, the one used in the experiments.
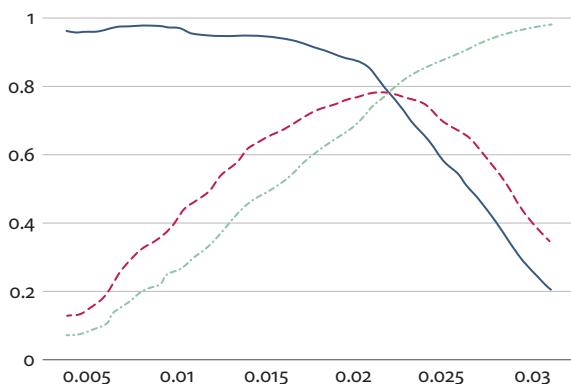
Figure 11.1 Precision (solid blue), recall (dash-dotted green), and F-score (dashed red) for a range of 200 putative thresholds examined with the PMI method The *x*-axis is pronunciation difference and the *y*-axis the fractional value of precision, recall and F-score. We settled on the threshold where the F-score peaks, around 0.022.

Similar sets of curves were determined for the other two techniques, the spectrogram-based method and the SCA method. Another advantage of determining

the threshold in this fashion is that it obviates the need to calibrate the three different scales used by the three pronunciation difference measures.

An alternative way of determining the threshold empirically is to use the entire distribution, examined in quartiles. Tukey (1977) suggests that all the points in a distribution that are more than 1.5 interquartile ranges below the first quartile be regarded as outliers (similarly, all the points 1.5 interquartile ranges above the third quartile are also outliers), but that will not interest us here). This definition is widely used in statistics packages, in particular in the well-known box-and-whisker diagrams, where the low outliers are those below the bottom whisker. Because the concept can be easily understood based on introductory texts, we will not present it in any further detail here.

7. Evaluation and results

As expected, all of the algorithms were able to identify a large number of loan words correctly when presented with a list of lexicalizations of the Swadesh concepts in unrelated languages. This is the "gold standard" we are ultimately interested in. But there were also interesting differences.

We first present the distribution of pronunciation difference scores together with the optimal threshold as determined by examining 200 potential thresholds (Figure 11.1). We also show the border below which outliers (in Tukey's sense) are found (Figure 11.2). Note, in particular, the "bump" on the left in the SCA distribution, which shows that this technique assigns low pronunciation difference scores to a rather large number of word pairs.

Based on the thresholds found, we also evaluated the algorithms on the basis of the same precision, recall and F-scores introduced above. Which algorithms detect loan words most effectively? Table 11.3 summarizes the success of the algorithms in detecting loan words, showing that the SCA method is clearly superior in this task.

Table 11.3 The performance of the algorithms based on edit distance using pointwise mutual information (PMI) and spectrograms as well as that of the SCA-based method in detecting loan words. The SCA method is clearly superior

	PMI	Spectrogram	SCA
Precision	0.84	0.75	0.85
Recall	0.74	0.74	0.88
F-score	0.78	0.74	0.85

Figure 11.2 The distribution of pronunciation differences based on the three techniques examined. In addition, the outlier boundary (see above) is presented with a dashed line and the optimal threshold with respect to F-score (see above) with a solid line. The *x*-axis is the pronunciation distance assigned by the algorithm, and the *y*-axis is the absolute frequency of the word pairs. The "bump" on the left of the SCA curve is promising.

We should note that, although we have focused our study on maximizing the F-score of recognition, there are perspectives which might, e.g., emphasize recall over precision. If the intention is to hand over a list of candidate loan words to a human expert for further review, then perhaps one might argue that it is better to err on the side of high recall and accept candidate lists of lower precision. As far as we can tell, however, this would not change our conclusion that the SCA method is superior in detecting loan words, since it is very clearly superior in recall.

In addition to the comparison based on setting an optimal threshold, we likewise wished to test the algorithms' performance on unseen data. This promises to provide a better view of how well the algorithms might work in a genuine research situation, where one could never know exhaustively what threshold to set.

In the cross-validation setup, we divide the entire data set into evenly sized subsets, using all but one of these to set the threshold, and then testing the procedures on the remaining data. In our case we applied 10-fold cross validation, which meant that we divided the data into ten subsets, first using nine of them (and therefore 90% of the data) to set the threshold parameter optimally. We note here that we did not continue the computationally expensive system of checking 200 candidate thresholds, but instead reduced the number of candidates to ten to keep running times manageable. We then tested the procedure on the remaining 10%. We repeated this ten times in order to avoid conclusions that might be based on a fortunate division into 90%/10%, and we report only the mean accuracy over all ten repetitions.

Table 11.4 Algorithm performance based on 10-fold cross validation

	PMI	Spectrogram	SCA
Precision	0.88	0.73	0.88
Recall	0.66	0.75	0.80
F-score	0.77	0.74	0.81

The size of the differences between SCA and the other two methods is lower in Table 11.4 than in Table 11.3, but SCA remains clearly superior.

We do not present the results that would be obtained using Tukey's outlier heuristic, but these were never superior to the results obtained by searching for an optimal threshold. We provide some examples of loan words detected by SCA but not by the other methods (Table 11.5). We discuss this further in the concluding section.

Table 11.5 Loan words detected by SCA but not by the other two measures of pronunciation difference

Concept	Turkic	Indo-Iranian
breast	kʊkrɛk	quqrak
correct	tuːra	tɔɣrɪ
fruit	miwe	mʲeva
sea	tʲeŋiẓ	dzingɨz
tree	tʲerɛk	daraxt

Examples of detected loans

A substantial number of correctly detected loans involve the insertion or deletion of segments, presumably to satisfy phonotactic constraints. Examples are given in Table 11.6.

Table 11.6 Some loans detected that involve the insertion or deletion of a segment

Concept	Turkic	Indo-Iranian
correct	tuːra	tøɣri
forest	wʊrmon	urmʊn
meat	gwʉʃt	guʃt'
lake	kwøl'	kul
old	kwønʲe	kʉjna

A second large number of detected loans depended on certain sounds being aligned, for example the vowels [a], [e], [ɔ] and [ɐ], as in "fruit", Turkic [miβe], Indo-Iranian [mev**a**] (where the corresponding vowels are printed in bold), or "breast", Turkic [køkrʲek], Indo-Iranian [kukr**a**k]. Another vowel set often exchanged was [i], [ɨ], [ɪ] and [ə], as in "smooth" Turkic [silləq], Indo-Iranian [siɬɨq'] or "fruit", Turkic [mʲ**i**va], Indo-Iranian [m**i**va]. Finally, the following also often corresponded: [ə], [u], [ɜ], [ʉ], [ø], [ʊ] and [ɔ]; see examples such as "egg", Turkic [tux**u**m] but also [tqʊm], Indo-Iranian [txəm]. Note that schwa [ə] appears in both of the last two lists in this paragraph.

We present some frequent consonant correspondences in Table 11.7. The sounds illustrating the correspondence are printed in bold. Although we have not tried to quantify their frequency, we suspect that it is the reason for SCA's superiority in this task.

Table 11.7 Some regular consonantal correspondences and illustrations from Turkic and Indo-Iranian

Consonant group	Swadesh concept	Turkic	Indo-Iranian
[k], [q], [g]	leaf	para**k**	bar**g**
	breast	køkrɛk	kukra**q**
[q], [x], [χ]	blood	**q**an	**x**un
	back	ɔr**q**a	ar**χ**a
[t], [d]	tree	**t**ʲerɛk	**d**arax
	sea	ʿ**t**ʲeŋɪz	**d**eŋiz
[β], [v], [w], [f]	animal	aj**β**an	haj**v**ɔn
		haj**β**an	haj**w**on
	to dig	ka**β**lɛm	kɔ**f**tan
	fruit	mi**w**e	mʲe**v**a
[ʧ],[ʃ], [ʤ], [j], [s]	dust	**ʃ**aŋ	**ʧ**aŋ
	to live	**ʤ**aʃa	**j**aʃam
	star	**ʤ**ɪldɨ̵z	**j**ɪldɨ̵z
	bird	qɯ**s**	ku**ʃ**
	star	**ʃ**ɯldɨ̵z	**j**ɪldɨ̵z

8. Discussion and prospects

We turn now to the conclusions we draw and some further research this work suggests.

Conclusions and a speculation

We can answer affirmatively that computational measures of pronunciation similarity can identify pairs of words in which a loan relationship is likely. Still, we also need to admit that F-scores of between 0.75 and 0.80 also mean that further work will be required. Further, we conclude that SCA is superior to the other edit distance-based methods we examined, and that it is the best algorithm available for identifying loans.

This leads to the interesting question of why SCA performs so much better, which is also more difficult to answer. SCA aligns based on sound *classes,* a very rough basis for similarity, but it also incorporates asymmetric substitution weights, which are lacking in the other approaches. Given that loan word detection is a symmetric task in our operationalization, it is unlikely that using asymmetric substitution costs is the key advantage. SCA also takes prosody into account, and it adds detailed segmental information in obtaining the final measures of pronunciation difference (distances). Either of these, but also the reduction to sound classes, might

be the key to SCA's success. In choosing among these, we shall hazard a speculation about the reason for SCA's superiority in the loan word recognition task.

It is common to note that sounds (and sound sequences) unavailable in a borrowing language are replaced quite regularly in loan words. For example, Spanish sequences involving stops followed by glides are commonly pronounced by native English speakers with stops followed by vowels. So, *Buenos días* "good day" [bʷe.nos. dias] is often rendered by these native English speakers as [bu.en.os.di.as]. But this replacement is regular, and can also be heard in English speakers' pronunciations of other words with stop-glide combinations, such as *igual* "same, equal" [i.gʷal], pronounced as [i.gu.al] or *agua* "water" [a.gʷa] as [a.gu.a], etc. It is plausible to assume that some of these replacements involve sounds in the same classes in the SCA algorithms, so that their alignment is likely to succeed. Table 11.5 shows examples of cases where the SCA algorithm alone was able to detect the loan word, and these examples confirm the suspicion that attributing very low costs to potentially very different sounds can be an advantage. If, in addition, the segmental differences do not add greatly to the sequence distance, this regularity of replacement will also explain the point of accumulation on the left side of the SCA frequency curve (Figure 11.2). We conjecture, then, that SCA is more successful because it assigns very little (perhaps no) weight to such regular correspondences. In any case, our empirical study confirms that the overall measure detects loan words more accurately. Naturally SCA assigns zero weight initially to the elements of various sound classes by design, and we do not mean to suggest that the observation is novel.

Future work

It is easy to imagine alternative technical approaches, since no algorithm was able to detect some of the loans. Table 11.8 indicates where there is room for improvement.

Table 11.8 Examples of loan word pairs undetected by any algorithm

Concept	Turkic	Indo-Iranian
fire	ɔlɔv	aɣaw
old	kwønʲɪ	kɨhna
egg	tqʊm	tuxm̩
meat	etʲ	jota
dust	ʃaq	tʃank

One promising avenue for further research would seek to incorporate more information in the comparison. Since the approach in this chapter compares word pairs

one pair at a time, we fail to exploit all the information in the data set. We might therefore try to compare not just one word pair at a time, but rather examine an aggregate measure of pronunciation difference. We might then compare the *mean* difference of the putative loan word to all the words in the sample, both those in its own language family and those in its putative source family. To be more precise, we might represent a given pronunciation j of a concept c in a Turkic language as t_j, suppressing the reference to c, which will be the same in any comparison of word pronunciations. Given a set of pronunciations of a given concept c from Turkic languages $\{t_1, t_2, \ldots, t_n\}$ and a second from Indo-Iranian $\{i_1, i_2, \ldots, i_m\}$, we compare t_j both to all the other pronunciations in Turkic t_j, as well as to all the pronunciations in Indo-Iranian i_k. In this way we obtain the mean distance of t_j to all other Turkic pronunciations of the same concept, as well as the mean distance to all Indo-Iranian pronunciations, i.e.,

$$D_{own} = \overline{d(t_j, i_j')},\ \forall j' \neq j \text{ and } D_{other} = \overline{d(t_j, i_k)},\ \forall k$$

It is clear that we should combine these somehow, but also that we should be particularly interested in t_j's, for which D_{own} is large and D_{other} small.

Prospects

In the rapidly evolving societies of our globalized times, social interaction – including interaction among speakers of different languages – is increasing apace, which will likely result in loan words becoming more frequent as well. This should delight students of contact linguistics, as it should provide the larger quantities of data needed to discern the patterns of new loan words, including what sorts of concepts are involved, the influence of the degree and nature of the contact on the likelihood of success, and the semantic and phonetic deformation that is involved. Mufwene has been among the first linguists to frame this phenomenon in terms of the *population structure* of the speakers involved (Collins & Mufwene 2005), bringing to contact linguistics a wider social and cultural dimension that profits from the analogy to a discipline proceeding through quantitative, mathematical models: ecology. Given the complexity of the contact situation, including the languages involved, but also the familial, material, social, economic and cultural relations among the speakers, it is only sensible to explore these situations using quantitative techniques.

Acknowledgements

an anonymous referee, who was unusually helpful.John Nerbonne was supported by the DFG Science Center "Words, Bones, Genes and Tools" at the University of Tübingen.

Supplementary data

Resources related to this work, such as data and code, are available at <https://github.com/jayliqinzhang/computational-loanword-detection>

References

Black, E., Lafferty, J. & Roukos, S. 1992. Development and evaluation of a broad-coverage probabilistic grammar of English-language computer manuals. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, 185–192. Shroudsburg PA: Association for Computational Linguistics. <https://www.aclweb.org/anthology/> (permanent archive). https://doi.org/10.3115/981967.981991

Collins, M. & Mufwene, S. S. 2005. What we mean when we say "Creole": An interview with Salikoko S. *Mufwene. Callaloo* 28: 425–462. https://doi.org/10.1353/cal.2005.0040

Dellert, J. 2017. Information-Theoretic Causal Inference of Lexical Flow. PhD dissertation, Eberhard-Karls-Universität Tübingen.

Dellert, J. 2018. Combining information-weighted sequence alignment and sound correspondence models for improved cognate detection. In *Proceedings of the 27th International Conference on Computational Linguistics*, 3123–3133. Shroudsburg PA: Association for Computational Linguistics. <https://www.aclweb.org/anthology/> (permanent archive).

Delz, M. 2013. A Theoretical Approach to Automatic Loanword Detection. Master's thesis, Eberhard-Karls-Universität Tübingen.

Dolgopolsky, A. B. 1964. Gipoteza drevnejšego rodstva jazykovych semej Severnoj Evrazii s verojatnostej točky zrenija (A probabilistic hypothesis concerning the oldest relationships among the language families of Northern Eurasia). *Voprosy Jazykoznanija* 2: 53–63.

Heeringa, W. 2004. Measuring Dialect Pronunciation Differences using Levenshtein Distance. PhD dissertation, University of Groningen.

IPA (International Phonetic Association). 1999. *Handbook of the International Phonetic Association*. Cambridge: CUP.

Jobling, M. A., Hurles, M. E. & Tyler-Smith, C. 2004. *Human Evolutionary Genetics: Origins, Peoples and Diseases*. New York NY: Garland.

Köllner, M. & Dellert, J. 2016. *Ancestral State Reconstruction and Loanword Detection*. Tübingen: Universitätsbibliothek.

Kondrak, G., Marcu, D. & Knight, K. 2003. Cognates can improve statistical translation models. In *Proceedings of the 2003 North American Association for Computational Linguistics*. Vol. 2, 46–48. Shroudsburg PA: Association for Computational Linguistics. <https://www.aclweb.org/anthology/> (permanent archive). https://doi.org/10.3115/1073483.1073499

Kondrak, G. & T. Sherif. 2006. Evaluation of several phonetic similarity algorithms on the task of cognate identification. In *Proceedings of the Workshop on Linguistic Distances*, J. Nerbonne & E. Hinrichs (eds), 43–50. Shroudsburg PA: Association for Computational Linguistics. <https://www.aclweb.org/anthology/> (permanent archive).
https://doi.org/10.3115/1641976.1641983

List, J.-M. 2012. SCA: Phonetic alignment based on sound classes. In *New Directions in Logic, Language and Computation* [LNCS 7415], 32–51. Berlin: Springer. https://doi.org/10.1007/978-3-642-31467-4_3

List, J.-M., Walworth, M., Greenhill, S. J., Tresoldi, T. & Forkel, R. 2018. Sequence comparison in computational historical linguistics. *Journal of Language Evolution* 2: 130–144. https://doi.org/10.1093/jole/lzy006

Manni, F. 2017. Linguistic Probes into Human History. PhD dissertation, University of Groningen.

Manni, F., Heeringa, W. & Nerbonne, J. 2006. To what extent are surnames words? Comparing geographic patterns of surname and dialect variation in the Netherlands. *Literary and Linguistic Computing* 21: 507–527. https://doi.org/10.1093/llc/fql040

Mayr, E. 2002. Interview with Ernst Mayr. *BioEssays* 24: 960–973. https://doi.org/10.1002/bies.10167

Mennecier, P., Nerbonne, J., Heyer, E. & Manni, F. 2016. A Central Asian language survey: Collecting data, measuring relatedness and detecting loans. *Language Dynamics and Change* 6: 57–98. https://doi.org/10.1163/22105832-00601015

Mufwene, S. S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511612862

Mufwene, S. S. 2005. Language evolution: The population genetics way. In *Gene, Sprachen, und ihre Evolution* [Schriftenreihe der Universität Regensburg 29], G. Hauska (ed.), 30–52. Regensburg: Universität Regensburg.

Mufwene, S. S. 2008. *Language Evolution: Contact, Competition and Change*. London: Bloomsbury.

Nerbonne, J. 2003. Linguistic variation and computation. In *Proceedings of the Tenth Conference of the European Chapter of the Association for Computational Linguistics*, 3–10. Shroudsburg PA: Association for Computational Linguistics. <https://www.aclweb.org/anthology/> (permanent archive). https://doi.org/10.3115/1067807.1067810

Nerbonne, J. 2009. Data-driven dialectology. *Language and Linguistics Compass* 3: 175–198. https://doi.org/10.1111/j.1749-818X.2008.00114.x

Nerbonne, J. & Siedle, C. 2005. Dialektklassifikation auf der Grundlage aggregierter Ausspracheunterschiede. *Zeitschrift für Dialektologie und Linguistik* 72: 129–147.

Shackleton Jr., R. G. 2010. *Quantitative Assessment of English-American Speech Relationships* [Groningen Dissertations in Linguistics 81]. Groningen: Center for Language and Cognition Groningen.

Sperber, D. 1996. *La Contagion des Idées*. Paris: Odile Jacob.

Swadesh, M. 1952. Lexicostatistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96: 452–463.

Tukey, J. W. 1977. *Exploratory Data Analysis*. Reading: Addison-Wesley.

Wieling, M., Leinonen, T. & Nerbonne, J. 2007. Inducing sound segment differences using pair hidden Markov models. In *Proceedings of 9th Meeting, Association for Computational Linguistics Special Interest Group in Computational Morphology and Phonology*, 48–56. Shroudsburg PA: Association for Computational Linguistics. <https://www.aclweb.org/anthology/> (permanent archive). https://doi.org/10.3115/1626516.1626523

Wieling, M., Margaretha, E. & Nerbonne, J. 2012. Inducing a measure of phonetic similarity from pronunciation variation. *Journal of Phonetics* 40: 307–314. https://doi.org/10.1016/j.wocn.2011.12.004

Zhang, L. 2016. A More Sensitive Edit-Distance for Measuring Pronunciation Distances and Detecting Loanwords. Master's thesis, University of Groningen & University of Malta.

Learnability and ecological factors as motivators of language change

Nour Efrat-Kowalsky

URPP Language and Space, University of Zurich

This chapter adopts Mufwene's (2001) framework according to which language acquisition is a process whereby competition and selection of linguistic features from the inputs create a unique idiolect. It examines the principles of selection of a given feature over others, and the mechanisms by which a feature spreads across a community (i.e., from one idiolect to another). These questions are addressed at the individual and population levels by analysing data from a twitter corpus. The frequency of a feature in the input as well as a suggested learnability factor are relevant to this discussion. Learnability is defined as the combination of the specific linguistic properties of a linguistic feature, and the universal human capacity to learn those specific properties.

Keywords: learnability, competition, selection, language-change, acquisition

1.   Introduction

This chapter follows the framework developed by Mufwene (2001, 2008) that views language acquisition as a process of re-analysis of linguistic features. Through a mechanism of competition and selection, linguistic features from the input are recombined to create a unique idiolect. The chapter investigates language change from this perspective: what causes a feature to be selected over others at the individual level, and how this feature spreads across a community. I do so by reviewing the scholarship on the acquisition of one linguistic feature in one language – grammatical gender assignment in Dutch – by different learner groups: monolinguals, bilinguals and second language learners. All learner groups use the same learning strategies when acquiring the Dutch gender assignment: they all overgeneralize the common definite determiner *de* with neuter nouns but not the other way around (see Section 3). However, the learner groups differ quantitatively:

bilinguals and second language learners overgeneralize *de* to a larger extent than monolingual learners do. This chapter further shows that even though the over-generalization of *de* is mainly thought of in the linguistic literature, and in Dutch society as a whole, as a characteristic of an immigrant variety of Dutch, mono-lingual L1 speakers unknowingly overuse *de* with neuter nouns in adulthood as well. It does so by presenting examples from a Twitter corpus. The use of Twitter for building a corpus has the advantage of being close to spoken language, even though it is typed; it also provides sociolinguistic information about the users thanks to their profiles. We also have access to their online network through their followers and the following functions afforded by the system. This allows us to track the speciation of linguistic features from one speaker-user to another and evaluate the success and failure of the spread of some features.

2.   Theoretical background

According to the competition and selection model developed by Mufwene (2001, 2008), the primary linguistic data each learner is exposed to during language acquisition consist of "a pool of linguistic features", i.e. varying features provided by different and competing idiolects, rather than a complete system. The "feature pool" (Mufwene 2001) contains linguistic features from all linguistic domains: phonetic, phonological, syntactic, semantic, pragmatic and lexical. The learner is in contact with different speakers during acquisition, each providing her with features from their specific idiolect. Some of these features may overlap, whereas others differ. Thus, language acquisition is formulated as a process of reconstruc-tion of linguistic features into a dynamic system. Figure 12.1, adapted from Aboh (2015: 114–115) shows this process.
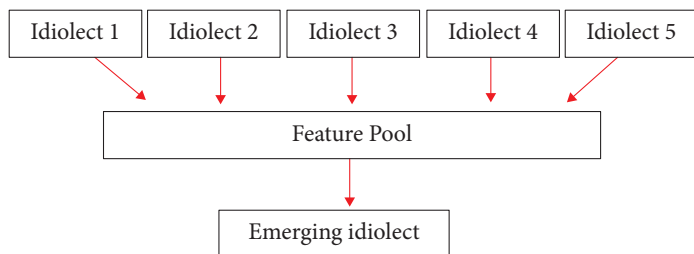
Figure 12.1

The learner is endowed with a cognitive system that weighs the competing features in the feature pool against each other and allows the selections of some features

over others. This selection is instantiated in the choices the learner makes in her speech. The exact same process happens when more than one language or dialect is available in the input (Figure 12.2). In such a case, the variation in the feature pool is expanded and includes less closely related features.

```
┌──────────────────┐   ┌──────────────────┐   ┌──────────────────┐
│ Standard French  │   │   Local French   │   │      Dutch       │
└──────────────────┘   └──────────────────┘   └──────────────────┘
         │                      │                      │
         └──────────────┐       │       ┌──────────────┘
                        ▼       ▼       ▼
              ┌────────────────────────────────┐
              │          Feature Pool          │
              └────────────────────────────────┘
                              │
                              ▼
                  ┌────────────────────────┐
                  │    Emerging idiolect    │
                  └────────────────────────┘
```
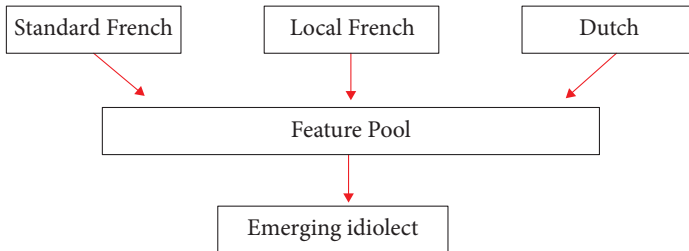
Figure 12.2

In both scenarios, when the features coming from the contributing idiolects diverge, they are in competition with each other. When they are similar we can speak of a situation of convergence: they reinforce each other (Ansaldo 2009). In such a case, the converging features may be favoured in a given ecology as dominant in the emerging idiolect. Note that features from the input might be selected as dominant, but other latent features may still exist in the new idiolect. This is, for instance, the case when some syntactic patterns are only used in the formal register, as opposed to its competing equivalent which is used in informal contexts. In French, for instance, clitic inversion (*Jean vient-il ce soir*? "Lit. John comes-he tonight") is more likely to be used in written and formal contexts than its semantic equivalent (*Jean vient ce soir*? "John comes tonight" with a final rising contour), the most common option in informal contexts. These two constructions require very different syntactic operations with different typological realities. While the first example is common to Romance languages, the second is found in Gungbe (Kwa) as well as other Niger Congo languages. These constructions therefore represent two different pieces of grammar, or two sub-grammars (cf. Aboh 2015, 2016).

Mufwene compares the relation between dominant and latent linguistic features to the selection of genes in an emerging organism:

> The selections an individual makes are similar to that between whether an off-spring will inherit brown or blue eyes from their parents, in the sense that very often selecting some particular variants does not entail complete exclusion of the seemingly latent/recessive alternatives … (Mufwene 2002: 47)

Even in cases in which learners have the same feature pool to draw from, the recombination of these features and those that end up being dominant in their

idiolects differs from learner to learner (Mufwene 2001). What makes each new idiolect distinct from its input sources is not only the particular combination of features, however. Indeed, during the process of recombination the features may be modified in a way that best fits the new emerging system, thus allowing the learner to adapt to her community's communicative needs.

Genetic linguistics takes language contact to be an incidental or "abnormal" circumstance in language change, rather than a catalytic one (Bailey 1982). But when language acquisition is framed as innovative re-analysis, always happening in contact conditions – in answer to the actuation problem (Weinreich et al. 1968) – there is no internally motivated language change. Language change is always externally motivated, as it is always brought about under conditions of contact.

The next section discusses the acquisition of the definite determiner system in Dutch, as reported in the literature, and how different learners acquire the gender feature.

3.   Dutch gender acquisition

Dutch has a two-way grammatical gender system that distinguishes between neuter and common gender nouns. The gender of a noun is seemingly arbitrary (Donaldson 1987), and there are no reliable phonological or semantic regularities to indicate its gender. The only evidence for a noun's gender comes from the definite, relative and demonstrative determiners and the adjectives, on which the gender is marked overtly. I focus here on the definite determiners. Singular common nouns are preceded by the definite determiner *de* (1a), while neuter singular nouns are preceded by the definite determiner *het* (1b).

(1)   a.   *De man* "the man"
       b.   *Het boek* "the book"

Both common and neuter nouns are preceded by the definite determiner *de* when they are plural (2a, b), and both common and neuter nouns are preceded by the definite determiner *het* in the diminutive form (3a, b).

(2)   a.   *De mannen* "the men"
       b.   *De boeken* "the books"

(3)   a.   *Het voetje* "the foot-DIM"
       b.   *Het kindje* "the child-DIM"

The distribution of the determiner system is repeated in Table 12.1.

Table 12.1 Determiner distribution in Dutch

	Indefinite	Definite singular	Definite plural	Definite diminutive
Neuter noun	*een*	*het*	*de*	*het*
Common noun	*een*	*de*	*de*	*het*

The distribution of common and neuter nouns in Dutch is skewed. An analysis of the CELEX database (compiled on the basis of two dictionaries and the most frequent lemmata from the text corpus of the Institute for Dutch Lexicology (160,000 lemmata)) shows that when comparing type frequencies, common nouns (i.e., *de*-N combinations) outnumber neuter nouns (i.e., *het*-N combinations) by a ratio of 2:1. The skewing is even greater in the token frequencies: common nouns outnumber neuter nouns by a ratio of 3:1 (van Berkum 1996). In addition, the neuter definite determiner *het* is often produced in its reduced form *'t* which is highly non-salient, and can be perceived as *de* in certain phonetic contexts (Unsworth & Hulk 2010). Finally, *het* is not only used as a determiner but can also function as a pronoun (e.g., *het is een mooi boek* "it is a beautiful book"). Altogether, the evidence for the neuter definite determiner in the Dutch input is weak and opaque compared to that of the common definite determiner or the pronominal usage.

These do indeed influence the acquisition of neuter nouns negatively, with two main distinctions: first, the acquisition of grammatical gender assignment in Dutch is a very lengthy process for all learners; and second, all learners overgeneralize the common determiner *de* with neuter nouns during acquisition, but not the other way around. While all learners master the common definite determiner quite quickly, this is not the case with the neuter definite determiner. It takes all learners a significant amount of time to fully acquire the gender assignment of neuter nouns, with monolingual children not mastering the system until they are at least six years old (Bol & Kuiken 1988; De Houwer & Gillis 1998; Van der Velde 2003, 2004; van Kampen & Wijnen 2000). This is in contrast to other languages where the acquisition of the gender system is completed by a much earlier age. In French, for example, monolingual children make no mistakes in gender assignment from age three onwards (Jakubowicz 2002).

The same is true for bilingual children growing up in the Netherlands in a household where one parent speaks Dutch natively and the other does not. For the bilingual children studied by Hulk (2007), for example, the dominant language is Dutch, and they acquire the gender assignment of neuter nouns at about the same rate as monolingual children do. A different type of bilingual is children born in the Netherlands to first- and second-generation immigrants, usually from Morocco or Turkey, who live in semi-closed immigrant communities. It is uncertain whether or

not these children are exposed to a substantial amount of Dutch input from birth, and if so, what type of Dutch they are exposed to (L1 versus L2). Experimental data point to the fact that these children do not master the use of the neuter definite determiner even in adolescence. Hulk & Cornips (2006a, 2006b) show that they reach only 43.3% accuracy at age 9;3–10;5, while Cornips et al. (2006) reports a 53.8% error rate for bilinguals at age 10;5–12;11 in comparison to the monolingual children in the same study who only have a 21.4% error rate at that age. Cornips and Hulk (2007) suggest a "fossilization" effect in these bilinguals: while monolinguals reach ultimate attainment, their bilingual age-matched peers seem to "fossilize" in a non-target stage. However, there is ample evidence that the target for each group is different and that this "fossilization" is in fact the ultimate attainment of a different variant of Dutch. In an interview with Nortier and Dorleijn (2008), an informant from Rotterdam reveals how the overgeneralization of common gender is an essential feature of linguistic and stylistic norms among adolescents of Moroccan descent. The overgeneralization of *de* is considered a marker of "allochthone" identity, in contrast to the "autochthone" identity (De Rooij 2005).

The acquisition of the neuter definite determiner is also a lengthy process for adult second language learners. Snow et al. (1981) report that Turkish and Moroccan immigrants who settled in the Netherlands in the 1960s frequently deleted the definite determiner and/or overgeneralized the common determiner *de* (see also Muysken 1984 for similar results). Blom et al. (2008) tested high-proficiency Moroccan second language of Dutch. These learners overgeneralized *de* with 71% of all definite singular neuter nouns. They overgeneralized the common definite determiner to such an extent that it is now considered a characteristic of an immigrant variety of Dutch, spoken by second- and third- generation immigrant children born in the Netherlands.

The reason for the significant amount of time taken by learners to acquire the Dutch neuter gender has to do with their learning strategies. It is widely accepted that the first acquisition stage of grammatical gender in Dutch involves lexical-statistical learning (4a, b) (Blom et al. 2008; Bol & Kuiken 1988; De Houwer 1990; Van der Velde 2003;Wijnen & Verrips 1998). When the input reaches a critical mass, learners assume abstract rules, like the ones shown in (5a, b). Notice that because of the nature of the input, the abstract rule in (5b) applies to a very limited set, whereas the abstract rule in (5a) is the default and applies elsewhere, thus essentially making the acquisition of the abstract rule in (5b) identical to the lexical list in (4b). This means that the acquisition of the neuter gender in Dutch is very sensitive to input effects and requires a large and consistent amount of input.

(4)   a.   /*de man*/ "the man", /*de vrouw*/ "the woman"
        b.   /*het boek*/ "the book", /*het paard*/, "the horse"

(5)   a.   *de* → [+def]
      b.   *het* → [+def, +neut, −plur]

To sum up, the acquisition of Dutch gender assignment is uneven. While learners of all types are able to formulate an abstract rule for the common gender and apply it to nouns, this is not the case with neuter nouns. Learners are forced to collect a word list of neuter nouns, a great cognitive burden. In order for them to compile the list, there is a need for exposure to enough diverse input, containing as many neuter nouns as possible. Then, in order to cement the status of the nouns in the list as neuter, there is a need for repeated positive evidence in the input for each noun. Therefore, mastering the gender assignment of neuter nouns in Dutch requires a sufficient amount of continuous quality input. Different types of learners receive different amounts and different quality of input, and some do not reach the threshold required to master the neuter gender assignment in Dutch. Instead, their grammars are shaped by the input that they hear, with many fewer nouns, or none at all, belonging to the neuter category.

4.   Selection

We have seen that the feature *de+neuterN* is present (as latent or dominant) in all Dutch learners' feature pools. We still need to answer the question of why it is selected over the standard feature *het+neuter* in some cases. The selection of one feature over its competitor in each interaction is determined by how readily available it is (i.e., the degree of dominance) and by the interactional context, for example, speaking to someone within or outside the speech community, speaking in front of a crowd of strangers, marking identity, and so on. The frequency of a feature in the input influences the dominance or latency of the feature in the feature pool.

However, ecological factors alone cannot explain some typological observations. Aboh (2015: 137) discusses how some features in Surinamese Creoles and Haitian Creole have not been selected despite the ecology favouring them. In those cases, the features coming from the substrate and the superstrate languages converge (DeGraff 2005; van den Berg 2007), making them heavily favoured by the ecology and, theoretically, having no competing features in the feature pool. Despite that, those dominant features were selected against in the emerging creoles. Aboh concludes that there must be a reason, independent of the ecology, which plays a role in the selection process. He proposes that inherent properties of a feature, such as whether it relates to an interface, contribute to its selection. In his system, features that relate to discourse-syntax interface are strong competitors and can be selected even in ecological settings that do not favour their selection.

In this chapter, I contribute to this discussion by proposing learnability as another factor that contributes to the mechanism of competition and selection. By learnability I mean the combination of the specific linguistic properties of a linguistic feature on the one hand, and the universal human capacity to learn those specific properties, on the other.

In the case of Dutch gender, there is an imbalance in the learnability of two competing features. The determiner expressing common gender, *de*, is very easy to acquire because learners are able to develop a general rule that captures the data, a very efficient learning strategy (cf. Blom et al. 2008; Ullman 2001a, 2001b. The development of a rule for the neuter gender appears more intricate, though. For this gender, there seems to be an ongoing competition between declarative and procedural learning, since it takes a while before learners master knowledge of neuter gender. Such competition between learning hypotheses arguably requires much cognitive effort on the part of the learner. The imbalance in the learnability of these two features leads to the overgeneralization of the feature that is easier to learn (and therefore mastered much earlier).

The learnability factor, together with ecological factors such as population structure or context and the frequency and salience of a feature in the input affecting its dominance in the learners' feature pools, influences the selection of one variant over another. Hence, what determines which feature will be selected in a certain grammar is threefold:

1. **Dominance**. The statistical distribution of the competing variants in the feature pool. If the frequency of one variant is much greater than the frequency of the other variant(s) in the input, it will contribute to the dominance of that feature in the feature pool.
2. **Ecological factors**. Other ecological factors include population structure, sociohistorical considerations, identity marking, prestige, location, age, gender, the interlocutor, the medium of conversation, and any other factor that may affect the selection of a linguistic form over another.
3. **Learnability**. A feature which is very difficult to learn gives space for a competing feature to rise and gain traction. The more difficult a feature is to learn, the longer it will be in competition during (first/second) language acquisition. During this competition the competing feature is being selected (as well as/instead of the original feature), making it more dominant in the learner's feature pool and more frequently present in the inputs of other learners they encounter.

5.   Feature spread

The mechanism of competition and selection is not restricted to language acquisition. This process continues for as long as the speaker is exposed to different inputs

presenting features that compete with the existing features in her idiolect. The formation of an individual's idiolect will not end until she dies, even though most of the linguistic system is formed by puberty (Mufwene 2001). Thus, one's language is continuously affected by the individuals she comes in contact with, be they monolingual or bilingual speakers or second language learners: speakers always adapt to their ecology. In their interactions, they adjust their communicative strategies to one another and to new communicative needs. Thus, if a speaker is interacting within a speech community where the non-standard variant *de+neuterN* is more widely accepted, she may end up producing it too, even if that feature is normally latent for her. This mutual accommodation causes the creation of new, non-identical linguistic innovations that in turn trigger a process of constant competition and selection (Mufwene 2001). Some of these innovations are picked up by other speakers in the speech community and become dominant properties in the communal language. Le Page and Tabouret-Keller (1985) identify this process as "focusing", whereby members of the same speech community sound more like one another than like non-members.

There are cases in which individual members of a speech community use a feature that is more typical of another group, even if the majority of these members rarely or never interact with one another (Mufwene 2001: 151). The speakers who come into contact with members outside their network act as "agents of transmission", propagating linguistic features outside their own speech community, while importing exogenous features.

In Dutch, overgeneralizing the common definite determiner *de* with neuter nouns is a feature most associated with second language learner communities (such as immigrant communities), and with the youth language of minorities (*straattaal*). Worth noting is that some monolingual Dutch speakers use it too, although they don't belong to a speech community that overgeneralizes it, and are not in regular contact with speakers who do.

One example of such an "agent of transmission" can be found in Example (6). The tweet in (6) contains the neuter noun *zout* "salt" in combination with the common determiner *de*. The tweet is written by a young woman from Rotterdam of Turkish descent. Her Twitter profile expresses her belonging to both Dutch and Turkish identities. She tweets regularly in both Dutch and Turkish. The majority of her 1,826 followers are Dutch of Turkish descent, but she also has many followers identified as autochthone L1-Dutch. Amongst them are a student from Utrecht, a professional from Amsterdam, and a politician from a major political party. They come from diverse age groups, genders, professions and locations across the Netherlands.

(6)   *Ik pakte de zout er al bij Willem;)*
         "I already picked up the salt from Willem;)"

This user is an agent of transmission. She propagates a linguistic feature from her own speech community, i.e. the use of *de* with neuter nouns, to speakers outside that community. By tweeting this, she increases the frequency of the feature *de+neuterN* in the inputs of her followers, thus making it more salient in their feature pools.

I argue that because *de+neuterN* is a latent feature in monolingual Dutch adult speakers' feature pools as a result of the intricate acquisition path of the neuter gender, each time they encounter it, its frequency in their feature pools increases, making it stronger and more likely to be selected thereafter (cf. Yang 2016). Henceforth a cycle begins: the more a feature is used in a community, the more its probability of being selected increases, and the more dominant it becomes in the individual's feature pool. As a result, this feature is favoured by the ecology and likely to be selected and used again, even if it is only latent in the speaker's grammar. This cycle of change essentially describes a process of intensification that may lead eventually to change at the population level (i.e., of the E-language (Chomsky 1986)).

Using social media as a tool for linguistic investigation, and Twitter in particular, allows us to track the spread of a feature. This is possible due to the nature of social media and the way Twitter is structured. Twitter users provide their own sociolinguistic information through their profiles. Although the public information users display about themselves varies and shouldn't be taken at face value, what they do or don't share also provides insight into their linguistic practices. For example, if they only provide profession-related information we may conclude that they are using this platform as a professional setting, and using language accordingly.

Thanks to the user's public profile we can access her social network., i.e., whom she follows and who follows her. These followers can be considered as members of the user's (virtual) speech community. When a Twitter user employs a certain linguistic feature, we can track who is at the other end of the input, what pool this feature is added to and what other features it is in competition with. Theoretically, we are then able to longitudinally track the input and output of each follower in order to identify the tipping point at which a latent non-productive feature will be selected as output, and in what communicative context. Moreover, a reader may retweet a tweet. By doing so, they repeat all the linguistic features in that tweet. If a tweet contains *de+neuterN*, a retweet of that feature confirms its status in the speech community and expands its reach to new members/followers.

5.1   Twitter

Uses on Twitter provide partial evidence of the overgeneralization of *de+N* in Dutch. Although Twitter is a platform of written communication, language use is closer to spoken language than to its formal written counterpart. Moreover,

Dorleijn (2016) notes that, in computer-mediated communication, online users rely on conventionalized forms to avoid ambiguities in the absence of phonetic cues. Therefore, I expect users to resort to the most transparent forms, easily interpretable by readers. Regarding the use of grammatical gender in Dutch, I expect that users would prefer a form not totally new to them and which can be interpreted by their followers. Hence, if a tweet contains the common article with a neuter noun, I suggest interpreting it as firstly, belonging to the user's repertoire, even if it may be latent in her overall speech; and secondly, a feature (even if latent) in the feature pools of the readers of this tweet, i.e. people belonging to the user's network.

5.1.1 *Methodology*

The Twitter data were collected by accessing the Twitter API with the programming language R (R Core Team 2016), using a methodology developed by Snyder (2017). To that end, I used the **twitteR** package (Gentry 2015), the **RJSONIO** package (Lang 2014) and the **stringr** package (Wickham 2015). Then, I connected R to the MySQL (Oracle 2016) database. For this, I used the RMySQL package (Ooms et al. 2016). MySQL acted as my server in which I stored my database. This database grew as the R scripts continued searching the Twitter API. I searched the Twitter API by running three word lists, each containing 115 neuter nouns in combination with one of three common articles: *de* 'the.COMMON', *die* 'that.COMMON', or *deze* 'this.COMMON'. The lists also contained all nouns in combinations with frequently occurring adjectives (with the common gender marking). This was done so that no possible result could be excluded because the ADJ appears between the DET and the noun head in Dutch. The word lists contained neuter nouns with a very high token frequency as well as low-frequency nouns. A total of 181 tweets were collected from October 5th to November 11th 2017. Sadly, most of the tweets had to be removed from the corpus due to false results, as a consequence of either the word list, idioms, or typing errors. For example, the tweet fragment in Example (7) shows a false result. The word list included the neuter noun *boek* "book". But in the NP [*de boek-auteur*] the determiner is agreeing in gender with the head *auteur*, which is common, and not with *boek*. The same is true for the tweet fragment in Example (8). The determiner agrees in gender with the head of the phrase which is *presentatie*, not *boek*. Therefore, these are not examples of overgeneralization of the feature *de+N*. More tweets had to be removed because they were falsely identified as Dutch by the code, but their source was in fact South Africa, meaning they were in Afrikaans, where the gender distinction has already disappeared. All tweets were double-screened by the author and by a native Dutch speaker.

(7) *Voor die drie dagen werk moet de boek-auteur betalen*
     "For those three days of work the book author has to pay"

(8)   *op weg naar Rotterdam voor de boek presentative*
      "on the way to Rotterdam for the book presentation"

5.1.2   *Results*

The examples below give a good indication that some autochthone L1 Dutch speakers use the common definite determiner *de* with neuter nouns. Example (9) presents a part of a tweet containing the neuter noun *hart* "heart" preceded by the common determiner *de*. The database I built for this study also contains the user's profile information. In this case, she is a nineteen-year-old woman from the Netherlands. As mentioned before, the overgeneralization of *de* is mostly associated with street language and urban identity. This user chose to give her location as "The Netherlands" rather than an urban environment such as Rotterdam or Amsterdam. She associates herself with the national identity, which in turn is associated with a more standard use of Dutch, rather than with an urban identity associated with the overuse of *de*. We can conclude that her use of *de+neuterN* is a "slip of the keyboard", a documented unconscious realization rather than a deliberate use aimed at performing an urban identity. Indeed, this is her only use of *de+neuterN*. The user mainly follows people from her own real-life community and celebrities who appeal to teens. Her followers are mostly people in her own real-life community or similar communities.

(9)   *pure vrouw met de hart op de juiste plaats!*
      "pure woman with the heart in the right place!"

The tweet fragment in (10) contains the neuter noun *werk* "work" with the common determiner *de*. It was tweeted by a young professional man from Rotterdam. This tweet was in a chain of responses to a tweet posted by the television documentary journalism program "*Tegenlicht*". Unlike the Example in (9), this tweet was authored in a professional, rather than a personal, capacity. His followers are mostly professionals working in start-ups or technicians and planners. He has a mixed following of Dutch and non-Dutch (English-speaking) profiles.

(10) *Overigens ga ik van de werk de reportage pas terugkijken.*
      "By the way, I will only look back over the report from work."

Example (11) illustrates the use of *de* with the neuter noun *boek* "book". The Twitter account is that of a high-school Dutch teenager who has 118 followers, mainly her schoolmates.

(11) *die vrouw zei tegen me dat ze de toetsen maakt met de vragen van de boek,*
      "that woman told me that she makes the tests with the questions from the book,"

Example (12), in which *boek* is again preceded by *de*, was not tweeted from a personal profile, but rather from an institution's profile – Winschoten library in the north of the Netherlands. This profile provided another tweet in the corpus that was deemed a typing error. This tweet however, does not seem to be a mistake.

(12)   *Dirkjan is de boek die ik heb gekozen.*
       "Dirkjan is the book I have chosen."

These examples show that monolingual Dutch speakers well beyond the age of acquisition use the common determiner with neuter nouns. As illustrated in Examples (10) and (12), this use is not restricted to colloquial language (as in Examples 9 and 11). The definite determiner-neuter noun combination is attested in professional settings as well. These examples act as positive evidence and proof of methodology for tracking the life cycle of a linguistic feature from output to input, and following its spread and increased dominance in a (virtual) network.

As mentioned earlier, using a Twitter corpus allows one to track the spread of this feature in the user's speech network, thanks to the sociogeographical information in the tweeters' profiles and those of their followers. Worth noting is that none of these tweeters either follows or is being followed by someone who belongs to the ethnic groups identified in the literature as users of *de* with neuter nouns. In other words, the make-up of their online speech communities appears to be ethnically and linguistically autochthone. Because these speakers don't seem to be in regular contact with the agents of transmission of this linguistic feature, one may ask how it emerges in their feature pools and why it is selected.

I argue that the feature *de+neuterN* is present in monolingual Dutch speakers' and learners' input due to the nature of the acquisition of the neuter gender. The cycle of change discussed earlier is therefore expected to obtain even in (speculative) instances of completely closed communities within which monolingual L1 speakers interact only with other monolingual L1 speakers, even though in such case it would be much slower than in more open communities containing second language learners. In the context of Dutch gender, second language learners increase the frequency of the feature *de+neuterN* in the input. This makes it more dominant in individual speakers' feature pools, and increases its favourability in more communicative contexts. In this way, second language learners accelerate the cycle of change, but do not initiate it.

6.   Conclusion

This chapter presents a case of ongoing language change and analyzes it from Mufwene's perspective of competition and selection. In order for language change

to take place at the population level, the innovation occurring at the individual level during language acquisition must either have happened similarly in many other individuals in the community, or have spread from one or a few individuals throughout the community. Most likely, and in the case presented here, it is both. Today, Dutch has a two-way gender system. For historical reasons, there are no reliable phonological or semantic regularities to indicate the gender of the noun to the learner. This, combined with the lower frequency of neuter nouns compared to common nouns in Dutch learners' input and the non-salience of the neuter definite determiner, makes the gender system extremely difficult to master. Learners need to compile a word list of neuter nouns, while applying the common gender elsewhere. This leads to the emergence of a competing variant to the neuter determiner, i.e. the use of the common definite determiner with neuter nouns (*de+neuterN*).

The more this competing feature is used, the more dominant it becomes in the learner's feature pool, making the standard variant even harder to learn. This is true for all learners of Dutch, as shown in Section 3. However, while some learners receive enough input of the type *het+neuterN* for this to eventually become dominant, other learners (particularly bilinguals and second language learners from ethnic minority communities in the Netherlands) never do. This leaves the non-standard feature, *de+neuterN*, dominant.

Section 5 made a case for using a Twitter corpus as an effective tool to study language change within a (virtual) speech community. In Section 5.1 my analysis of a small-scale Twitter corpus showed how we can track the spread of the feature *de+neuterN* through an online speech community, as well as between speech communities. Tweets from L1 Dutch speakers using the feature *de+neuterN* in various contexts were presented, supporting the claim that the feature *de+neuterN* is becoming more dominant within the autochthone speech community. Moreover, the corpus presents an example of a bilingual speaker from an ethnic minority community who overgeneralizes the common definite determiner. The analysis of her followers indicates that her tweet, carrying this feature, reached not only speakers from a speech community where this feature is common, but also those from the autochthone speech community, in which this feature is much less accepted. By doing so, this user increases the frequency of this latent feature in her followers' feature pools, and by extension makes it more dominant at the population level.

Social media give linguists the unique opportunity to track not only the use of certain linguistic forms, but also their spread. Unlike other written language, which provides biographical information only about the writer (the output), social media give us access to (self-disclosed) information about the reader as well (the input). It is undeniable that without studying the input, it is impossible

to accurately characterize any acquisition path, or even discuss a learner's target language. In the case of bilingual learners' acquisition of grammatical gender in Dutch, one cannot make any claims regarding the "unsuccessful" ultimate attainment of the target language, without studying the input.

Lastly, this chapter discussed the selection of one feature over another. Ecological factors alone do not fully explain the process of selection and spread of a feature. While they provide an explanation of which factors influence the dominance and the likelihood for a variant to be selected, they do not explain why a latent variant is selected over its standard counterpart, and why some of the innovations die with the individual while others spread within a community. In this chapter, I have suggested that the missing factor is learnability. A feature which is hard to learn will likely produce misuse during acquisition, i.e. a competing variant. This competing variant will not disappear quickly, but will linger in the learner's feature pool, as in the case of the overgeneralization of the common gender determiner in Dutch. One of the consequences of the lingering of the competing variant is that it is present in more contexts for a longer period of time. The learnability of a feature together with ecological factors (that influence the dominance of a feature) give a holistic explanation of the competition and selection model of language acquisition and change. Taken together, they lead me to predict that the feature *de+N* will eventually become dominant for a significant proportion of Dutch speakers, leading to the erosion of the two-way gender system in Dutch.

References

Aboh, E. O. 2015. *The Emergence of Hybrid Grammars: Language Contact and Change*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139024167

Aboh, E. O. 2016. Information structure: A cartographic perspective. In *The Oxford Handbook of Information Structure*, C. Féry & S. Ishihara (eds), 147–164. Oxford: OUP.

Ansaldo, U. 2009. *Contact Languages: Ecology and Evolution in Asia*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511642203

Bailey, C. N. 1982. The garden path that historical linguistics went astray on. *Language & Communication* 2: 151–160. https://doi.org/10.1016/0271-5309(82)90005-2

Blom, E., Polisenská, D. & Weerman, F. 2008. Articles, adjectives, and age of onset: The acquisition of Dutch grammatical gender. *Second Language Research* 24: 297–332. https://doi.org/10.1177/0267658308090183

Bol, G. W. & Kuiken, F. 1988. Grammaticale Analyse van Taalontwikkelingsstoornissen (Grammatical Analysis of Language Development Disorders). PhD dissertation, University of Amsterdam.

Chomsky, N. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York NY: Praeger.

Cornips, L. & Hulk, A. 2007. Diminutives and gender awareness in bilingual acquisition. Talk presented at the 6th International Symposium on Bilingualism, University of Hamburg, June.

Cornips, L., Van der Hoek, M. & Verwer, R. 2006. The acquisition of grammatical gender in bilingual child acquisition of Dutch (by older Moroccan and Turkish children). In *Linguistics in the Netherlands*, B. Los & J. van de Weijer (eds), 40–51. Amsterdam: John Benjamins. https://doi.org/10.1075/avt.23.07cor

DeGraff, M. 2005. Morphology and word order in "creolization" and beyond. In *The Oxford Handbook of Comparative Syntax*, G. Cinque & R. S. Kayne (eds), 293–372. Oxford: OUP.

De Houwer, A. 1990. *The Acquisition of Two Languages from Birth: A Case Study*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511519789

De Houwer, A. & Gillis, S. 1998. *The Acquisition of Dutch* [Pragmatics & Beyond New Series 52]. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.52

de Rooij, V. 2005. De studie van Etnolecten: Overdenkingen vanuit een etnografisch perspectief [The study of ethnolects: Reflections from an ethnographic perspective]. Talk presented at the Workshop Etnisch Nederlands, Utrecht University, November.

Donaldson, B. C. 1987. *Dutch Reference Grammar*. Leiden: Martinus Nijhoff.

Dorleijn, M. 2016. Can internet data help to uncover developing preferred multilingual usage patterns? An exploration of data from Turkish-Dutch bilingual internet fora. *Journal of Language Contact* 9: 130–162. https://doi.org/10.1163/19552629-00901006

Gentry, J. 2015. twitteR: R based Twitter client. R package version 1.1.9. <https://CRAN.Rproject.org/package=twitteR> (16 December 2020).

Hulk, A. 2007. Deviance in early child bilingualism. In *Romance Linguistics 2006: Selected Papers from the 36th Linguistic Symposium on Romance Languages* [Current Issues in Linguistic Theory 287], J. Camacho, N. Flores-Ferrán, L. Sánchez, V. Déprez & M. J. Cabrera (eds), 177–98. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.287.14hul

Hulk, A. & Cornips, L. 2006a. Neuter gender and interface vulnerability in child L2/2L1 Dutch. In *Paths of Development in L1 and L2 Acquisition* [Language Acquisition and Language Disorders 39], S. Unsworth, T. Parodi, A. Sorace & M. Young-Scholten (eds), 107–134. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.39.06hul

Hulk, A. & Cornips, L. 2006b. *Between 2L1- and child L2 acquisition: An experimental study of bilingual Dutch*. In *Interfaces in Multilingualism: Acquisition and Representation* [Hamburg Studies on Multilingualism 4], C. Lleo (ed.), 115–137. Amsterdam: John Benjamins. https://doi.org/10.1075/hsm.4.06hul

Jakubowicz, C. 2002. Functional categories in (ab)normal language acquisition. In *The Process of Language Acquisition*, I. Lasser (ed.), 165–202. Berlin: Peter Lang.

Lang, D. T. 2014. RJSONIO: Serialize R objects to JSON, JavaScript. <https://CRAN.Rproject.org/package=RJSONIO> (16 December 2020).

Le Page, R. B. & Tabouret-Keller, A. 1985. *Acts of Identity: Creole-Based Approaches to Language and Identity*. Cambridge: CUP.

Mufwene, S. S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511612862

Mufwene, S. S. 2002. Competition and selection in language evolution. *Selection* 3: 45–56. https://doi.org/10.1556/Select.3.2002.1.5

Mufwene, S. S. 2008. *Language Evolution: Contact, Competition and Change*. London: Continuum.

Muysken, P. 1984. Attitudes and experiences of discrimination: The Netherlandic of Moroccan foreign workers. In *Sociolinguistics in the Low Countries* [Studies in the Sciences of Language Series 5], K. Deprez (ed.), 333–356. Amsterdam: John Benjamins. https://doi.org/10.1075/ssls.5.17muy

Nortier, J. & Dorleijn, M. 2008. A Moroccan accent in Dutch: A sociocultural style restricted to the Moroccan community? *International Journal of Bilingualism* 12: 125–142. https://doi.org/10.1177/1367006908012010801

Ooms, J., James, D., DebRoy, S., Wickham, H. & Horner, J. 2016. RMySQL: Database interface and "MySQL" driver for R. <https://CRAN.R-project.org/package=RMySQL> (16 December 2020).

Oracle. 2016. MySQL Server 5.7. <https://www.mysql.com> (16 December 2020).

R Core Team. 2016. *R. A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/> (16 December 2020).

Snow, C., Van Eeden, R. & Muysken, P. 1981. The interactional origins of foreigner talk: Municipal employees and foreign workers. *International Journal of the Sociology of Language* 28: 81–92. https://doi.org/10.1515/ijsl.1981.28.81

Snyder, K. 2017. Exploring the Development of Stylistic Practices with a Sociolinguistic Twitter Corpus. Master's dissertation, University of Amsterdam.

Ullman, M. 2001a. A neurocognitive perspective on language: The declarative/ procedural model. *Nature Reviews (Neuroscience)* 2: 717–726. https://doi.org/10.1038/35094573

Ullman, M. 2001b. The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition* 4: 105–122. https://doi.org/10.1017/S1366728901000220

Unsworth, S. & Hulk, A. 2010. L1 acquisition of neuter gender in Dutch: Production and judgement. In *Proceedings of Generative Approaches to Language Acquisition 2009*, J. Costa, A. Castro, M. Lobo & F. Pratas (eds), 483–492. Cambridge: Cambridge Scholars.

van Berkum, J. 1996. *The Psycholinguistics of Grammatical Gender: Studies in Language Comprehension and Production*. Nijmegen: Nijmegen University Press.

van den Berg, M. 2007. *A Grammar of Early Sranan*. Zetten: Drukkerij Manta.

Van der Velde, M. 2003. Déterminants et Pronoms en Néerlandais et en Français: Syntaxe et Acquisition. PhD dissertation, Université Paris 8.

Van der Velde, M. 2004. L'acquisition des déterminants en L1: Une étude comparative entre le français et le néerlandais. *Acquisition et Interaction en Langue Étrangère* 21: 9–46. https://doi.org/10.4000/aile.1723

van Kampen, J. & Wijnen, F. 2000. Grammaticale ontwikkeling (Grammatical development). In *Kindertaalverwerving: Een Handboek voor het Nederlands* (*Child Language Acquisition: A Handbook for Dutch*), S. Gillis & A. M. Schaerlaekens (eds), 225–285. Groningen: Martinus Nijhoff.

Weinreich, U., Labov, W. & Herzog, M. I. 1968. *Empirical Foundations for a Theory of Language Change*. Austin TX: University of Texas Press.

Wickham, H. 2015. stringr: Simple, consistent wrappers for common string operations. <https://CRAN.R-project.org/package=stringr> (16 December 2020).

Wijnen, F. & M. Verrips. 1998. The acquisition of Dutch syntax. In *The Acquisition of Dutch* [Pragmatics and Beyond New Series 52], S. Gillis & A. De Houwer (eds), 223–300.Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.52.07wij

Yang, C. 2016. *The Price of Linguistic Productivity: How Children Learn to Break Rules of Language*. Cambridge MA: The MIT Press. https://doi.org/10.7551/mitpress/9780262035323.001.0001

CHAPTER 13

The restructuring of Salikoko Mufwene through competition and selection

A conversation between Salikoko Mufwene and Michel DeGraff

Michel DeGraff
Massachusetts Institute of Technology

The goal of this postcript is to celebrate the transformative journey of a dear colleague and the still-evolving ecology of his transformation as a scholar who has doubled as mentor for me (Michel DeGraff) and for my work both in linguistics proper (especially in the field of Creole studies) and in the realm of linguistics and education.[1] Perhaps a better term to describe Salikoko Mufwene's intellectual path is one of "restructuring" – a process that can be characterized as exceptional, "abnormal" even – through competition and selection (all puns intended). This transformation has taken Salikoko (hereafter, Sali) from boarding school pupil in the Democratic Republic of Congo (former Republic of Zaire) to creolist extraordinaire and a long career at the University of Chicago, including a stint as Chair of its Linguistics Department (1995–2001), Faculty Director of the University of Chicago Centre in Paris (2013–2014), and Interim Faculty Director of the University's Centre for the Study of Race, Politics, and Culture (2018–2020). Sali's contributions to Creole studies, especially his uniformitarian stance, together with his steadfast intellectual camaraderie and friendship, have greatly inspired my own work against "Creole Exceptionalism" – both in linguistics and on the ground in Creole-speaking societies such as my native Haiti.

---

1. This is a revised and updated version of a Focus article on Salikoko Mufwene by Michel DeGraff published in *Carrier Pidgin* 29, 2001 <http://mufwene.uchicago.edu/interviewWith-Michel.html> (16 December 2020).

From founder principle to funder principle

In 1996, Salikoko Mufwene published a very influential paper in which he adopted and adapted the "founder principle" concept from the field of population genetics (Mufwene 1996). In that article, the concept was used to explain the earliest settlers' crucial influence on the linguistic traits of the Creole languages emerging in plantation settlement colonies. Sali's basic insight was that these early settlers' speech varieties would be the ones having the greatest impact on the eventual shape of the contact language, no matter the ultimate size of the colony's population.

Now, let's say we extrapolate Sali's "founder principle" from Creole studies to the *history* thereof. As we consider the fact that the *founders* of Creole studies (i.e., the first scholars to describe Creole languages in the seventeenth through nineteenth centuries) were, directly or indirectly, *funded* by the colonial enterprise, it seems reasonable to also evoke a "*funder* principle" as one key factor in the emergence and transmission of Creole Exceptionalism (DeGraff 2020). In turn, this extrapolation – from theories of Creole formation to the historical and political context for the production of these theories – forces me to reflect on Sali's own biography, and how it may have provided a key context for the history of his currently held positions vis-à-vis Creole Exceptionalism and other foundational issues in our field.

This seems a good point at which to start the conversation with Sali.

On Creole studies as the making and unmaking of exceptionalist myths about the colonized and their languages

MICHEL: Growing up in Haiti, for thirteen years I attended a school run by a French religious order of Catholic brothers, Les Frères de l'Instruction Chrétienne (F.I.C.). These brothers were passionately francophile and brutally creolophobe – to the point of exerting physical violence as punishment against children caught speaking Haitian Creole (known as "Kreyòl" in Haiti). Like many among the Haitian elites, the F.I.C. brothers overtly despised cultural phenomena that were not of unambiguous European pedigree, from linguistic to religious practice. Not very "Christian," it seems to me.

From that experience, and from what I have read and analyzed since, I've come to a rather pessimistic conclusion: certain modes of (mis-)education in neo-colonial societies like Haiti often turn students, along with their professors, into conformist and elitist non-thinkers – uncritical consumers of pre-established myths, obsequious and self-serving upholders of that European "normative gaze". And you won't be surprised that I take the term "neo-colonial" to apply beyond the

scope of the Caribbean, Latin America, and Africa, and to include certain sectors of the North American and European intelligentsia who still uphold dogmas that undermine the intellectual and cultural worth of racialized or minoritized groups.

Since you yourself grew up in the Congo toward the end of Belgian rule, you may have received a thoroughly colonial and faith-based education like I did. As compared to my own, your education in the Belgian Congo was perhaps even closer to the self-alienating indoctrination that Frantz Fanon so soberly analyzed. Yet, you are among the most reflexive and most critical creolists I can think of. Over the years, you have consistently refused to take any orthodoxy for granted. You often refer jocularly to your positions as "heresies".

For example, you were among the very first creolists to question the traditional correlation of Creole continua (from so-called "basilectal" to "acrolectal" varieties) with so-called decreolization, namely the notion that, over time, Creole languages, in the Caribbean, tend to lose their putative "Creole" features to become structurally closer to the nearest European languages. In the traditional correlation, decreolization happens when a Creole is spoken as a vernacular alongside a prestigious European language with official status; as a consequence, the Creole will, over time, undergo a structural shift from basilectal to acrolectal, as if it were pulled toward the European language by some sort of sociolinguistic attraction, due to the prestige of those who speak that European language. But you and others have argued that, contrary to this decreolization narrative, the Creole varieties closer to the acrolects, the so-called "upper-mesolects," would have come into existence earlier than the basilectal ones. This would imply that the most "decreolized" varieties came into existence in advance of the most "creolized" varieties. Such a contradiction would then suggest that the classic decreolization story is logically and sociohistorically flawed!

I also remember that, as early as your 1986 article "*Les langues créoles peuvent-elles être définies sans allusion à leur histoire?*" ("Can Creole languages be defined without reference to their history?") in *Études Créoles*, you demystified traditional attempts to define Creoles as a type according to their structural features (Mufwene 1986). In a later publication, *The Ecology of Language Evolution* (Mufwene 2001), you boldly claim, right on page 1, that "Creoles have developed by the same restructuring processes that mark the evolutions of non-Creole languages". This, in a nutshell, is the anti-exceptionalist uniformitarian position on Creole languages that I myself have been defending in my own work.

Elsewhere in the same book you state that one of your goals is "to prevent creolistics from being a consumer subdiscipline which espouses gratuitously, without questions asked, some still unjustified working assumptions and theoretical models accepted in other subdisciplines of linguistics" (2001: xiv). You won't be surprised, then, that in my 2001 essay "On the origin of Creoles: A Cartesian critique

of neo-Darwinian linguistics," I give you thanks for your "always 'heretical' inspiration" (DeGraff 2001: 300).

Now, what (if anything) did your childhood and education in colonial, then post-colonial, Africa contribute to your later intellectual interests and, in particular, to your "heretical" approaches to language contact and language evolution? More specifically, what aspects of your growing up in the (post-)colonial Congo may have prevented you from falling, later on, into the all-too-seductive trap of (post-)colonial Creole Exceptionalism?

SALI: It is true that the Belgian colonial school system in the Congo was set up primarily to train colonial auxiliaries, who typically perpetuated a view of Africa, or the Third World in general, from a non-indigenous perspective. However, many of us who grew up during the transition from colonial rule to political independence – or, more accurately, economic neo-colonialism – learned to question not only pre-independence European rule, but also its post-independence replacements.

I left home early, at the age of twelve, for boarding school, likely to be easily influenced by older kids. My parents constantly advised me to be critical and not to be a sheepish follower of bigger kids. What they didn't anticipate was that I would also take pride in questioning authority.

Four years later, I was expelled from the junior seminary – my first boarding school – for "insubordination" (as it was formulated in my dismissal letter). I just disputed things that did not make sense to me.

At the next boarding school, I was sometimes chastised, if not altogether dismissed from some class sessions for challenging my teachers' confusing explanations of some facts. My teachers, then, were Belgians who could not make it at home and had a chance in the ex-colony as "technical assistants". Sound familiar? I remember having problems with a priest, one of my teachers, because he often said that a practice was good because "that's how [he] did it". And I was quick to retort, disrespectfully of course, that his reason was not good enough to consider the particular practice as good.

During my time in college, I had to control myself, wait until I earned my *Licence* [i.e., Bachelor of Arts (BA) degree] and got access to graduate education in order to better nurture my critical thinking. (A teaching assistant advised me against pointing out to my phonetics professor that his notation had an inconsistency, as he used different symbols for the vowel of *bon* and *botte* in French – ignoring the need for a tilde for the first word. The reason was that I should not "embarrass" the professor, who would otherwise flunk me. No kidding! The advice reminded me of a Belgian history teacher of mine in high school who had threatened to flunk me after, one day, I produced a list of dates that were not the same

for the same events in his *exposés*!) Aside from the fact that I was learning a lot of interesting things, including the discovery of linguistics through English philology, the reason for this transitional conformism is that the university system was more competitive than high school and its selection process more peremptory. While we were all funded by the state, only a small fraction – not more than fifty per cent – of those who began university could finish the four-year programme. Yet, the beginning class represented the cream of the crop – at best, the top five per cent – from the high school system of the time.

This is one story in my past that makes me believe in luck as a necessary factor for success, in addition to all other qualifications.

On biography, geography and bibliography

MICHEL: Well, Sali, I myself am convinced that you are blessed with both talent and luck. Plus you do work very, very hard, don't you? No need to answer that. The answer is in your biography and in your bibliography.

On the topics of biography and bibliography: abstracting away from the extraordinary thematic diversity in your research and publications, one can detect a remarkable consistency in your bibliography, a thread of Ariadne, or a few such threads, if you will. What's also remarkable – to me, in particular, as a creolist who hails from Haiti – is the extent to which your research themes seem to consistently track, mainly via language and linguistics, questions of identity formation and transformation within the African Diaspora's struggles and successes in the Americas.

Some core questions raised in your research seem singularly connected to one theme that both you and I face on a daily basis, namely the theme of exile in the context of our personal and intellectual questioning of (mythical theories of) origins and homelands. Let me illustrate with a couple of titles taken from your prolific bibliography: "How African is Gullah, and why?" (Mufwene & Gilman 1987) and "The ecology of Gullah's survival" (Mufwene 1997). Compare these actual titles with the following biographical titles that I've made up, tongue-in-cheek but not in an arbitrary way: "How African is Sali's life in America, and why?" and "The ecology of Sali's survival in Africa then in America". Levity aside, can you help us identify any of your various scientific quests that may be substantially related to deeply personal aspects of your life, be they emotional, spiritual, socioeconomic, ideological, etc.?

But, before you answer, let me clarify a bit:

One can – of course, with hindsight and with imagination, lots of it! – link the above (real and fictional) titles to aspects of your biography-cum-geography. One

could creatively link these titles to your, or any migrant's, efforts at adapting to new "ecologies" while trying to preserve cultural roots and avoid various sorts of imagined and all-too-real attempts at extinction. Understanding that this dialogue is neither confession, nor post-modern literary criticism, one could nevertheless ask about Sali's biography the same questions that Sali asked of Gullah and its (original) speakers: how is identity defined and/or transformed in exile? Or: how "African" is Sali in America?

SALI: Starting my professional career in Jamaica after graduating from the University of Chicago was a blessing, though it did not feel that way at the time. I realized how many brilliant minds there are in so-called "Third World" countries, especially among students, who either will never get an academic voice or won't even have opportunities to fulfill their intellectual dreams. That's part of the burden of being a "minority scholar" or rather a "minoritized scholar", namely, having to speak not just to express one's own views but also those shared with other disenfranchised thinkers who have no access to the platform that one has reached, especially regarding unproven and puzzling assumptions about their languages and cultures.

It is in this spirit that the dedication of my book *Créoles, écologie sociale, évolution linguistique* (Mufwene 2005) says: "*A tous mes collègues francophones, surtout les plus défavorisés, et à ceux qui sont privés de la plateforme d'expression dont je dispose.*" After a few years on some councils of the Organisation Internationale de la Francophonie (OIF) in the 1990s (up to 2002), I was reminded how economic poverty may enable the silencing of intellectual innovations that are considered subversive by "*les bailleurs de fonds*" in the Global North. Consistent with your "funder principle," the latter are often more interested in seeing their funding serve their own interests rather than doing the right thing. Voices of many African scholars on these councils would not be considered seriously unless they were supported by some powerful Western members of the relevant councils and made more consistent with their neo-colonial ideologies. I had the privilege of counting as an American scholar from a prestigious university (luckily nominated by a couple of those powerful colleagues who could live with heresies which were not dangerous to their agendas). In any case, needless to say I became annoying, especially regarding whether French should be considered an endangered language and whether a "partnership" should be set up between [vernacular speakers of] French and those of indigenous African languages to fight the spread of English. (I use the square brackets because OIF spoke of "*langues partenaires*," consistent with the bizarre metaphor of "killer language" in Anglophone literature.) Well, I have no longer been invited to sit on OIF's advisory councils, but I have published a few papers on language endangerment in which I hope people can recognize that

I represent the Global South, whose dynamics of language coexistence have not been like they are in the Global North.

Anyway, let's return to my Jamaican débuts. Working with senior colleagues like the late Mervyn Alleyne and the late Dennis Craig, as well as with my peer Hubert Devonish at the University of the West Indies at Mona, Jamaica, helped me pay more attention to the linguistic and political realities around me. Interactions with these Caribbean colleagues made me more interested in Creole languages. None of these scholars can be considered a conformist. None of them expected me to fully agree with them, either. So, as you see, little could have been more nurturing than such a setting for a heretical mind.

I went to Jamaica as a theoretical linguist and left it fascinated by the myriads of challenges that the study of Creole languages presents to general linguistics. The more I know about Creoles (strictly, those lexified by European languages), the more questionable I find several assumptions about them, and the more I believe that creolistics, in return, should contribute to general linguistics.

As you noted above, I suggested in my book *The Ecology of Language Evolution* that creolistics has been too much of a consumer discipline: the instances are all too rare where basic assumptions in general linguistics are questioned because of facts observed about Creoles. For instance, one could ask what makes them, or perhaps what does *not* make them, peculiar regarding speech continuum, non-monolithic grammatical structures (or "coexistent systems" in the way William Labov prefers to discuss this aspect of African American English), and genetic classifications.

Speech continua are everywhere, regardless of whether you focus on regional or social variation. The boundaries posited by dialectologists are conveniences for academic discourse. Work by William Kretzschmar (University of Georgia) and Dennis Preston (emeritus at Michigan State University and now at Oklahoma State University) reveals that it is naive to reify those boundaries. Also, the stratification of lects into basilect, mesolect and acrolect could apply anywhere. What has received the least attention in this context is the role of idiolects, which reflect the fact that speakers vary among themselves anatomically (thus in their perceptions and ability to reproduce what they perceive) and mentally (every teacher should know this!). When you add the fact that naturalistic language learning/development proceeds by inference, and the speakers' interactional histories are not identical (as I pointed out in *Language Evolution: Contact, Competition and Change*, 2008), continua should have been treated as normal and common phenomena that need not be associated, exceptionally, with "decreolization," which I prefer to call "debasilectalization" ("[m]oreover, from a social ecological perspective, no two speakers have had identical experiences of social interaction, hence of being exposed to identical primary linguistic data, which have influenced their language

'acquisition' processes" [2008: 27]). In my opinion, at the risk of being offensive, since Creoles cannot be defined structurally, perhaps the whole spectrum from the basilect to the acrolect (which was not the lexifier, by the way!) should be identified as Creole.

Creole languages have been disenfranchised socially and "exceptionalized" in linguistics because of the arbitrary and almost exclusive focus on their basilects, the most non-standard varieties (contrasted mistakenly with the standard varieties of their lexifiers – thus distorting the actual language contact history). This bias seems to have been enticed by the non-European identity of the subaltern populations that produced them. (Your absolutely well-documented publications on and against "Creole Exceptionalism" [DeGraff 2005, 2020] trace this all the way back to the nineteenth century!) The practice suggests that, contrary to the reality we all know, the non-Creole colonial restructured varieties of the same European languages are all standard. From an evolutionary perspective, I see more theoretical significance in Mervyn Alleyne's identification of what evolved in the Caribbean as "Haitian," "Jamaican," "Bajan," etc., without the term "Creole," although the natives have every right to break the continuum into as many distinctions as they think appropriate, for reasons that satisfy their politics. From an evolutionary perspective, Alleyne's labels cover the full continua, from the basilect to the acrolect, as new, colonial phenomena diverging from the corresponding European linguistic continua, just as "American (English)" covers the full regional and social continuum. Therefore, if non-standard Haitian or Jamaican are Creole, so are the acrolectal counterparts, consistent with my position that it is not structural features that account for disenfranchising the basilectal varieties as Creole.

I will not elaborate further, but this is an issue I raised in passing already in *The Ecology of Language Evolution*. Now, as I suggest above, the politics of whether the basilect and the acrolect are the same language is one that only the natives of the relevant polities can handle among themselves. I expressed this subversion from a uniformitarian, evolutionary linguistics perspective. There is a sense in which basilectal varieties of Creoles are the counterparts of non-standard English and French in the former settlement colonies of the Americas and elsewhere, but it would be a colonial attitude on my part me as an outsider to impose this position on the natives! It should not matter whether the varieties are mutually intelligible or not, unless the natives want to capitalize on this factor. There are languages that are claimed to be mutually intelligible, just as there are dialects that are said not to be mutually intelligible.

Regarding non-monolithic systems, there are, in normal Creole speech, overlaps and frequent alternations between structures preserved intact from the lexifier (a set-theory union of non-standard varieties to which the Africans were exposed on Atlantic and Indian Ocean plantation colonies) and innovations (features not

recognizable by speakers of the lexifiers). The latter are what the debate on the development of Creoles has preferred to focus on. (That's part of the bias that has overemphasized divergence over inheritance, and has precluded Creoles from serving as windows onto earlier stages of their colonial lexifiers.) However, rule overlap is true of non-Creole systems too. They dispute Antoine Meillet's slogan that "*la langue est un système où tout se tient*" or Ferdinand de Saussure's claim that the components of a language are integrated like pieces of a mosaic. If one must apply that Saussurean metaphor, reality reveals language to be a sloppy mosaic in which the pieces are far from being mutually delineating. In reality, these pieces of language overlap almost everywhere. (This is part of what produces the kind of variation so central to Labovian linguistics.)

Regarding genetic classification, it is high time the kinds of external history brought to bear in discussions of the development of Creoles were applied to the evolution of other languages. Some have claimed that the comparative method cannot apply to Creoles, hence they cannot be classified genetically. That's an a priori conclusion, especially given an embarrassing practice that has compared Creoles with standard varieties that had little to do with their emergence. The more we seem to grasp about the complexity of the development of Creoles, the more I believe the books should be reopened and rewritten regarding certain positions in genetic linguistics, and regarding the explanatory significance of Stammbaums. Those trees show the end results of processes not fully accounted for. Genetic creolistics is crying out for the missing explanation. I have attempted this in a couple of papers, one of which is published in the second edition of the *Handbook of Historical Linguistics*, edited by Brian Joseph, Richard Janda, and Barbara Vance, though you undoubtedly recognized the beginnings of this evolution of my work in *The Ecology of Language Evolution* (2001). Meanwhile, the dogmatic distinction between Creoles and non-Creoles becomes more and more artificial and fades away.

When I started my career in Jamaica, the dominant trend worldwide was to discuss Creoles as aberrations or deviations of some sort, as languages that still needed to be dignified as "normal". Some creolists still characterize Creoles, even in some of our most prestigious journals, as "unnatural" or "irregular". Unfortunately, and as you yourself have been documenting in your articles on and against Creole Exceptionalism (DeGraff 2005, 2020) and in your much appreciated work in the MIT-Haiti Initiative advocating the use of Kreyòl in Haitian schools and universities, misconceptions about Creoles have not changed much since the nineteenth century.

Not having been trained in creolistics in graduate school, I felt odd trying to analyze Creoles according to the standard training I had received in general linguistics (especially regarding the semantics of time reference) and figuring out

how to enrich the theoretical framework I was using. Yet I saw no sound alternative to this uniformitarian approach. One blessing in my life is that Caribbean scholars paid attention to what I was doing, while the rest of the world did not seem to care, including those who had trained me at Chicago. Perhaps I should not be that unfair. Some of my former professors did see the value of my scholarship and eventually brought me back to Chicago; and it now looks like many scholars elsewhere did, considering how often I have been invited to contribute to special volumes, to participate in specialized workshops, and to give public lectures and keynote addresses.

In any case, the fact that the scholars the closest to the Creoles that interested me did not think I was insane was reassuring enough for me. Today, I also feel that more attention should be paid to scholars who evolve in places where Creoles are spoken. For instance, the papers of Yves Dejean that you have shared with some of us pointedly expose some unjustified assumptions about Creole communities and about the coexistence of Creoles and their acrolects, especially where the actual non-standard lexifiers are no longer spoken, such as in Haiti and Jamaica. Please don't get me wrong. I'm not saying that French and English are no longer spoken in these polities. It's that the non-standard colonial varieties to which the Africans were exposed are no longer spoken. Those were the real lexifiers. It's by paying attention to colonial non-standard French and English in places like Louisiana, St Barths, Smith, and Ocracoke Islands and the like that one realizes, through some structural similarities with Creole basilects, what features the colonial lexifiers appear to have shared. Naturally, the varieties I have mentioned here are not snapshots of those early colonial varieties either. They too have evolved.

Oops! "Acrolect" is a term I too should use with caution. In 2003 I read an insightful paper by a Jamaican scholar who questions the colonial way in which the term has been defined, based on a foreign standard. In the paper, which has now evolved into a book (from a transitional stage as a PhD dissertation), *The Acrolect in Jamaica: The Architecture of Phonological Variation* (2018), G. Alison Irvine-Sobers perceptively questions whether the notion can be defined structurally, especially on a pre-defined battery of features dictated by a foreign norm (a "normative gaze" from the outside). She then goes on to reveal embarrassing inconsistencies in our scholarly practice. I hope creolists can adequately appreciate the value of this scholarship inspired by local sociolinguistic facts observed by a local investigator.

You ask: "How African is Sali in America?" It is difficult to answer that question. When I returned home to the Congo for the first time in 1984 (after ten years of absence), even members of my own family thought I was not fully "African" any more. Of course, I was not and am not – except in my phenotype. I am now an epitome of culture contact in an exogenous setting and have been both partly

deculturated from my ancestral cultural background and acculturated to other ecologies since I left home. If one could define "Creole" in terms of cultural hybrid-ization as done by the authors of *Éloge de la créolité*, then I have "creolized" too, albeit anachronistically. I have been influenced by the places where I have lived and the many good friends I have socialized with, by my (self-)training in a variety of disciplines, true to the interdisciplinary orientation of my scholarship, and by my interactions and occasional collaborations with some colleagues from different intellectual and cultural backgrounds, including yourself and Enoch Aboh, both generativists. I also developed a friendly relationship with the late Robert Chauden-son, as I had done with Guy Hazaël-Massieux, despite some differences regarding details of the emergence of Creoles. I could mention more names, but then this section would be much longer than need be. (I have often spoken of "polyploidy" in the emergence of idiolects, with each individual speaker selecting inputs into their emergent idiolects from diverse speakers and recombining them into their own systems. My intellectual and social development instantiates the phenomenon!)

Nonetheless, I must acknowledge my gratitude to lay people of African descent in Jamaica and the Sea Islands of South Carolina, as they have contributed significantly to strengthening my Africanity. I have always felt the most natural as a human being, rather than as an intellectual machine, among them. They remind me of my very modest background and the pride that we all should have in our cultural ancestries. It is the natural cultural behaviour of these people, and their folk interpretations of how they see themselves, that have shaped part of my schol-arly "restructuring," as you may express it.

However, there is also another side of the coin in this trajectory of my life. I could not help noticing that the Africa I saw in 1984 was no longer the one I had frozen in my memory. There is no static African culture any more than there is any static culture anywhere on this planet. This is one of the misunderstandings in the literature on language endangerment, where scholars forget that members of a population make their culture *and* their languages as they evolve from day to day.

I also know that I am not fully Americanized, not just because I still have an accent but also because I often react un-Americanly to situations. I am just an instance of culture contact in North America, absorbing a new culture against the backdrop of an Africa I brought with me in the 1970s, influenced by several other cultures that I have appreciated in my travels around the world, and eclectic in my behaviour as in my thinking.

You can see the kinds of experience and attitude that must have contributed to my "feature pool" theoretical idea, and to the centrality of the behaviours of *individuals* in my model of language evolution. This model is not driven only by my familiarity with the literature on population genetics, but also by my meta-perception of my varied life experience.

I do feel that I am a world citizen with a strong and nostalgic attachment to part of the Africa in which I grew up, and with aspects of the West that I have selected into my present personality. The ecology of my survival lies in that eclecticism, being adaptive to new living conditions and not losing a sense of who I am – better yet, of what I do not want to be – constantly evolving, with continuities and changes.

In exile, in a new setting, the balancing act in the give-and-take game of life is difficult, though easier when one can live without the impositions of self-conscious behaviour. The evolution of one's personality is in some ways like language evolution, by competition and selection, which takes place largely without engaging one's own awareness.

On symbolic markets and personal investments in Creole studies and in Creole communities

MICHEL: I remember reading somewhere that identity may be less about who we are than about who we feel we aren't or (are made to feel) we can't be. But, don't worry, I won't ask you who it is you don't want to be like!

Here's one other question about the (implicit) biography in your bibliography. Besides accidents of history and geography (e.g., your fieldwork on Jamaican Creole in Jamaica and Gullah in South Carolina), are there any particular salient events – any clearly identifiable turning points – that have attracted you from one research topic or one theoretical framework to the next?

Here I am thinking about the title of one of your articles on (dis)similarities between creolization and language acquisition, "Hints from Tazie" – Tazie being your now adult daughter born in the USA. This is one of the titles that suggest that for you the personal is never too far from the intellectual. In your internal and external ecologies, the private and the public are perhaps not in binary opposition, but overlapping regions in a continuum. Of course, you should feel free to draw the line as you see fit for this conversation.

Let's take another topic, perhaps more germane to questions of identity, migration, transformation and extinction, namely your interest in language endangerment. This interest is not so recent. Witness your 1991 article "Some reasons why Gullah is not dying yet". Then, you revisited the topic in your book on *The Ecology of Language Evolution*. There you further articulate your account of language evolution by competition and selection, subject to factors defined by specific ecologies, towards an understanding of how the socioeconomic activities of a population influence the fates of their languages.

From your answer above, one could reasonably speculate that your approach to language evolution is ultimately connected to a wider probe about personal

transformation (via, e.g., migration, career moves and investments in symbolic markets à la Bourdieu). One could also speculate about the metaphysical implications of your work vis-à-vis the intrinsic impermanence of individuated cultural phenomena, personal relationships, life, and so on.

I realize that this line of questioning puts me far out on a new-age limb. Do not hesitate to bring me back to earth – or in line. Better yet, just ignore anything you don't want to discuss.

SALI: Oh, dear! You have been asking me tough questions! I have not spent much time putting my life in perspective. I think that typically I have reacted to explanations that I consider implausible or downright outrageous while working with others that I find adequate. The state of the art in creolistics in the early 1980s is really what brought me to the field.

I am grateful that some common threads have emerged in the ways I have approached issues, and that I could integrate some of the discussions in my books. For the longest time, I considered myself primarily a critic of the scholarship. But it is now more obvious to me that I have been doing more than just critiquing what others have done. I have actually outlined a research programme of my own, and generated a particular uniformitarian approach to language change and speciation that can bridge creolistics and other diachronically-oriented research areas in linguistics and even beyond. We must just bear in mind that uniformitarianism does not entail the same explanation for every situation – only that similar ecological factors may entail similar explanations, with a great deal of the actual outcome of language contact also depending on what exact contingencies these ecological factors apply to and when.

Originally, I was interested in morphosyntactic characteristics of Creoles. The trigger was really some dissatisfaction with the state of the art in the early 1980s. Then I was appalled by the unnecessarily special explanations proposed for the development of Creoles – what you've been calling "Creole Exceptionalism" in your own work (DeGraff 2005, 2020). Since I was not trained in creolistics and thus started without a global view of the field, I basically have continued to react with consternation to some of the accounts I have read.

The decreolization hypothesis seemed outrageous to me, because everything I learned about the history of the relevant territories suggested a different language evolution trajectory, more consistent with the late Robert Chaudenson's views on the development of Creoles, though not identical. These views have too hurriedly been dismissed by some as "superstratist". The very suggestion that people from the lower class aspire to speak like those of the upper class is so contrary to the sociolinguistic reality around us. The suggestion that the factors associated with the putative decreolization would have worked on African Americans but not on

white speakers of similar vernaculars in the USA is preposterous. I have also been struck by the fact that little attention has been paid to how speakers of Creoles and African American English comment, sometimes derisively or mockingly, on the acrolects of their multilectal and multiethnic continua. Speakers of Creoles and African American English know they are disenfranchised socioeconomically. However, while it is evident that they would be happy to improve their economic conditions, far from their minds is the idea that they would also like to behave linguistically and in other cultural ways exactly like speakers of the corresponding acrolects. Their vernaculars have a social, often also ethnic, indexical value they are not ashamed of, contrary to what one may infer from the decreolization hypothesis.

Perhaps another biographical footnote is in order here, in the spirit of your "biography-cum-bibliography" focus: I myself started my life at the bottom of the socioeconomic ladder, and have often wanted things that the upper strata could afford, but never could I think of wanting to be like them.

In the context of the USA, you can see that at some point linguists must have misinterpreted African Americans' struggle for equality as some desire to (fully) adopt European-American values or become European-American cultural clones. Anybody who has observed affluent African Americans should know much better. There is cultural diversity even in upper middle-class America.

The decreolization hypothesis is an unfortunate misinterpretation in a linguistics that has been primarily exercised by scholars from the white middle class. These scholars seem to not realize that the underprivileged populations whose language varieties they have investigated have, by and large, no social identity problem, and have not wanted to be like them or speak a vernacular like the scholars', though some of the Creole speakers have felt the need to speak another lingua franca for socioeconomic reasons – a lingua franca that happens to be similar to that spoken by the scholars. Speakers can accumulate a wide range of varieties, depending on access and learning skills, but the additional varieties need not replace the earlier ones. The decreolization hypothesis is a pathological interpretation that has little to do with Creole speakers.

Having dealt with such issues in creolistics, it was only natural for me to voice my opinion in the ongoing expressions of concern about language endangerment. The vast majority of linguists who have expressed opinions on the subject matter are theoretical linguists, who have done relatively little research on language ecology. Their opinions seem to reflect more guilt about European colonization of the past 400 years than a real understanding of language vitality, the broader context in which language endangerment must be discussed. Worst of all, they have expressed more concern about languages as commodities for linguistic analysis than about the costs and benefits to the populations who have shifted languages. I couldn't believe my eyes when I read in a publication by a linguist whom I need

not name that indigenous languages are a form of wealth, presumably comparable to the material wealth of polities which are poor in linguistic diversity. The most charitable interpretation I could make of it was that the intention was right but poorly expressed in a book (*Language and Poverty*, 2008) that could have been expected to show the extent to which the language used in the school system, and in disseminating useful information, can contribute to preventing economic development if it is not what the relevant population speaks as its vernacular.

Even a well-documented book such as *Vanishing Voices* (by Daniel Nettle & Suzanne Romaine 2000) fails to address this socioeconomic aspect of the subject matter, at least to my satisfaction. In several ways, as compared to environmentalists' research into the situation of endangered species, linguists concerned with endangered languages pay too little attention to language ecology, the dynamics of which are poorly understood. The very confusion of language "maintenance" with language "preservation", which is evident in much of the literature, is embarrassing for the field. It's not clear to me how a language can be "preserved" even now, at the time of excellent recordings, since utterances are not language itself. Those documenting languages are certainly not preserving them, not even in ways comparable to lifeless fruits or brains in jars, because narratives, dictionaries, and grammars are not languages, though they make available materials out of which languages can be (re-)created.

It's interesting that you invoke Bourdieu's notion of the "linguistic market". Anybody who is consistent with that approach should realize that it is inevitable that Creole speakers will preserve their Creole while they are still marginalized socioeconomically. And, still following Bourdieu's linguistic-market model, there will be an array of symbolic and real – and sometimes conflicting – interests to be derived from the maintenance of the Creole. Popular culture in the Caribbean has certainly contributed to it, which politicians remember during political campaigns. It helps to reconnect with one's constituency in their vernacular rather than in the elite variety!

It's sad, isn't it, that the overwhelming majority of a country's population is crowded at the margins, especially in socioeconomic and political terms. The linguistic marginalization is but a reflection of these margins. It did not strike me until I spent time in Jamaica that the vitality of Patois arises from that marginalization, which leads the population to exploit linguistic differences and identify with Patois. The same must be true of Haiti; I remember you quoting the phrase "Linguistic Apartheid" from Paul Dejean's (1993: 123–124) work. (Things are of course more complex than the simplification I am presenting here.)

A whole lot of the misunderstanding now has to do with misconceptions about how globalization works, but we'll leave that alone. I have said plenty about this in *Language Evolution: Contact, Competition and Change* (2008) and several other publications on language vitality.

Let me reach again for a biographical footnote – and I am glad that these footnotes can be put in "focus"!

The adaptive pressures that you and I as immigrants to North America experience in our linguistic and economic lives differ in significant ways from those of Creole speakers in Jamaica or Haiti, because the socioeconomic structures are not the same, and challenges to an immigrant are not the same as those to a native.

The situation in Jamaica reminded me of socioeconomic conditions in the Congo and other sub-Saharan African nations, where a foreign language is used in the small sector of the economy that participates in, or interfaces with, the global economy of the world. An important difference is that in these other places the languages of the masses are not genetically related to the official languages. Another difference is that the Congo is heavily multilingual, with lots of ethnic diversity, whereas places like Haiti are virtually monolingual and mono-ethnic. I think in Jamaica ethnic stratification has taken a back seat to economic stratification. But that's a rather complex topic, which I would prefer to discuss elsewhere.

While in Jamaica, it was easy for me to see why, from a political point of view, some have wanted to identify Patois as a separate language, though its speakers are really ambivalent about whether or not it is. My students have also made me more ambivalent about this issue.

One thing that was evident to me was that the worsening economic conditions seemed to have led speakers of Patois to use it with more pride as a marker of identity. In the Caribbean, as in sub-Saharan Africa, the scarcity of, or lack of growth in, job opportunities requiring the privileged variety has created little motivation for shifting to the latter. Popular culture, especially music, in which artists (many of them little schooled) often articulate the frustrations of the disenfranchised majority, has nurtured the Creole vernaculars (and indigenous languages in the case of Africa). Note also that the minoritized majority socialize among themselves, in their vernaculars of course. The elites of their families, if any, also reconnect with them in these heritage vernaculars, not wanting to be treated as snobs or haughty, especially when interacting with the elderly. It could be that I had failed to notice it twenty to thirty years earlier, but it was now more obvious to me that just the opposite of the mythical decreolization was happening in Jamaica. This is one thing that one may want to read about in Velma Pollard's work on Dread Talk.

On the journey ahead

MICHEL: Is there anything that you haven't found time to work on, and that you fear you may never have time to work on because of your other "competing" (pun

intended) interests and obligations? In other words, what other academic pursuits might you have "selected" if your own "ecology" had been different?

SALI: Several years ago now, my daughter and I once bought two books on serendipity, which underscore accident as an important factor in the development of my research questions and hypotheses. There are several possible ramifications of what I have done to date. I believe that my main challenge is to discipline myself so that I can continue to address the many questions that arise from my book *The Ecology of Language Evolution*, which I consider the summary of an ambitious research programme. I should refrain from other interesting pursuits, except social ones.

My visits to Jamaica and East Asia caused me to think more about the differing ways in which colonization, decolonization, and globalization have taken place around the world, and the concurrent diversity of their influences on language evolution, which is differential (producing non-identical outcomes in different ecologies) but still consistent with the uniformitarian approach to the subject matter, as I noted above. My accidental discovery of the separate social identity of the Peranakan Chinese in the Straits of Malacca (well known to local historians) and the role they played in language evolution in the region has caused me to rethink some assumptions about the development of Creoles, for instance, the mischaracterization of the roles of social status and practicality (of costs and benefits to speakers) in language shift.

Learning about settlement patterns in colonial Singapore has also led me to wonder whether the mixing of Africans on Atlantic and Indian Ocean plantations really had a similar counterpart in Hawaii, and whether we have been correct in assuming that the ecologies of the development of Caribbean Creoles are so similar to that of Hawaiian Creole. Since 2004, I have argued in some publications, including *Language Evolution: Contact, Competition and Change* (2008), that the late Derek Bickerton and others were incorrect in assuming that Hawaiian Creole had evolved in the same kind of contact ecology as Caribbean and Indian Ocean Creoles. I single out Bickerton because of his forcefulness in proposing a unified account of the emergence of Creoles, however controversial the Language Bioprogram Hypothesis has been.

Reading more about population movements in the history of mankind, including the work of Cavalli-Sforza and André Martinet on human dispersals from the Eastern European steppes towards the Atlantic Ocean, has caused me to wonder why genetic linguistics has ignored, or downplayed, the role of contact in language diversification. Can one investigate language speciation without being informed by population genetics, or studies of populations in general? I was more encouraged into this kind of thinking after reading David Reich's *Who We Are and How We Got*

Here: Ancient DNA and the New Science of the Human Past (2018). It leads to a different understanding of *Homo sapiens*'s dispersal out of Africa and the emergence and speciation of Indo-European languages. The dispersal routes as suggested by genetics dispute traditional hypotheses, and make allowance for population contacts as actuators of change and speciation. The emergence of Creole languages appears to be the tail end (from our temporal perspective now!) of something that has recurred several times in history, minus chattel slavery of course. Contrary to Creole Exceptionalism, can genetic creolistics help us reopen the books for a more adequate interpretation of externally versus internally motivated language change, the actuation of change, language shift, and language speciation, among other things? Instead of thinking that, because Creole languages are allegedly exceptional, we cannot answer so many questions about how their grammars emerged, we should be asking whether some of the assumptions we have inherited from historical and genetic linguistics are not being disputed by the kinds of data we are dealing with in Creole studies. And we probably know more about the ecologies of the emergence of these new vernaculars than we do about those of their lexifiers. There are a host of other questions on my mind, including the following:

– Does it make sense to study language change without considering patterns of interaction among speakers?
– Even if contact of languages is not factored in, how about the variation inherent in a language among its idiolects and among its dialects?
– Can one address the actuation question without addressing these factors in language evolution?

I obviously have more research questions than I can pursue in the remaining time of my life, and I hope there are junior scholars around interested in addressing some of them. However, there is one that I am not ready to relegate to the younger generation: may Pidgins have emerged later than Creoles and in territorial complementary distribution with them? There is a growing number of publications on trade between European mercantile companies and African rulers that suggests that the transactions took place through interpreters, who became important power brokers all the way into the exploitation colonization of the African continent, Asia, and the Pacific since the late nineteenth century. Even the "exploitation colonization" itself proceeded though interpreters, to whom the late William Samarin referred as "colonial auxiliaries". No wonder the phenomenon of "elite closure" is so deeply rooted in sub-Saharan Africa! The vast majority of Africans socialize in their indigenous vernaculars or lingua francas.

I am dying to get my hands dirty with field research and structural analysis again, an anticipated healthy break from my present concerns with ecological

aspects of language evolution. Every time I read your papers, for instance, I wish I could set some time aside to explore just another aspect of Creole structures, especially those presumed to be unproblematic. There's usually a lot of excitement to derive from those unexpected discoveries, and a lot of unsuspected challenges to face, which invite us to reopen books that were closed too soon.

Creolistics is such a poorly exploited gold mine for general linguistics!

References

DeGraff, M. 2001. On the origin of Creoles: A Cartesian critique of neo-Darwinian linguistics. *Linguistic Typology* 5(2–3): 213–310.

DeGraff, M. 2005. Linguists' most dangerous fallacy: The myth of Creole Exceptionalism. *Language in Society* 34: 533–591. https://doi.org/10.1017/S0047404505050207

DeGraff, M. 2020. Toward racial justice in linguistics: The case of Creole studies. *Language* 96(4): e292–e306. https://doi.org/10.1353/lan.2020.0080

Dejean, P. 1993. *Haiti: Alerte, on tue!* Montréal: CIDHICA.

Irvine-Sobers, G. A. 2018. *The Acrolect in Jamaica: The Architecture of Phonological Variation*. Berlin: Language Science Press.

Mufwene, S. S. 1986. Les langues créoles peuvent-elles être définies sans allusion à leur histoire? (Can Creole languages be defined without reference to their history?). *Études Créoles* 9(1): 135–150.

Mufwene, S. S. & Gilman, C. 1987. How African is Gullah, and why? *American Speech* 62(2): 120–139. https://doi.org/10.2307/455274

Mufwene, S. S. 1991. Some reasons why Gullah is not dying yet. *English World-Wide* 12(2): 215–243. https://doi.org/10.1075/eww.12.2.04muf

Mufwene, S. S. 1996. The founder principle in Creole genesis. *Diachronica* 13(1): 83–134. https://doi.org/10.1075/dia.13.1.05muf

Mufwene, S. S. 1997. The ecology of Gullah's survival. *American Speech* 72(1): 69–83. https://doi.org/10.2307/455608

Mufwene, S. S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511612862

Mufwene, S. S. 2005. *Créoles, ecologie sociale, evolution linguistique*. Paris: L'Harmattan.

Mufwene, S. S. 2008. *Language Evolution: Contact, Competition and Change*. London: Continuum Press.

Nettle, D. & Romaine, S. 2000. *Vanishing Voices: The Extinction of the World's Languages*. Oxford: OUP.

Reich, D. 2018. *Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past*. New York NY: Pantheon Books.

Language Index

Subject Index

Variation Rolls the Dice: A worldwide collage in honour of Salikoko S. Mufwene aims to celebrate Mufwene's ground-breaking contribution to linguistics in the past four decades. The title also encapsulates his approach to language as both systemic and socio-cultural practices, and the role of variation in determining particular evolutionary trajectories in specific linguistic ecologies. The book therefore focuses on variation within and across languages, within and across speakers, and how this fundamental aspect of human behavior can affect language structure in time and space. Mufwene has been instrumental in putting creole languages on the map of General Linguistics and connecting their analysis to issues of language acquisition, multilingualism, language contact, language evolution, and language typology. Thanks to the diversity of topics and the wide-ranging theoretical persuasions of the contributors, this volume aims at a large readership including both scholars and advanced students interested in cutting-edge research in the aforementioned domains.

JOHN BENJAMINS PUBLISHING COMPANY